LNAI 6521

Mohua Banerjee
Anil Seth (Eds.)

Logic and Its Applications

4th Indian Conference, ICLA 2011
Delhi, India, January 2011
Proceedings

FoLLI

Springer

Lecture Notes in Artificial Intelligence 6521

Edited by R. Goebel, J. Siekmann, and W. Wahlster

Subseries of Lecture Notes in Computer Science

FoLLI Publications on Logic, Language and Information

Mohua Banerjee Anil Seth (Eds.)

Logic and Its Applications

4th Indian Conference, ICLA 2011
Delhi, India, January 5-11, 2011
Proceedings

Springer

Volume Editors

Mohua Banerjee
Indian Institute of Technology Kanpur
Department of Mathematics and Statistics
Kanpur 208016, India
E-mail: mohua@iitk.ac.in

Anil Seth
Indian Institute of Technology Kanpur
Department of Computer Science and Engineering
Kanpur 208016, India
E-mail: seth@iitk.ac.in

Library of Congress Control Number: 2010941366

CR Subject Classification (1998): I.2.3, I.2, F.1-4, D.2.4

LNCS Sublibrary: SL 7 – Artificial Intelligence

ISSN	0302-9743
ISBN-10	3-642-18025-6 Springer Berlin Heidelberg New York
ISBN-13	978-3-642-18025-5 Springer Berlin Heidelberg New York

springer.com

Printed in Germany

Typesetting: Camera-ready by author, data conversion by Scientific Publishing Services, Chennai, India
Printed on acid-free paper 06/3180 5 4 3 2 1 0

Preface

This volume contains extended abstracts of some papers presented at ICLA 2011: the 4th Indian Conference on Logic and Its Applications held during January 9–11, 2011, at Delhi University.

ICLA is a biennial conference organized under the auspices of the Association for Logic in India. Its scope includes pure and applied formal logic as well as the history of logic with emphasis on relations between traditional Indian systems and modern logic.

In response to the call for papers for ICLA 2011, there were 34 submissions. Each submission was reviewed by at least two, and on average three Programme Committee (PC) members. Some PC members chose to consult external reviewers whose names are listed herein. The committee decided to accept 14 papers for presentation and publication in this volume and another ten papers for presentation only. The programme also included three invited talks.

We are grateful to the PC members for their efforts in reviewing and selecting the papers. We also thank all the external reviewers for their help. Special thanks are due to the invited speakers for contributing their papers to the proceedings at a short notice. Finally, we thank all those who submitted their papers to ICLA.

The EasyChair system was of great help in the submission stage, the PC meeting and finally in preparing the proceedings. We also thank the Editorial Board of the FoLLI series, and Ursula Barth of Springer for overseeing production of the final volume.

November 2010

Mohua Banerjee
Anil Seth

Conference Organization

Programme Chairs

Mohua Banerjee	IIT Kanpur, India
Anil Seth	IIT Kanpur, India

Programme Committee

Dietmar Berwanger	ENS Cachan, France
Achim Blumensath	TU Darmstadt, Germany
Richard Booth	University of Luxembourg, and Mahasarakham University, Thailand
Didier Dubois	IRIT, Paul Sabatier University, Toulouse, France
Valentin Goranko	Technical University of Denmark
Wilfrid Hodges	United Kingdom
Peter Jipsen	Chapman University, Orange, CA, USA
Steven Lindell	Havenford College, PA, USA
Benedikt Löwe	ILLC Amsterdam, The Netherlands
Shahid Rahman	University of Lille 3, France
R. Ramanujam	IMSc Chennai, India
K. Ramasubramanian	IIT Bombay, India
Sundar Sarukkai	Manipal University, India
S.P. Suresh	Chennai Mathematical Institute, India
Andrzej Szałas	Linköping University, Sweden
Katherine Thompson	University of Vienna, Austria
Jouko Väänänen	University of Helsinki, Finland
Scott Weinstein	University of Pennsylvania, USA
Hans van Ditmarsch	University of Sevilla, Spain

Organizing Committee

Mohua Banerjee	IIT Kanpur, India
Amitabha Gupta (Chair)	Former Professor, IIT Bombay, India
Kamal Lodaya	IMSc Chennai, India
Godabarisha Mishra	Indian Council of Philosophical Research, Delhi, India
R. Ramanujam	IMSc Chennai, India
S.P. Suresh	Chennai Mathematical Institute, India
Ashok Vohra	Delhi University, India

External Reviewers

Vince Barany
Arina Britz
Martin Caminada
Louise Dennis
Christian Fermueller
Vassilios Gregoriades
Davide Grossi
Joost Joosten
Alexander Kartzow
Alexander Kreuzer
Kamal Lodaya
Thomas Müller
Salvador Mascarenhas
Eric Pacuit
Steffen Reith
Floris Roelofsen
Sonja Smets
Ivan Varzinczak
James Worrell

Table of Contents

Semantics Based on Conceptual Spaces

Peter Gärdenfors

Lund University Cognitive Science, Kungshuset, S-22222 Lund, Sweden
Peter.Gardenfors@lucs.lu.se

Abstract. The overall goal is to show that conceptual spaces are more promising than other ways of modelling the semantics of natural language. In particular, I will show how they can be used to model actions and events. I will also outline how conceptual spaces provide a cognitive grounding for *word classes*, including nouns, adjectives, prepositions and verbs.

1 Introduction

Within traditional philosophy of language, semantics is seen as mapping between language and the world (or several "possible worlds"). This view has severe problems. For one thing, it does not involve the users of the language. In particular, it does not tell us anything about how individual users can grasp the meanings determined by such a mapping (Harnad 1990, Gärdenfors 1997).

Another tradition, *cognitive semantics*, brings in the language user by focusing on the relations between linguistic expressions and the user's mental representation of their meanings. According to cognitive semantics, the meanings of words are represented specifically as *image schemas*. These schemas are abstract mental pictures with an inherent spatial structure, constructed from elementary topological and geometrical structures like "container," "link", and "source-path-goal." Such schemas are commonly assumed to constitute the representational form common to perception, memory, and semantic meaning.

Although there have been some attempts to construct computational models of image schemas (e.g. Holmqvist 1993), they are not well suited for formal modelling. In particular, they are not well developed for handling dynamic entities, such as actions and events. In this article, I will model actions and events using conceptual spaces (Gärdenfors 2000). My goal is to show that conceptual spaces show more promise than other ways of modeling the semantics of natural language (see also Gärdenfors (1996)). I will further show how they can provide a cognitive grounding for *word classes*. In linguistics, word classes are defined by syntactic criteria. However, a theory of cognitive semantics worthy of its name should at least be able to explain the main categories of words – i.e., nouns, adjectives, prepositions, and verbs -- in terms of cognitive mechanisms. I will outline such an account.

2 Conceptual Spaces as a Semantic Framework

A given conceptual space consists of a number of *quality dimensions*. Examples of quality dimensions are temperature, weight, brightness, pitch, and force, as well as the

M. Banerjee and A. Seth (Eds.): ICLA 2011, LNAI 6521, pp. 1–11, 2011.

three ordinary spatial dimensions of height, width, and depth. Some quality dimensions are of an abstract non-sensory character. One aim of this article is to argue that force dimensions are essential for the analysis of actions and events.

Quality dimensions correspond to the different ways stimuli can be judged similar or different. For example, one can judge tones by their pitch, and that will generate a certain ordering of the auditory perceptions. As a general assumption, the smaller the distance between the representations of two objects, the more similar they are. The coordinates of a point within a conceptual space represent particular instances along each dimension: for example, a particular temperature, a particular weight, etc. For simplicity, I assume that the dimensions have some metric, so that one can talk about distances in the conceptual space. Such distances indicate degrees of similarity between the objects represented in the space.

It is further assumed that each of the quality dimensions can be described in terms of certain geometrical shapes. A psychologically interesting example is colour. Our cognitive representation of colour can be described along three dimensions. The first is hue, represented by the familiar colour circle going from red to yellow to green to blue, then back to red again. The topology of this dimension is thus different from the quality dimensions representing time or weight, which are isomorphic to the real number line. The second dimension is saturation, which ranges from grey at the one extreme, to increasingly greater intensities of colour at the other. This dimension is isomorphic to an interval of the real number line. The third dimension is brightness, which varies from white to black, and thus is also isomorphic to a bounded interval of the real number line. Together, these three dimensions---one circular, two linear---constitute the colour domain as a subspace of our perceptual conceptual space. It is typically illustrated by the so-called colour spindle.

The primary function of the dimensions is to represent various qualities of objects in different domains. Since the notion of a domain is central to the analysis, I should give it a more precise meaning. To do this, I will rely on the notions of separable and integral dimensions, which I take from cognitive psychology (Garner 1974, Maddox 1992, Melara 1992). Certain quality dimensions are *integral*: one cannot assign an object a value on one dimension without giving it a value on the other(s). For example, an object cannot be given a hue without also giving it a brightness (and a saturation). Likewise the pitch of a sound always goes with a particular loudness. Dimensions that are not integral are *separable*: for example, the size and hue dimensions. Using this distinction, a *domain* can now be defined as a set of integral dimensions that are separable from all other dimensions.

A *conceptual space* can then be defined as a collection of quality dimensions divided into domains. However, the dimensions of a conceptual space should not be seen as fully independent entities. Rather, they are correlated in various ways, since the properties of those objects modelled in the space co-vary. For example, in the fruit domain, the ripeness and colour dimensions co-vary.

It is impossible to provide any complete listing of the quality dimensions involved in the conceptual spaces of humans. Learning new concepts often means expanding one's conceptual space with new quality dimensions (Smith 1989).

3 Properties and Concepts

Conceptual spaces theory will next be used to define a *property*. The following criterion was proposed in Gärdenfors (1990, 2000), where the geometrical characteristics of the quality dimensions are used to introduce a spatial structure to properties:

Criterion P: A *natural property* is a convex region in some domain.

The motivation for Criterion *P* is that, if some objects located at *x* and *y* in relation to some quality dimension(s) are both examples of a concept, then any object that is located between *x* and *y* with respect to the same quality dimension(s) will also be an example of the concept.

Properties, as defined by criterion *P*, form a special case of *concepts*. I define this distinction in Gärdenfors (2000) by saying that a property is based on a single domain, while a concept is based on one *or more* domains. This distinction has been obliterated in both symbolic and connectionist accounts, which have dominated the discussions in cognitive science. So for example, both properties and concepts are represented by predicates in first-order logic. However, the predicates of first-order logic correspond to several different grammatical categories in natural language, most importantly those of adjectives, nouns, and verbs.

A paradigm example of a concept that is represented in several domains is "apple" (compare Smith et al. 1988). One of the first problems when representing a concept is to decide which are the relevant domains. When we encounter apples as children, the first domains we learn about are, presumably, those of colour, shape, texture, and taste. Later, we learn about apples as fruits (biology), about apples as things with nutritional value, etc.

The next problem is to determine the geometric structure of the domains: i.e., which are the relevant quality dimensions. Taste space can be represented by the four dimensions of sweet, sour, salty, and bitter; the colour domain by hue, saturation, and brightness. Other domains are trickier. For example, it is difficult to say much about the topological structure of "fruit space", in part because fruits (such as apples) can be described relative to several domains. Some ideas about how such "shape spaces" should be modelled have been discussed in e.g. Marr and Nishihara (1978), Edelman (1999), and Gärdenfors (2000). Instead of offering a detailed image of the structures of the different domains, let me represent the "apple" regions in the domains verbally, as follows:

Domain	*Region*
colour	red-yellow-orange
shape	roundish
texture	smooth
taste	regions of the sweet and sour dimensions
nutrition	values of sugar content, fibre content, vitamins, etc.
fruit	specification of seed structure, fleshiness, peel type, etc.

Concepts are not just bundles of properties. They are also *correlations* between regions from different domains that are associated with the concept. The "apple" concept has a strong positive correlation between sweetness in the taste domain and sugar content in the nutrition domain, and a weaker positive correlation between redness and sweetness.

Such considerations motivate the following definition for concepts. (For a more precise definition, see Chapter 4 in Gärdenfors 2000.)

Criterion C: A *concept* is represented as a set of convex regions in a number of domains, together with information about how the regions in different domains are correlated.

Elements from theories in psychology and linguistics contribute to the analysis of concepts I present here. The kind of representation intended by Criterion *C* is, on the surface, similar to *frames*, with slots for different *features* (sometimes called slots, attributes, or roles; see for example Noy and McGuinness (2001)). Frames have been very popular within cognitive science as well as in linguistics and computer science. However, Criterion *C* is richer than frames, since it allows representing concepts as more or less *similar* to each other and objects (instances) as more or less representative of a concept. Conceptual spaces theory can be seen as combining frame theory with prototype theory, although the geometry of the domains makes possible inferences that cannot be made in either of those theories (Gärdenfors 1990, 2000).

4 The Semantics of Adjectives, Nouns, and Prepositions

Next I will outline how analysing properties and concepts in terms of conceptual spaces can provide a *cognitive grounding* for different word classes. In this section I discuss adjectives, nouns, and prepositions. I will discuss verbs later.

The main semantic difference between adjectives and nouns is that adjectives (e.g., “red,” “tall,” “round”) normally refer to a single domain and represent properties; while nouns (e.g., “dog,” “apple”, “town”) normally relate to several domains. (Verbs, unlike nouns or adjectives, are characterized by their dynamic content, which I will analyze in terms of actions based on the force domain.)

Most properties expressed by adjectives in natural languages are natural properties according to Criterion *P*. For instance, in Gärdenfors (2000) I conjectured that all colour terms in natural languages express natural properties with respect to the the colour dimensions of hue, saturation, and brightness. This means that there should be no language which has a single word for the colours denoted by “green” and “orange” in English (and which includes no other colours), since such a word would represent two disjoint areas in the colour space. Sivik and Taft (1994) and Jäger (2009) have provided strong support for this conjecture. Their studies follow up on the investigations of basic colour terms by Berlin and Kay (1969), who compared and systematized terms from a wide variety of languages. Jäger (2009) studied colour classification data from 110 languages and found a median value of 93.6% correct classifications in an optimally convex partitioning of the colour space. Given the statistical aberrations in the data, this is a very high figure, which gives strong support to Criterion *P* at least within the domain of colours.

A paradigmatic example of the semantics for nouns is the analysis of "apple" from the previous section. Nouns do not only denote physical objects as located within a limited spatial region: consider, for example, "thunder", "family", and "language," let alone more abstract nouns. A noun typically denotes a phenomenon with *correlations* across a number of domains: in other words, nouns are represented by clusters in the conceptual space. Not all potential such clusters will be named by nouns in a language; an important factor is whether the correlations are pragmatically significant: that is, whether they are helpful in choosing the right actions.

Prepositions have likewise been studied extensively within cognitive semantics (for example, Herskovits 1986, Lakoff 1987, Landau and Jackendoff 1993, Zwarts 1995, Zwarts and Winter 2000, Zwarts to appear). A locative preposition (e.g., "in front of") combines with a noun phrase (e.g., "the castle") that refers to a spatially located object. The preposition maps the reference object to a region that is related to the object. (This criterion is put forward by e.g. Jackendoff 1983 and Landau and Jackendoff 1993, p. 223). Zwarts (1995) proposes to analyse this region as a set of vectors radiating from the reference object.

The basic semantic function of prepositions is to express spatial relations; but they are also used in a number of metaphorical and metonymic ways. Landau and Jackendoff (1993) offer a neuro-linguistic explanation. They propose two distinct cognitive systems: one for *objects* (the "what" system), and one for *places* (the "where" system). These systems relate to two different pathways in the visual cortex. The separation of the systems results in the separation between the *nominal* and *prepositional* systems in language. Zwarts (to appear) argues that, in some contexts, the force domain is also necessary to analyse the meaning of prepositions, meaning that their semantics cannot be handled solely as spatial relations. To what extent one can find neuro-scientific support for the representation of force remains to be seen.

5 Modelling Actions

One idea for a model of actions comes from Marr and Vaina (1982) and is explored further in Vaina (1983). Marr and Vaina extend Marr and Nishihara's (1978) cylinder models of objects to an analysis of actions. In Marr and Vaina's model, an action – say, a person walking – is described via differential equations for the movements of the implicated body parts.

It is clear that these equations can be derived, by application of Newtonian mechanics, from the forces that are applied to the legs, arms, and other moving body parts. Our cognitive apparatus is not precisely built for thinking in terms of Newtonian mechanics, but I hypothesize that, nevertheless, our brains successfully extract the forces that lie behind different kinds of movement-involving action. I will present some support for this hypothesis below. More precisely, I submit, building on Gärdenfors (2007), that the fundamental cognitive representation of any action consists of the *pattern of forces* that generates it. I speak of patterns of forces, since, for bodily motions, several body parts are involved; and thus, several force vectors are interacting (in analogy with Marr and Vaina's differential equations). It should be emphasized, however, that the "forces" represented by the brain are psychological and not the scientific construct introduced by Newton.

The patterns of forces can be represented in principally the same way as the patterns of shapes discussed earlier. For example, the pattern of force of a person running is different from the pattern of a person walking; and the pattern for saluting is different from that of throwing (Vaina and Bennour 1985).

The best source of empirical support for my hypothesis comes from psychophysics. During the 1950's, Gunnar Johansson developed his patch-light technique for analyzing biological motion without any direct information about shape. (For a survey, see Johansson 1973.) He attached light bulbs to the joints of actors who were dressed in black and moved in a black room. The actors were filmed performing various actions, such as walking, running, and dancing. Subjects watching the films, in which only the dots of light could be seen, recognized the actions within tenths of a second. Furthermore, the movements of the dots were immediately interpreted as coming from the actions of a human being. Further experiments by Runesson and Frykholm (1981, 1983) showed that subjects were able to extract subtle details about the actions, such as the gender of walkers or the weight of lifted objects (where the objects, like the actors themselves, were not visible).

One lesson to be learned from the experiments by Johansson and his followers is that the kinematics of a movement contains sufficient information for identifying the underlying dynamic patterns of force. Runesson (1994, pp. 386-387) claims we can directly perceive the forces that control various kinds of motion. He calls this principle the *kinematic specification of dynamics*, according to which the kinematics of a movement contains sufficient information to identify the underlying dynamic patterns of force. It is obvious that his principle accords well with the representation of actions that is proposed here. (Note however that Runesson takes a Gibsonian perspective on the perceptual information available, which means he would find it methodologically unnecessary to consider such mental constructions as conceptual spaces.)

Even though the empirical evidence is incomplete, my proposal is that, by adding force dimensions to a conceptual space, one obtains the basic tools for analyzing the dynamic properties of actions. The forces involved need not only be physical forces, but also *emotional* or *social* forces.

To identify the structure of the action space, one should investigate similarities between actions. This can be done with basically the same methods as for investigating similarities between objects: e.g., "walking" is more similar to "running" than to "throwing". Little is known about the geometrical structure of the action space. I make the weak assumption that the notion of betweenness remains meaningful. This allows me to formulate the following criterion, in analogy with Criterion *C*:

Criterion A: An *action category* is represented as a convex region in action space.

Of course, the more forces are involved, the more multi-dimensional the action space will be and the more complicated it will be to identify the relevant convex regions. One way to support the connection between Criterion *C* and Criterion *A* is to establish that action categories share a similar structure with object categories, as Hemeren (2008, p. 25) has suggested. In a series of experiments (1996, 1997, 2008), he showed that action categories have a similar hierarchical structure and show similar typicality effects to object concepts. Overall, there are strong reasons to believe that actions exhibit many of the *prototype effects* that Rosch (1975) described for object categories.

6 A Two-Vector Model of Events

In keeping with Gärdenfors and Warglien (submitted), I want briefly to present a model of events that is likewise based on conceptual spaces. Events are treated as complex structures that build on conceptual spaces, in particular the action space. The starting point is that all events involve an *agent* and a *patient*.

Agents and patients are modelled as (material or non-material) objects, and can therefore be represented as points in conceptual spaces. The domains of the spaces determine the relevant properties of the agent and the patient. An agent is an animate or inanimate object. Even though I am not providing any analysis of causation here, the common understanding is that the agent is the one causing something to happen. (of course, one should allow that the action can be null, in the case of an event that is a state). An event is individuated by the further understanding that the agent causes the event to happen independently of other events.

An agent is described by an *agent space* that at minimum contains a force domain in which the action performed by the agent can be represented (this is the assumption of *agency*). Following the analysis from the previous section, I will model an action as a force vector (or, more particularly, as a pattern of forces). The agent space may also contain a physical space domain that assigns the agent a location. In particular, in the special case when patient = agent -- i.e., the agent is doing something to itself -- the properties of the agent must be modelled.

A *patient* is again an animate or inanimate object. The patient is described by a *patient space* that contains the domains needed to account for those properties of the patient relevant to the event that is modelled. The properties often include the location of the patient and sometimes its emotional state. A force vector is associated with the patient and represents the (counter-)force exerted by the patient in relation to the agent's action. This can be an intentionless physical force, as when a door does not open when pushed; or it can be an intentionally generated force, as when a person pushes back upon being pushed. For many events, the representation of the patient's force vector is unknown and can be left unspecified; or else it can be taken as prototypical, entailing that the consequences of the agent's action are left open to various degrees.

The force exerted by the agent's action will change one or more properties of the patient. The elementary operations possible on vectors provide a reasonable account for how changes can result from compositions of forces from the agent and the patient. The resultant force vector is the $r = f + c$, where f is the force generated by the agent's action and c is the counter-force of the patient. We then define an event as a mapping between an action in an agent space and a resulting change in a patient space that is the result of applying r. Central to the event are the changes to properties in other domains of the patient space. For example, the location of the patient may change; or its colour may change, if the event involves the action of painting.

This way of representing things makes an explicit difference between an action that is mapped into the patient space and the force exerted by such action. Two different actions (e.g. kicking and punching) might produce the same force vector r in the patient space. It might not be sufficient for characterizing an event to represent the force composition of f and c; the initiating action from the agent must also be represented. This will become even more relevant as event categories are introduced below.

As a simple example, consider the event of *pushing a table*. In such an event, the agent (a person) applies a physical force to a patient (a table). The result is a change in the location of the patient and thus a change in its properties (unless there are balancing counter-forces present, such that the resulting change vector is zero). The change vector depends on the properties of the patient and other aspects of the surrounding world (for example, friction). Another example is an event of walking. In this case, the agent and the patient are identical, so the agent applies a force to itself.

Some actions are ongoing: the agent exerts the force for an unbounded period of time, for example by walking or pushing an object, with the consequence that there may be no definite end point to the changes in the patient space. This is a special case of a more general type of event: *processes*. In bounded events, the agent's force vector is applied for a limited time period, and the change vector in patient space has a clear end point. In many languages, the difference between processes and bounded events is reflected in the syntax: for example, by various forms of aspect. The focus here is on bounded events, but most of the elements of the event representations will apply to corresponding representations of (unbounded) processes.

In general, events should be represented not only as single spatiotemporally located instances, but also as event categories, like "climbing a mountain". I will next provide a formal framework for analysing event categories, of which single events can be considered as instances.

The earlier description of the change vector can be generalized to that of a *change vector field.* The change vector field associates to each point in the patient space that vector change induced by a particular action, taking into account, if necessary, the (counter-)force exerted by the patient. An event category then represents how the agent space potentially affects the patient vector field. For example, the event category of pushing a table should represent the effect of different, albeit similar, patterns of force on the different points in the table patient space.

Event categories can be represented at different levels: there are subcategories of events just as there are of objects. For example, "pushing a door open" is a subcategory of "pushing a door" where the agent force exceeds the counter-force of the patient. "Pushing but failing to open a door" is another subcategory, one where the counter-force cancels out the agent force.

For many kinds of events where the focus is on the changes in the patient, the identity of the agent can be and often is ignored. For example, in the event of somebody falling ill, the cause of illness is often not considered. Similarly, if the force vector is null (i.e., the event is a state), the identity of the agent is irrelevant.

This model of events and event categories is presented in greater detail in Gärdenfors and Warglien (submitted). What is new here, apart from using conceptual spaces as the general supporting framework, is the introduction of the two vectors as forming the core of an event.

7 The Role of Events in the Semantics of Verbs

The fundamental connection between the semantics of natural language and events is that a simple sentence *typically expresses an event*. For this reason, events are central units in any theory of semantics. For this reason, a typical sentence contains the basic building blocks of subject, object and verb, corresponding to the agent, patient and vectors of my model. A single verb can never completely describe an event, but only bring out one aspect of it. I propose that a verb represents *one of the vectors in the model of an event.* In linguistics, a distinction is often made between *manner* and *result* verbs. I suggest that if the verb focuses on the force vector of the agent, as for example in “push” or “hit”, then it is a *manner* verb; while if it focuses on the change vector of the patient, as for example in “move” or “stretch”, it is a *result* verb.

In the cognitive semantics tradition of Lakoff (1987) and Langacker (1987), the focus has been on the spatial structure of the image schemas only (the very name suggests this), with no attempt to represent the forces involved in the event. This is an essential departure from the way I have proposed modelling actions and events.

Talmy (1988) presents an alternative model of action and interaction. Talmy emphasizes the role of forces and dynamic patterns in image schemas through what he calls *force dynamics*. He develops a schematic-based formalism that allows him to represent the difference of force patterns in expressions like “the ball kept rolling because of the wind blowing on it” and “the ball kept rolling despite the stiff grass”. Interactions between agent (what he calls the *agonist*) and Patient (the *antagonist*) are central to his framework as they are to the one presented here. However, some important differences should be highlighted. First is the role that spaces and mappings between spaces play. While Talmy’s force dynamics are situated in generic spaces, I am grounding the semantics of events in a theory of conceptual spaces and of the mappings between them. This creates a more flexible and comprehensive framework, one that can take into account the qualitative dimensions of the agent’s actions and the changes in the patient. Second, my framework makes a natural distinction between single events and event categories, and is able to account for those events in which the reaction of the patient is not specified.

Finally, an important advantage of the spatial representation of the event structure I have presented is that it allows one to map from one event type to another, comparing their structure and making it possible to address, for example, the metaphorical use of an action or event. In Warglien and Gärdenfors (submitted), we explore the use of topological tools to model such metaphorical mappings.

8 Conclusion

This paper has been of a programmatic nature, advocating an approach to the semantics of natural languages based on conceptual spaces. I have outlined how properties, concepts, actions, and events can all be modelled and a cognitive semantics for different word classes generated.

The model of events combines the analysis of nouns, in terms of concepts for representing agents and patients, with the analysis of actions. Such an analysis shows the intimate connection between how we cognitively represent actions and objects, One that is reflected in the linguistic tools we use for expressing actions and events. I propose that the force vector produced by the agent and the result vector induced in the patient reflect the distinction between manner and result verbs. More abstractly, events represent what philosophers call propositions: that is, the semantic contents of basic sentences.

I have tried to show how conceptual spaces in general, and their application to the force domain in particular, can be useful tools for sharpening cognitive semantics. With the aid of the topological and geometric structure of the various domains, a better foundation for the concept of image schemas is obtained. This applies in particular to dynamic schemas.

Acknowledgements

I gratefully acknowledge support from the VAAG project under the ESF framework LogiCCC and the Swedish Research Council for support under a Senior Individual Grant and the Linnaeus environment *Thinking in Time: Cognition, Communication and Learning*.

References

1. Berlin, B., Kay, P.: Basic Color Terms: Their Universality and Evolution. University of California Press, Berkeley (1969)
2. Edelman, S.: Representation and Recognition in Vision. MIT Press, Cambridge (1999)
3. Gärdenfors, P.: Induction, conceptual spaces and AI. Philosophy of Science 57, 78–95 (1990)
4. Gärdenfors, P.: Conceptual spaces as a basis for cognitive semantics. In: Clark, A., et al. (eds.) Philosophy and Cognitive Science, pp. 159–180. Kluwer, Dordrecht (1996)
5. Gärdenfors, P.: Does semantics need reality?. In: Does Representation Need Reality? Proceedings of the International Conference New Trends in Cognitive Science, Austrian Society of Cognitive Science Technical Report 97-01, Vienna, pp. 113–120 (1997)
6. Gärdenfors, P.: Conceptual Spaces: The Geometry of Thought. MIT Press, Cambridge (2000)
7. Gärdenfors, P.: Representing actions and functional properties in conceptual spaces. In: Ziemke, T., Zlatev, J., Frank, R.M. (eds.) Body, Language and Mind. Embodiment, vol. 1, pp. 167–195. Mouton de Gruyter, Berlin (2007)
8. Gärdenfors, P., Warglien, M.: Using conceptual spaces to model actions and events. Manuscript (submitted)
9. Garner, W.R.: The Processing of Information and Structure. Erlbaum, Potomac (1974)
10. Harnad, S.: The symbol grounding problem. Physica D 42, 335–346 (1990)
11. Hemeren, P.E.: Frequency, ordinal position and semantic distance as measures of cross-cultural stability and hierarchies for action verbs. Acta Psychologica 91, 39–66 (1996)
12. Hemeren, P.E.: Typicality and context effects in action categories. In: Proceedings of the 19th Annual Conference of the Cognitive Science Society, p. 949. Lawrence Erlbaum Associates, Stanford (1997)
13. Hemeren, P.E.: Mind in Action. Lund University Cognitive Studies 140, Lund (2008)
14. Herskovits, A.: Language and Spatial Cognition: An Interdisciplinary Study of the Prepositions in English. Cambridge University Press, Cambridge (1996)
15. Holmqvist, K.: Implementing Cognitive Semantics. Lund University Cognitive Studies 17, Lund (1993)
16. Jackendoff, R.: Semantics and Cognition. MIT Press, Cambridge (1983)
17. Jäger, G.: Natural colour categories are convex sets. In: Aloni, M., Bastiaanse, H., de Jager, T., Schulz, K. (eds.) Logic, Language and Meaning. LNCS (LNAI), vol. 6042, pp. 11–20. Springer, Heidelberg (2010)

18. Johansson, G.: Visual perception of biological motion and a model for its analysis. Perception and Psychophysics 14, 201–211 (1973)
19. Lakoff, G.: Women, Fire, and Dangerous Things. The University of Chicago Press, Chicago (1987)
20. Landau, B., Jackendoff, R.: “What” and “where” in spatial language and spatial cognition. Behavioral and Brain Sciences 16, 217–238 (1993)
21. Langacker, R.W.: Foundations of Cognitive Grammar. Theoretical Prerequisites, vol. I. Stanford University Press, Stanford (1987)
22. Maddox, W.T.: Perceptual and decisional separability. In: Ashby, G.F. (ed.) Multidimensional Models of Perception and Cognition, pp. 147–180. Lawrence Erlbaum, Hillsdale (1992)
23. Marr, D., Nishihara, H.K.: Representation and recognition of the spatial organization of three-dimensional shapes. Proceedings of the Royal Society in London B 200, 269–294 (1978)
24. Marr, D., Vaina, L.: Representation and recognition of the movements of shapes. Proceedings of the Royal Society in London B 214, 501–524 (1982)
25. Melara, R.D.: The concept of perceptual similarity: from psychophysics to cognitive psychology. In: Algom, D. (ed.) Psychophysical Approaches to Cognition, pp. 303–388. Elsevier, Amsterdam (1992)
26. Noy, N.F., McGuinness, D.L.: Ontology development 101: A guide to creating your first ontology, Stanford Knowledge Systems Laboratory Technical Report, Stanford, CA (2001)
27. Rosch, E.: Cognitive representations of semantic categories. Journal of Experimental Psychology: General 104, 192–233 (1975)
28. Runesson, S.: Perception of biological motion: The KSD-principle and the implications of a distal versus proximal approach. In: Jansson, G., Bergström, S.-S., Epstein, W. (eds.) Perceiving Evens and Objects, pp. 383–405. Lawrence Erlbaum Associates, Hillsdale (1994)
29. Runesson, S., Frykholm, G.: Visual perception of lifted weights. Journal of Experimental Psychology: Human Perception and Performance 7, 733–740 (1981)
30. Runesson, S., Frykholm, G.: Kinematic specification of dynamics as an informational basis for person and action perception. Expectation, gender recognition, and deceptive intention. Journal of Experimental Psychology: General 112, 585–615 (1983)
31. Sivik, L., Taft, C.: Color naming: a mapping in the NCS of common color terms. Scandinavian Journal of Psychology 35, 144–164 (1994)
32. Smith, L.: From global similarities to kinds of similarities – the construction of dimensions in development. In: Vosniadou, S., Ortony, A. (eds.) Similarity and Analogical Reasoning, pp. 146–178. Cambridge University Press, Cambridge (1989)
33. Smith, E.E., Osherson, D.N., Rips, L.J., Keane, M.: Combining prototypes: a selective modification model. Cognitive Science 12, 485–527 (1988)
34. Talmy, L.: Force dynamics in language and cognition. Cognitive Science 12, 49–100 (1988)
35. Vaina, L.: From shapes and movements to objects and actions. Synthese 54, 3–36 (1983)
36. Vaina, L., Bennour, Y.: A computational approach to visual recognition of arm movement. Perceptual and Motor Skills 60, 203–228 (1985)
37. Warglien, M., Gärdenfors, P.: Semantics, conceptual spaces and the meeting of minds. Manuscript (submitted)
38. Zwarts, J.: The semantics of relative position. In: Proceedings of Semantics and Linguistic Theory, vol. 5, pp. 405–422 (1995)
39. Zwarts, J.: Forceful prepositions. In: Evans, V., Chilton, P. (eds.) Language, Cognition and Space: The State of the Art and New Directions. Equinox Publishing (to appear)
40. Zwarts, J., Winter, Y.: Vector Space Semantics: A model-theoretic analysis of locative prepositions. Journal of Logic, Language and Information 9, 169–211 (2000)

Four Corners—East and West

Graham Priest

Departments of Philosophy,
Universities of Melbourne and St Andrews,
and the Graduate Center, City University of New York

Abstract. In early Buddhist logic, it was standard to assume that for any state of affairs there were four possibilities: that it held, that it did not, both, or neither. This is the *catuskoti*. Classical logicians have had a hard time making sense of this, but it makes perfectly good sense in the semantics of various paraconsistent logics, such as First Degree Entailment. Matters are more complicated for later Buddhist thinkers, such as Nagarjuna, who appear to suggest that none or these options, or more than one, may hold. These possibilities may also be accommodated with contemporary logical techniques. The paper explains how.

Keywords: *catuskoti*, Buddhist logic, Nagarjuna, First Degree Entailment, many-valued logic, relational semantics.

1 Introduction

Western Logic has been dominated by the Principles of Excluded Middle and Non-Contradiction. Given any claim, there are two possibilities, *true* and *false*. These are exhaustive and exclusive. Contemporary Western logic has come to realise that this may be far too narrow-minded. There may well be situations where we need to countenance things that are neither true nor false, or both true and false.[1] Indeed, these possibilities are built into the semantics of various logics (many-valued, relevant, paraconsistent). The technology of deploying such techniques is now relatively well understood.

Western logic might well have learned its lesson from India. Though the traditional schools of Indian logic never had the mathematical tools to articulate their positions into anything like modern Western formal logics, a much more open-minded attitude was present from the earliest years. According to a principle of Buddhist logic clearly pre-dating the Buddha, given any claim, there are four possibilities, *true* (only), *false* (only), *both* or *neither*. This was called the *catuskoti*[2] (literally: 'four corners'). Western philosophers and logicians, armed only with their knowledge of bivalent Western logic, have had a hard time of making sense of the *catuskoti*, but by deploying the techniques of modern many-valued logic, this is simple, as we will see.

[1] See, e.g., Priest (2008), ch. 7.

[2] Actually, *catuṣkoṭi*, but I ignore the diacriticals in writing Sanskrit words, except in the bibliography.

M. Banerjee and A. Seth (Eds.): ICLA 2011, LNAI 6521, pp. 12–18, 2011.

Matters became more complex as Buddhist thought developed in the early centuries of the Common Era. Here we find the great philosopher Nagarjuna, and those who followed him in the Madhyamaka school, appearing to say that none of the four *kotis* (corners) may hold, or sometimes that more than one—even all—of them may hold. How to accommodate this possibility with the techniques of modern logic is less obvious. However, it also can be done, and we will see this too.[3]

The following paper is therefore another illustration of the possibility of the history of logic and contemporary logic informing each other, to their mutual benefit—and one, moreover, that illustrates the fruitful interplay between Eastern and Western thought.

2 A Little History

The *catuskoti* is illustrated at the very beginning of Buddhist thought, when some of the Buddha's followers asked him to answer various difficult metaphysical questions, such as what happens to an enlightened person after death. The Buddha is explicitly presented with four possibilities, that the enlightened person exists, that they do not exist, that they both exist and do not exist, that they neither exist nor do not exist—the four corners of the *catuskoti*. The Buddha does not balk at the way things are presented. True, he refuses to answer the question, but the normal reason given is that thinking about such things is a waste of time, time better spent on matters more conducive to awakening. Just occasionally, there is a hint that there is something else going on, possibly a false presupposition to all four possibilities. This thought was perhaps to be taken up later, but nothing further is made of the matter at this point in Buddhist thought. At this stage, then, the *catuskoti* functions something like a *principle of excluded fifth*: there are exactly four exclusive possibilities, *quintum non datur*.[4]

3 Making Sense of the *Catuskoti*

Philosophers who know only classical or traditional logic have a hard time making sense of the *catuskoti*. The natural way for them to formulate the four possibilities concerning some claim, A, are:

(a) A
(b) $\neg A$
(c) $A \wedge \neg A$
(d) $\neg(A \vee \neg A)$

[3] It should be pointed out that not all Buddhists subscribed to the *catuskoti*. It was not endorsed by the Dignaga-Dharmakirti school of Buddhist logic. Like the Nyaya, this school of logic endorsed both the Principles of Non-Contradiction and Excluded Middle. See Scherbatsky (1993), pt. 4, ch. 2.

[4] For a more extended discussion of the history, including textual sources and quotations, see Priest (2010). See also Ruegg (1977) and Tillemans (1999).

(c) will wave red flags to anyone wedded to the Principle of Non-Contradiction—but the texts seem pretty explicit that you might have to give this away. There are worse problems. Notably, assuming De Morgan's laws, (d) is equivalent to (c), and so the two *kotis* collapse. Possibly, one might reject the Principle of Double Negation, so that (d) would give us only $\neg A \wedge \neg\neg A$. But there are worse problems. The four cases are supposed to be exclusive; yet case (c) entails both cases (a) and (b). So the corners again collapse.

The obvious thought here is that we must understand (a) as saying that A is true and not false. Similarly, one must understand (b) as saying that A is false and not true. Corners (a) and (b) then become: $A \wedge \neg\neg A$ and $\neg A \wedge \neg A$ (i.e., $\neg A$). Even leaving aside problems about double negation, case (c) still entails case (b). We are no better off.[5]

There is, however, a way of understanding the *catuskoti* that will jump out at anyone with a passing acquaintance with the foundations of relevant logic. First Degree Entailment (FDE) is a system of logic that can be set up in many ways, but one of these is as a four-valued logic whose values are t (true only), f (false only), b (both), and n (neither). The values are standardly depicted by the following Hasse diagram:

$$\begin{tikzcd}
 & t & \\
b \arrow[ur] & & n \arrow[ul] \\
 & f \arrow[ul] \arrow[ur] &
\end{tikzcd}$$

Negation maps t to f, vice versa, n to itself, and b to itself. Conjunction is greatest lower bound, and disjunction is least upper bound. The set of designated values, D, is $\{t, b\}$. Validity is defined in terms of the preservation of designated values in all interpretations.[6] The four corners of the *catuskoti* and the Hasse diagram seem like a marriage made for each other in a Buddhist heaven.[7]

Proof theoretically, FDE can be characterised by the following rule system. (A double line indicates a two-way rule, and overlining indicates discharging an assumption.)[8]

$$\frac{A, B}{A \wedge B} \qquad \frac{A \wedge B}{A\ (B)}$$

$$\frac{A\ (B)}{A \vee B} \qquad \frac{A \vee B \quad \begin{matrix}\overline{A}\\ \vdots \\ C\end{matrix} \quad \begin{matrix}\overline{B}\\ \vdots \\ C\end{matrix}}{C}$$

[5] A full discussion of the unsuccessful ways that people have tried to get around these problems within the confines of classical—or at least intuitionist—logic, can be found in Priest (2010). See also Westerhoff (2009), ch. 4.

[6] See Priest (2008), ch. 8.

[7] As observed in Garfield and Priest (2009).

[8] See Priest (2002), 4.6.

$$\frac{\neg(A \wedge B)}{\neg A \vee \neg B} \quad \frac{\neg(A \vee B)}{\neg A \vee \neg B} \quad \frac{\neg\neg A}{A}$$

We see, then, how the four corners of the *catuskoti* can be accommodated in ways very standard in contemporary non-classical logic.

4 Rejecting All the *Kotis*

So far so good. Things get more complicated when we look at the way that the *catuskoti* is deployed in later developments in Buddhist philosophy—especially in the way it appears to be deployed in the writings of Nagarjuna and his Madhyamaka successors. We have taken the four corners of truth to be exhaustive and mutually exclusive. A trouble is that we find Nagarjuna appearing to say that sometimes none of the four corners may hold.[9] Why he says this, and what he means by it, are topics not appropriate for this occasion.[10] The question here is simply how to accommodate the possibility using the techniques of contemporary (non-classical) logic.

The easiest way of doing so is by taking there to be a fifth possibility:

(e) none of the above.

The most obvious way to proceed is now to take this possibility as a fifth semantic value, and construct a five-valued logic. Thus, we add a new value, e, to our existing four (t, f, b, and n).[11] Since e is the value of things that are neither true nor false (and so not true), it should obviously not be designated. Thus, we still have that $D = \{t, b\}$. How are the connectives to behave with respect to e? Both e and n are the values of things that are neither true nor false, but they had better behave differently if the two are to represent distinct alternatives. The simplest suggestion is to take e to be such that whenever any input has the value e, so does the output: e-in/e-out.[12]

The logic that results by modifying FDE in this way is obviously a sub-logic of it. It is a proper sub-logic. It is not difficult to check that all the rules of FDE are designation-preserving except the rule for disjunction-introduction, which is not, as an obvious counter-model shows. However, replace this with the rules:

$$\frac{\varphi(A) \quad C}{A \vee C} \qquad \frac{\varphi(A) \quad C}{\neg A \vee C} \qquad \frac{\varphi(A) \quad \psi(B) \quad C}{(A \wedge B) \vee C}$$

[9] Just to make matters confusing, some people refer to this denial (the 'four-cornered negation') itself as the *catuskoti*. The Buddhist tradition is, in fact, not alone in sometimes denying the four *kotis*. See Raju (1953).

[10] Again, for a fuller discussion of the matter, together with textual sources and quotations, see Priest (2010). See also the pages indexed under 'Tetralemma' in Garfield (1995), and Garfield and Priest (2003).

[11] As in Garfield and Priest (2009). Happily, e, there, gets interpreted as emptiness.

[12] We will see that this behaviour of e falls out of a different semantics for the language in section 6.

where $\varphi(A)$ and $\psi(B)$ are any sentences containing A and B.[13] Call these the φ *Rules*, and call this system FDE_φ. FDE_φ is sound and complete with respect to the semantics.[14]

5 Accepting More than One *Koti*

Again, so far so good. There is a harder challenge to be faced, though. Forget the fifth possibility for the moment; we will return to it again later. The problem is that Nagarjuna sometimes seems to say that more than one of the *kotis* may hold—even all of them. Again, this is not the place to discuss what is going on here philosophically.[15] The question is how to accommodate the view in terms of modern logic.

In classical logic, evaluations of formulas are functions which map sentences to one of the values 1 and 0. In one semantics for FDE, evaluations are thought of, not as functions, but as relations, which relate sentences to some number of these values. This gives the four possibilities represented by the four values of our many-valued logic.[16]

We may do exactly the same with the values t, b, n, and f themselves. So if P is the set of propositional parameters, and $V = \{t, b, n, f\}$, an evaluation is a relation, ρ, between P and V. In the case at hand, we want to insist that every formula has at least one of these values, that is, the values are exhaustive:

Exh: for all $p \in P$, there is some $v \in V$, such that $p\rho v$.

If we denote the many-valued truth functions corresponding to the connectives $\neg$, $\vee$, and $\wedge$ in FDE, by $f_\neg$, $f_\vee$, and $f_\wedge$, then the most obvious extension of ρ to all formulas is given by the clauses:

- $\neg A\rho v$ iff for some x such that $A\rho x$, $v = f_\neg(x)$
- $A \vee B\rho v$ iff for some x, y, such that $A\rho x$ and $B\rho y$, $v = f_\vee(x, y)$
- $A \wedge B\rho v$ iff for some x, y, such that $A\rho x$ and $B\rho y$, $v = f_\wedge(x, y)$

One can show, by a simple induction, that for every A there is some $v \in V$ such that $A\rho v$. I leave the details as an exercise.

Where, as before, $D = \{t, b\}$, we may simply define validity as follows: $\Sigma \vDash A$ iff for all ρ:

- if for every $B \in \Sigma$, there is a $v \in D$ such that $B\rho v$, then there is a $v \in D$ such that $A\rho v$

That is, an inference is valid if it preserves the property of relating to *some* designated value.

[13] Instead of $\varphi(A)$ (etc.), one could have any sentence that contained all the propositional parameters in A.
[14] Details of the proof may be found in Priest (2010).
[15] A full discussion can be found in Priest (2010).
[16] See Priest (2008), 8.2.

Perhaps surprisingly, validity on this definition coincides with validity in FDE.[17] This is proved by showing that the rules of FDE are sound and complete with respect to the semantics.[18]

6 None of the *Kotis*, Again

Let us, finally, return to the possibility that none of the *kotis* may hold. In Section 4, we handled this possibility by adding a fifth value, e. The relational semantics provides a different way of proceeding. We simply drop the exhaustivity condition, **Exh**, so allowing the possibility that an evaluation may relate a parameter (and so an arbitrary formula) to none of the four values. The logic this gives is exactly FDE_{φ}.[19]

In fact, if we require that every formula relates to *at most* one value, then it is easy to check that we simply have a reformulation of the 5-valued semantics, since taking the value e in the many-valued semantics behaves in exactly the same way as not relating to any value does in the relational semantics.

7 Conclusion

We have now seen how the ideas of the *catuskoti* and its developments can be made sense of using the techniques of many-valued and relational semantics. FDE does justice to the four possibilities. This has, as we have noted, a many-valued and a relational semantics. If none of the four *kotis* may obtain, we have FDE_{φ}. Again, this has a many-valued and a relational semantics, the second of which allows for more that one of the *kotis* obtaining, as well as none.

Of course, there are important questions about what all this *means*. Some of these questions are familiar from the contemporary philosophy of logic, such as ones concerning the possibility of truth value gaps and gluts. Some of them concern Buddhist philosophy, and specifically the metaphysical picture which informs (and may be informed by) the technical machinery. This is obviously not the place to discuss such matters. Suffice it for the present to have shown some interesting connections between Buddhist thought and the techniques of contemporary non-classical logic.

References

[1] Garfield, J.: The Fundamental Principles of the Middle Way: Nāgārjuna's Mūlamadhyamikakārikā. Oxford University Press, New York (1995)
[2] Garfield, J., Priest, G.: Nāgārjuna and the Limits of Thought. Philosophy East and West 53, 1–21 (2003); reprinted as Priest, G.: Beyond the Limits of Thought, 2nd (extended) edn., ch.16. Oxford University Press, Oxford (2002)

[17] Perhaps not. See Priest (1984).
[18] A proof of this fact can be found in Priest (2010).
[19] Again, a proof of this fact can be found in Priest (2010).

[3] Garfield, J., Priest, G.: Mountains are Just Mountains. In: D'Amato, M., Garfield, J., Tillemans, T. (eds.) Pointing at the Moon: Buddhism, Logic, Analytic Philosophy, ch. 7. Oxford University Press, New York (2009)
[4] Priest, G.: Hypercontradictions. Logique et Analyse 107, 237–243 (1984)
[5] Priest, G.: Paraconsistent Logic. In: Gabbay, D., Guenther, D. (eds.) Handbook of Philosophical Logic, 2nd edn., vol. 6, pp. 287–393. Kluwer Academic Publishers, Dordrecht (2002)
[6] Priest, G.: Introduction to Non-Classical Logic: from If to Is. Cambridge University Press, Cambridge (2008)
[7] Priest, G.: The Logic of the Catuṣkoṭi. Comparative Philosophy 1, 32–54 (2010)
[8] Raju, P.: The Principle of Four-Cornered Negation in Indian Philosophy. Review of Metaphysics 7, 694–713 (1953)
[9] Ruegg, D.: The Uses of the Four Positions of the Catuṣkoṭi and the Problem of the Description of Reality in Mahāyāna Buddhism. Journal of Indian Philosophy 5, 1–71 (1977)
[10] Scherbatsky, T.: Buddhist Logic, vol. 1. Matilal Banarsidass, Delhi (1993)
[11] Tillemans, T.: Is Buddhist Logic Non-classical or Deviant? In: Scripture, Logic, Language: Essays on Dharmakīrti and his Tibetan Successors, ch. 9. Wisdom Publications, Boston (1999)
[12] Westerhoff, J.: Nāgārjuna's Metaphysics: a Philosophical Introduction. Oxford University Press, Oxford (2010)

Infinite Games and Uniformization

Wolfgang Thomas

RWTH Aachen University, Lehrstuhl Informatik 7, 52056 Aachen, Germany
thomas@informatik.rwth-aachen.de

Abstract. The problem of solvability of infinite games is closely connected with the classical question of uniformization of relations by functions of a given class. We work out this connection and discuss recent results on infinite games that are motivated by the uniformization problem.

The fundamental problem in the effective theory of infinite games was posed by Church in 1957 ("Church's Problem"; see [2,3]). It refers to two-player games in the sense of Gale and Stewart [6] in which the two players 1 and 2 build up two sequences $\alpha = a_0 a_1 \ldots$, respectively $\beta = b_0 b_1 \ldots$, where a_i, b_i belong to a finite alphabet Σ. Player 1 picks a_0, then Player 2 picks b_0, then Player 1 picks a_1, and so forth in alternation. A play is an ω-word over $\Sigma \times \Sigma$ of the form $\binom{a_0}{b_0}\binom{a_1}{b_1}\binom{a_2}{b_2}\ldots$; we also write $\alpha^\frown\beta$. A game is specified by a relation $R \subseteq \Sigma^\omega \times \Sigma^\omega$, or equivalently by the ω-language $L_R = \{\alpha^\frown\beta \mid (\alpha, \beta) \in R\}$. Player 2 wins the play $\alpha^\frown\beta$ if $(\alpha, \beta) \in R$.

Church's Problem asks for a given relation R ("specified in some suitable logistic system" [2]) whether Player 2 has a winning strategy in the game defined by R, i.e., whether there is a corresponding function f mapping finite play prefixes $\binom{a_0}{b_0}\binom{a_1}{b_1}\ldots\binom{a_n}{*}$ to the set Σ, providing the information which letter to pick next (and – if there is a winning strategy f – to provide a definition of f). A strategy f induces a function $f' : \Sigma^\omega \to \Sigma^\omega$; for a winning strategy f we then have $(\alpha, f'(\alpha)) \in R$ for all $\alpha \in \Sigma^\omega$.

A prominent class of games is given by the regular (or equivalently: the MSO-definable) relations R, which we identify here with the associated regular ω-languages L_R. In this case a complete solution is known by the "Büchi-Landweber Theorem" ([1]; see e.g. [7]): Given a regular game (specified, e.g., by a Büchi automaton or an MSO-formula),

– either of the players has a winning strategy (the game is "determined"),
– one can decide who is the winner,
– and one can construct a finite-state strategy for the winner (i.e., a strategy realizable by a Mealy automaton).

The extraction of a function f from a relation R such that for each argument x we have $(x, f(x)) \in R$ is called "uniformization". More precisely, a class $\mathcal{R}$ of binary relations $R \subseteq D \times D'$ is *uniformizable* by functions in a class $\mathcal{F}$ if for each $R \in \mathcal{R}$ there is a function $f \in \mathcal{F}$ such that the graph of f is contained in R, and the domains of R and f coincide (see Fig. 1 for an illustration).

M. Banerjee and A. Seth (Eds.): ICLA 2011, LNAI 6521, pp. 19–21, 2011.

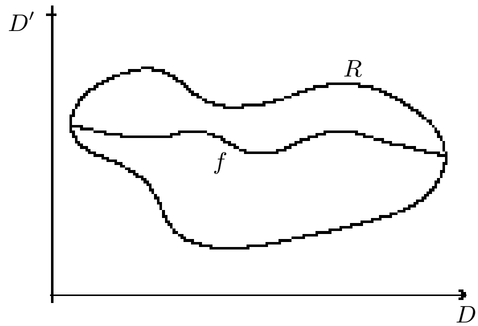


Fig. 1. Uniformization

Two well-known examples from recursion theory and from automata theory are concerned with the recursively enumerable, respectively the rational relations; here we have the "ideal" case that the graphs of the required functions are precisely of the type of the given relations.

For the first example, let us recall that a partial function from $\mathbb{N}$ to $\mathbb{N}$ is recursive iff its graph is recursively enumerable. The Uniformization Theorem of recursion theory says that a binary recursively enumerable relation R is uniformizable by a function whose graph is again recursively enumerable, i.e. by a (partial) recursive function f. (A computation of $f(x)$ works as follows: Enumerate R until a pair (y, z) is reached with $x = y$, and in this case produce z as output.)

For the second example, recall that a binary rational word relation is defined (for instance) in terms of a finite nondeterministic two-tape automaton that scans a given word pair (u, v) asynchronously, i.e. with two reading heads that move independently from left to right over u, respectively v. Rational relations are uniformizable by rational functions, defined as the functions whose graph is a rational relation (see e.g. [4,10]).

In the task of uniformization as it appears in Church's Problem, there are two special features: First, one looks for functions that are computed "online" (step by step in terms of the argument). Secondly, determinacy results constitute a very special situation where non-existence of a function f with $(x, f(x)) \in R$ for all x implies the existence of a function g such that $(g(y), y) \notin R$ for all y.

The first requirement can be weakened in the sense that the functions f used for uniformization could use more information than just $a_0 \ldots a_i$ for producing the letter b_i. Different types of "look-ahead" can be studied, the extreme case being that the whole sequence $\alpha = a_0a_1 \ldots$ is given when $\beta = b_0b_1 \ldots$ is to be formed. Finite look-ahead corresponds to the condition that the i-th letter b_i of β depends only on a finite prefix of $\alpha = a_0a_1 \ldots$; one calls such functions "continuous" (in the Cantor topology over Σ^ω). We discuss recent results of [8] for regular games: If the uniformization of a regular relation R is possible with a continuous function, then a function of bounded look-ahead k suffices, where, for all i, the letter b_i depends only on the prefix $a_0 \ldots a_{i+k}$.

The Büchi-Landweber Theorem can be phrased as saying that MSO-definable strategies suffice for solving MSO-definable games, or – in other words – that MSO-definable relations can be uniformized by MSO-definable functions. (Here we use a notion of definability of functions in terms of arguments given as finite words.)

This motivates a study of Church's Problem for other classes of relations. We start with relations that are first-order definable rather than MSO-definable. There are two natural versions of first-order logic, denoted FO(+1) and FO(<), where the items in brackets indicate the available arithmetical signature. We recall results of [9] where it is shown that a determinacy theorem holds for the FO(+1)-, respectively the FO(<)-definable relations, and that appropriate winning strategies exist which are again FO(+1)-, respectively FO(<)-definable. Continuing this track, we exhibit cases where this transfer fails (Presburger arithmetic is an example), and then address corresponding results of Fridman [5] for non-regular games that are defined by various types of ω-pushdown automata.

These investigations are small steps towards a more comprehensive understanding of uniformization problems in the context of infinite games. So far, general conditions are missing that "explain" the known scattered results.

References

1. Büchi, J.R., Landweber, L.H.: Definability in the monadic second-order theory of successor. J. Symb. Logic 34, 166–170 (1969)
2. Church, A.: Applications of recursive arithmetic to the problem of circuit synthesis. In: Summaries of the Summer Institute of Symbolic Logic, vol. I, pp. 3–50. Cornell Univ., Ithaca (1957)
3. Church, A.: Logic, arithmetic, and automata. In: Proc. Int. Congr. Math. (1962); Inst. Mittag-Leffler, Djursholm, Sweden, pp. 23–35 (1963)
4. Eilenberg, S.: Automata, Languages, and Machines, vol. A. Academic Press, New York (1974)
5. Fridman, W.: Formats of winning strategies for six types of pushdown games. In: Montanari, A., et al. (eds.) Proc. 1st Symp. on Games, Automata, Logic, and Formal Verification, GandALF 2010. Electr. Proc. in Theor. Comput. Sci., vol. 25 (2010)
6. Gale, D., Stewart, F.M.: Infinite games with perfect information. Ann. Math. Studies 28, 245–266 (1953)
7. Grädel, E., Thomas, W., Wilke, T. (eds.): Automata, Logics, and Infinite Games. LNCS, vol. 2500. Springer, Heidelberg (2002)
8. Holtmann, M., Kaiser, L., Thomas, W.: Degrees of Lookahead in Regular Infinite Games. In: Ong, L. (ed.) FOSSACS 2010. LNCS, vol. 6014, pp. 252–266. Springer, Heidelberg (2010)
9. Rabinovich, A., Thomas, W.: Decidable theories of the ordering of natural numbers with unary predicates. In: Ésik, Z. (ed.) CSL 2006. LNCS, vol. 4207, pp. 562–574. Springer, Heidelberg (2006)
10. Sakarovitch, J.: Elements of Automata Theory. Cambridge Univ. Press, Cambridge (2009)

On Fuzzy Sets and Rough Sets from the Perspective of Indiscernibility

Mihir K. Chakraborty

UGC Visiting Professor,
Center for Cognitive Science,
Jadavpur University, Kolkata 700032, India
Honorary Visiting Professor,
Indian Statistical Institute, Kolkata
mihirc4@gmail.com

Abstract. The category theoretic approach of Obtułowicz to Pawlak's rough sets has been reintroduced in a somewhat modified form. A generalization is rendered to this approach that has been motivated by the notion of rough membership function. Thus, a link is established between rough sets and L-fuzzy sets for some special lattices. It is shown that a notion of indistinguishability lies at the root of vagueness. This observation in turn gives a common ground to the theories of rough sets and fuzzy sets.

1 Introduction

Two mathematical theories viz., theory of fuzzy sets and theory of rough sets have gained importance during the past few decades in addressing the issue of vagueness be it in reality (i.e. ontological) or be it in perception (i.e. epistemological). A good amount of mathematical-mass has gathered around both concepts. Similarly, there are claims of major applications of both the theories. I am of the opinion that some quite significant mathematical results have been developed within the ambit of both theories. As mentioned at the outset, from the angle of dealing with the issue of vagueness, both the theories have made important contributions.

Right from the inception of rough set theory by Z. Pawlak in 1982 [19], there had been efforts to compare it with the then existing fuzzy set theory [20, 30] which came into being in 1965 through a paper by L. A. Zadeh [32]. Since that time many research articles have been published dealing with the interconnection of the two theories. One direction of research has been to combine the two theories e.g. [4, 5, 25] and the other direction endeavours to explain one theory in terms of the others e.g. [3, 17, 30, 31]. In [31], the author shows that from one standpoint rough sets reduce to fuzzy sets, but from another these are different. Yet in another direction of research many-valued extensions of classical rough sets is sought [29], but in it there is no characterization of classical rough set. Except in [17] and partially in [3], the comparative studies are not based on category theory.

M. Banerjee and A. Seth (Eds.): ICLA 2011, LNAI 6521, pp. 22–37, 2011.

In this paper we shall delve into some kind of comparison of the two theories from the category theoretic angle and investigate into their interdependency. In particular, we shall investigate whether one of the concepts may be reduced to the other in some categorical sense. Secondly, it will also be investigated how the theories address vagueness. It will be argued finally that for this purpose both the theories depend on some notion of indistinguishability of the elements in the universe of discourse and that is reflected in the identity arrows of the objects.

In the first part we shall take up the mathematical investigation while in the second part the issue of vagueness shall be discussed.

2 Fuzzy Sets

A fuzzy subset $\widetilde{A}$ of a universe X is represented by a mapping (also denoted by $\widetilde{A}$), from X to a suitable lattice L. The value $\widetilde{A}(x)$ for any $x \in X$ is interpreted as the degree of membership of x in the fuzzy subset $\widetilde{A}$. Usually L is taken as the unit interval $[0, 1]$. In this latter case the connection between a fuzzy subset and a vague concept is almost immediate. A vague concept C may be represented by a fuzzy subset $\widetilde{A}$ where $\widetilde{A}(x)$ denotes the extent to which x falls under the concept C. C is vague if and only if there are some $x \in X$ for which $\widetilde{A}(x) \neq 0$, 1. In other words, it is a borderline case of the vague concept C. But as is clear, there may be varying degrees in the borderline instances relative to C. We shall come to this issue in the second part. Sometimes, in the general case that is when the value set is a lattice L, the fuzzy subset $\widetilde{A}$ is called an L-fuzzy set [8] or L-valued set. Our ultimate interest in this paper is in L-fuzzy sets where L is a finite linear lattice with certain other operators.

Higgs [10] introduced and studied Heyting algebra-valued sets in full detail in an unpublished but celebrated paper in 1973. This seminal work is based on the category theoretic approach. The $n+1$-element linear lattice $L_{n+1} = \{0 < 1 < 2 < 3 < \ldots < n\}$ is a complete Heyting algebra. So, for a fixed n, L_n-valued sets on X form a subclass of all Heyting algebra valued sets over X. In particular, L_4-valued sets constitute one such subclass. Obtułowicz in 1987 [17] established a connection between L_4-valued sets over X and Pawlak's rough sets over X. In Section 3 we shall discuss this connection.

In this section we present Higgs' category of Heyting algebra valued sets.

Let X be a universe and L a complete Heyting algebra. Let ϵ: $X \times X \to L$ be an L-valued fuzzy binary relation, satisfying the following conditions.

(H_1) $\qquad \epsilon(x, y) = \epsilon(y, x)$
(H_2) $\qquad \epsilon(x, y) \wedge \epsilon(y, z) \leq \epsilon(x, z)$, $x, y, z \in X$.
It then follows that
(H_0) $\qquad \epsilon(x, y) \leq \epsilon(x, x)$ for any x, y.

$(H_1)$and (H_2) are conditions that generalize the notions of symmetry and transitivity properties of a binary relation in the fuzzy case. If, additionally, $\epsilon(x, x) > 0$ for all x, (H_0) may be taken as the generalization of the notion of reflexivity.

So, with this additional condition ϵ may be interpreted as a fuzzy equivalence relation on X. X endowed with an ϵ, satisfying $(H_1)$and (H_2) viz., (X, ϵ) is an object of the category of Higgs that shall be denoted by Set(H).

Morphisms from the object (X, ϵ) to (Y, ϵ') is an L-valued mapping $f : X \times Y \to L$ satisfying conditions

$$(M_1) \qquad \epsilon(x, x') \wedge f(x', y) \leq f(x, y).$$
$$f(x, y) \wedge \epsilon'(y, y') \leq f(x, y').$$
$$(M_2) \qquad f(x, y) \wedge f(x, y') \leq \epsilon'(y, y').$$
$$(M_3) \qquad Sup_y(x, y) = \epsilon(x, x).$$

The identities of the category are the ϵ's and the morphisms are composed as follows:

if $f : X \times Y \to L$ and $g : Y \times Z \to L$ then $f \circ g : X \times Z \to L$ where $f \circ g(x, z) = sup_{y \in Y}(f(x, y) \wedge g(y, z))$.

One can notice that conditions (M_1), (M_2) and (M_3) together constitute a generalization of the notion of function when the fuzzy identities on the universes X and Y are ϵ and ϵ' respectively.

An L-fuzzy subset "within" an object (X, ϵ) may then be defined as a mapping $\mu : X \to L$ given by $\mu(x) = \epsilon(x, x)$.
It is immediate that $\mu(x) \wedge \epsilon(x, y) \leq \epsilon(y, y) = \mu(y)$.
$\mu(x)$ thus defined is (the description of) a sub-object of the object (X, ϵ).

It should be recalled that after the advent of fuzzy set theory there was a spurt of category theoretic interpretation of vagueness [7, 24]. The work of Higgs, though originating from a different motivation, has turned out to be one of the most significant category theoretic studies of fuzzy sets. Subsequent most significant studies of fuzzy sets in this direction include [7, 11–13]. For a brief survey on the various categorical approaches and the interrelation among these categories one may be referred to [1].

3 Rough Sets

The starting point of Pawlak's rough set theory is a non-empty universe X with an equivalence relation R arising out of an information system that shows values of each object of the universe with respect to a prefixed set of attributes. Two objects of the universe become indistinguishable (indiscernible) if they have the same values with respect to each attribute. Thus an equivalence relation is obtained. The pair (X, R) has been called an approximation space. There are several attitudes of defining a rough set (in fact, a rough subset) (cf. [2]). But for the present purpose, as a starter, we take the triple $< X, R, A >$ where X is the universe, R an equivalence relation on X and A, a subset of X, as a rough set. Afterwards, we shall shift a bit away from this description.

The lower approximation $\underline{A}$ of the set A is defined as $\cup\{[x]_R : [x]_R \subset A\}$ and the upper approximation $\overline{A}$ by $\cup\{[x]_R : [x]_R \cap A \neq \phi\}$. $\overline{A} \setminus \underline{A}$ is called the boundary of A in (X, R). Thus the universe is divided into three components, the lower approximation or the interior (I_A) of A, the boundary (B_A) of A and the exterior (E_A) that is the set $X \setminus (I_A \cup B_A)$. The boundary may be empty, but if it is not, the equivalence classes in B_A must not be singleton sets. It may be noted that this division of the space is not one to one with the subsets of X. Two subsets A and A' may give rise to the same interior, exterior and boundary.

Obtułowicz in [17] proposed a representation of rough sets in terms of L_4-valued sets in the following way.

Given a rough set $< X, R, A >$, an L_4-fuzzy subset f_A is defined by

$f_A(x) = 3$ if $x \in \underline{A} = I_A$,
$f_A(x) = 2$ if $x \in \overline{A} \setminus \underline{A} = B_A$,
$f_A(x) = 1$ if $x \in X \setminus \overline{A} = E_A$.

In order to see this L_4-fuzzy subset as a sub-object of some object Obtułowicz defines a fuzzy equivalence relation ϵ_A by

$\epsilon_A(x, x) = f_A(x)$,
and if $x \neq y$,
$\epsilon_A(x, y) = 1$ if xRy holds,
$\qquad\qquad = 0$ if xRy does not hold.

It is an easy exercise to check that ϵ_A satisfies all the axioms (H_0), (H_1), (H_2), and thus (X, ϵ_A) forms an object in Higgs' category with respect to the Heyting algebra L_4. The function f_A gives a subobject of (X, ϵ_A).

One can also check that ϵ_A satisfies the following conditions (so called *roughness conditions*) too.

(R_1) $1 \leq \epsilon_A(x, x)$.
(R_2) If $2 \leq \epsilon_A(x, y)$ then $x = y$ or, in other words, $\epsilon_A(x, y) = 1$ or 0 for $x \neq y$.
(R_3) If $\epsilon_A(x, y) = 1$ then $\epsilon_A(x, x) = \epsilon_A(y, y)$.
(R_4) If $\epsilon_A(x, x) = 2$ then $\epsilon_A(x, y) = 1$ for some y.

Thus the rough set $< X, R, A >$ carries with it a fuzzy identity relation ϵ_A such that

(i) the degree of self-identity of x in X is greater than or equal to 1 (3 if $x \in \underline{A}$, 2 if $x \in \overline{A} \setminus \underline{A}$ and 1 if $x \in (\overline{A})^c$),
(ii) for $x \neq y$, the fuzzy identity is either 1 or 0, and x, y belong to the same equivalence class with respect to R if and only if $\epsilon_A(x, y) = 1$,
(iii) if x, y are different and belong to the same class then their degrees of self-identity are the same,
(iv) if x belongs to the boundary then there is at least another element y in its class.

Obtułowicz also proved a converse theorem.

Let (X, ϵ) be an object of Higgs' category on L_4 with the additional conditions (R_1), (R_2), (R_3), (R_4), then there exists an equivalence relation R on X and a subset $A \subset X$ such that $\epsilon = \epsilon_A$. Definitions of R and A are:

- xRy if and only if $\epsilon(x, y) \geq 1$.
- $A = \{x : \epsilon(x, x) = 3\} \cup B$, where B is the union of proper subsets taken one from each equivalence class $[x]$ such that $\epsilon(x, x) = 2$. By condition (R_4), it is always possible to choose such a proper subset.
- the set $\{x : \epsilon(x, x) = 3\}$ and B are closed relative to the relation R.

From the construction it is clear that A is not uniquely determined. Any subset A' of X which is roughly equal to A, that is $\underline{A} = \underline{A'}$, $\overline{A} = \overline{A'}$, would be such that $\epsilon_A = \epsilon_{A'}$ and conversely.

This non-uniqueness is quite justified since rough sets $< X, R, A >$ and $< X, R, A' >$ should be considered as the 'same' provided A is roughly equal to A'. Subsets A and A' are roughly equal in the context (X, R), but 'rough sets' $< X, R, A >$ and $< X, R, A' >$ are equal. The unimportance of a particular A is also evident from the following description of a rough set [19].

Given (X, R) let us denote by $\approx$, the rough equality of subsets A and A'. Now $\approx$ being an equivalence relation in $P(X)$, the power set of X, one gets the quotient set $P(X)/\approx$. An element of $P(X)/\approx$ is also called a rough set. Thus if $A \approx A'$, the classes $[A]_\approx$ and $[A']_\approx$ are the same rough sets. In this description a particular A has no importance, its class matters.

From the standpoint of first order logic, one can see that one needs a 4-valued linear lattice as the (truth-)value set viz., $\{0 < 1 < 2 < 3\}$ not simply a three-valued set $\{1 < 2 < 3\}$.

One needs 0-ary predicates $\overline{0}$,$\overline{1}$, $\overline{2}$, unary predicate symbols p'_1, p'_2 ...; and binary predicate symbols $\overline{\epsilon_1}$, $\overline{\epsilon_2}$, ... and equality. The interpretations of p'_i's are L_4-fuzzy subsets f_A of a domain X and those of $\overline{\epsilon_i}$'s are fuzzy relations ϵ_A. The values of f_A lie in the ordered subset $\{1 < 2 < 3\}$ of L_4 and the values of ϵ_A over the whole set. In fact, the monadic predicates are redundant as we will see later while treating the general case in the next section. Although A appears in the suffix, it is really unimportant for the interpretation.

4 Rough Membership Function

Given an approximation space (X, R) and subset A, in Pawlak's rough set approach, all elements on the boundary are treated alike. But one equivalence class of the boundary may have more overlap with A than the other. Thus, conceptually, one can visualize this feature as a notion of some object of the universe being more near to the core (the interior) of the vague concept than the other. This idea has an elegant mathematical representation.

Let μ_A be the function from X to $[0, 1]$ defined by $\mu_A(x) = \frac{card([x]_R \cap A)}{card([x]_R)}$, where card denotes the cardinality. If X is assumed to be finite, we see that

$\mu_A(x) = 1$ if $x \in \underline{A} = I_A$,
$\mu_A(x) = 0$ if $x \in E_A$,
$0 < \mu_A(x) < 1$ if $x \in \overline{A} \setminus \underline{A} = B_A$.

Thus all objects in the boundary are not treated equally. The more an object's class overlaps A, the more it is considered to be closer to the interior. All objects clustered in one equivalence class however share the same extent of closeness to the core. From the perspective of intuition this seems to be quite natural if only the extreme two-valued attitude is not adhered to. μ_A has been termed as *rough membership function* in literature [23]. Clearly, two sets A, A' may have the same membership function. This notion may be extended to the case when X is not finite by adopting a suitable measure function on the power set $P(X)$.

For further development, it would be convenient to take only integral values to denote the extent of closeness of an object to the core. So from the function μ_A another function χ_A is defined by
$\chi_A = \mu_A \times$ (l.c.m. of the denominators of all $\mu_A(x)$ such that $0 < \mu_A(x) < 1$).
Thus, $\chi_A(x) = 0$ if $x \in E_A$.
$\chi_A(x) =$ the l.c.m. if $x \in I_A$.
$1 < \chi_A(x) <$ l.c.m. if $x \in B_A$.

For a finite set X, the power set of X is finite and hence the collection $\{\mu_A(x) : 0 < \chi_A(x) < 1, A \subset X\}$ is also finite. One can take the l.c.m. of the denominators of all such $\mu_A(x)$'s and use this number as the multiplier to define $\chi_A(x)$ for all A. Let this l.c.m. or multiplier be taken as n.

Example 1. Let $X = \{a, b, c, d, e, f, g, h\}$ and the partition be given by $\{a, b\}$, $\{c, d, h\}$, $\{e, g\}$, $\{f\}$.
All possible membership values that lie between 0 and 1 are $\frac{1}{2}, \frac{1}{3}, \frac{2}{3}$. So, the l.c.m is 6. Now, if $A = \{b, e, g\}$, $B = \{b, d, e, g, h\}$, $C = \{b, d, e, g\}$, the functions χ_A, χ_B and χ_C are given by

	a	b	c	d	e	f	g	h
χ_A	3	3	0	0	6	0	6	0
χ_B	3	3	4	4	6	0	6	4
χ_C	3	3	2	2	6	0	6	2

We would now like to place this concept of gradation of boundary elements in Obtułowicz's framework by extending the value set to the set $L_{n+1} \equiv \{0 < 1 < 2 < 3 < \ldots < n\}$. A preliminary informal attempt in this direction was made in [3].

Firstly one should observe that L_{n+1} is a Heyting algebra. When a fuzzy equivalence relation ϵ is defined from $X \times X$ to L_{n+1} such that ϵ satisfies condition (H_1) and (H_2) and hence (H_0), (X, ϵ) is an object of the category of Higgs where the Heyting algebra is taken to be L_{n+1} for an arbitrary but fixed n.

Let the following procedure be adopted now.

For any $A \subset X$, define a function F_A by

$$F_A(x) = \begin{cases} 1, & \chi_A(x) = 0; \\ \chi_A(x), & \text{otherwise.} \end{cases}$$

So, it follows that for each $A \subset X$,

$F_A(x) = n$ if $x \in \underline{A} = I_A$.
$F_A(x) = 1$ if $x \in E_A$.

The intermediate values are distributed between 2 to $n-1$. The function F_A is constructed only to maintain Obtułowicz's structure in the extended scenario also. From such an $F_A(x)$, $\mu_A(x)$ can be retrieved. Thus the function F_A can be considered as codifying all the information of the rough membership function μ_A.

We now state one half of the representation theorem in this general context.

Proposition 1. *Let* $< X, R, A >$ *be a rough set. Let* n *be determined as the l.c.m. as mentioned before. Let a fuzzy relation* ϵ_A*:* $X \times X \to L_{n+1}$ *be defined by* $\epsilon_A(x,x) = F_A(x)$*, and for* $x \neq y$*,*

$$\epsilon_A(x,y) = \begin{cases} 1, & \text{if } xRy \text{ holds;} \\ 0, & \text{otherwise.} \end{cases}$$

Then ϵ_A *satisfies the conditions* (H_0)*,*(H_1)*,* (H_2)*,* (R_1)*,* (R_2)*,* (R_3)*,* (R_4') *where* (R_4') *is the condition: if* $\epsilon_A(x,x) \in \{2, \ldots n\}$ *then* $\epsilon_A(x,y) = 1$ *for some* y*.*

Proof. We first observe that $\epsilon_A(x,x) \in \{1, 2, \ldots, n\}$ and $\epsilon_A(x,y) = 0$ or 1. So, $\epsilon_A(x,y) \leq \epsilon_A(x,x)$ for any x, y. Thus the condition (H_0) holds.
(H_1) follows from the symmetry of R.
(H_2) follows from the transitivity of R.
Conditions (R_1) and (R_2) are immediate from the definition.

To establish (R_3), let $\epsilon_A(x,y) = 1$. Then two cases, $x = y$, or $x \neq y$. In the first case $\epsilon_A(x,x) = 1 = \epsilon_A(y,y)$. In the second case, xRy. So, $[x]_R = [y]_R$. So, $F_A(x) = F_A(y)$. Hence $\epsilon_A(x,x) = \epsilon_A(y,y)$.

To establish (R_4'), let $\epsilon_A(x,x) \in \{2, \ldots, n-1\}$. Then $F_A(x) \in \{2, \ldots, n-1\}$. So, $x \in \overline{A} \setminus \underline{A}$. So there is some $y \in [x]_R$, $x \neq y$ and $F_A(x) = F_A(y)$. That is $\epsilon_A(x,x) = \epsilon_A(y,y)$. □

Two points are to be noted here: firstly, condition (R_4') takes care of the fact that the blocks (equivalence classes) appearing on the boundary are now not treated alike. Secondly, if the value of an object lies between 2 and $n-1$, the object is on the boundary and in its block there is more than one object. Thus, a rough set (given by its rough membership value) can be treated as a L_{n+1}-valued fuzzy set.

As two rough sets $< X, R, A >$ and $< X, R, A' >$ may give rise to the same rough membership function, they may be represented by the same L_{n+1}-valued fuzzy set. The non-uniqueness of A stands in support of the presentation of a rough set by giving a partition to the universe X and by colouring the blocks in varying degrees of intensity representing the extent of belongingness of the objects of a particular block to the concept.

To establish the converse, however, a further conceptual advance is to be made.

The rough membership function is just one way of assigning values other than the greatest and the least in the linear scale to the objects belonging to the boundary. This should not be taken as the only way. One may not use the cardinality measure but use other kind of measure or no measure function at all. What is important here is that a block receives a grade for each element in it and these grades are in some relationship with the grade of the fuzzy equivalence relation defined in the universe. Secondly the dependency on the subset A is to be given up since after the blocks are formed, even in the classical case (Pawlak) the blocks belonging to the lower approximation and the boundary of a set only count. Also from the previous proposition viz., Proposition 1, we have noticed that ultimately an L_{n+1}-valued fuzzy subset F_A and an L_{n+1}-valued equivalence relation ϵ_A satisfying some more conditions arise and A is then redundant in the sense that it may be replaced by some other A' roughly equal to A and giving the same rough membership function.

So we propose to ignore A altogether and define a rough set by $< X, R, I, B >$ as was done by Obtułowicz in the case of classical rough set theory, I, a union of blocks being the interior and B another union of blocks as the boundary. It is to be marked that no mention of any subset A is there in this approach. In Obtułowicz's formulations elements in I receive the value 3, and those in B receive 2. Accordingly in our general case we just wish to see the highest value n being given to the objects in I, the value 1 given to the objects in E and values between 2 to $n-1$ to the elements in the boundary. The boundary is now layered as $B = B_2 \cup B_3 \cup \cdots \cup B_{n-1}$ where each B_i is the union of some blocks and such that every element in B_i receives the value i. Some B_i may be empty. If this modified definition is admitted we can obtain a converse of the earlier proposition in the following form.

Proposition 2. *Let X be a universe and ϵ be an $\{0, 1, \ldots, n\}$-valued fuzzy relation satisfying conditions (H_0),(H_1), (H_2), (R_1), (R_2), (R_3), (R'_4). Then there exists an equivalence relation R in X and sets I, $B = B_2 \cup B_3 \cup \cdots \cup B_{n-1}$ forming the interior and the layered boundary respectively of a rough set in $< X, R >$ such that I and B are disjoint and each subsets I, $B_2, \ldots, B_{n-1}$ are R closed.*

Proof. We define a function $F : X \to L_{n+1}$ by $F(x) = \epsilon(x, x)$ and a binary relation R in X by xRy iff $\epsilon(x, y) \geq 1$.
Conditions $(R_1), (H_1)$ and (H_2) establish that R is an equivalence relation.
We then define sets
$I = \{x \in X : F(x) = n\}$
and $B_i = \{x \in X : F(x) = i, \quad i = 2, \ldots, n-1\}$.
So, $F(x) = 1$ if $x \in X \setminus I \cup \{B_i\}$.
Conditions (R_2) and (R_3) establish that I and each B_i are closed with respect to R. (R'_4) shows that for each $x \in B_i$ there exists a $y \in B_i$ such that $x \neq y$ and xRy holds. □

The above two propositions constitute a complete representation of rough sets in terms of some L_{n+1}-valued fuzzy set. Since, at this final stage we are not referring to a subset A of X to define a rough set, the number n may be taken arbitrarily, the least being 2 when we get Obtułowicz's lattice.

In the approach of Wygralak [30] and Yao [31] too, a rough set is shown to be special types of fuzzy sets. But neither of these approaches has attempted to see a rough set as a fuzzy equivalence relation that gives rise to a crisp partition of the universe along with a fuzzy subset of the universe tied up with the partition.

In the present approach, a collection of rough sets on an approximation space (X, R) is given by $(< X, \{\epsilon_i\}_{i\in I})$, where ϵ_i is an L_{n+1}-fuzzy equivalence relation satisfying the aforesaid conditions, $\epsilon_i(x, y) \geq 1$ if and only if $\epsilon_j(x, y) \geq 1$ for each $x, y \in X$ and $i, j \in I$, and xRy holds if and only if $\epsilon_i(x, y) \geq 1$. Such a representation shall be called a *rough set model*.

To obtain a first order logic, the language should have the equality predicate $(=)$, and binary predicate symbols $\overline{\epsilon_1}$, $\overline{\epsilon_2}$, ..., (as many as would be necessary), and 0-ary predicates $\overline{0}, \overline{1}, \ldots \overline{n}$, depending upon the requirement / choice.

We shall need the following proper axioms to be accepted for each $\overline{\epsilon}_i$:

(H_0) $\overline{\epsilon_i}(x, y) \rightarrow \overline{\epsilon_i}(x, x)$.
(H_1) $\overline{\epsilon_i}(x, y) \leftrightarrow \overline{\epsilon_i}(y, x)$.
(H_2) $(\overline{\epsilon_i}(x, y) \wedge \overline{\epsilon_i}(y, z)) \rightarrow \overline{\epsilon_i}(x, z)$.
(R_1) $\overline{1} \rightarrow \overline{\epsilon_i}(x, x)$.
(R_2) $\sim (x = y) \rightarrow (\overline{\epsilon_i}(x, y) \leftrightarrow \overline{0} \vee \overline{\epsilon_i}(x, y) \leftrightarrow \overline{1})$.

To accommodate (R_3), (R_4') we have to take the following rules of inference.

(R_3) $\frac{\vdash\ \overline{\epsilon_i}(x,y)\ \leftrightarrow\ \overline{1}}{\vdash\ \overline{\epsilon_i}(x,x)\ \leftrightarrow\ \overline{\epsilon_i}(y,y)}$.

(R_4') $\frac{\vdash\ (\overline{\epsilon_i}(x,y)\ \leftrightarrow\ \overline{2})\ \vee \ldots \vee\ (\overline{\epsilon_i}(x,y)\ \leftrightarrow\ \overline{n-1})}{\vdash\ \exists y\ (\overline{\epsilon_i}(x,y)\ \leftrightarrow\ \overline{1})}$.

(R_5) $\frac{\vdash\ \overline{1}\ \rightarrow\ \overline{\epsilon_i}(x,y)}{\vdash\ \overline{1}\ \rightarrow\ \overline{\epsilon_j}(x,y)}$.

$\overline{\epsilon_i}$'s are interpreted as L_{n+1}-valued fuzzy equivalence relations satisfying the roughness conditions. Since the value set L_{n+1} is a Heyting algebra the conjunction $(\wedge)$, the disjunction $(\vee)$ are computed by 'min' and 'max' operators respectively. The implication $\rightarrow$ is computed by the operator $\Rightarrow$ given by
$a \Rightarrow b = n$ if $a \leq b$.
$\quad = b$ otherwise.
The negation $(\sim)$ is computed by $\neg$ given by $a \Rightarrow 0$ i.e.
$\neg 0 = n$
$\neg x = 0$ if $x = 1, \ldots, n$.

Axioms (H_0), (H_1), (H_2), (R_1) and (R_2) receive the value n for any valuation and rules (R_3), (R_4') and (R_5) are sound in any rough set model.

Thus a logic for rough sets is also defined. However we are not satisfied with the negation operator. We would rather prefer the operator $\dot{-}$ given by $\dot{-}(x) = n - x$.

The set $\{0 < 1 < \ldots < n < n\}$ is obviously closed with respect to the operator $\dot{-}$. In this value set $\dot{-}\dot{-}\ x = x$ but $x \vee \dot{-}(x) \neq n$ always. In this algebra $\Rightarrow$ is the residuation with respect to 'min' but $\dot{-}(x) \neq x \Rightarrow 0$. We may call this algebra $L_{n+1}(\dot{-})$. So the structure $L_{n+1}(\dot{-})$ with the operators min, max, $\Rightarrow, \dot{-}$ now no longer remains a Heyting algebra. But the claims made so far remain valid. Since the objects, morphisms, identities and compositions of morphisms can be defined as before, a category of $L_{n+1}(\dot{-})$-valued sets is formed.

We could also take $\Rightarrow$ as
$a \Rightarrow b = n$ if $a \leq b$
$\quad = \dot{-} a \vee b$ if $a > b$.
By so doing $\Rightarrow$ no longer remains a residuation with respect to min but $\dot{-} a = a \Rightarrow 0$. We may call this algebra $L'_{n+1}(\dot{-})$.

The reason for dissatisfaction with the Heyting (or intuitionistic) negation $\sim$ and hence with the operator $\neg$ is called for. The intuitionistic negation being defined as $\neg a = a \Rightarrow 0$ and $\Rightarrow$ being the residuum with respect to the lattice meet $\wedge$, one gets $a \wedge \neg a = 0$ (although $a \vee \neg a \neq$ the top element always). But in the case of the borderline instances of a vague predicate, we are not in favour of this property. We are rather in support of Zadeh's original idea regarding fuzziness, viz. if "x is tall" is true to the extent .6 then "x is tall and x is not tall" is true to the extent $.6 \wedge (1 - .6) = .4$, which is not 0. We are aware of the fact that a huge amount of literature on fuzzy sets has been produced taking an MV-algebra as the truth set in which the truth value of the above conjunctive sentence would be 0. We are also aware of the fact that among the philosophers of vagueness, there is a strong group that thinks it proper to accept the above conjunction to be crisply false even though the predicate 'tall' is vague and admits degree of truth. We, however, wish to stick to Zadeh's tradition in this respect and that is why we prefer the operator $\dot{-}$ which is the counterpart of $1 - x$, the standard fuzzy set negation, in the present finite linear truth set. Keeping max, min and $\dot{-}$ for conjunction, disjunction and negation respectively, we have proposed two different implications, the former one is the residuation with respect to $\wedge$ and the latter one defined in the usual way in terms of $\dot{-}$ and $\vee$. The main intention is to stick to the philosophy that in a vague context, for borderline cases the law of contradiction in general fails. Secondly, we have tried to deviate as little as possible from Obtułowicz's construction.

Thus we propose a little drift from Heyting algebra in order to obtain more intuitively acceptable results with negation while retaining the major categorical content of the approach of Obtułowicz.

In the following section we shall discuss the relation between vagueness, indiscernibilities and categories formed out of the indiscernibilities.

5 Vagueness

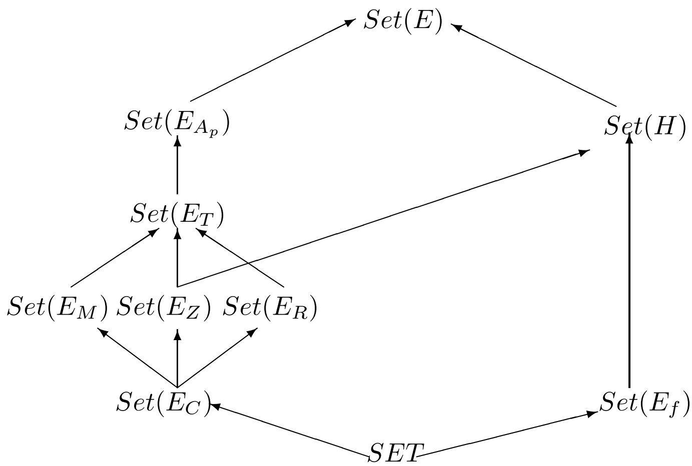
Without entering the debate whether vagueness resides in 'reality' or not, everybody will, hopefully, accept that the language used for communication can in no way avoid it. Parikh [18] rightly observed that vagueness is an essential feature not only of ordinary languages but also of 'precise', 'artificial' languages used in "physics and in fact, any science that attempts to correlate observation with words and numbers". Fuzzy sets and rough sets are two main mathematical models to address vagueness. Early Pawlak in the introduction to his short communication [20], declares, "we compare this concept (*the concept of rough sets: present author*) with that of the fuzzy set and we show that these two concepts are different". At that period Pawlak was firm in his belief that rough set is properly addressing vagueness since it talks about 'boundary' of a set and the property 'rough' is ascribed to a set. On the other hand, although the qualifier 'fuzzy' has been ascribed to set too, actually fuzzy set theory deals with degree of membership of an object to a set and hence is dealing with some kind of uncertainty of belongingness of objects. While vagueness is the property of sets uncertainty is the property of an object. In later years however Pawlak changed his position as is evident from his following remark, "Both fuzzy and rough set theory represent two different approaches to vagueness. Fuzzy set theory addresses gradualness of knowledge, expressed by fuzzy membership – whereas rough set theory addresses granularity of knowledge expressed by indiscernibility relation" [21, 22]. This opinion, however, is not accepted by all important researchers in this field. For example in [5] Dubois and Prade opine, "We argue that fuzzy sets and rough sets aim to different purposes and that it is more natural to try to combine the two models of uncertainty (vagueness for fuzzy sets and coarseness for rough sets) in order to get a more accurate account of imperfect information."

In the opinion of the present author, the boundary of a set is fuzzy/vague, since there are uncertainties in the case of some objects about their belongingness to the set. At the root of vagueness lies an indiscernibility of some kind or other. The gradualness in fuzzy set theory blurring belongingness and non-belongingness in a concept or in other words allowing for borderline instances, arises due to indiscernibility, be it epistemic or ontic. Similarly the indistinguishability arising out of initial information system within the universe generates the thick boundary of a rough set – the elements in this region whether within A or outside are all borderline instances of the vague concept whose one possible extension is A. An interesting contribution of rough set theory in this regard lies in the acceptance, in principle, that the set A is not the unique extension of the concept. Any set A' roughly equal to A may also be taken as another extension. This variability of extension renders vagueness to the concept. In [27], Shapiro enlists four reasons for the indeterminacy of a borderline case 'a' in a concept P viz.

(i) a is either P or non-P, but it is not known which, or even not knowable.
(ii) a is actually neither P nor non-P.
(iii) a is partially P and partially non-P.
(iv) Dependency on the context (perspective), a is sometimes P and sometimes non-P.

Fuzzy set theorists take the view-point (iii) and assign an intermediate 'truth' value to the sentence 'a is P'. On the other hand, a rough set theorist's account clearly indicates the borderline instances of a concept in a given context that is the partioning of the universe. They are neither P nor non-P so that view-point (ii) is adopted. But a change in the context may bring such an instance within P or push it to non-P. Moreover, through rough membership functions, grades of belongingness to the concept are assigned to these elements. Thus in the rough set approach to vagueness, we find traces of the view-points (iii) and (iv) also. In classical fuzzy set theory the degree of membership of an object is directly given

– the indiscernibility hidden behind is not usually brought to the fore. However, in the category theoretic studies explicit mention of the indiscernibility relation is made by taking this relation as the identity morphism of the object. In Higgs' category Set(H) the identity morphism on the object $< X, \epsilon >$ is the mapping $\epsilon : X \times X \to L$ itself. Similar is the case for other formulation of fuzzy sets with truth-value sets different from the Heyting algebra L. A detailed study in this respect is made in [1]. We present here a summary in the following diagram.



The diagram depicts categories, the lowest one being the category SET of sets with the crisp identity. The arrows indicate that the category at the tail is identifiable with some subcategory of that at its head. E's with or without suffixes and H within the braces denote the indiscernibility (or fuzzy equivalence or approximate identity) involved in the making of the objects. The identity arrows and the morphisms of the categories a prototype of which is the Higgs' category Set(H), are to be defined as in Section 2. We now state below the other equivalences and the objects therein.

Set(E_C).
$E_C : X \times X \to \{0, 1\}$ such that
(I)$E_C(< x, y >) \leq E_C(< x, x >)$.
(II)$E_C(< x, y >) = E_C(< y, x >)$.
(III)$E_C(< x, y >) \wedge E_C(< y, z >) \leq E_C(< x, z >)$.
The objects are of the form $< X, E_C, A >$ where $A(x) = E_C(< x, x >)$.

Set(E_{Ap}).
$E_{Ap} : X \times X \to L$ (a residuated lattice [9] with $*$ as the product operation which is commutative, 0, 1 being the lower and upper bounds) such that
(I)$E_{Ap}(< x, y >) \leq E_{Ap}(< x, x >)$, $E_{Ap}(< x, x >) = 1$ or 0.
(II)$E_{Ap}(< x, y >) = E_{Ap}(< y, x >)$.
(III)$E_{Ap}(< x, y >) * E_{Ap}(< y, z >) \leq E_{Ap}(< x, z >)$.
An object is of the form $< X, E_{Ap}, A >$ where $A(x) = E_{Ap}(< x, x >)$. E_{Ap} stands for 'approximate equality'.

Set(E_T).
This category is obtained as a special case of Set(E_{Ap}) when $L = [0, 1]$ and * is taken to be any semi-continuous t-norm on $[0, 1]$ [15].

Categories Set(E_M) (Menger, [16]), Set(E_Z) (Zadeh, [8]) and Set(E_R) (Ruspini, [26]) are all obtained by taking particular instances of t-norms.

Set(E_f) (Eytan, [7]).
$E_f : X \times X \to [0, 1]$ such that
(I)$E_f(< x, y >) \leq E_f(< x, x >)$.
(II)$E_f(< x, y >) = 0$ for $x \neq y$.
It follows that $E_f(< x, y >) \wedge E_f(< y, z >) \leq E_f(< x, z >)$.
The objects are $< X, E_f, A >$ where $A(x) = E_f(< x, x >)$.

Set(H) is Higg's category which has been already defined in more detail.

Set(E).
Here $E : X \times X \to L$ (a residuated lattice) has been defined in a slightly different way.
(I) $E(< x, y >) = E(< y, x >)$.
(II) $E(< x, y >) * E(< y, z >) \leq E(< x, z >)$.
(III)$E(< x, y >) * E(< x, x >) = (< x, y >)$.
The objects are as before $< X, E, A >$ where $A(x) = E(< x, x >)$.

Most of the well known fuzzy sets are objects or sub-objects of objects in either of the categories depicted in the figure or similar one.

The underlying philosophy in all the aforesaid examples of categorical objects is the acceptance of the principle that the degree of belongingness of an element to a concept (either crisp or vague) is the degree to which it is indiscernible with itself relative to the underlying indiscernibility relation.

This point needs some clarification. It is a deep issue, but a rather simplistic explanation is presented here. We know objects in terms of their properties. If a property is vague, its applicability to an object may be partial or graded. Hence the existence of the object relative to the property becomes graded too. Extensionally, a property is represented by a set. When the property is vague, the corresponding set is fuzzy and the belongingness or existence of the object in the set becomes graded. We are in agreement with Dubois and Prade [6] in that the converse is not always the case. A membership degree is not always the representation of uncertainty or vagueness in a concept. Now, a property induces an indiscernibility relation in the universe of objects. Two objects having the same property are indistinguishable relative to that property. When the property is vague, the indiscernibility induced by it is fuzzy. If P is a vague property and ϵ_P is the indiscernibility induced then a measure of indiscernibility between objects x and y may be given by

$$\epsilon_P(x, y) = (P(x) \to P(y)) \wedge (P(y) \to P(x)),$$

$P(x)$ and $P(y)$ being grades of applicability of P to x and y respectively and $\to$ being a suitable implication. (It should, however, be mentioned that this is not the only way of measuring the indiscernibility. See for instance [24, 28].) So $\epsilon_P(x, x) = P(x) \rightarrow P(x)$. Usual implication operators give $P(x) \rightarrow P(x)$ the highest value. But there are others for which the value may not be so. Examples of such implications are plenty (Zadeh, Kleene-Dienes, Reichenbach, cf. [15]). It would be, however, natural to assume that the degree of indiscernibility of x with itself is greater than or equal to the degree of its indiscernibility with other objects, i.e. $\epsilon_P(x, x) \geq \epsilon_P(x, y)$. This is precisely the reflexivity criterion (cf. Section 2). In the cases of $\rightarrow$ by Zadeh and Kleene-Dienes, this inequality holds for ϵ_P. But these should be taken only as instances. We can ignore the way ϵ_P is defined, i.e. by using implication, and take it as a reflexive fuzzy relation on the fuzzy set P. Hence, along with the above condition, it should also satisfy the condition that $\epsilon_P(x, x) \leq P(x) \wedge P(x) = P(x)$. In fact it is assumed to be the same as $P(x)$, i.e. $\epsilon_P(x, x) = P(x)$. The objects of the categories are formed in this way in terms of the identity morphisms. In the logics, ϵ's are taken as primitives whose models satisfy this condition. Thus the degree of existence of an object in a concept is taken to be the same as the degree of its indiscernibility with itself or degree of self-identity relative to the concept. This philosophical view-point gets support even from practical angle – for example in [14] the authors state, "In the case $E(x, x) < 1$, the value $E(x, x)$ would be interpreted as the degree to which x exists in X or belongs to X, i.e $E(x, x)$ reflects a membership degree".

Viewed as in this paper – that is by extending Obtułowicz's approach from 4-valued to $n+1$-valued Heyting algebra, rough sets turn out to be sub-objects of the subcategories Set $(L_{n+1}), n \geq 1$, of Higg's category Set(H).

However, since we prefer to have negation defined in a different way and the algebraic structure then being $L_{n+1}(\dot{-})$ or $L'_{n+1}(\dot{-})$, the corresponding categories of rough sets will no longer be a subcategory of Set(H) generally, but will lie in its 'vicinity'.

It should be mentioned that in the paper [1], a category ROUGH having rough sets $< X, R, A >$ as objects and another kind of morphisms was defined. The identity morphism in this category was taken to be the identity map from $\overline{A}$ to $\overline{A}$. In our present diagram this category is not depicted. It lies between two topoi viz., category $Set(E_C)$ and category $Set(H)$.

It should also be mentioned that in this diagram the categories presented in [11–13] are not depicted. In fact, it would be an interesting and important project to set up a general framework in which a proposed new category with fuzzy equality as the identity within the objects may be placed appropriately.

6 Concluding Remarks

We are aware of the fact that the philosophical standpoint about the degree of existence of an object in connection with its self-identity may be debated upon. What we would like to stress here is that a nice mathematical framework has been developed. Moreover, this framework gives some sort of foundation to both fuzzy set theory and rough set theory so far as representation of vagueness is concerned. But even as a piece of mathematics, several points still remain open for investigation:

(1) completeness theorem for the three different logics given by the truth structures L_{n+1} (a Heyting algebra), $L_{n+1}(\dot{-})$ and $L'_{n+1}(\dot{-})$,
(2) location of the categories that have been proposed relative to the algebraic structures $L_{n+1}(\dot{-})$ and $L'_{n+1}(\dot{-})$,
(3) generalization of the case when the universe is infinite.

We hope these issues may be taken up in the future. The interconnection between vague concepts and indiscernibility hopefully will become clearer as the mathematical representation of these notions become more and more elegant. At least history supports this viewpoint: a nice mathematical model reveals reality better.

References

1. Banerjee, M., Chakraborty, M.K.: Foundations of vagueness: a category-theoretic approach. Electronic Notes in Theoretical Computer Science 82(4) (2003)
2. Banerjee, M., Chakraborty, M.K.: Algebras from rough sets. In: Pal, S.K., Polkowski, L., Skowron, A. (eds.) Rough-Neural Computing, pp. 157–184. Springer, Heidelberg (2004)
3. Chakraborty, M.K.: Pawlak's landscaping with rough sets. Transactions on Rough Sets VI, 51–63 (2007)
4. Dubois, D., Prade, H.: Rough fuzzy sets and fuzzy rough sets. International Journal of General Systems 17, 191–209 (1990)
5. Dubois, D., Prade, H.: Putting rough sets and fuzzy sets together. In: Slowiniski, R. (ed.) Intelligent Decision Support, pp. 203–232. Kluwer, Dordrecht (1992)
6. Dubois, D., Prade, H.: The three semantics of fuzzy sets. Fuzzy Sets and Systems 90(2), 141–150 (1997)
7. Eytan, M.: Fuzzy sets: a topos-logical point of view. Fuzzy Sets and Systems 5, 47–67 (1981)
8. Goguen, J.: L-fuzzy sets. Journal of Math. Anal. Appl. 18, 145–174 (1967)
9. Hájek, P.: Metamathematics of Fuzzy Logic. Kluwer Acadenic Publishers, Dordrecht (1998)
10. Higgs, D.: A category approach to Boolean-valued set theory. Preprint, University of Waterloo (1973)
11. Höhle, U.: Monoidal closed categories, weak topoi and generalized logics. Fuzzy Sets and Systems 42, 15–35 (1991)
12. Höhle, U., Stout, L.N.: Foundations of fuzzy sets. Fuzzy Sets and Systems 40, 257–296 (1991)
13. Klawonn, F.: Fuzzy sets and vague environments. Fuzzy Sets and Systems 66, 207–221 (1994)
14. Klawonn, F., Kruse, R.: Equality relations as a basis for fuzzy control. Fuzzy Sets and Systems 54, 143–156 (1993)
15. Klir, G.J., Yuan, B.: Fuzzy Sets and Fuzzy Logic: Theory and Applications. Prentice Hall, India (2006)

16. Menger, K.: Probabilistic theory of relations. Nat. Acad. Sci. U.S.A. 37, 178–180 (1951)
17. Obtułowicz, A.: Rough sets and Heyting algebra valued sets. Bulletin of the Polish Academy of Sciences (Mathematics) 35(9-10), 667–671 (1987)
18. Parikh, R.: The problem of vague predicates. In: Cohen, R.S., Wartofeny, M.W. (eds.) Language, Logic and Method, pp. 241–261. D. Reidel Publishing Co., Dordrecht (1983)
19. Pawlak, Z.: Rough sets. International Journal of Computer and Information Sciences 11, 341–356 (1982)
20. Pawlak, Z.: Rough sets and fuzzy sets. Fuzzy Sets and Systems 17, 99–102 (1985)
21. Pawlak, Z.: Vagueness and uncertainty: a rough set perspective. Computational Intelligence: An International Journal 11, 217–232 (1995)
22. Pawlak, Z.: Vagueness - a rough set view. In: Mycielski, J., Rozenberg, G., Salomaa, A. (eds.) Structures in Logic and Computer Science. LNCS, vol. 1261, pp. 106–117. Springer, Heidelberg (1997)
23. Pawlak, Z., Skowron, A.: Rough membership functions. In: Yager, R., Fedrizzi, M., Kacprzyk, J. (eds.) Advances in Dempster-Shafer Theory of Evidences, pp. 251–271. Wiley, New York (1994)
24. Pultr, A.: Fuzziness and fuzzy equality. In: Skala, H.J., Termini, S., Trillas, E. (eds.) Aspects of Vagueness, pp. 119–135. D. Reidel Publishing Co., Dordrecht (1984)
25. Radzikowska, A.M., Kerre, E.E.: A comparative study of fuzzy rough sets. Fuzzy Sets and Systems 126, 137–155 (2002)
26. Ruspini, E.: Recent developments in fuzzy clustering. In: Yager, R.R. (ed.) Fuzzy Set and Probability Theory: Recent Developments, pp. 133–147. Pergamon, Oxford (1982)
27. Shapiro, S.: Vagueness in Context. Clarendon Press, Oxford (2006)
28. Valverde, L.: On the structure of F-indistinguishability operators. Fuzzy Sets and Systems 17(3), 313–328 (1985)
29. Vitoria, A., Maluszynski, J., Szalas, A.: Modeling and reasoning with paraconsistent rough sets. Fundamentae Informaticae 97, 405–438 (2009)
30. Wygralak, M.: Some remarks on rough and fuzzy sets. BUSEFAL 21, 43–49 (1985)
31. Yao, Y.Y.: A comparative study of fuzzy sets and rough sets. Information Sciences 109, 227–242 (1998)
32. Zadeh, L.A.: Fuzzy sets. Information and Control 18, 338–353 (1965)

The Logic of Campaigning

Walter Dean[1] and Rohit Parikh[2,⋆]

[1] Philosophy Department, University of Warwick
[2] Brooklyn College and the Graduate Center of CUNY
rparikh@gc.cuny.edu

Abstract. We consider a political candidate campaigning to be elected. Her chances of being elected will depend on how various groups of voters perceive her, and how they perceive her will depend on what she has said. Different groups of voters may have different preferences and a statement preferred by one group of voters may be disliked by another. Moreover, voters may be optimistic (willing to think the best of the candidate), pessimistic (inclined to expect the worse), or expected value voters, who average over various possibilities which may come about if she is elected. Given these considerations, what should she say? We formalize this problem in propositional logic with certain utility values, and certain intensities of preference for various groups of voters, and show that *if the voters are expected value voters*, then she is best off being as explicit as possible. Thus a reluctance to be explicit may come about as a result of the presence of optimistic voters.

***

After Barack Obama's comments last week about what he typically eats for dinner were criticized by Hillary Clinton as being offensive to both herself and the American voters, the number of acceptable phrases presidential candidates can now say is officially down to four. "At the beginning of 2007 there were 38 things candidates could mention in public that wouldn't be considered damaging to their campaigns, but now they are mostly limited to 'Thank you all for coming,' and 'God bless America"' [said] George Stephanopoulos.

The Onion, 1 May, 2008
NB: The Onion is a satirical weekly newspaper in the US.

1 Introduction

A very important part of elections is campaigning. In the US for instance, the actual election takes only one day. However, preparations for the election tend to begin about a year in advance, and campaigning and raising funds for access to media are important stages in the weeks and months preceding the election itself.

⋆ Research supported in part by a grant from PSC-CUNY.

M. Banerjee and A. Seth (Eds.): ICLA 2011, LNAI 6521, pp. 38–49, 2011.

This paper is at the very beginning stages of this important area – what candidates should say in order to do better in the election.

Note that this paper is **not** about social choice theory. We will not speak about *Arrow's theorem* [Arrow], or the *Gibbard Satterthwaite theorem* [Gibbard, Satter], or *approval voting*, studied by Brams and Fishburn [BF]. These are of course important topics about which a lot of research has been done. However, this paper is entirely about *campaigning* and the effect of a candidate's statements on the voters.

2 The Formalism

When a candidate utters a sentence A, she is evaluating its effect on several groups of voters, $G_1, ..., G_n$, with one group, say G_1 being her primary target *at the moment.*

Thus when Clinton[1] speaks in Indiana, the Indiana voters are her primary target but she is well aware that other voters, perhaps in North Carolina, are eavesdropping. Her goal is to increase the likelihood that a particular group of voters will vote for her, but without undermining the support she enjoys or hopes to enjoy from other groups. If she can increase *their support* at the same time as wooing group G_1, so much the better, but at the very least, she does not want to undermine her support in G_2 while appealing to G_1. Nor does she want to be caught in a blatant contradiction. She may not always succeed, as we all know, but remaining consistent, or even truthful, is surely part of her strategy. Lies are expensive.

We will represent a particular *group* of like minded voters as *one* formal voter, but since the groups are of different sizes, these formal voters will not all have the same influence. A formal voter who represents a larger group of actual voters will have a larger *size.* We will assume that each voter has a preferred ideal world – how that voter would like the world to be as a result of the candidate's policies, should she happen to be elected.

Thus suppose the main issues are represented by $\{p, q, r\}$, representing perhaps, policies on the Afghan war, energy, and taxes. If the agent's ideal world is $\{p, q, \neg r\}$, then that means that the voter wants p, q to be true, and r to be false. But it may be that p is more important to the voter than q. Then the world $\{\neg p, q, \neg r\}$ which differs from the ideal world in just p will be worse for the voter than the one, $\{p, \neg q, \neg r\}$, which differs in just q.

We represent this situation by assigning a utility of 1 to the ideal world, and assigning weights to the various issues, adding up to at most 1. If the weights of p, q, r are .4, .2, and .4 respectively and the ideal world is $p, q, \neg r$, then a world in which p, q, r are all true will differ from the ideal world in just r. It will thus have a utility of (1 - .4), or .6.

[1] Some of our examples are taken from the 2008 US election which seems suitable now that the heat is gone.

Each voter also has a theory T_c of the candidate, and in the first pass we will assume that the theory is simply generated by things which the candidate has said in the past. If the candidate has uttered (presumably consistent) assertions $A_1, ..., A_5$, then T_c will be just the logical closure of $A_1, ..., A_5$. If the candidate is truthful, then T_c will be a subtheory of T_a which is the candidate's own theory of the world.

The voter will assume that if the candidate is elected, then one of the worlds which model T_c will come to pass. The voter's utility for the candidate will be obtained from the utilities of these worlds, perhaps by calculating the expected utility over the (finitely many) models of T_c.

We are implicitly assuming that all the worlds are equally likely, something which is not always true, but even such a simple setting turns out to be rich enough for some insights. A different probability distribution will not change the results for optimistic or pessimistic voters, who do not average but take the best or the worst respectively. For those voters who average according to some distribution other than the flat one, the result will still hold that the average of some quantity α over two disjoint sets Y and Z lies in between the average over Y and over Z respectively. Thus proposition 1, below, will still go through.

A similar consideration will apply if some worlds are considered impossible by all voters. Say the truth assignment which assigns value 1 to all of p, q, r is considered impossible by the candidate as well as the voters. Then we can just take the formula $\neg(p \wedge q \wedge r)$ to be in the theory T_c.

Suppose now that the candidate (who knows all this) is wondering what to say next to some group of voters. She may utter some formula A, and the perceived theory T_c will change to $T_c' = T_c + A$ (the logical closure of T_c and A) if A is consistent with T_c, and $T_c * A$ if not. Here the $*$ represents an AGM-like revision operator [AGM].
(Note: The AGM operator $*$ accommodates the revision of a theory T by a formula A which is *inconsistent* with T. For the most part we will assume here that A is in fact something which the candidate believes and is consistent with T_c which is a subtheory of T_a (her actual beliefs), and thus $T_c * A$ really amounts to $T_c \dot{+} A$, i.e., the closure of $T_c \cup \{A\}$.)

Thus the candidate's utterance of A will change her perceived utility in the minds of the voters and her goal is to choose *that* A which will maximize her utility summed over all groups of voters. We can now calculate the utility to her of the utterance of a particular formula A.

Each group of voters will revise their theory of the candidate by including the formula A, and revising their utility evaluation of the candidate.

Let the old utility to group G_i calculated on the basis of T_c be U_i and the new utility calculated on the basis of $T_c * A$ be U_i'. Let w_i be the weight of the group G_i calculated on the basis of size, likelihood of listening to A which is greater for the current target group, and the propensity to actually vote. Then the change in utility – for group G_i will be $w_i(U_i' - U_i)$. The total for all groups on the basis of uttering A, or the *value* of A, will be

$$val(A) = val(A, T_c) = \Sigma w_i(U'_i - U_i)$$

The rational candidate should utter *that* A which will have the largest value for $val(A)$.

3 Some Examples

3.1 What Happens When Someone Says Something?

How does the information state of other people change?

We start with an informal example[2]

In the Coffee Shop

Three people, A, B, C walk into a coffee shop. One of them orders cappuccino, one orders tea, and one orders icecream. The waiter goes away and after ten minutes *another* waiter arrives with three cups. "Who has the cappuccino?" "I do," says A. "Who has the tea?" "I do," says C.

Will the waiter ask a *third* question?"

Waiter's Deduction

Consider the possible situations for waiter 2. They are

1) CTI 2) CIT
3) TCI 4) TIC
5) ICT 6) ITC

Here CTI indicates that A has cappuccino, B has tea, and C has icecream.

When A says that he has the cappuccino, 3,4,5,6 are eliminated. The waiter now has,

1) CTI 2) CIT

When C says that he has the tea, 1 is eliminated.

Now 2 alone is left and the waiter knows that B has the icecream.

Thus the waiter need not ask a third question.

3.2 Learning from Communication

Observation: Suppose a group of people are commonly aware of a number of possibilities (states) among which they are uncertain. They commonly know some fact B if B is true of all these possibilities. Now, if a public announcement of some true formula A is made, then the new situation is obtained by *deleting* all states s where A is false. The coffee shop example illustrated this fact.

This moral is apparent in the sequels to Aumann's "Agreeing to disagree" paper [Aum]. The first such sequel is by Geanakoplos and Polemarchakis [GP], and another one is due to Parikh and Krasucki [PK]. Of course there are many others. A logic for dealing with such issues was initially developed by Jan Plaza [Pla], and the book [DEL] deals entirely with this issue. See also, [Bacharach,Cave].

We shall be concerned in this paper with the change in knowledge (belief) which comes about when a *candidate for election* says something and how her statements affect the voters' view of her.

[2] Thanks to Johan van Benthem for this example which is a variant of one he used with Dutch children.

4 The Use of Language in Campaigning

4.1 Towards a Formal Model: Languages and Theories

- We begin by considering a single candidate C.
- C's views about the issues are formulated in a propositional language $\mathfrak{L}$ containing **finitely many** atomic propositions $At = \{P_1, \ldots, P_n\}$. These propositions indicate that certain actions will be or will not be taken.
- For instance:
  - $P_1 =$ `We should withdraw from Iraq.`
  - $P_2 =$ `I will impose no new taxes.`
  - ...
  - $P_n =$ `We should bail out the banks.`
- $T_a = C$'s **actual theory** (i.e. the entirety of her views)
- $T_c = C$'s **current theory** (i.e. what's she's said thus far)
- Typically (but not always) $T_c \subseteq T_a$.

Here we imagine the language $\mathfrak{L}$ as obtained from the primitive statements (members of At) like P_i using truth functional symbols, perhaps $\neg, \vee, \wedge$. When truth values are assigned to the primitive statements, then the truth values of compound statements can be easily computed using the appropriate truth tables. Thus for instance the value of $\neg A$ is *true* just in case the value of A is *false*, and the value of $A \wedge B$ is *true* just in case *both* the values of A and B are *true*. We use the colorful word *world* as an alternative to the more prosaic *truth assignment*. If w is a world, and A is a formula, then we write $w \models A$ to mean that A is true in w. If T is a theory (a logically closed set of formulas) then we write $w \models T$ to mean that all members of T are true in w. If the candidate's current theory is T_c then we will write X_c for the corresponding set of worlds $X_c = \{w | w \models T_c\}$.

4.2 Worlds and Preferences

- We conflate propositional valuations and worlds $w \in 2^{At}$.
- We also define $w[i] = \begin{cases} 1 & w \models P_i \\ -1 & w \not\models P_i \end{cases}$
- We initially consider a single group of voters V (think of this as a constituency).
- The voters in V are characterized by their preference for a set of ***ideal worlds***.
- This is formalized via two functions p_v, x_v:
  - $p_v(i) = \begin{cases} 1 & V \text{ would prefer } P_i \text{ to be true} \\ 0 & V \text{ is neutral about } P_i \\ -1 & V \text{ would prefer } P_i \text{ to be false} \end{cases}$
  - $x_v : At \to [0, 1]$ the weight which V assigns to P_i s.t. $\sum w_v(i) \leq 1$.

NB: We have used -1 rather than 0 for *false* since it makes the arithmetic simpler.

4.3 Utilities of Worlds and Theories

– The utility of a world for V is defined as

$$u(w) = \sum_{1 \leq i \leq n} p_v(i) \cdot x_v(i) \cdot w[i]$$

– Note that a candidate's current theory T_c is likely to be **incomplete** – i.e. she may not have expressed a view on some P_i.
– To calculate the utility of an arbitrary T we need to know how V will "fill in the blanks," i.e., extend the evaluation from single worlds to a *set* of worlds.

4.4 Voter Types

– We suggest that there are least three types of voters:
  - **Optimistic voters** (who assume the best about C given T_c)
  - **Pessimistic voters** (who assume the worst about C given T_c)
  - **Expected value voters** (who average across possibilities compatible with T_c).

We are using the flat probability distribution here, but another one would work as well, provided that it is common knowledge.

Mathematically:

– optimistic voters: $ut^o(T) = \max\{u(w) : w \models T\}$
– pessimistic voters: $ut^p(T) = \min\{u(w) : w \models T\}$
– expected value voters: $ut^e(T) = \frac{\sum_{w \models T} u(w)}{|\{w : w \models T\}|}$

4.5 The Value of a Message

– Suppose T is the logical closure C of T_c .
– What's the best thing for her to say next?
– Roughly: $val(A, T) = ut(T \circ A) - ut(T)$
Here $T \circ A$ indicates a revision of the theory T after the formula A is received.
– But the precise definition will depend on
  - the kind of voter we're assuming (i.e. **o** vs. **p** vs. **e**)
  - the set from which A is selected
    (Which will affect which is the *best* A.)
– Wrt the latter, consider A from
  - $\mathcal{X}_a = T_a$ (i.e. only "true convictions")
  - $\mathcal{X}_t = \mathcal{L} - \{\neg A : T_a \vdash A\}$ (i.e. anything consistent with "true convictions" = **tactical**)
  - $\mathcal{X}_m = \mathcal{L} - \{\neg A : T_c \vdash A\}$ (i.e. anything consistent with the current theory = **Machiavellian**)
  - $\mathcal{X}_\ell = \mathcal{L}$ (i.e. any sentence in the language, allowing for contradictions and deception)

- Note: $\mathcal{X}_a \subseteq \mathcal{X}_t \subseteq \mathcal{X}_m \subseteq \mathcal{X}_\ell$

- If we have $\mathcal{X} = \mathcal{X}_\ell$ then T_c may become **inconsistent** since a formula A in $\mathcal{X}_\ell$ might be inconsistent with T_c.
- In this case, $\circ = *$ (i.e. an AGM-like update operation).
- In the other cases, $\circ = \dot{+}$ (logical closure of $T \cup \{A\}$).
- If $\mathcal{X} = \mathcal{X}_a, \mathcal{X}_t$ or $\mathcal{X}_m$, then we let

$$val(A,T) = ut(T \dot{+} A) - ut(T)$$

where ut is one of ut^o, ut^p or ut^e.
- We can now define **best statements** for C given T from $\mathcal{X}$ as follows:

$$best(T, \mathcal{X}) = argmax_A val(A,T) : A \in \mathcal{X}$$

Note: The paper [AGM] considered the issue of how a theory T is to be revised when a formula A *inconsistent with* T is received. While we mention this case as relevant, we do not carry out a detailed study of it.

Single voter

- Suppose $T_c = \{P_1 \vee P_2, P_1 \to P_3, P_2 \to \neg P_3\}$
- There are two assignments satisfying T_c:
  $w = \langle 1, -1, 1\rangle$, $w' = \langle -1, 1, 1\rangle$.
- Consider a single voter V_1 with the following preferences:
  - $p_1(1) = 1, p_1(2) = -1, p_1(3) = -1$
  - $x_1(1) = .5, x_1(2) = 0, x_1(3) = 0$
- What should C say?
- (Note: she only needs to consider P_1, P_2.)
- $ut_1^e(T) = \frac{\sum_{w \models T} u_1(w)}{|\{w : w \models T\}|} = (.5 + -.5)/2 = 0$
- $ut_1^e(T \dot{+} P_1) = \frac{\sum_{w \models T \dot{+} P_1} u(w)}{|\{w : w \models T \dot{+} P_1\}|} = .5/1 = .5$
- $ut_1^e(T \dot{+} P_2) = \frac{\sum_{w \models T \dot{+} P_2} u(w)}{|\{w : w \models T \dot{+} P_\}|} = -.5/1 = -.5$
- So $best_1(T_c, \mathcal{X}_m) = P_1$.

Multiple voters

- Consider a *second* voter V_2 with the following preferences:
  - $p_2(1) = -1, p_2(2) = 1, p_2(3) = 1$
  - $x_2(1) = .5, x_2(2) = .25, x_2(3) = 0$
- $ut_2^e(T) = \frac{\sum_{w \models T} u(w)}{|\{w : w \models T\}|} = (-.75 + .75)/2 = 0$
- $ut_2^e(T \dot{+} P_1) = \frac{\sum_{w \models T \dot{+} P_1} u(w)}{|\{w : w \models T \dot{+} P_1\}|} = -.75/1 = -.75$
- $ut_2^e(T \dot{+} P_2) = \frac{\sum_{w \models T \dot{+} P_2} u(w)}{|\{w : w \models T \dot{+} P_2\}|} = .75/1 = .75$
- So $best_2(T_c, \mathcal{X}_t) = P_2$.
- Since $.75 > .5$, if V_1 and V_2 are the entire audience, C should say P_2.
- In general, $Best_{\mathcal{V}}(T, \mathcal{X}) = argmax_A \sum_{i \in \mathcal{V}} val_i(A,T) : A \in \mathcal{X}$.

5 Complex Statements

Proposition 1. Assume **e**-voters. For all A, B s.t. $A, B, A \wedge B \in \mathcal{X}_m$, (i.e., $A, B, A \wedge B$ consistent with T_c) there exist $a, ..., f \in [0,1]$ s.t.

1) $a \cdot val(A,T) + b \cdot val(\neg A,T) = 0$
2) $val(A \wedge B,T) = val(A,T) + val(B, T \dotplus A) = val(B,T) + val(A, T \dotplus B)$
3) $c \cdot val(A \vee B,T) + d \cdot val(A \wedge B,T) = e \cdot val(A,T) + f \cdot val(B,T)$

Proof: For 1), $ut(T) = a \cdot ut(T + A) + (1 - a) \cdot ut(T + \neg A)$

where $a = \frac{\{w \mid w \models T \dotplus A\}}{\{w \mid w \models T\}}$ Here we make use of the two facts that a) the (change in) utility over the union of two disjoint sets must be an average of the change in utility over each separately and b) the change in utility over the whole set (all models of T) is clearly 0 since things have been left as they are. □

5.1 Moving to Complete Theories

Proposition 2. If all voters are e-voters, then there is a **complete** $T \supseteq T_c$ s.t. $ut^e(T) \geq ut^e(T_c)$.

Proof: From the above, we must have exactly one of

i) $val(P_i, T) = val(\neg P_i, T) = 0$
ii) $val(P_i, T) > 0$ and $val(\neg P_i, T) < 0$
iii) $val(P_i, T) < 0$ and $val(\neg P_i, T) > 0$

Suppose $Q_i, \ldots, Q_k$ $(k \leq n)$ are all the atoms not in T_c.

Let $T_0 = T_c$ and $T_{i+1} = \begin{cases} T_i \cup Q_i & val(Q_i, T_i) \geq 0 \\ T_i \cup \neg Q_i & \text{else} \end{cases}$

Let $T = Cn(T_k)$.

This T is complete and has a value at least as great as that of T_c. □

Note: The result will still hold if all voters are e-voters or p-voters. When a statement is made to pessimistic voters, then their opinion will either improve or remain the same. For e-voters, we saw that at least one of the formulas A and $\neg A$ will have non-negative value. Hence the formula $a \cdot val(A,T) + b \cdot val(\neg A,T) = 0$ in proposition 1 will be replaced by $a \cdot val(A,T) + b \cdot val(\neg A,T) \geq 0$ for suitable non-negative a, b. It is optimists who need to be kept in the dark.

Proposition 3. One of the best extensions of T_c is a complete theory $T \supseteq T_c$

Proof
Suppose T' is a best extension of T_c and T' is incomplete. By the previous corollary, there is $T'' \supseteq T'$ which is a complete extension of T' (and thus of T_c) such that $ut^e(T'') \geq ut^e(T')$. □

The previous result suggests that if C assumes **e**-voters, then it will never be to C's disadvantage to move towards a complete theory. But why then do we have the *Onion* phenomenon? I.e. why do candidates state vacuities like "`God bless America`" or "`9/11 was a tragedy`."

– **Conjecture:** They might be assuming that there are at least some **o**-voters (who 'always assume the best').

  (For an optimistic voter, any additional information will either leave his best world *in* or take it *out*, and so he can never feel *better* after additional information. It is best to keep such voters in a state of uncertainty.)
– $T \supseteq T' \implies \max\{u(w) \mid w \models T'\} \leq \max\{u(w) \mid w \models T\}$
– I.e. $T \supseteq T' \implies ut^o(T') \leq ut^o(T)$
– $T \supseteq T' \implies \min\{u(w) \mid w \models T'\} \geq \min\{u(w) \mid w \models T\}$
– I.e. $T \supseteq T' \implies ut^p(T') \geq ut^p(T)$
– Another possibility is that when you remove one dollar from voter V and give one dollar to voter V', then research shows that the anger which voter V feels will exceed the pleasure which voter V' feels. This is a case of "minimizing regret". Thus it could be that a voter feels more passion about some proposition which 'goes wrong' than about the same proposition which 'goes right'. In that case a candidate would be making a mistake being explicit about it since her likely gain with some voters may well be less than her loss with others.

  The current election in the US (two years after the one on which we are drawing for examples) is much richer than the last one since we have some 'fascinating' candidates. Hopefully the current observations will lead to greater insight for the next incarnation of this work.

5.2 Does Order Matter?

– Does the order in which C says A and B matter?
– Proposition 1.2 suggests "no" in the case A, B are consistent with T_c.
– This may seem like a counter-intuitive result:
  • $A =$ `Read my lips: 'no new taxes.'`
  • $B =$ `We must institute user fees.`
– $\boldsymbol{A}; \boldsymbol{B}$ allowed Bush senior to seem as if he favored low taxes and small government.
– $\boldsymbol{B}; \boldsymbol{A}$ might have had the opposite result.
– Our current model doesn't account for this.
– Planned extensions
  • extend with a formal model of implicature
  • type dynamics: after hearing A, maybe some voters **change type** from expected value to optimistic

6 Independent Topics

– Suppose that A and B are in disjoint languages (and hence about unrelated topics).
– e.g. $A \in \mathcal{L}_1$ is about abortion, $B \in \mathcal{L}_2$ is about Iraq.
– Our intuition is that order will not matter in this case.

– Then we should have

$$val(A, T) + val(B, T * A) = val(B, T) + val(A, T * B)$$

even if any of $A, B, A \wedge B$ are **inconsistent** with T_c.
– I.e. even if $T_c \vdash \neg A$, then updating $A; B$ should have the same effect as update $B; A$.
– The next result addresses this point ...

Definition 1. *Let T be a theory in the language $\mathcal{L}$, $\langle \mathcal{L}_1, \mathcal{L}_2 \rangle$ a partition of $\mathcal{L}$ into disjoint sublanguages.*

– *We say that $\mathcal{L}_1, \mathcal{L}_2$ **split** T if there are $A \in \mathcal{L}_1$, $B \in \mathcal{L}_2$ s.t. $T = Cn(A, B)$.*
– *Similarly we say that pairwise disjoint languages $\mathcal{L}_1, .., \mathcal{L}_n$ split T if there are $A_i \in \mathcal{L}_i$ s.t. $T = Cn(A_1, ..., A_n)$.*
– *In such a case, we say that $\langle L_1, ..., L_n \rangle$ is a T-**splitting**.*

Proposition 4. ([Pa'99], [KM'07]) Every first order theory has a unique finest splitting.

Thus a theory can be seen uniquely as consisting of a number of subtheories, each about its own subject matter. A numerical notion of information can be defined for the propositional case (Pa'09) and it can be shown that T splits into T_1 and T_2 iff the information in T is no more than the information in T_1 plus T_2.

Example: Suppose T is generated by the two axioms, $P, Q \vee R$. Then T splits into T_1 generated in the sublanguage $\{P\}$ by P, and T_2 generated in the sublanguage $\{Q, R\}$ by $Q \vee R$. T cannot be split further. The information that one of Q, R is true is *shared* between Q and R, and is not information about either separately.

Proposition 5. Suppose

– C's current theory is T over language $\mathcal{L}$.
– $\mathcal{L}$ can be split into $\mathcal{L}_1, \mathcal{L}_2$.
– Let $A \in \mathcal{L}_1$ and $B \in \mathcal{L}_2$ be **any** statements that the candidate could make.

Then $val(A, T) + val(B * A) = val(B, T) + val(A, T * B)$ where $*$ is an update operator satisfying $T * A * B = T * B * A$.

7 Implicature

Motorist: My car is out of gasoline.
Passerby: There is a gasoline station around the corner.

In this example from Paul Grice, the passerby has not *said* but has *implicated* that as far as she knows, the gasoline station is open. As Savage and Austin have pointed out, each statement made is also an *action* and is evaluated as a move in a game. In this case, the game is cooperative as the passerby (presumably) only wants to help the motorist.

In other situations, there could be an element of opposition between a speaker and a listener, and the listener will learn to read *between the lines.*

A candidate who is aware of the fact that her words are being *interpreted* will speak in awareness of this fact.

Paul Grice [Grice] discusses the issue of extra information, technically called *implicature*, which is conveyed when a person speaks to another. Grice assumed, as in the gasoline station example, that the two parties have the same interest, namely to get the motorist going. But the assumption of common interests can be dropped and there is a whole line of research about cheap talk [CS, FR] where the interests are no longer *wholly shared* but some information can still be conveyed. For instance, Hillary Clinton, campaigning in Indiana, said that she had once shot a duck as a child. This fact in itself did not express any political action directly, but nonetheless there would be a Gricean effect and a conservative voter could assume that she would not push gun control too strongly.

We will pursue this line in subsequent research.

8 Future Work

- candidates address multiple groups of voters with *partial* knowledge of their relative sizes
- multiple candidates (their statements can interact and they can speak about or reply to each other)
- outside events (i.e. "nature" sequentially makes certain propositions true with probabilities either known or unknown to the candidates – e.g. hurricanes, bank failures)
- enriching the language used by the candidates
- e.g. with a conditional operator to formalize

```
If Israel attacks Iran, then the US must ...
```

- extend formal theory of implicature [after [Benz et al, [BJR] or Parikh & Ramanjuan [PR'03]]

References

[AGM] Alchourron, C.E., Gardenfors, P., Makinson, D.: On the Logic of Theory Change: Partial Meet Contraction and Revision Functions. The Journal of Symbolic Logic 50(2), 510–530 (1985)

[Arrow] Arrow, K.J.: Social Choice and Individual Values. Wiley, New York (1951b); 2nd ed. (1963)

[Aum] Aumann, R.: Agreeing to Disagree. Annals of Statistics 4, 1236–1239 (1976)

[Austin] Austin, J.L.: How to Do Things with Words, Oxford (1962)

[Bacharach] Bacharach, M.: Some Extensions of a Claim of Aumann in an Axiomatic Model of Knowledge. J. Economic Theory 37, 167–190 (1985)

[BF] Brams, S., Fishburn, P.: Approval Voting. Springer, Heidelberg (2007); Originally published in 1983

[Cave] Cave, J.A.K.: Learning to Agree. Economics Letters 12, 147–152 (1983)

[BJR] Benz, A., Jäger, G., Van Rooij, R. (eds.): Game Theory and Pragmatics. Palgrave Studies in Pragmatics, Language and Cognition. Macmillan, Palgrave (2005)

[CS] Crawford, V.P., Sobel, J.: Strategic information transmission. Econometrica 50(6), 1431–1451 (1982)

[DEL] van Ditmarsch, H., van der Hoek, W., Kooi, B.: Dynamic Epistemic Logic. Springer's Synthese Library Series (2008)

[FR] Farrell, J., Rabin, M.: Cheap Talk. The Journal of Economic Perspectives 10(3), 103–118 (Summer 1996)

[GP] Geanakoplos, J., Polemarchakis, H.: We Can't Disagree Forever. J. Economic Theory 28, 192–200 (1982)

[Gibbard] Gibbard, A.: Manipulation of voting schemes: a general result. Econometrica 41(4), 587–601 (1973)

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

[KM'07] Kourousias, G., Makinson, D.: Parallel interpolation, splitting, and relevance in belief change. Journal of Symbolic Logic 72(3), 994–1002 (2007)

[Pa'99] Parikh, R.: Belief revision and language splitting. In: Moss, Ginzburg, de Rijke (eds.) Proc. Logic, Language and Computation, CSLI 1999, pp. 266–278 (1999); Earlier version appeared in 1996 in the prliminary proceedings

[Pa'09] Parikh, R.: Beth definability, Interpolation and Language Splitting. In: To appear in the proceedings of the Beth symposium, Amsterdam (2009)

[PK] Parikh, R., Krasucki, P.: Communication, Consensus and Knowledge. Journal of Economic Theory 52, 178–189 (1990)

[PR'03] Parikh, R., Ramanujam, R.: A Knowledge based Semantics of Messages. J. Logic, Language and Information 12 (2003)

[Pla] Plaza, J.A.: Logics of Public Communications, in Knowledge, Rationality & Action, A Special Section of: Synthese. An International Journal for Epistemology, Methodology and Philosophy of Science 158(2), 165–179 (2007)

[Satter] Satterthwaite, M.A.: Strategy-proofness and Arrow's Conditions: Existence and Correspondence Theorems for Voting Procedures and Social Welfare Functions. Journal of Economic Theory 10, 187–217 (1975)

[Sav] Savage, L.: The Foundations of Statistics. Dover, New York (1972); Originally published in 1954 by John Wiley and Sons

A Stochastic Interpretation of Propositional Dynamic Logic: Expressivity

Ernst-Erich Doberkat*

Chair for Software Technology
Technische Universität Dortmund
doberkat@acm.org

Abstract. We propose a probabilistic interpretation of Propositional Dynamic Logic (PDL). We show that logical and behavioral equivalence are equivalent over general measurable spaces. Bisimilarity is also discussed and shown to be equivalent to logical and behavioral equivalence, provided the base spaces are Polish spaces. We adapt techniques from coalgebraic stochastic logic and point out some connections to Souslin's operation $\mathcal{A}$ from descriptive set theory.

1 Introduction

Propositional Dynamic Logic (PDL) is a modal logic originally proposed for modelling program behavior. Its basic operators are of the form $\langle\pi\rangle$, where π is a non-deterministic program; a formula $\langle\pi\rangle\varphi$ holds in a state s iff some terminating execution of π in s may lead to a state in which φ holds. Programs are composed from basic programs by sequential composition, iteration, and by non-deterministic choice. Usually a test operator is available as well: if φ is a formula, then program $\varphi?$ tests whether φ is true; if it is, the program continues, if not, it fails. This dynamic logic is interesting from an application point of view, see [1] for an overview from a semantic perspective. It has attracted attention as a possible model for two-person games, where the programs are thought of as games for modelling the behavior of the players, see [9].

We are proposing an interpretation of PDL through stochastic Kripke models. This appears to be new, it is motivated by two observations. First, games and economic behavior are successfully modelled through probabilistic models, and PDL permits capturing games and their semantics, it might be interesting to know what probabilistic properties are reflected by the logic. For example, the notion of bisimilarity is of some interest in modelling the equivalence of games [9, Section 4], it has also been extensively studied in the area of coalgebras, modal logics and their probabilistic interpretations [3,4], so it is worthwhile to dwell on this common interest. The second observation addresses the dynamic nature of PDL. When interpreting modal or coalgebraic logics, each modal operator is assigned a relation or a predicate lifting which is associated with its interpretation. This property has to be addressed for a probabilistic interpretation of PDL. Closely connected with the interpretation of logics is the question of expressivity

* Research funded in part by Deutsche Forschungsgemeinschaft, grant DO 263/12-1, *Koalgebraische Eigenschaften stochastischer Relationen.*

M. Banerjee and A. Seth (Eds.): ICLA 2011, LNAI 6521, pp. 50–64, 2011.

of their models. This problem is addressed in the rest of the paper, performing some transformations from one model class to others, for which the characterization of expressivity has been undertaken already.

After proposing a probabilistic interpretation – what do we do with it? First, we have a look at the structure of the sets of states for which a formula holds. These set are usually measurable in probabilistic interpretations of modal logics, but this is not automatically the case in the present interpretation. We require the Souslin operation from descriptive set theory in order to make sure that these sets are well-behaved. We then show how different models for the logic do compare to each other. Logically equivalent models are behaviorally equivalent, thus having the same theory is for two models equivalent to finding a model onto which both can be mapped by a morphism. This result is well-known in "static" modal logics. Bisimulations are considered next, and here we need to restrict the generality of the models under consideration to those working over Polish spaces, i.e., over complete and separable metric spaces. A fair amount about models for logics of the type FRAG over Polish spaces, so we transform the problem again, solve it in the sphere of FRAG, and transform the solution back to PDL. We show that bisimilarity is equivalent to logical and behavioral equivalence, provided the models are defined over Polish spaces.

Organization. Section 2 collects some notions from probability. Section 3 defines the set of programs we are working with, and defines some logics of interest. The interpretation proposed in Section 4 is analyzed in terms of infinite well-founded trees, and some results on measurability are derived. Morphisms for comparing models are introduced in Section 5. Section 6 defines issues of expressivity formally and derives the main result in two steps, isolating topological requirements from general questions of measurability. Section 7 wraps it all up and proposes further research.

2 Preliminaries

The interpretation of logics through stochastic Kripke models requires some tools from measure theory. For more information on these topics the reader is referred to Srivastava's treatise [12] on Borel sets, or to the tutorial Chapter 1 in [4].

Measurable Spaces. Let $(X, \mathfrak{C})$ be a measurable space, i.e., a set X with a σ-algebra $\mathfrak{C}$ of subsets; the elements of $\mathfrak{C}$ are called $\mathfrak{C}$-*measurable sets* (or just *measurable sets*, if no confusion arises). Denote by $\sigma(\mathfrak{C}_0)$ the smallest σ-algebra which contains the family $\mathfrak{C}_0$ of sets. $\mathfrak{S}(X, \mathfrak{C})$ denotes the set of all subprobabilities on $(X, \mathfrak{C})$. Let $(Y, \mathfrak{D})$ be another measurable space, a map $f : X \to Y$ is called $\mathfrak{C}$-$\mathfrak{D}$-*measurable* iff $f^{-1}[D] \in \mathfrak{C}$ for all $D \in \mathfrak{D}$. This implies that a real-valued map f on X is $\mathfrak{C}$-*measurable* iff the set $\{x \in X \mid f(x) < r\}$ is a member of $\mathfrak{C}$ for each $r \in \mathbb{R}$.

Let $\mu \in \mathfrak{S}(X, \mathfrak{C})$ be a subprobability and $f : X \to Y$ be $\mathfrak{C}$-$\mathfrak{D}$-measurable. Put $\mu^f(D) := \mu(f^{-1}[D])$ whenever $D \in \mathfrak{D}$, then μ^f is the *image of measure* μ *under* f; apparently $\mu^f \in \mathfrak{S}(Y, D)$. The integral with respect to an image measure can be computed through the original measure, as the *change of variables formula* shows.

Lemma 1. *Let $(X, \mathfrak{C})$ and $(Y, \mathfrak{D})$ be measurable spaces, $f : X \to Y$ be a $\mathfrak{C}$-$\mathfrak{D}$-measurable map, and $\mu \in \mathfrak{S}(X, \mathfrak{C})$. Then $\int_Y g(y)\, \mu^f(dy) = \int_X (g \circ f)(x)\, \mu(dx)$ for each $\mathfrak{D}$-measurable and bounded $g : Y \to \mathbb{R}$.* $\dashv$

Stochastic Relations. A *stochastic relation* $K : (X, \mathfrak{C}) \rightsquigarrow (Y, \mathfrak{D})$ between the measurable spaces $(X, \mathfrak{C})$ and $(Y, \mathfrak{D})$ is a map $K : X \to \mathfrak{D} \to [0, 1]$ with these properties: $K(x)$ is for each $x \in X$ a subprobability on $(Y, \mathfrak{D})$, and the map $x \mapsto K(x)(D)$ is $\mathfrak{C}$-measurable for each $D \in \mathfrak{D}$. In the parlance of probability theory, stochastic relations are called transition probabilities. We note in particular that $K(x)(Y) \leq 1$ for $x \in X$, hence $K(x)(Y) < 1$ may occur, so that mass may vanish. This caters for the observation that, e.g., programs sometimes do not terminate.

If $L : (Y, \mathfrak{D}) \rightsquigarrow (Z, \mathfrak{E})$ is another stochastic relation, then the *convolution* $L{*}K$ *of* L *and* K is defined through $\big(L{*}K\big)(x)(E) := \int_Y L(y)(E)\ K(x)(dy)$ $(x \in X, E \in \mathfrak{E})$. Standard arguments show that $L{*}K : (X, \mathfrak{C}) \rightsquigarrow (Z, \mathfrak{E})$ is a stochastic relation between $(X, \mathfrak{C})$ and $(Z, \mathfrak{E})$. Note that the convolution has identities I_X, I_Y with $K{*}I_X = I_Y{*}K = K$: put $I_X(x)(C) := (x \in C?1 : 0)$ for $x \in X, C \in \mathfrak{C}$; this is the *indicator function for* $C \subseteq X$.

Completion And Operation $\mathcal{A}$. Given $\mu \in \mathfrak{S}\,(X, \mathfrak{C})$, a set $N \subseteq X$ is called a μ*-null set* iff there exists $N \subseteq N_0 \in \mathfrak{C}$ with $\mu(N_0) = 0$; $\mathfrak{N}_\mu$ is the set of all μ-null sets. Define the μ*-completion* $\overline{\mathfrak{C}}^\mu$ of $\mathfrak{C}$ as $\sigma(\mathfrak{C} \cup \mathfrak{N}_\mu)$, thus $M \in \overline{\mathfrak{C}}^\mu$ iff there exists $M_1 \subseteq M \subseteq M_2$ with $M_1, M_2 \in \mathfrak{C}$ and $\mu(M_2 \setminus M_1) = 0$. The *universal completion* $\overline{\mathfrak{C}}$ is defined as $\overline{\mathfrak{C}} := \bigcap\{\overline{\mathfrak{C}}^\mu \mid \mu \in \mathfrak{S}\,(X, \mathfrak{C})\}$. A measurable space is called *complete* iff it coincides with its completion.

V^w denotes for a set V the set of all finite words with letters from V including the empty string ϵ. Let $\{A_s \mid s \in \mathbb{N}^w\}$ be a collection of subsets of a set X indexed by all finite sequences of natural numbers, then the *Souslin operation* $\mathcal{A}$ on this collection is defined as $\mathcal{A}\big(\{A_s \mid s \in \mathbb{N}^w\}\big) := \bigcup_{\alpha \in \mathbb{N}^\mathbb{N}} \bigcap_{n \in \mathbb{N}} A_{\alpha|n}$, where $\alpha|n$ are just the first n letters of the sequence α. This operation is intimately connected with the theory of analytic sets [12].

Proposition 1. *The completion* $\overline{\mathfrak{B}}^\mu$ *of* σ*-algebra* $\mathfrak{B}$ *is closed under the Souslin operation* $\mathcal{A}$ *whenever* $\mu \in \mathfrak{S}\,(X, \mathfrak{B})$*. Thus a complete measurable space is closed under this operation.* $\dashv$

3 The Logic

The logic under consideration is a modal logic, the modal operators of which are given through a set of programs. The programs in turn are composed from a set of basic statements, which cannot be decomposed further. This section defines programs and the logic. We also give here a fragment of the logic which models straight line programs, and we define for comparison, reference and motivation a classical variant of the logic which is to be interpreted by set theoretic relations.

Programs. Let $\mathbb{U}$ be a set of ur-programs, i.e., of programs that cannot be decomposed further. This set is fixed. Given $\mathbb{U}$, we define the set $\mathbb{P}$ of programs through this grammar $\pi ::= \upsilon \mid \pi_1; \pi_2 \mid \pi_1 \cup \pi_2 \mid \pi_1^*$ with $\upsilon \in \mathbb{U}$ an ur-program. Thus a program is an ur-program, the *sequential composition* $\pi_1; \pi_2$ which executes first π_1 and then π_2, the *nondeterministic choice* $\pi_1 \cup \pi_2$, which selects nondeterministically among π_1 and π_2 which one to execute, or the (indefinite) *iteration* π^* of program π_1 which executes π_1 a finite number of times, possibly not at all.

sPDL *and the Fragment* FRAG. A formula in sPDL is given through this grammar

$$\varphi ::= \top \mid p \mid \varphi_1 \wedge \varphi_2 \mid [\pi]_q \varphi.$$

Here p is an atomic proposition, taken from a fixed set $\mathbb{A}$ of atomic propositions, $\pi \in \mathbb{P}$ is a program, and $q \in \mathbb{Q} \cap [0, 1]$ is a rational number from the interval $[0, 1]$. The informal meaning of formula $[\pi]_q \varphi$ being true in state s is that after executing program π in state s, the probability of reaching a state in which formula φ holds is not greater than q. Denote by $\mathbf{F}_{\mathsf{sPDL}}$ the set of all formulas of sPDL.

We will also consider the fragment FRAG of sPDL in which the set of programs is restricted to members of $\mathbb{U}^w$, hence to sequential compositions of ur-programs. Denote the set of formulas of FRAG by $\mathbf{F}_{\mathsf{FRAG}}$.

These logics do have only the bare minimum of logical operators, negation and disjunction are missing. It will turn out that for discussing expressivity of these logics, negation or disjunction need not be present, but — strange enough — conjunction must be there. The technical reasons for preferring conjunction over disjunction will be discussed after stating Proposition 9.

The second remark addresses the intended meaning of the modal operator $[\pi]_q$ which specifies a probability **at most** q. Usually a modal operator of the form $\langle a \rangle_q$ is defined with the intended meaning that $\langle a \rangle_q \varphi$ holds in state s iff after executing action a in state s the system will be brought into a state in which φ holds with probability **at least** q, see [3,4]. Since we will recursively collect probabilities along different paths, it is intuitively more satisfying to argue with an upper bound than with a lower bound.

Vanilla PDL. Propositional dynamic logic PDL is defined in modal logic through this grammar

$$\varphi ::= \top \mid p \mid \neg\varphi_1 \mid \varphi_1 \wedge \varphi_2 \mid \varphi_1 \vee \varphi_2 \mid \langle \pi \rangle \varphi.$$

with $p \in \mathbb{A}$ and $\pi \in \mathbb{P}$. The set $\mathbb{P}$ of programs is unchanged [1, Example 1.15]. Note that we admit as operations disjunction as well as negation to PDL, and that the modal operator $\langle \pi \rangle$ associated with program π is not decorated with a numeric argument. The intuitive interpretation of $\langle \pi \rangle \varphi$ holding in a state s is that upon executing program π in state s a state can be reached in which formula φ holds.

We remind the reader of this logic, because we will draw some motivation from its interpretation when defining the interpretation of sPDL.

4 Interpretation of sPDL

sPDL will be interpreted through Kripke pre-models. Whereas a Kripke model assigns to each modal operator a relation through which the operator is to be interpreted, or, in coalgebraic logics, a predicate lifting, we cannot do that in the present context. This is due to the fact that the programs, i.e., the modal operators, have a specific structure which is not matched by the stochastic relations. This is possible, however, for Kripke models of the fragment FRAG, and these models will be used for generating the interpretation of the full logic. The correspondence of Kripke pre-models for sPDL and Kripke models for FRAG will be observed further when investigating expressivity in Section 6.

Interpreting PDL. We recall the interpretation of Vanilla PDL first (see [1, Example 1.26]). Let $\mathcal{R} := \big(S, \{R_\pi \mid \pi \in \mathbb{P}\}, (V_p)_{p\in\mathbb{A}}\big)$ be a Kripke model, i.e., S is a set of possible worlds, R_π is for each program $\pi \in \mathbb{P}$ a relation on S, and V_p is for each atomic proposition $p \in \mathbb{A}$ a subset of S. The family of relations satisfies these conditions: $R_{\pi_1;\pi_2} = R_{\pi_1} \circ R_{\pi_2}, R_{\pi_1\cup\pi_2} = R_{\pi_1} \cup R_{\pi_2}, R_{\pi^*} = \bigcup_{n\in\mathbb{N}_0} R_{\pi^n}$ with $A \circ B$ as the relational composition of relations A and B, and $R_{\pi^0} = \varepsilon$, the identity relation. Thus the relational structures on R match the algebraic structure on the set of all programs: sequential composition of programs corresponds to relational composition, and nondeterministic choice of programs corresponds to the union of the corresponding relations.

The interpretation of $\top$ and the Boolean operations in PDL is straightforward, and we put $\mathcal{R}, s \models p \Leftrightarrow s \in V_p; \mathcal{R}, s \models \langle\pi\rangle\varphi \Leftrightarrow \exists t \in S : \langle s, t\rangle \in R_\pi$ and finally $\mathcal{R}, t \models \varphi$. Hence we find, e.g., $\mathcal{R}, s \models \langle\pi^*\rangle\varphi \Leftrightarrow \exists n \in \mathbb{N}_0 : \mathcal{R}, s \models \langle\pi^n\rangle\varphi$.

Stochastic Kripke Pre-Models. A *stochastic Kripke pre-model* $\mathcal{K} = \big((S, \mathfrak{B}), (K_\pi)_{\pi\in\mathbb{U}}, (V_p)_{p\in\mathbb{A}}\big)$ for the logic sPDL is a measurable space $(S, \mathfrak{B})$, the space of states, a family of stochastic relations $K_\pi : (S, \mathfrak{B}) \rightsquigarrow (S, \mathfrak{B})$, the transition law, indexed by the ur-programs, and a family of measurable sets $V_p \in \mathfrak{B}$, indexed by the atomic propositions. $K_\pi(s)(D)$ is the probability for the new state to be an element of D after executing program π in state s.

We assume that we have transition laws for the programs' building blocks only, and not, as in the case of relational Kripke models, for each program, thus we use the term "pre-model" rather than "model". Whereas the algebraic structure of $\mathbb{P}$ can be modelled in the set of relations, this is not the case for stochastic relations: there is a natural composition operator, given by the convolution, but there does not seem to be an intuitively satisfying way of modelling the non-deterministic choice or the indefinite iteration.

All the same, the relational approach will be used as a source for guidelines for the probabilistic approach. Fix a stochastic Kripke pre-model $\mathcal{K}$. Let $A \in \mathfrak{B}$ be a measurable set; we define recursively a set-valued map $\mathcal{I}_q^A$ from the set of programs to the subsets of S, indexed by the rationals on the interval $[0, 1]$. For the time being, the informal interpretation of $\mathcal{I}_q^A(\pi)$ is the characterization of all those states which through executing program π bring the system into a state in A with probability less than q.

As auxiliary sets we define for $q \in \mathbb{Q} \cap [0, 1]$ and $n \in \mathbb{N}$

$$Q^{(n)}(q) := \{a \in (\mathbb{Q} \cap [0, 1])^n \mid a_1 + \cdots + a_n \leq q\},$$
$$Q^{(\infty)}(q) := \{a \in (\mathbb{Q} \cap [0, 1])^\infty \mid a_0 + a_1 + \cdots \leq q\}.$$

Put $K_\varepsilon := I_S$ for simplicity and for uniformity: the empty program does not do anything.

a. If $\pi \in \mathbb{U} \cup \{\varepsilon\}$, then $\mathcal{I}_q^A(\pi) := \{s \in S \mid K_\pi(s)(A) < q\}$.This yields all states for which the execution of program π leads to A with probability less that q.
b. Let $s \in \mathcal{I}_q^A(\pi_1; \ldots; \pi_n)$ iff $\big(K_{\pi_n} * K_{\pi_{n-1}} * \ldots * K_{\pi_1}\big)(s)(A) < q$, for $\pi_1, \pi_2, \ldots, \pi_n \in \mathbb{U}$, thus executing programs $\pi_1, \ldots, \pi_n$ sequentially is modelled through the convolution of the corresponding transition probabilities, and we determine all states for which the combined programs yield a probability less that q to be in set A.

Consider the case $n = 2$. By the definition of the convolution,

$$(K_{\pi_2} * K_{\pi_1})(s)(A) < q \text{ iff } \int_S K_{\pi_2}(t)(A)\ K_{\pi_1}(s)(dt) < q,$$

thus upon executing program π_1 in state s the system goes into an intermediate state t, and executing π_2 in this intermediate state the probability of entering A is determined. Since the transitions happen at random, averaging through the corresponding transition probability yields the desired probability, which is then tested against q.

c. Let $\pi_1, \pi_2 \in \mathbb{P}$, then $\mathcal{I}_q^A(\pi_1 \cup \pi_2) := \bigcup\{(\mathcal{I}_{a_1}^A(\pi_1) \cap \mathcal{I}_{a_2}^A(\pi_2)) \mid a \in Q^{(2)}(q)\}$. Selecting nondeterministically one of the programs π_1 or π_2, $\mathcal{I}_{a_1}^A(\pi_1)$ accounts for all states which are lead by executing π_1 to a state in the set A with probability at most a_1, similarly, $\mathcal{I}_{a_2}^A(\pi_2)$ for π_2. Since we want to bound the probability from above by q, we require $a_1 + a_2 \leq q$.
d. Let $\pi_1, \pi_2, \pi_3 \in \mathbb{P}$, then $\mathcal{I}_q^A(\pi_1; (\pi_2 \cup \pi_3)) := \mathcal{I}_q^A(\pi_1; \pi_2 \cup \pi_1; \pi_3)$, similarly, $\mathcal{I}_q^A((\pi_1 \cup \pi_2); \pi_3)$ is defined. This corresponds to the distributive laws.
e. If $\pi_1, \pi_2 \in \mathbb{P}$, define $\mathcal{I}_q^A(\pi_1^*) := \bigcup\{\bigcap_{n\in\mathbb{N}_0} \mathcal{I}_{a_n}^A(\pi_1^n) \mid a \in Q^{(\infty)}(q)\}$, similarly, $\mathcal{I}_q^A(\pi_1; \pi_2^*)$ and $\mathcal{I}_q^A(\pi_1^*; \pi_2)$ are defined. If executing program π_1 exactly n times results in a member of A with probability not exceeding a_n, then executing π_1 a finite number of times (including not executing it at all) results in a member of A with probability at most $a_0 + a_1 + \ldots$, which should be bounded above by q for the resulting state to be a member of A with probability at least q.

The definition of $\mathcal{I}_q^A(\pi)$ shows that the elementary building blocks from which these sets are computed are the sets $\mathcal{I}_q^A(\pi_1; \ldots; \pi_n)$ for ur-programs $\pi_1, \ldots, \pi_n$. These building blocks are combined through elementary set operations, they in turn are determined by the stochastic relations which come with the Kripke pre-model, either directly ($n = 1$) or through convolutions ($n > 1$).

The intuition says that executing π^* is somewhat akin either doing nothing at all or to execute π followed by executing π^*, hence to executing $\varepsilon \cup \pi; \pi^*$.

Example 1. Let $\pi \in \mathbb{P}$, then

$$\begin{aligned}
\mathcal{I}_q^A(\varepsilon \cup \pi; \pi^*) &= \bigcup_{a\in Q^{(2)}(q)} \left(\mathcal{I}_{a_1}^A(\varepsilon) \cap \mathcal{I}_{a_2}^A(\pi; \pi^*)\right) \\
&= \bigcup_{a\in Q^{(2)}(q)} \Big(\mathcal{I}_{a_1}^A(\pi^0) \cap \bigcup_{b\in Q^{(\infty)}(a_2)} \bigcap_{n\in\mathbb{N}_0} \mathcal{I}_{b_n}^A(\pi^{n+1})\Big) \\
&= \bigcup_{c\in Q^{(\infty)}(q)} \bigcap_{n\in\mathbb{N}_0} \mathcal{I}_{c_n}^A(\pi^n) = \mathcal{I}_q^A(\pi^*).
\end{aligned}$$

$\dashv$

Define the *interpretation order* on the set of programs through $\pi_1 \sqsubseteq \pi_2$ iff $\mathcal{I}_q^A(\pi_1) \subseteq \mathcal{I}_q^A(\pi_2)$ for all q and all A, then $\sqsubseteq$ is reflexive and transitive. Put $\equiv := \sqsubseteq \cap \sqsupseteq$ as the associated equivalence relation, then $\pi_1 \equiv \pi_2$ iff π_1 and π_2 have the same interpretation. We will consider programs rather than classes for convenience.

Proposition 2. *The interpretation order has the following properties:*

a. $\pi_1 \sqsubseteq \pi_2$ *iff* $\pi_1 \cup \pi_2 \equiv \pi_2$*, and* $\pi_1 \sqsubseteq \pi_2$ *implies* $\pi; \pi_1 \sqsubseteq \pi; \pi_2$ *for all* π.
b. $\pi^* = \sup_{n \in \mathbb{N}_0} \pi^n$.
c. π^* *is the smallest fixed point of the monotone map* $\tau \mapsto \varepsilon \cup \pi; \tau$.

Proof. Property a. is trivial. For establishing property b. one proves first by induction that $\pi^n \sqsubseteq \pi^*$ for all $n \in \mathbb{N}_0$. Moreover, if $\pi^n \sqsubseteq \pi'$ for all $n \in \mathbb{N}_0$, then it is not difficult to see that $\pi^* \sqsubseteq \pi'$. Thus π^* is the smallest upper bound to $\{\pi^n \mid n \in \mathbb{N}_0\}$. Finally, the map $\tau \mapsto \varepsilon \cup \pi; \tau$ is monotone, and π^* is a fixed point of $\tau \mapsto \varepsilon \cup \pi; \tau$ by Example 1. If $\widetilde{\tau}$ is another fixed point of this map, then $\widetilde{\tau} \equiv \varepsilon \cup \pi^1 \cup \cdots \cup \pi^n \cup \pi^{n+1}; \widetilde{\tau}$, so that by part a. $\pi^n \sqsubseteq \widetilde{\tau}$ for all $n \in \mathbb{N}_0$. Thus π^* is in fact the smallest fixed point. $\dashv$

Consequently, the semantics for the iteration construct defined through $\mathcal{I}$ uses actually a fixed point with the order adapted to the programs' effects. This appears to be a sensible alternative to associating the set-theoretically smallest fixed point (in the fashion of the μ-calculus) to this construct, see the approach proposed in, e.g., [9].

Associate with each program $\pi \in \mathbb{P}$ a tree $\mathcal{T}(\pi)$. We will be using the following format for writing down a tree with root node labelled ν and at most countable offsprings $\sigma_0, \sigma_1, \ldots$ from left to right: $\lceil\, \nu \;\|\; \sigma_0 \mid \sigma_1 \mid \ldots \rfloor$.

A. If $\pi \in \mathbb{U}$, then $\mathcal{T}(\pi) := \pi$, thus ur-programs constitute the leaves of the tree.
B. If $\pi_1, \ldots, \pi_n \in \mathbb{U}$ $(n > 1)$, then $\mathcal{T}(\pi_1; \ldots; \pi_n) := \lceil\, \mathfrak{comp} \;\|\; \pi_1 \mid \cdots \mid \pi_n \rfloor$,thus the tree associated with a finite sequence of ur-programs has the composition symbol $\mathfrak{comp}$ as the label of the root, and the ur-programs as offsprings.
C. Define recursively $\mathcal{T}(\pi_1 \cup \pi_2) := \lceil\, \mathfrak{union} \;\|\; \mathcal{T}(\pi_1) \mid \mathcal{T}(\pi_2) \rfloor$, for the programs $\pi_1, \pi_2, \pi_3 \in \mathbb{P}$, similarly, $\mathcal{T}(\pi_1; (\pi_2 \cup \pi_2))$ and $\mathcal{T}((\pi_1 \cup \pi_2); \pi_3)$ are defined. The root has the label $\mathfrak{union}$, the left and right offspring correspond to the operands.
D. Given the programs $\pi_1, \pi_2 \in \mathbb{P}$, define recursively $\mathcal{T}(\pi_1^*) := \lceil\, \mathfrak{star} \;\|\; \varepsilon \mid \mathcal{T}(\pi_1) \mid \cdots \mid \mathcal{T}(\pi_1^n) \mid \ldots \rfloor$, the trees $\mathcal{T}(\pi_1; \pi_2^*)$ and $\mathcal{T}(\pi_1^*; \pi_2)$ are defined similarly. Thus the root node is decorated with the annotation $\mathfrak{star}$, and it has countably many offsprings, each corresponding to executing the program exactly n times. Note that ε serves to indicate that the program is not executed at all.

Fig. 1 gives an impression what the partially expanded tree $\mathcal{T}((\pi_0 \cup \pi_1; \pi_2^*)^*)$ looks like.

Lemma 2. *The tree* $\mathcal{T}(\pi)$ *for each program* $\pi \in \mathbb{P}$ *is well-founded.* $\dashv$

The set operations for determining $\mathcal{I}_q^A$ are not always countable, so one might suspect that they go beyond what can be represented through a σ-algebra. If the underlying state space is a measurable space which is closed under the Souslin operation, however, we can establish that we stay within the realm of measurable sets. The proof of the following Proposition proceeds by induction on the program tree $\mathcal{T}(\pi)$.

Proposition 3. *Let* $(S, \mathfrak{B})$ *be a measurable space which is closed under the Souslin operation* $\mathcal{A}$*, then* $\mathcal{I}_q^A(\pi) \in \mathfrak{B}$ *for each program* $\pi \in \mathbb{P}$*, provided* $A \in \mathfrak{B}$. $\dashv$

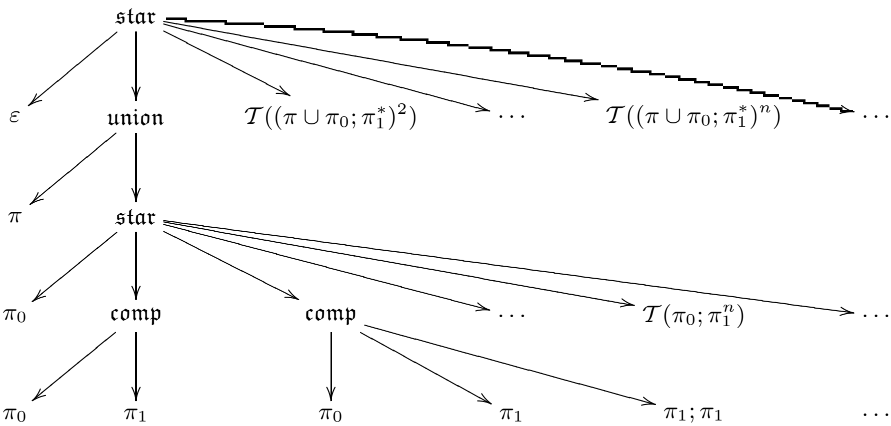


Fig. 1. Partially expanded tree for $(\pi_0 \cup \pi_1; \pi_2^*)^*$

The Semantics for sPDL. The semantics for the formulas is defined now. For this, define the *extent* $[\![\varphi]\!]_{\mathcal{K}}$ *of formula* φ as the set of all worlds $\{s \in S \mid \mathcal{K}, s \models \varphi\}$ of the Kripke pre-model $\mathcal{K}$ in which formula φ is true. Define the semantics of formula φ through $\mathcal{K}, s \models \varphi \Leftrightarrow s \in [\![\varphi]\!]_{\mathcal{K}}$,where $[\![\varphi]\!]_{\mathcal{K}}$ is defined recursively in this way.

$$\begin{aligned}
[\![\top]\!]_{\mathcal{K}} &:= S, \\
[\![p]\!]_{\mathcal{K}} &:= V_p, \text{ if } p \in \mathbb{A}, \\
[\![\varphi_1 \wedge \varphi_2]\!]_{\mathcal{K}} &:= [\![\varphi_1]\!]_{\mathcal{K}} \cap [\![\varphi_2]\!]_{\mathcal{K}}, \\
[\![[\pi]_q \varphi]\!]_{\mathcal{K}} &:= \mathcal{I}_q^{[\![\varphi]\!]_{\mathcal{K}}}(\pi), \text{ if } \pi \in \mathbb{P}, q \in \mathbb{Q} \cap [0,1].
\end{aligned}$$

Concerning the structure of these sets, we state

Proposition 4. *If the state space* $(S, \mathfrak{B})$ *of the Kripke pre-model* $\mathcal{K}$ *is closed under operation* $\mathcal{A}$*, then* $[\![\varphi]\!]_{\mathcal{K}} \in \mathfrak{B}$ *for all formulas* φ*.* $\dashv$

Thus if the state space $(S, \mathfrak{B})$ of the Kripke pre-model $\mathcal{K}$ equals the completion $\overline{\mathfrak{C}}^{\mu}$ for some $\mu \in \mathfrak{S}(S, \mathfrak{C})$, or if it equals the universal completion $\overline{\mathfrak{C}}$ for some σ-algebra $\mathfrak{C}$, then $[\![\varphi]\!]_{\mathcal{K}} \in \mathfrak{B}$ for all formulas φ, see Proposition 1. If, however, the state space is not closed under operation $\mathcal{A}$, then for some formulas φ the set $[\![\varphi]\!]_{\mathcal{K}}$ may not be representable as an event.

The Semantics of Fragment FRAG. We cannot associate with each program a stochastic relation for sPDL, but we are able to do this for fragment FRAG. A *Kripke model* $\mathcal{M} = \big((S, \mathfrak{B}), (M_\pi)_{\pi \in \mathbb{U}^w}, (V_p)_{p \in \mathbb{A}}\big)$ for the fragment FRAG of sPDL is defined just as a pre-models for sPDL, with the exception that $M_\pi : (S, \mathfrak{B}) \rightsquigarrow (S, \mathfrak{B})$ is a stochastic relation for each $\pi \in \mathbb{U}^w$, hence for each finite sequence of ur-programs.

We define the semantics of formulas in FRAG in an obvious way, the interesting case is the treatment of the modal operator$[\pi]_q$: $[\![[\pi]_q \varphi]\!]_{\mathcal{M}} := \{s \in S \mid M_\pi(s)\big([\![\varphi]\!]_{\mathcal{M}}\big) < q\}$. Structural induction establishes measurability of the formulas' extension.

Proposition 5. *Let $\mathcal{M} = \big((S, \mathfrak{B}), (M_\pi)_{\pi\in\mathbb{U}^w}, (V_p)_{p\in\mathbb{A}}\big)$ be a Kripke model for* FRAG*, then $[\![\varphi]\!]_\mathcal{M} \in \mathfrak{B}$ for all formulas φ of* FRAG. $\dashv$

Measurability of the sets $[\![\varphi]\!]_\mathcal{M}$ is easier to obtain for models than for pre-models, because the structure of the formulas is considerably simpler, in particular we may restrict our attention to programs that have a predetermined number of components, in contrast to iterations which are finite but of indefinite length.

Given a stochastic Kripke pre-model $\mathcal{K}$ over the state space $(S, \mathfrak{B})$ with stochastic relations $(K_\pi)_{\pi\in\mathbb{U}}$, we construct a stochastic Kripke model $\mathcal{K}^\dagger$ for the fragment FRAG with stochastic relations $(M_\pi)_{\pi\in\mathbb{U}^w}$ such that $M_{\pi_1;\ldots;\pi_n} := K_{\pi_n} * K_{\pi_{n-1}} * \ldots * K_{\pi_1}$, all other components are inherited from $\mathcal{K}$. Conversely, a stochastic Kripke model $\mathcal{M}$ for FRAG with relations $(M_\pi)_{\pi\in\mathbb{U}^w}$ yields a pre-model $\mathcal{M}_\ddagger$ for sPDL with stochastic relations $(K_\pi)_{\pi\in\mathbb{U}}$ upon setting $K_\pi := M_\pi$ for the ur-program $\pi \in \mathbb{U}$, again all other components remaining the same. Then $\big(\mathcal{K}^\dagger\big)_\ddagger = \mathcal{K}$ for each pre-model $\mathcal{K}$. Structural induction on the formula shows that

Lemma 3. *Let $\mathcal{K}$ be a stochastic Kripke pre-model, then $[\![\varphi]\!]_\mathcal{K} = [\![\varphi]\!]_{\mathcal{K}^\dagger}$ holds for each formula $\varphi \in \mathbf{F}_{\mathsf{FRAG}}$.* $\dashv$

The interplay of pre-model $\mathcal{K}$ with model $\mathcal{K}^\dagger$, and of model $\mathcal{M}$ with the pre-model $\mathcal{M}_\ddagger$ will be of considerable interest when investigating expressivity in Section 6.

5 Morphisms

Morphisms are used to relate pre-models to each other. A morphism preserves the structure of an interpretation: the states in which atomic propositions are true are related to each other, and the probabilistic structure, i. e., the transition laws, are compared against each other.

Let $\mathcal{K} = \big((S, \mathfrak{B}), (K_\pi)_{\pi\in\mathbb{U}}, (V_p)_{p\in\mathbb{A}}\big)$ and $\mathcal{L} = \big((T, \mathcal{C}), (L_\pi)_{\pi\in\mathbb{U}}, (W_p)_{p\in\mathbb{A}}\big)$ be stochastic Kripke pre-models. A $\mathfrak{C}$-$\mathfrak{D}$-measurable map $f : S \to T$ is called a *pre-model morphism* $f : \mathcal{K} \to \mathcal{L}$ iff $f^{-1}[W_p] = V_p$ for each atomic proposition $p \in \mathbb{A}$, and $K_\pi^f = L_\pi \circ f$ for each ur-program $\pi \in \mathbb{U}$. The first condition states that $s \in V_p$ iff $f(s) \in W_p$ for each atomic proposition. The second condition states that $L_\pi\big(f(s)\big)(D) = K_\pi(s)\big(f^{-1}[D]\big)$ for each $s \in S$, each measurable set $D \in \mathfrak{D}$, and each ur-program $\pi \in \mathbb{U}$. Thus the probability in $\mathcal{L}$ of bringing state $f(s)$ into D through executing ur-program π is the same as executing π in state s and ending up in $f^{-1}[D]$ in $\mathcal{K}$.

This property is preserved through sequential program composition.

Proposition 6. *Let $\mathcal{K}$ and $\mathcal{L}$ be stochastic Kripke pre-models as above, and $f : \mathcal{K} \to \mathcal{L}$ be a morphism. Then we have for each $s \in S, D \in \mathfrak{D}$ and ur-programs $\pi_1, \ldots, \pi_n \in \mathbb{U}$* $\big(L_{\pi_1} * \ldots * L_{\pi_n}\big)(f(s))(D) = (K_{\pi_1} * \ldots * K_{\pi_n})(s)\big(f^{-1}[D]\big)$. $\dashv$

Comparing Morphisms. Dealing with Kripke models $\mathcal{M}$ and $\mathcal{N}$, the definition of a *model morphism* $f : \mathcal{M} \to \mathcal{N}$ remains essentially the same. To be specific, if the stochastic relations $(M_\pi)_{\pi\in\mathbb{U}^w}$ and $(N_\pi)_{\pi\in\mathbb{U}^w}$ govern the probabilistic behavior, then the first condition on atomic propositions remains as it is, and the second one is replaced by $M_\pi^f = N_\pi \circ f$ for each finite sequence $\pi \in \mathbb{U}^w$ of ur-programs. We obtain as an immediate consequence.

Corollary 1. *Consider for a measurable map $f : S \to T$ these statements, where $\mathcal{K}$ and $\mathcal{L}$ are pre-models, and $\mathcal{M}$ and $\mathcal{N}$ are models with state spaces S resp. T. Then $f : \mathcal{K} \to \mathcal{L}$ is a pre-model morphism iff $f : \mathcal{K}^\dagger \to \mathcal{L}^\dagger$ is a model morphism, and if $f : \mathcal{M} \to \mathcal{N}$ is a model morphism, then $f : \mathcal{M}_\ddagger \to \mathcal{N}_\ddagger$ is a pre-model morphism.* ⊣

Morphisms preserve and reflect the validity of formulas both in sPDL and in the fragment FRAG.

Proposition 7. *Let $f : \mathcal{K} \to \mathcal{L}$ be a pre-model morphism, then $\mathcal{K}, s \models \varphi \Leftrightarrow \mathcal{L}, f(s) \models \varphi$ for each state s in $\mathcal{K}$ and each formula $\varphi \in \mathbf{F}_{\mathsf{sPDL}}$. If $g : \mathcal{M} \to \mathcal{N}$ a model morphism, then $\mathcal{M}, s \models \varphi \Leftrightarrow \mathcal{N}, g(s) \models \varphi$ for each state s in $\mathcal{M}$ and each formula $\varphi \in \mathbf{F}_{\mathsf{FRAG}}$.* ⊣

The Test Operator. The usual definition of PDL includes a test operator. With each formula φ a program $\varphi?$ is associated which tests whether φ holds. Intuitively, if the test succeeds, then the program continues. Given a Kripke model $\mathcal{R}$ with S as the set of possible worlds, the set of relations for the interpretation of the logic is extended by the relation $R_{\varphi?} := \{\langle s, s\rangle \mid s \in S, \mathcal{R}, s \models \varphi\}$, see [1, p. 23].

The test operator is integrated into sPDL in this way. Given a Kripke pre-model $\mathcal{K}$ over state space $(S, \mathfrak{B})$ and a formula φ, we define $K_{\varphi?}(s)(B) := I_S(s)(B \cap [\![\varphi]\!]_{\mathcal{K}})$, and $K_{\overline{\varphi}?}(s)(B) := I_S(s)\big(B \cap (S \setminus [\![\varphi]\!]_{\mathcal{K}})\big)$ for $B \in \mathfrak{B}$. $K_{\varphi?}$ is the relation associated with testing for φ, and $K_{\overline{\varphi}?}$ tests whether φ does not hold — recall that the logic is negation free.

Lemma 4. *Let $\mathcal{K}$ be a Kripke pre-model over state space $(S, \mathfrak{B})$ which is closed under the Souslin operation $\mathcal{A}$. Then both $K_{\varphi?}$ and $K_{\overline{\varphi}?}$ are a stochastic relations for each formula φ.* ⊣

Thus, if the set $[\![\varphi]\!]_{\mathcal{K}}$ is not measurable, the test operator associated with φ is not representable within the framework of the Kripke model; consequently, we may not be able to test *within this model*, e.g., whether an iteration has terminated. This emphasizes again the importance of operation $\mathcal{A}$.

Example 2. Let $p \in \mathbb{A}$ be an atomic proposition, $\pi \in \mathbb{U}$ an ur-program and φ a formula. Then $[\![[p?;\pi]_q\varphi]\!]_{\mathcal{K}} = \{s \in S \mid K_\pi(s)(V_p \cap [\![\varphi]\!]_{\mathcal{K}}) < q\}$, and $[\![[\pi;\overline{p}?]_q\varphi]\!]_{\mathcal{K}} = \big(S \setminus V_p \cap [\![[\pi]_q\varphi]\!]_{\mathcal{K}}\big) \cup V_p$. ⊣

Tests are compatible with morphisms.

Proposition 8. *Let $f : \mathcal{K} \to \mathcal{L}$ be a morphism between pre-models. Then $K^f_{\varphi?} = L_{\varphi?} \circ f$ for every formula $\varphi \in \mathbf{F}_{\mathsf{sPDL}}$.* ⊣

This is the reason why we are not interested in this operator when looking into the expressivity of Kripke pre-models is that these operators do not contribute to the relevant properties of this problem.

6 Expressivity

We are able to compare the behavior of two (pre-) models once we know that there exists a morphism between then. This leads to the notion of behavioral equivalence: there exists a reference system onto which the pre-models may be mapped. On the other hand, we can compare pre-models through their theories: two pre-models are logically equivalent iff for each state in one model there exists a state in the other model in which renders exactly the same formulas are true. It will be shown that these notions of expressivity are equivalent, and it is easy to see that behaviorally equivalent models are logically equivalent. The proof for the other direction proceeds technically as follows: we show that logically equivalent pre-models give rise to logically equivalent models, and we know for models that logical and behavioral equivalence coincide. This result is then carried over to the world of pre-models. There is a small catch, though: the equivalence mentioned holds for models of another, albeit closely related, logic. This has to be taken care of.

Define for a state s in pre-model $\mathcal{K}$ the sPDL-*theory* $Th_{\mathcal{K}}(s)$ *of* s as the set of all formulas in $\mathbf{F}_{\mathsf{sPDL}}$ which are true in s, formally $Th_{\mathcal{K}}(s) := \{\varphi \in \mathbf{F}_{\mathsf{sPDL}} \mid \mathcal{K}, s \models \varphi\}$. Similarly, the FRAG-theory $Th_{\mathcal{M}}(s)$ of state s in model $\mathcal{M}$ is defined as all formulas in the fragment FRAG which are true in s; we use the same notation for both logics, trusting that no confusion arises. These sets are closely related. Just for the record:

Lemma 5. *Let $\mathcal{K}$ be a pre-model, then $Th_{\mathcal{K}^\dagger}(s) = Th_{\mathcal{K}}(s) \cap \mathbf{F}_{\mathsf{FRAG}}$ for each state s in $\mathcal{K}$.* $\dashv$

Logical vs. Behavioral Equivalence. The pre-models $\mathcal{K}$ and $\mathcal{L}$ are called *logically equivalent* iff given a state in $\mathcal{K}$, there exists a state in $\mathcal{L}$ which has the same theory, and vice versa. This translates to $\{Th_{\mathcal{K}}(s) \mid s \text{ is a state in } \mathcal{K}\} = \{Th_{\mathcal{L}}(t) \mid t \text{ is a state in } \mathcal{L}\}$. Logical equivalence for models is defined in the same way. We infer from Lemma 5.

Corollary 2. *If the pre-models $\mathcal{K}$ and $\mathcal{L}$ are logically equivalent, so are the models $\mathcal{K}^\dagger$ and $\mathcal{L}^\dagger$.* $\dashv$

The pre-models $\mathcal{K}$ and $\mathcal{L}$ are called *behaviorally equivalent* iff there exists a model $\mathcal{Q}$ and surjective pre-model morphisms $\mathcal{K} \xrightarrow{f} \mathcal{Q} \xleftarrow{g} \mathcal{L}$. In fact, let s be a state in $\mathcal{K}$ and φ be a formula, then there exists a state t in $\mathcal{L}$ such that $f(s) = g(t)$, consequently by Proposition 7 $\mathcal{K}, s \models \varphi \Leftrightarrow \mathcal{Q}, f(s) \models \varphi \Leftrightarrow \mathcal{Q}, g(t) \models \varphi \Leftrightarrow \mathcal{L}, t \models \varphi$.

This accounts for the name, and the argumentation shows

Lemma 6. *Behaviorally equivalent pre-models are logically equivalent.* $\dashv$

Behavioral equivalence is defined for Kripke models in the same way through the existence of a pair of morphisms with the same target.

Lemma 7. *Let $\mathcal{K}$ and $\mathcal{L}$ be behaviorally equivalent pre-models, then $\mathcal{K}^\dagger$ and $\mathcal{L}^\dagger$ are behaviorally equivalent models.* $\dashv$

Logical equivalence and behavioral equivalence are closely related for Kripke models, as the following proposition shows. We have to show that we can find a pre-model for logically equivalent models which serves as a target for morphisms which are defined on the given models. This construction is technically somewhat involved, but we can fortunately draw on the analogous results for another, closely related logic; this result is massaged into being suitable for the present scenario.

Proposition 9. *If the Kripke models $\mathcal{M}$ and $\mathcal{N}$ are logically equivalent, then they are behaviorally equivalent as well.*

Proof. 1. $\mathcal{M} = \big((S, \mathfrak{B}), (M_\pi)_{\pi \in \mathbb{U}^w}, (V_p)_{p \in \mathbb{A}}\big)$ and $\mathcal{N} = \big((S, \mathfrak{C}), (N_\pi)_{\pi \in \mathbb{U}^w}, (W_p)_{p \in \mathbb{A}}\big)$ are the Kripke models under consideration. Construct as in [4, Section 2.6] a measurable space $(H, \mathfrak{E})$ and surjective maps $f : S \to H$ and $g : T \to H$ together with stochastic relations $H_\pi : (H, \mathfrak{E}) \rightsquigarrow (H, \mathfrak{E})$ for $\pi \in \mathbb{U}^w$ with these properties:

i. f is $\mathfrak{B}$-$\mathfrak{E}$-measurable, g is $\mathfrak{C}$-$\mathfrak{E}$-measurable,
ii. $f\big[[\![\varphi]\!]_\mathcal{M}\big] = g\big[[\![\varphi]\!]_\mathcal{N}\big]$ for all formulas $\varphi \in \mathbf{F}_{\mathsf{sPDL}}$,
iii. $[\![\varphi]\!]_\mathcal{M}$ is f-invariant, $[\![\varphi]\!]_\mathcal{N}$ is g-invariant for all formulas $\varphi \in \mathbf{F}_{\mathsf{sPDL}}$ (thus, e.g., $s \in [\![\varphi]\!]_\mathcal{M}$ and $f(s) = f(s')$ implies $s' \in [\![\varphi]\!]_\mathcal{M}$),
iv. $H_\pi^f = M_\pi \circ f$ and $H_\pi^g = N_\pi \circ g$ for all $\pi \in \mathbb{U}^w$,
v. $\mathfrak{E} = \sigma\big(\{f\big[[\![\varphi]\!]_\mathcal{M}\big] \mid \varphi \in \mathbf{F}_{\mathsf{sPDL}}\}\big) = \sigma\big(\{g\big[[\![\varphi]\!]_\mathcal{N}\big] \mid \varphi \in \mathbf{F}_{\mathsf{sPDL}}\}\big)$.

The actual construction in [4, Section 2.6] is carried out, however, for a logic given by the grammar

$$\psi ::= \top \mid \psi_1 \wedge \psi_2 \mid \langle a \rangle_q \psi,$$

where a is taken from an arbitrary non-empty set of actions, $q \in \mathbb{Q} \cap [0, 1]$, and the interpretation of formula $\langle a \rangle_q \psi$ is that $\langle a \rangle_q \psi$ is true in state s iff action a in s leads the model to a state in which ψ holds **with probability at least** q. A careful analysis of the proofs in [4, Section 2.6] (in particular of, resp., Proposition 2.6.8 and Lemma 2.6.15), however, shows that the latter condition may be replaced everywhere by the definition of validity for $[\pi]_q \varphi$ proposed in the present paper, without changing the proofs' substance.

2. Define $X_p := f\big[V_p\big]\big(= g\big[W_p\big]\big)$ for the atomic proposition $p \in \mathbb{A}$, then $X_p \in \mathfrak{E}$. Due to the invariance property of f resp., g, we conclude that both $f^{-1}\left[X_p\right] = V_p, g^{-1}\left[X_p\right] = W_p$. hold. In fact, if $f(s) \in X_p = f\big[V_p\big]$, there exists $s' \in V_p$ with $f(s) = f(s')$. Since V_p is f-invariant, we conclude $s \in V_p$, so that $f^{-1}\left[X_p\right] \subseteq V_p$. The inclusion $f^{-1}\left[X_p\right] \supseteq V_p$ is obvious. 3. Consequently, $\mathcal{H} := \big((H, \mathfrak{E}), (H_\pi)_{\pi \in \mathbb{U}^w}, (X_p)_{p \in \mathbb{A}}\big)$ is a Kripke model with morphisms $\mathcal{M} \xrightarrow{f} \mathcal{H} \xleftarrow{g} \mathcal{N}$. Thus the logically equivalent Kripke models $\mathcal{M}$ and $\mathcal{N}$ are behaviorally equivalent. $\dashv$

Some remarks are in order. ❶ The proofs from [4, Section 2.6] are adapted, they require the logical equivalence of the underlying Kripke models. They factor the Kripke models according to the logic, saying that two states are equivalent iff they satisfy exactly the same formulas in the respective models. The factor spaces are related to each other by glueing the equivalence classes for s in $\mathcal{M}$ and for t in $\mathcal{N}$ iff $Th_\mathcal{M}(s) = Th_\mathcal{N}(t)$; logical equivalence renders this construction permissible. The combined space is essentially the state space for model $\mathcal{H}$. A suitable σ-algebra on $\mathcal{H}$ is constructed, and suitable stochastic relations are defined. For these constructions the underlying logic is required to be closed under conjunctions for measure theoretic reasons. ❷ The observation just made is the technical reason for requiring the logic being closed under conjunctions; the alternative of closing under disjunctions, however, is for measure theoretic purposes not equally attractive. ❸ Incidentally, since we work in a σ-algebra, we do not need negation, which tells us when a formula is *not* true. On the level of models we can state that formula φ does not hold in state s iff $s \in S \setminus [\![\varphi]\!]_{\mathcal{M}}$, which is a member of the σ-algebra over which we are working, hence a member to our universe of discourse, provided $[\![\varphi]\!]_{\mathcal{M}}$ is. ❹ An alternative to the proof presented would have been through a coalgebraic approach by defining for each modal operator $[\pi]_q$ a suitable predicate lifting, and by investigating the corresponding logic over a coalgebra, see [4, Section 4.4] or [5]. Balancing the — considerable — effort of doing so against modifying a construction which works with conventional methods resulted in the proof presented above.

We are ready to prove the first installment of our main result

Proposition 10. *Let $\mathcal{K}$ and $\mathcal{L}$ be Kripke pre-models, then $\mathcal{K}$ and $\mathcal{L}$ are behaviorally equivalent if and only if $\mathcal{K}$ and $\mathcal{L}$ are logically equivalent.* $\dashv$

It may be noted that the result above holds for any measurable space, independent of whether or not the validity sets of the formulas are measurable (which require, by Proposition 3, closedness under the Souslin operation $\mathcal{A}$).

In contrast, relating bisimilarity to the logical and behavioral equivalence makes fairly strong assumptions on the base space.

Bisimilarity. Call the stochastic Kripke pre-models $\mathcal{K}$ and $\mathcal{L}$ *bisimilar* iff there exists a pre-model $\mathcal{M}$ such that $\mathcal{K} \xleftarrow{f} \mathcal{M} \xrightarrow{g} \mathcal{L}$ for suitable pre-model morphisms f and g. Pre-model $\mathcal{M}$ is sometimes called *mediating*. Bisimilarity is a key concept in modal logics as well as in the theory of coalgebras. The reader is referred to [1] and to [10] for comprehensive discussions, where also the relationship to the coalgebraic aspects of Milner's original concept of bisimilar concurrent systems [7] is discussed. An immediate observation is (cp. Proposition 7 and Lemma 6).

Lemma 8. *Bisimilar pre-models are logically equivalent.* $\dashv$

We cannot show in general measurable spaces that logically equivalent models are bisimilar; for this, we have to specialize the base spaces in which we are working to Polish spaces. A topological space (X, τ) is called *Polish* iff it is second countable, and there exists a metric for τ which is complete, see [12]. The σ-algebra $\mathfrak{B}(X, \tau)$ of *Borel sets* for a topological space (X, τ) is the smallest σ-algebra on X which contains the open sets, hence $\mathfrak{B}(X) = \sigma(\tau)$; as usual, we omit the Borel sets, when we talk about a topological space. A map from a topological space (X, τ) to a measurable space $(S, \mathfrak{C})$ is called *Borel* iff it is $\mathfrak{B}(X)$-$\mathfrak{C}$-measurable.

For Polish spaces we can establish the equivalence of all three notions of expressivity.

Theorem 1. *Let $\mathcal{K}$ and $\mathcal{L}$ be Kripke pre-models, and consider these statements.*

a. $\mathcal{K}$ and $\mathcal{L}$ are logically equivalent.
b. $\mathcal{K}$ and $\mathcal{L}$ are behaviorally equivalent.
c. $\mathcal{K}$ and $\mathcal{L}$ are bisimilar.

Then (c) ⇒ (a) ⇔ (b), and if the state spaces of $\mathcal{K}$ and $\mathcal{L}$ are Polish spaces, then all three statements are equivalent.

Proof. 1. The implications (c) ⇒ (a) ⇔ (b) are Proposition 10 together with Lemma 8. Hence the implication b ⇒ c remains to be established.

2. Let $\mathcal{K} = (S, (K_\pi)_{\pi\in\mathbb{U}}, (V_p)_{p\in\mathbb{A}})$ and $\mathcal{L} = (T, (L_\pi)_{\pi\in\mathbb{U}}, (W_p)_{p\in\mathbb{A}})$ be behaviorally equivalent Kripke pre-models over the Polish spaces S and T. Then the Kripke models $\mathcal{K}^\dagger$ and $\mathcal{L}^\dagger$ are behaviorally equivalent models by Lemma 7, so there exists a model $\mathcal{M} = \big((Q,\mathcal{H}), (M_\pi)_{\pi\in\mathbb{U}^w}, (X_p)_{p\in\mathbb{A}}\big)$ and model morphisms $\mathcal{K}^\dagger \xleftarrow{f} \mathcal{M} \xrightarrow{g} \mathcal{L}^\dagger$. We infer from [5, Proposition 6.18] that we may assume Q to be a separable metric space with $\mathcal{H} = \mathfrak{B}(Q)$. Put $Y := \{\langle s,t\rangle \mid s \in S, t \in T, f(s) = g(t)\}$, and let $\beta : Y \to S, \gamma : Y \to T$ be the projections. Since f and g are surjective and Borel, β and γ are surjective and Borel. We infer from [6, Theorem 3.8] that we can find for each $\pi \in \mathbb{U}^w$ a stochastic relation $N_\pi : Y \rightsquigarrow Y$ such that $K_\pi^\dagger \circ \beta = N_\pi^\beta, L_\pi^\dagger \circ \beta = N_\pi^\gamma$. Define $Z_p := Y \cap V_p \times W_p$ for the atomic proposition $p \in \mathbb{A}$, then $\mathcal{N} := \big(Y, (N_\pi)_{\pi\in\mathbb{U}^w}, (Z_p)_{p\in\mathbb{A}}\big)$ is a Kripke model with $\mathcal{K}^\dagger \xleftarrow{\beta} \mathcal{N} \xrightarrow{\gamma} \mathcal{L}^\dagger$. Consequently, the pre-model $\mathcal{N}_\ddagger$ is a mediating pre-model for $\mathcal{K}$ and $\mathcal{L}$. ⊣

The crucial step in the proof is the existence of the stochastic relations N_π for each $\pi \in \mathbb{U}^w$. This actually requires some heavy machinery from measurable selection theory which is available in Polish spaces, but not in general measurable spaces.

Thus we have carried over the result of the equivalence of logical equivalence, bisimilarity and behavioral equivalence. It has been established for the logics mentioned in the proof of Proposition 9 (they are sometimes called *Hennessy-Milner logics*) for the case that the underlying space is Polish, and that there is a countable set of actions, see [4, Section 2.3]. The latter assumption is made in order to make sure that the factor spaces which are needed for the constructions are well-behaved. Note that by separating concerns we did not need an assumption on countability, and that bisimilarity only required the assumption on a Polish base space.

7 Conclusion

We propose a probabilistic interpretation for sPDL, the modal operators of which are given by $[\pi]_q$ for a program π with the intended meaning that $[\pi]_q\varphi$ holds in a state s if a terminating execution of program π in state s will reach a state in which formula φ holds has a probability not greater than q. This deviates slightly from the usual probabilistic interpretation of modal logics, see [3,4], because the proposed interpretation is more adequate for the present logic. It is shown that behavioral and logical equivalence are the same, they are equivalent to bisimilarity in the case of models based on Polish spaces.

The models we investigate do not show an interpretation for each modal operator. We have to generate the interpretation from the model for a fragment. The technique seems to be interesting in itself. The question arises whether it can be applied to dynamic coalgebraic logics, i.e., to dynamic logics in which the modal operators are given through predicate liftings. Coalgebraic logic has recently attracted some interest [11,4,5] as a unifying and powerful generalization of modal logics. Another promising line is the exploration of ideas pertaining to the probabilistic interpretations of games.

The interpretation proposed here shows some rough edges: First, the semantics of formulas $[\pi^*]_q\varphi$ is defined through an iterative fixed point rather than a smallest one, cp. Example 1 and Proposition 2. This keeps it in line with the interpretation of sPDL in modal logics [2, Sections 6.6, 6.7], but in contrast to a very similar approach in game logics [9], see also the discussion of belief structures in [8]. This topic will have to be investigated further. Second, we establish bisimulations only for Kripke models on Polish spaces; on the other hand, the extent of formula $[\pi^*]\varphi$ can only be shown to be measurable if the underlying measurable space is closed under the Souslin operation $\mathcal{A}$. But non-discrete Polish spaces are *never* closed under this operation. This follows from the observation that in such spaces there exist analytic sets which are not Borel measurable; the collection of analytic sets is closed under operation $\mathcal{A}$, however [12, Theorems 4.1.5, 4.1.13]. Third, we propose an interpretation of FRAG through a stochastic Kripke model, but then we continue in a nondeterministic fashion. We will investigate how this nondeterminism can be replaced by a purely stochastic approach.

Acknowledgements. The referees' remarks and suggestions are appreciated; the author is grateful for discussions with Ingo Battenfeld and Christoph Schubert.

References

1. Blackburn, P., de Rijke, M., Venema, Y.: Modal Logic. Cambridge Tracts in Theoretical Computer Science, vol. 53. Cambridge University Press, Cambridge (2001)
2. Blackburn, P., van Benthem, J.: Modal logic: A semantic perspective. In: Blackburn, P., et al. (eds.) Handbook of Modal Logic, pp. 1–84. Elsevier, Amsterdam (2007)
3. Desharnais, J., Edalat, A., Panangaden, P.: Bisimulation of labelled Markov-processes. Information and Computation 179(2), 163–193 (2002)
4. Doberkat, E.-E.: Stochastic Coalgebraic Logic. EATCS Monographs in Theoretical Computer Science. Springer, Heidelberg (2009)
5. Doberkat, E.-E., Schubert, C.: Coalgebraic logic over general measurable spaces - a survey. Math. Struct. Comp. Science (in print, 2011) (Special issue on coalgebraic logic)
6. Doberkat, E.-E., Srivastava, S.M.: Measurable selections, transition probabilities and Kripke models. Technical Report 185, Chair for Software Technology, Technische Universität Dortmund (May 2010)
7. Hennessy, M., Milner, R.: On observing nondeterminism and concurrency. In: de Bakker, J.W., van Leeuwen, J. (eds.) ICALP 1980. LNCS, vol. 85, pp. 395–409. Springer, Heidelberg (1980)
8. Moderer, D., Samet, D.: Approximating common knowledge with common belief. Games and Economic Behavior 1, 170–190 (1989)
9. Pauly, M., Parikh, R.: Game logic — an overview. Studia Logica 75, 165–182 (2003)
10. Rutten, J.J.M.M.: Universal coalgebra: a theory of systems. Theor. Comp. Sci. 249(1), 3–80 (2000) (Special issue on modern algebra and its applications)
11. Schröder, L., Pattinson, D.: Modular algorithms for heterogeneous modal logics. In: Arge, L., Cachin, C., Jurdziński, T., Tarlecki, A. (eds.) ICALP 2007. LNCS, vol. 4596, pp. 459–471. Springer, Heidelberg (2007)
12. Srivastava, S.M.: A Course on Borel Sets. Graduate Texts in Mathematics. Springer, Berlin (1998)

Contextual Coalitional Games

Patrick Doherty[1], Tomasz Michalak[3], Jacek Sroka[2], and Andrzej Szałas[1,2]

[1] Department of Computer and Information Science, Linköping University, SE-581 83 Linköping, Sweden
[2] Institute of Informatics, Warsaw University 02-097 Warsaw, Poland
[3] Electronics and Computer Science, University of Southampton SO17 7JB Southampton, UK

Abstract. The study of cooperation among agents is of central interest in multi-agent systems research. A popular way to model cooperation is through coalitional game theory. Much research in this area has had limited practical applicability as regards real-world multi-agent systems due to the fact that it assumes *deterministic* payoffs to coalitions and in addition does not apply to multi-agent environments that are *stochastic* in nature. In this paper, we propose a novel approach to modeling such scenarios where coalitional games will be contextualized through the use of logical expressions representing environmental and other state, and probability distributions will be placed on the space of contexts in order to model the stochastic nature of the scenarios. More formally, we present a formal representation language for representing contextualized coalitional games embedded in stochastic environments and we define and show how to compute *expected Shapley values* in such games in a computationally efficient manner. We present the value of the approach through an example involving robotics assistance in emergencies.

1 Introduction

The study of cooperation among agents is of central interest in multi-agent systems research. The reason for this is that more often than not, agents working together perform tasks more efficiently than agents that do not. Although our intuitions tell us this is so, formal models provide a basis for actually proving when and when not this is the case, in addition to providing a basis for efficient implementation of cooperative multi-agent systems.

A popular way to model cooperation is through coalitional games. The key questions in coalitional game theory are related to division of payoff from cooperation so that stability and/or fairness are achieved. Although many of these issues have already been extensively studied in the AI/MAS context [12], most of the research has had limited practical applicability as regards real-world multi-agent systems. The reasons for this, from the modeling point of view, are twofold. Firstly, most work to date assumes deterministic payoffs to coalitions, which is clearly not achievable in many multi-agent systems which are embedded in stochastic environments. Secondly, although some recent work in AI has proposed game models that account for uncertainty, these new developments are highly theoretical and do not take into account computational issues.

M. Banerjee and A. Seth (Eds.): ICLA 2011, LNAI 6521, pp. 65–78, 2011.

Specifically, from the computational point of view, the paramount question is how to *concisely* represent a coalitional game when the number of potential coalitions is exponential, precisely $2^n - 1$ where n is the number of agents. Important measures used to asses representations are: *expressiveness*, i.e., does it allow representation of a broad class of games, and *efficiency*, i.e., does it allow for efficient computing of solutions to games from a considered class.

As a motivating example of a real-world multi-agent system, consider a scenario in the emergency services application domain where, as support for rescue missions, one wants to hire configurations of autonomous ground robots (UGVs) and Unmanned Aerial Vehicles (UAVs) from a number of suppliers. Each of the robots, may take on different roles, based on particular sensor capability. In addition, operational efficiency may be affected by particular environmental characteristics which in turn influence the payoffs to coalitions in a contextual manner. Since future environmental characteristics are unknown, the contexts are stochastic in nature.

For instance, in cases where there is wind and rain in the catastrophe areas, it may be the case that only one type of UAV can be used. When the wind is very strong such as during a typhoon, UAVs are useless and one has to depend more on the use of UGVs or other types of vehicles. Use of different configurations of UGVs and UAVs contribute to different hiring costs and differences in resulting quality of usage. One also has to pay a certain fixed fee for keeping equipment ready for immediate use. In cases such as this, one of the main issues is how to distribute the total budget among different suppliers, when a long term contract is being negotiated and many such missions are expected to be carried out under various circumstances. The key factors involved here are dynamic coalition formation, dynamic contexts in which coalitions form, and the stochastic nature in which these contexts occur in the long run.

The research topic is to develop general and computationally efficient frameworks to be able to model such scenarios. Although such scenarios can be modeled to some extent by a number of existing theoretical frameworks that account for uncertainty [8], even for a relatively small number of agents and states in the environment, these models become impractical from a computational point of view and perhaps even a modeling point of view. Thus simplicity is one of our important goals.

Therefore, in this paper we propose a novel approach to modeling such scenarios where coalitional games will be contextualized through the use of logical expressions representing environmental and other states. Probability distributions are placed on the space of contexts in order to model the stochastic nature of the scenarios. More specifically:

- we define contextual coalitional games embedded in stochastic environments and show how to efficiently translate contextual coalitional games into linear combinations of traditional coalitional games;
- we propose a family of formalisms for representing contextualized coalitional games, where each specific formalism is obtained from the general pattern instantiated by fixing a specific representation of traditional coalitional games and a specific logic;
- we define and show how to compute *expected Shapley values* in such games in a computationally efficient manner; and

– we instantiate our general representation and exemplify its use by modeling the informal scenario involving UGVs and UAVs.

The paper is structured as follows. The first section contains notation and preliminary definitions. In the second section we introduce and discuss contextual coalitional games and the Shapley value for such games. In the third section we introduce and motivate our representational viewpoint as well as demonstrate its flexibility and conciseness. We also show that the new representation differs in complexity from conventional games by a factor of the number of states. Next, we instantiate our general definition to Marginal Contribution Nets [7]. Then, we consider computational aspects related to our representation and discuss related work.

2 Preliminaries

A game-theoretical convention for modeling coalitional games is a characteristic function game (CFG) representation. In this approach values of all non-empty coalitions are explicitly listed. Formally, a *coalitional game* is described by a tuple $\mathcal{G} = \langle A, v \rangle$, where $A = \{a_1, \ldots, a_n\}$, is a set of $n = |A|$ agents, and a function $v : 2^A \longrightarrow \Re$ maps any coalition, i.e., a set of agents, to a real value, where it is assumed that $v(\emptyset) = 0$. The coalition of all the agents in the game is called the *grand coalition*.

Example 2.1 (Characteristic function). For $A = \{a_1, a_2, a_3\}$, a sample characteristic function is:

$$
\begin{array}{lll}
v(\{a_1\}) = 0 & v(\{a_2\}) = 0 & v(\{a_3\}) = 1 \\
v(\{a_1, a_2\}) = 1 & v(\{a_1, a_3\}) = 1 & v(\{a_2, a_3\}) = 1 \\
v(\{a_1, a_2, a_3\}) = 2. & &
\end{array}
\qquad \lhd
$$

The majority of the best-known solution concepts used with coalitional games have been developed building upon the above CFG representation. Arguably the most famous normative solution concept is the Shapley value. Assuming that the grand coalition is optimal and eventually will form, the Shapley value shows what is the fair division of payoff between agents.[1] Any agent is reimbursed, not only for its performance in the grand coalition, but for its potential marginal contribution to every other coalition. It is assumed that agents join the coalitions in random order and thus all permutations of agents are equally likely. More formally, let $\Pi(A)$ be the set of all permutations of agents in A. For $\pi \in \Pi(A)$ denote by $C_\pi(a_i) \stackrel{\text{def}}{=} \{a_j \mid \pi(a_j) < \pi(a_i)\}$, where $\pi(a_j) < \pi(a_i)$ denotes the fact that agent a_j occurs in π before agent a_i. The *Shapley value of agent* a_i in a game $\mathcal{G} = \langle A, v \rangle$, denoted by $\phi_{\mathcal{G}}(a_i)$, is given by the following expression:

$$\phi_{\mathcal{G}}(a_i) = \frac{1}{n!} \sum_{\pi \in \Pi(A)} \big[v(C_\pi(a_i) \cup \{a_i\}) - v(C_\pi(a_i)) \big].$$

[1] The grand coalition is optimal if its value is at least as large as the sum of the values of any partition of agents into smaller coalition. This assumption ensures that it is a rational choice to form the grand coalition, as is required by the Shapley value as well as many other solution concepts. Nevertheless, the formal analysis is meaningful without the assumption.

Example 2.2 (Shapley value). For the game $\mathcal{G} = \langle A, v\rangle$ defined in Example 2.1, the Shapley values of successive agents are $\phi_{\mathcal{G}}(a_1) = \phi_{\mathcal{G}}(a_2) = \frac{1}{2}$ and $\phi_{\mathcal{G}}(a_3) = 1$. $\lhd$

The importance of the Shapley value comes from the fact that it is the only payoff division scheme that satisfies the following natural "fairness" axioms:

1. *efficiency*: it fully distributes the total payoff available to the agents:
$$\sum_{a\in A} \phi_{\mathcal{G}}(a) = v(A) \tag{1}$$
2. *symmetry*: if agents a_i and a_j are interchangeable, then they have the same payoff:
$$\begin{array}{l}\text{if, for any } C \subseteq A \setminus \{a_i, a_j\},\ \text{one has } v(C \cup \{a_i\}) = v(C \cup \{a_j\}) \\ \text{then } \phi_{\mathcal{G}}(a_i) = \phi_{\mathcal{G}}(a_j)\end{array} \tag{2}$$
3. *dummy*: if an agent a_i does not contribute to any coalition then its value is 0:
$$\text{if, for any } C \subseteq A \setminus \{a_i\},\ \text{one has } v(C) = v(C \cup \{a_i\}) \text{ then } \phi_{\mathcal{G}}(a_i) = 0 \tag{3}$$
4. *linearity*: for any two coalitional games $\mathcal{G} = \langle A, v\rangle$ and $\mathcal{G}' = \langle A, v'\rangle$:
$$\phi_{a*\mathcal{G}+b*\mathcal{G}'}(a_i) = a * \phi_{\mathcal{G}}(a_i) + b * \phi_{\mathcal{G}'}(a_i) \tag{4}$$
where $a, b \in \mathfrak{R}$ and $a * \mathcal{G} + b * \mathcal{G}' \stackrel{\text{def}}{=} \langle A, a * v + b * v'\rangle$.

If a coalitional game is modeled using a CFG representation, computation of the Shapley value as well as many other solution concepts becomes problematic. This is because the number of feasible coalitions grows exponentially in the number of agents. It means that the size of the input renders the computational insights regarding those solution concepts meaningless for larger n.

We will say that a given representation of coalitional games is *fully expressive* iff it allows to represent a characteristic function of any coalitional game. Clearly, the CFG representation is fully expressive.

3 Formalization of Contextual Coalitional Games

In this section, we formally introduce coalitional games with stochastic contexts and their representations.

Definition 3.1. A *contextual coalitional game* (CCG, in short) is a tuple: $\langle A, \mathcal{S}, \vartheta, \mathcal{P}\rangle$, where:

- A is a set of agents in the game;
- $\mathcal{S} \stackrel{\text{def}}{=} \{\sigma_1, \ldots, \sigma_k\}$ is a finite set of *states of the environment* in which the game is played; it is assumed that, in a given moment, the environment is in exactly one state;
- $\vartheta : \mathcal{S} \times 2^A \longrightarrow \mathfrak{R}$ is a mapping, which associates payoffs to coalitions in states;

- $\mathcal{P} = \{p_\sigma \mid \sigma \in S\}$ is a probability distribution on states, where p_σ denotes the probability that state σ materializes. $\lhd$

As before, we assume the payoff 0 for the empty coalition, i.e., for all $s \in \mathcal{S}$ it holds that $\vartheta(s, \emptyset) = 0$.

Definition 3.2. Let $\mathcal{G} = \langle A, \mathcal{S}, \vartheta, \mathcal{P}\rangle$. The *expected value $v_{\mathcal{G}}(C)$ of a coalition $C \subseteq A$ in game $\mathcal{G}$* is defined by:[2]

$$v_{\mathcal{G}}(C) \stackrel{\text{def}}{=} \sum_{\sigma \in \mathcal{S}} p_\sigma * \vartheta(\sigma, C). \qquad \lhd$$

The following example illustrates the idea of CCGs.

Example 3.3. A sample contextual coalitional game $\mathcal{G}$ can be given by setting $A = \{a_1, a_2\}$, $\mathcal{S} = \{\sigma_1, \sigma_2\}$, $p_{\sigma_1} = 0.4$ and $p_{\sigma_2} = 0.6$ and

$$\begin{array}{lll} \vartheta(\sigma_1, \{a_1\}) = 2 & \vartheta(\sigma_1, \{a_2\}) = 3 & \vartheta(\sigma_1, \{a_1, a_2\}) = 4 \\ \vartheta(\sigma_2, \{a_1\}) = 2 & \vartheta(\sigma_2, \{a_2\}) = 1 & \vartheta(\sigma_2, \{a_1, a_2\}) = 3. \end{array}$$

Consider coalitional games $\mathcal{G}_1 = \langle A, v_1\rangle, \mathcal{G}_2 = \langle A, v_2\rangle$, where:

$$\begin{array}{lll} v_1(\{a_1\}) = 2 & v_1(\{a_2\}) = 3 & v_1(\{a_1, a_2\}) = 4 \\ v_2(\{a_1\}) = 2 & v_2(\{a_2\}) = 1 & v_2(\{a_1, a_2\}) = 3. \end{array}$$

The intuition behind the contextual coalitional game $\mathcal{G}$ is that the coalitional game $\mathcal{G}_1$ takes place when the environment is in the state σ_1 (with probability 0.4) and the coalitional game $\mathcal{G}_2$ takes place when the environment is in the state σ_2 (with probability 0.6). $\lhd$

We can generalize this in the following proposition, showing that contextual coalitional games can be represented as linear combinations of traditional coalitional games. This can be proved by a direct application of Definition 3.2.

Proposition 3.4. Let $\mathcal{G} = \langle A, \mathcal{S}, \vartheta, \mathcal{P}\rangle$ be a CCG. Then $\mathcal{G} = \sum_{\sigma \in \mathcal{S}} p_\sigma * \mathcal{G}_\sigma$, where

$$\mathcal{G}_\sigma \stackrel{\text{def}}{=} \langle A, \vartheta_\sigma\rangle \text{ with } \vartheta_\sigma \stackrel{\text{def}}{=} \vartheta(\sigma, C). \qquad \lhd$$

We then have the following definition of the expected Shapley value for CCGs.

Definition 3.5. Let $\mathcal{G} = \langle A, \mathcal{S}, \vartheta, \mathcal{P}\rangle$ be a CCG. Then *the expected Shapley value for $\mathcal{G}$* is

$$\Phi_{\mathcal{G}}(a_i) \stackrel{\text{def}}{=} \Phi_{\sum_{\sigma \in \mathcal{S}} p_\sigma * \mathcal{G}_\sigma}(a_i). \qquad \lhd$$

[2] Throughout the paper we omit the expectation symbol for notational convenience.

4 Representations of Contextual Coalitional Games

A general representation for CCGs considered in this paper is composed of rules of the form:

$$\textit{prerequisite}\ (\alpha) \mid \textit{coalitional game representation}\ (\varrho) \tag{5}$$

where the prerequisite α is a formula expressed in some logical language $\mathcal{L}$. We do not fix any particular representation type used for ϱ. CFG is one such conventional game representation type, although in what follows, we will not restrict ourselves to only CFG representations. Intuitively, rule (5) reads as

> "in the states where the prerequisite α is true, the coalitional game is represented by ϱ".[3]

If multiple rules are true at the same time, then coalition values are to be computed additively.

The game consisting of no rules is called the *empty game*. In the empty game the payoff for all coalitions is 0.

This representation is intended to take into account influences or circumstances external to a coalitional game. Such influences are expressed by the "α parts" of rules (5). The formal meaning of α formulas is given by states, where each state materializes with a given probability. Such probability distributions are often given on the basis of statistical data and from other sources (see, e.g., [14]).

Let us now formally define our representation.

Definition 4.1. A *CCG representation* is a tuple $\langle A, \mathcal{S}, \mathcal{P}, \mathcal{R}, \mathcal{F}\rangle$, where:

- A, $\mathcal{S}$ and $\mathcal{P}$ are as in Definition 3.1;
- $\mathcal{R}$ is a finite set of rules of the form (5) such that for each $(\alpha|\varrho) \in \mathcal{R}$, for the game $\mathcal{G} = \langle A', v\rangle$ that ϱ represents, it holds that $A' \subseteq A$;
- $\mathcal{F} = \{\alpha \mid \text{ there is } (\alpha|\varrho) \in \mathcal{R}\}$, i.e., $\mathcal{F}$ is the set of formulas appearing as prerequisites in rules of $\mathcal{R}$. $\lhd$

Definition 4.2. An *interpretation* of a CCG representation $\langle A, \mathcal{S}, \mathcal{P}, \mathcal{R}, \mathcal{F}\rangle$ is a tuple $\langle A, \mathcal{S}, \mathcal{P}, \mathcal{R}, \mathcal{F}, f\rangle$, where:

- A, $\mathcal{S}$, $\mathcal{P}$, $\mathcal{R}$ and $\mathcal{F}$ are as in Definition 4.1;
- $f : \mathcal{F} \longrightarrow 2^{\mathcal{S}}$ is a mapping, which associates to formulas sets of states where they are TRUE. $\lhd$

Remark 4.3. Observe that f appearing in Definition 4.2 should reflect the semantics of a particular logic chosen for expressing prerequisites of rules.

Note also that f provides truth values of formulas in states. Namely, a formula $\alpha \in \mathcal{F}$ is TRUE in a state $\sigma \in \mathcal{S}$ iff $\sigma \in f(\alpha)$. $\lhd$

Representations and their meanings are defined as follows.

---

[3] For notational convenience, we assume that for an instance where the prerequisite α is omitted, this rule should be treated as having α being TRUE.

Definition 4.4. Given a state $\sigma \in \mathcal{S}$ and an interpretation $\mathcal{I} = \langle A, \mathcal{S}, \mathcal{P}, \mathcal{R}, \mathcal{F}, f \rangle$, the *meaning* of a rule of $\mathcal{R}$ is defined by

$$(\alpha|\varrho)^{\mathcal{I}}_{\sigma} \stackrel{\text{def}}{=} \begin{cases} \varrho \text{ if } \sigma \in f(\alpha) \\ \emptyset \text{ otherwise,} \end{cases} \tag{6}$$

where $\emptyset$ is the game given by a representation consisting of no rules. $\triangleleft$

Now we define the value of a coalition C and the Shapley value in a state $\sigma \in \mathcal{S}$ as follows.

Definition 4.5. Let $R = \langle A, \mathcal{S}, \mathcal{P}, \mathcal{R}, \mathcal{F} \rangle$ be a CCG representation with the set of rules $\mathcal{R} = \{\alpha_1|\varrho_1, \ldots, \alpha_m|\varrho_m\}$ and $\mathcal{I} = \langle A, \mathcal{S}, \mathcal{P}, \mathcal{R}, \mathcal{F}, f \rangle$ be an interpretation of R.

- For $\sigma \in \mathcal{S}$, by the $(\sigma, \mathcal{I})$-*reduct* of R we understand game $\mathcal{G}^{\mathcal{I}}_{\sigma}$ represented by the set of conventional rules $\{(\alpha_i|\varrho_i)^{\mathcal{I}}_{\sigma} \mid 1 \le i \le m\}$.
- The *value of a coalition* $C \subseteq A$ *in state* $\sigma \in \mathcal{S}$ under interpretation $\mathcal{I}$, is defined as $v_{\mathcal{G}^{\mathcal{I}}_{\sigma}}(C)$.
- The *Shapley value* for a_i over R, $\mathcal{I}$ and $\sigma \in \mathcal{S}$, denoted as $\phi^{\mathcal{I}}_{R,\sigma}(a_i)$, is defined as the Shapley value $\phi_{\mathcal{G}^{\mathcal{I}}_{\sigma}}(a_i)$. $\triangleleft$

Definition 4.6. We say that $R = \langle A, \mathcal{S}, \mathcal{P}, \mathcal{R}, \mathcal{F} \rangle$ *represents* a CCG $\mathcal{G} = \langle A, \mathcal{S}, \vartheta, \mathcal{P} \rangle$ over an interpretation $\mathcal{I} = \langle A, \mathcal{S}, \mathcal{P}, \mathcal{R}, \mathcal{F}, f \rangle$ provided that for any coalition $C \subseteq A$ and $\sigma \in \mathcal{S}$ we have that $v_{\mathcal{G}^{\mathcal{I}}_{\sigma}}(C) = \vartheta(\sigma, C)$, where $\mathcal{G}^{\mathcal{I}}_{\sigma}$ is the $(\sigma, \mathcal{I})$-*reduct* of R. $\triangleleft$

We have the following lemma showing that CCGs can be represented as traditional coalitional games.

Lemma 4.7. Let $R = \langle A, \mathcal{S}, \mathcal{P}, \mathcal{R}, \mathcal{F} \rangle$ represent a CCG $\mathcal{G} = \langle A, \mathcal{S}, \vartheta, \mathcal{P} \rangle$ over an interpretation $\mathcal{I} = \langle A, \mathcal{S}, \mathcal{P}, \mathcal{R}, \mathcal{F}, f \rangle$. Then:

$$\mathcal{G} = \sum_{\sigma \in \mathcal{S}} p_{\sigma} \mathcal{G}^{\mathcal{I}}_{\sigma}. \tag{7}$$

Proof. According to Definition 3.2, $v_{\mathcal{G}}(C) = \sum_{\sigma \in \mathcal{S}} p_{\sigma} * \vartheta(\sigma, C)$. By Definition 4.5, for any $C \subseteq A$ and $\sigma \in \mathcal{S}$ we have that $\vartheta(\sigma, C) = v_{\mathcal{G}^{\mathcal{I}}_{\sigma}}(C)$. Therefore,

$$v_{\mathcal{G}}(C) = \sum_{\sigma \in \mathcal{S}} p_{\sigma} * v_{\mathcal{G}^{\mathcal{I}}_{\sigma}}(C),$$

which completes the proof. $\triangleleft$

Similarly, in the broader contextual coalitional context, we compute the Shapley value for players in state $\sigma \in \mathcal{S}$ using the additivity axiom met by the Shapley value.

Having defined the Shapley value for a game in state $\sigma \in \mathcal{S}$, we are now interested in the value for a contextual coalitional game as a whole. In our stochastic environment this value will be a mapping which takes as input a tuple $\langle \mathcal{I}, R, a_i \rangle$, where $\mathcal{I}$ is an interpretation, R is a CCG representation and $a_i \in A$ is an agent, and returns the expected Shapley value of a_i in the game represented by R over $\mathcal{I}$. This value will be denoted by $\Phi_{R,\mathcal{I}}(a_i)$ and formalized as follows.

Definition 4.8. The *expected Shapley value of a contextual coalitional game represented by* $R = \langle A, \mathcal{S}, \mathcal{P}, \mathcal{R}, \mathcal{F}\rangle$ over an interpretation $\mathcal{I} = \langle A, \mathcal{S}, \mathcal{P}, \mathcal{R}, \mathcal{F}, f\rangle$ for player $a_i \in A$ is given by:

$$\Phi_{R,\mathcal{I}}(a_i) \stackrel{\text{def}}{=} \sum_{\sigma \in \mathcal{S}} p_\sigma * \phi^{\mathcal{I}}_{R,\sigma}(a_i). \qquad \lhd$$

Our contextual coalitional game representation is intended to reflect games which are repeated over a longer time period in order to make the stochastic nature of the expected Shapley values practically acceptable. For example, rather then considering a single rescue mission relative to the generic scenario described in the introduction, we would consider a time period where there might be many such missions. The equipment/services' suppliers need to have equipment and staff ready on demand, so they have to know in advance whether their income will be satisfactory. It is reasonable to assume that they receive a fixed fee covering fixed costs such as equipment amortization, maintenance, etc., independently of the number of missions actually carried out. For each mission carried out they then receive additional fees covering resources used, e.g., gas, electricity, repairing, etc. In such scenarios we mainly focus on the distribution of fixed fee, which reflects the importance of equipment and services supplied.

Definition 4.9. Let $\mathbb{P} = \langle \mathbb{R}, \mathbb{I}\rangle$, where $\mathbb{R}$ is a set of representations and $\mathbb{I}$ is a set of interpretations. We say that $\mathbb{P}$ is *fully expressive for CCGs* iff for any CCG $\mathcal{G}$ there is $R \in \mathbb{R}$ and $\mathcal{I} \in \mathbb{I}$ such that R represents $\mathcal{G}$ over $\mathcal{I}$. $\lhd$

Remark 4.10. Recall that any rule of the form $\textsc{True}|\varrho$ represents ϱ itself. Therefore $\mathbb{P} = \langle \mathbb{R}, \mathbb{I}\rangle$ is fully expressive if the representation type used for righthand sides of rules is fully expressive for conventional games. $\lhd$

By Definition 4.8, the expected Shapley value $\Phi_{R,\mathcal{I}}(a_i)$ is given by $\sum_{\sigma \in \mathcal{S}} p_\sigma * \phi^{\mathcal{I}}_{R,\sigma}(a_i)$. An algorithm for computing $\Phi_{R,\mathcal{I}}(a_i)$ directly from this formula provides the following complexity result.

Theorem 4.11. The complexity of computing the expected Shapley value for the CCG representation $R = \langle A, \mathcal{S}, \mathcal{P}, \mathcal{R}, \mathcal{F}\rangle$ over an interpretation $\mathcal{I} = \langle A, \mathcal{S}, \mathcal{P}, \mathcal{R}, \mathcal{F}, f\rangle$ is

$$O\left(|\mathcal{S}| \;*\; \max_{\rho \in \left\{(\alpha|\varrho)^{\mathcal{I}}_{\sigma} \;\middle|\; (\alpha|\varrho) \in \mathcal{R}\right\}} \{g(\rho), h(f)\}\right)$$

where $g(\varrho)$ is the complexity of computing the Shapley value for the representation ϱ and $h(f)$ is the complexity of checking whether a given formula is true in a given state. $\lhd$

5 Contextual Marginal Contribution Nets

The representation described in the previous section is general in the sense that the context α in a rule can denote a formula of a *given logic* and ϱ denotes a *conventional game representation* CFG.

6 Contextual Marginal Contribution Nets

In this section, we will instantiate the general representation in the following manner by choosing propositional logic as the given logic for α and by using basic Marginal Contribution Nets (abbreviated by MC-nets) of [7] for ϱ. The choice of MC-Nets for ϱ is useful due to the computationally efficient manner in which Shapely values can be computed and also due to the fact that the representation is in logical form.

Formally, given a set of agents, an MC-net is defined as a finite set ϱ of rules of the form

$$\mathbf{P} \rightarrow Value$$

where $Value$ is a real number and *pattern* $\mathbf{P}$ is a Boolean expression with agents as atoms. A coalition C of agents is said to *meet the requirements of* (or shortly *meet*) a given $\mathbf{P}$ (denoted by $C \models \mathbf{P}$) if $\mathbf{P}$ evaluates to TRUE when the values of all Boolean variables that correspond to agents in C are set to TRUE, and the values of all Boolean variables that correspond to agents not in C are set to FALSE. The value $v(C)$ is equal to the sum of all values from rules of which the requirements are met by C. More formally,

$$v(C) \quad = \sum_{\mathbf{P}\rightarrow Value\in\varrho:\ C\models\mathbf{P}} Value$$

Example 6.1 (MC-nets representation for Example 2.1). The coalitional game from Example 2.1 can be represented with only two rules $a_3 \rightarrow 1$ and $a_1 \wedge a_2 \rightarrow 1$. $\triangleleft$

Such rules have an interesting interpretation, as they show the marginal contribution to all the coalitions agents can form. The advantages of MC-nets are twofold. Firstly, they allow for representing many important classes of games in a number of rules that is polynomial in n. Secondly, they allow for computing the Shapley value in time linear in the number of rules. However, although the definition of MC-nets is quite general, this latter computational result, as discussed in [7], is limited only to patterns which are conjunctions of literals. More formally, patterns taking the form:

$$a_{i_1} \wedge \ldots \wedge a_{i_m} \wedge \neg a_{j_1} \wedge \ldots \wedge \neg a_{j_k} \tag{8}$$

Following [6] we will call them *basic patterns* and the representation *basic MC-nets*.[4] It will be formally denoted $\langle A, \varrho\rangle$ where ϱ is the set of all the rules. Note, that all the patterns in Example 6.1 are, in fact, basic.

MC-nets are fully expressive [7] even when limited to conjunctions of literals. The linear method of computation of the Shapley value from rules of the form shown in (8) is explained in Figure 1.

Let $\mathcal{V}_0 = \{p_0, \ldots, p_l\}$ be a finite *set of propositional variables*. Variables specify atomic properties of a context by means of a mapping:

$$f_0 : \mathcal{V}_0 \longrightarrow 2^{\mathcal{S}} \tag{9}$$

where $f_0(p)$ is the set of states in which p is TRUE.

[4] In the rest of the paper we will assume that every pattern has distinct literals, i.e., $|\{i_1, \ldots, i_m, j_1, \ldots, j_k\}| = m + k$. It can be easily seen that if a conjunction of literals cannot be normalized to this form, i.e., $i_u = j_w$ for some u, w, then removing it does not change the represented game.

Due to the linearity of the Shapley value, every basic rule

$$a_{i_1} \wedge \ldots \wedge a_{i_m} \wedge \neg a_{j_1} \wedge \ldots \wedge \neg a_{j_k} \to Value$$

can be considered as a separate game. The Shapley values for any agent a_{i_u} and a_{j_w} are respectively:

$$\frac{Value}{m\binom{m+k}{k}} \quad \text{and} \quad \frac{-Value}{k\binom{m+k}{m}} \tag{10}$$

Fig. 1. Ieong and Shoham's method for computing Shapley value

Propositional formulas over $\mathcal{V}_0$ are built using $\mathcal{V}_0$ and connectives $\neg, \vee, \wedge, \to, \equiv$. The set of propositional formulas is denoted by $\mathcal{F}_0$. The mapping f_0 is extended to $\mathcal{F}_0$ in the standard way:

$$\begin{array}{ll} f_0(\neg\alpha) & \stackrel{\text{def}}{=} \mathcal{S} - f_0(\alpha) \\ f_0(\alpha \vee \beta) & \stackrel{\text{def}}{=} f_0(\alpha) \cup f_0(\beta) \\ f_0(\alpha \wedge \beta) & \stackrel{\text{def}}{=} f_0(\alpha) \cap f_0(\beta) \\ f_0(\alpha \to \beta) & \stackrel{\text{def}}{=} f_0(\neg\alpha) \cup f_0(\beta) \\ f_0(\alpha \equiv \beta) & \stackrel{\text{def}}{=} f_0(\alpha \to \beta) \cap f_0(\beta \to \alpha). \end{array} \tag{11}$$

Consequently, the rule representation for contextual MC-nets consists of rules of the form:

$$\alpha | \{\mathbf{p} \to \mathit{Value}\} \tag{12}$$

where α is a propositional formula over the set of propositional variables $\mathcal{V}_0$, $\mathbf{p}$ is a pattern of the form (8) and *Value* is a real number.

Definition 6.2. By a *contextual MC-net* we understand any finite set of rules of the form (12). ◁

Definition 6.3. An interpretation of *contextual MC-nets* is a tuple $\langle A, \mathcal{S}, \mathcal{P}, \mathcal{R}, \mathcal{F}, f_0 \rangle$, where

- A is a finite set of agents in the game;
- $\mathcal{S}, \mathcal{P}$ are as in Definition 4.2;
- $\mathcal{R}$ is a finite set of rules of the form (12);
- $\mathcal{F} \subseteq \mathcal{F}_0$ are propositional formulas over $\mathcal{V}_0$, appearing as prerequisites in $\mathcal{R}$;
- f_0 is defined by (9) and (11). ◁

Since MC-nets are fully expressive, we have the following corollary (cf. Remark 4.10).

Corollary 6.4. The representation of contextual MC-nets is fully expressive. ◁

The complexity of computing the Shapley value for MC-nets is PTIME. Therefore, by Theorem 4.11 we have the following corollary.

Corollary 6.5. The complexity of computing the expected Shapley value for contextual MC-nets is in PTIME in the maximum of size of the representation and the number of states. $\triangleleft$

7 An Example Using Contextual MC-Nets

In the following example, we will show how contexts and uncertainty associated with contexts can be used to model stochastic contextual coalitional games.

Example 7.1. Using the scenario considered in the introduction, one can assume that there are states providing values for the propositional variables r, w, s standing for *rain*, *moderate wind* and *strong wind*. Assume that for the rescue missions considered, there is a probability distribution on weather conditions.

The CCG modeling our scenario is $\langle A, \mathcal{S}, \vartheta, \mathcal{P}\rangle$, where

- $A = \{uav_1, uav_2, ugv_1, ugv_2\}$;
- $\mathcal{S} = \{\sigma_1, \sigma_2, \sigma_3, \sigma_4\}$;
- $\vartheta(\sigma_1, \{uav_1\}) = 6$, $\vartheta(\sigma_1, \{uav_1, uav_2\}) = 13$, etc.;[5]
- $\mathcal{P}$ is provided in Table 1.

Table 1. Probability of various weather conditions

State	Weather	Literals true in the state	Probability
σ_1	rain and moderate wind	$r, w, \neg s$	0.20
σ_2	rain without wind	$r, \neg w, \neg s$	0.10
σ_3	rain with strong wind	$r, \neg w, s$	0.35
σ_4	no rain with strong wind	$\neg r, \neg w, s$	0.35

The following contextual MC-net rules are used to model our scenario:

$$\mid \{uav_1 \to 6, uav_2 \to 7, ugv_1 \to 3, ugv_2 \to 2, ugv_1 \wedge ugv_2 \to 1\} \tag{13}$$

$$r \wedge w \mid \{uav_1 \to -6, ugv_1 \to 2, ugv_2 \to 2\} \tag{14}$$

$$\neg r \wedge s \mid \{uav_1 \to -6, uav_2 \to -7, ugv_1 \to 4, ugv_2 \to 3.5\} \tag{15}$$

The first rule defines the basic game which applies in all contexts. The values of this game can be amended by other rules if specific weather condition contexts occur. Specifically, in the case of rain and moderate wind, uav_1 becomes useless and the importance of ground robots, ugv_1, ugv_2 increases. If the wind becomes strong, both UAVs are grounded and the importance of both ground robots increases even more.

[5] We avoid here listing values for all $4 * 15 = 60$ state–coalition pairs. The values are actually given by rules (13)–(15).

Using formulas from Figure 1, it is easy to check that the Shapley values in the game described by rules (13), (14) and (15) are respectively:

$$\begin{array}{llll} \phi_1(uav_1) = 6 & \phi_1(uav_2) = 7 & \phi_1(ugv_1) = 3.5 & \phi_1(ugv_2) = 2.5 \\ \phi_2(uav_1) = -6 & \phi_2(uav_2) = 0 & \phi_2(ugv_1) = 2 & \phi_2(ugv_2) = 2 \\ \phi_3(uav_1) = -6 & \phi_3(uav_2) = -7 & \phi_3(ugv_1) = 4 & \phi_3(ugv_2) = 3.5. \end{array}$$

By referring to prerequisites of the rules and Table 1, one observes that the rule (13) always holds, the rule (14) holds with probability 0.20, whereas the rule (15) holds with probability 0.35. Thus, the expected Shapley values for the entire game are:

$$\phi(uav_1) = 2.7, \phi(uav_2) = 4.55, \phi(ugv_1) = 5.3, \phi(ugv_2) = 4.125.$$

This means that ugv_1 contributes most value to the coalitional game, while uav_1 contributes the least value. $\triangleleft$

8 Related Work

Two main streams in the literature on coalitional games are relevant to the ideas contained in this paper. Firstly, there is a body of research where uncertainty is modeled probabilistically and secondly, there is a body of research which focuses on concise representations of coalitional games which enhances computational efficiency in their use.

Regarding the modeling of uncertainty in the context of coalitional games, a short but informative literature review is provided in [8]. Important and relevant recent contributions include [13], [9], [1,2,3] and [8] itself. We focus on [8], where Bayesian Coalitional Games are introduced as a tuple of agents, set of possible worlds (i.e., states), common prior over these worlds, each agent's information partition of the worlds, and their preferences over the distribution of payoffs. An information partition is composed of agents' information sets — subsets of worlds that are undistinguishable from the individual agent's point of view, but where the real world actually resides.[6] If the agent-specific elements are added to our model, we will obtain Bayesian Coalitional Games.

Nevertheless, the crucial difference between our approach and the others is related to the representation of a coalitional game. As all the other approaches build upon the conventional game theoretical method of representing games (i.e., characteristic function), the number of values to be defined is exponential in the number of agents. This prohibits efficient computation of solution concepts even for a moderate number of agents. Using our approach it is possible to represent many games in a polynomial number of rules.

In this respect our work is related to the literature on alternative representations of coalitional games. The aim of this research is to develop representations for coalitional games that are compact, but still allow for the efficient computation of solution concepts such as Shapley value and coalitional game cores [7], [6], [4], or for finding an optimal arrangement of coalitions in a system [10].

[6] For more details about the information partition method for modeling uncertainty in the non-cooperative game area, see [11].

For instance, [5,15] give the characteristic function a specific interpretation in terms of combinatorial structures. The advantage of this method is that the representation can always be guaranteed to be succinct. The disadvantage is that the representation is not fully expressive being incapable of expressing the full space of characteristic function game instances. Many of the other papers propose representations that are fully *expressive* but are not always guaranteed to be succinct [7]. Our work falls under this latter class.

9 Conclusions

In this paper, we proposed a representation for coalitional games which takes into account the stochastic nature of real-world multi-agent scenarios and which relaxes the need for a deterministic payoff to coalitions. The representation is based on the idea of contextualizing coalitional games through the use of logical expressions representing environmental and other state and placing probability distributions on the space of contexts in order to model the stochastic nature of the scenarios. The representation is succinct and intuitive and takes advantage of representational features of logic and its relation to probability. Additionally, we define and show how to compute *expected Shapley values* in such games in a computationally efficient manner. We show the value of the approach through a generic example involving robotics assistance in emergencies.

Acknowledgments

We would like to thank anonymous reviewers for helpful comments and remarks.

This work has been supported in part by grant N N206 567040 of Polish Ministry of Science and Higher Education and by grants from the Swedish Research Council (VR) Linnaeus Center CADICS, VR grant 621-2009-3857, the ELLIIT Excellence Center at Linköping-Lund for Information Technology, and the Swedish National Aviation Engineering Research Programme Vinnova grant NFFP5.

References

1. Chalkiadakis, G., Boutilier, C.: Bayesian reinforcement learning for coalition formation under uncertainty. In: Proc. of AAMAS 2004, pp. 1090–1097 (2004)
2. Chalkiadakis, G., Boutilier, C.: Coalitional bargaining with agent type uncertainty. In: Proc. IJCAI, pp. 1227–1232 (2007)
3. Chalkiadakis, G., Markakis, E., Boutilier, C.: Coalition formation under uncertainty: Bargaining equilibria and the Bayesian core stability concept. In: Proc. AAMAS, pp. 1–8 (2007)
4. Conitzer, V., Sandholm, T.: Complexity of determining nonemptiness in the core. In: Proc. IJCAI, pp. 219–225 (2004)
5. Deng, X., Papadimitriou, C.: On the complexity of cooperative solution concepts. Mathematical Oprerational Research 19(2), 257–266 (1994)
6. Elkind, E., Goldberg, L.A., Goldberg, P.W., Wooldridge, M.: A tractable and expressive class of marginal contribution nets and its applications. Mathematical Logic Quarterly 55(4), 362–376 (2009)

7. Ieong, S., Shoham, Y.: Marginal contribution nets: A complact representation scheme for coalitional games. In: Proc. ACM EC, pp. 170–179 (2006)
8. Ieong, S., Shoham, Y.: Bayesian coalitional games. In: Proc. AAAI, pp. 95–100 (2008)
9. Myerson, R.B.: Virtual utility and the core for games with incomplete information. In: Econometric Society 2004 North American Summer Meetings 659, Econometric Society (2004)
10. Ohta, N., Conitzer, V., Ichimura, R., Sakurai, Y., Iwasaki, A., Yokoo, M.: Coalition structure generation utilizing compact characteristic function representations. In: Gent, I.P. (ed.) CP 2009. LNCS, vol. 5732, pp. 623–638. Springer, Heidelberg (2009)
11. Osborne, M.J., Rubinstein, A.: A Course in Game Theory. MIT Press Books, vol. 1. The MIT Press, Cambridge (December 1994)
12. Sandholm, T., Lesser, V.R.: Coalitions among computationally bounded agents. Artificial Intelligence 94, 99–137 (1997)
13. Suijs, J., Borm, P., De Waegenaere, A., Tijs, S.: Cooperative games with stochastic payoffs. European Journal of Operational Research 113(1), 193–205 (1999)
14. Thrun, S., Burgard, W., Fox, D.: Probabilistic Robotics (Intelligent Robotics and Autonomous Agents). The MIT Press, Cambridge (September 2005)
15. Wooldridge, M., Dunne, P.E.: On the computational complexity of coalitional resource games. AI 170(10), 853–871 (2006)

Sensible Semantics of Imperfect Information*

On a Formal Feature of Meanings

Pietro Galliani

Universiteit van Amsterdam, Institute for Logic, Language and Computation,
P.O. Box 94242, 1090 GE Amsterdam, The Netherlands

Abstract. In [1] Cameron and Hodges proved, by means of a combinatorial argument, that no compositional semantics for a logic of Imperfect Information such as Independence Friendly Logic ([2]) or Dependence Logic ([3]) may use sets of tuples of elements as meanings of formulas.

However, Cameron and Hodges' theorem fails if the domain of the semantics is restricted to infinite models only, and they conclude that[1]

> *Common sense suggests that there is no sensible semantics for CS on infinite structures A, using subsets of the domain of A as interpretations for formulas with one free variable. But we don't know a sensible theorem along these lines* ([1]).

This work develops a formal, natural definition of "sensible semantics" according to which the statement quoted above can be proved.

1 Introduction

1.1 Logics of Imperfect Information

Logics of imperfect information are generalizations of First Order Logic which allow for more general patterns of dependence of independence between variables.

The study of these logics started with [5], which introduced the *branching quantifiers*

$$\begin{pmatrix} \forall x & \exists y \\ \forall z & \exists w \end{pmatrix} \phi(x, y, z, w)$$

whose interpretation corresponds to that of $\exists f \exists g \forall x \forall z \phi(x, f(x), z, g(z))$. As mentioned in the same paper, Ehrenfeucht was able to prove that Branching Quantifier Logic is strictly more expressive than First Order Logic: indeed, the sentence

$$\exists p \begin{pmatrix} \forall x & \exists y \\ \forall z & \exists w \end{pmatrix} (x = z \leftrightarrow y = w \wedge y \neq p)$$

* This work was supported by the European Science Foundation Eurocores programme LogICCC [FP002 - Logic for Interaction (LINT)].

[1] Here the acronym "CS" stands for CS-Logic, that is, for the variant of IF-Logic introduced by Wilfrid Hodges in [4].

M. Banerjee and A. Seth (Eds.): ICLA 2011, LNAI 6521, pp. 79–89, 2011.

holds in a model M if and only if there exists a function $f : M \to M$ that is injective but not surjective, that is, if and only if M is an infinite model.

A significant progress in the study of logics of imperfect information was made with the introduction of *Independence-Friendly Logic* ([6], [2], [7], [8]). This logic simplifies the syntax of Branching Quantifier Logic by introducing the *slashed quantifiers* $\exists x/W\phi$, interpreted as "there exists an x, chosen independently from the values of the variables in W, such that ϕ" and proved itself especially useful for the study of the game-theoretic interpretation of logics of imperfect information ([2]).

Later, Väänänen ([3]) introduced *Dependence-Friendly Logic*, which takes dependence (rather than independence) between quantifiers as the primary concept, and *Dependence Logic*, which separates the notion of dependency from that of quantification by introducing the *dependence atomic formulas* $=(t_1 \ldots t_n)$, whose interpretation is "the value of the term t_n is a function of the values of the terms $t_1 \ldots t_{n-1}$".

The formal definition of the language of Dependence Logic can be given as follows ([3]):

Definition 1 (Dependence Logic: Syntax). *Let Σ be a First Order signature. Then a Dependence Logic formula with signature Σ is an expression built according to the grammar*

$$\phi ::= Rt_1 \ldots t_n \mid t = t' \mid\ =(t_1 \ldots t_n) \mid \neg\phi \mid \phi \vee \phi \mid \exists x\phi$$

where n ranges over $\mathbb{N}$, R ranges over all n-ary relations in Σ, and $t_1 \ldots t_n, t$ and t' range over all terms[2] with signatures included in Σ.

As usual, we will write $\phi \wedge \psi$ as a shorthand for $\neg(\neg\phi \vee \neg\psi)$ and $\forall x\phi$ as a shorthand for $\neg(\exists x(\neg\phi))$.

Given a Dependence Logic formula ϕ, the set $FV(\phi)$ of all free variables *of ϕ is defined inductively as follows:*

- *$FV(Rt_1 \ldots t_n)$ is the set of all variables occurring in $t_1 \ldots t_n$;*
- *$FV(t = t')$ is the set of all variables occurring in t and t';*
- *$FV(=(t_1 \ldots t_n))$ is the set of all variables occurring in $t_1 \ldots t_n$;*
- *$FV(\neg\phi) = FV(\phi)$;*
- *$FV(\phi \vee \psi) = FV(\phi) \cup FV(\psi)$;*
- *$FV(\exists x\phi) = FV(\phi)\backslash\{x\}$.*

A formula ϕ is said to be a sentence *if and only if $FV(\phi) = \emptyset$.*

The semantics of Dependence Logic sentences can be defined in terms of semantic games, which are imperfect information generalizations of the usual semantic games for First Order Logic: in brief, for every First Order model M of signature Σ and for every Dependence Logic formula ϕ of the same signature Väänänen defined in [3] a two-player, zero-sum, finite, imperfect information game $H^M(\phi)$,

[2] IF-Logic terms are defined precisely as First Order Logic terms.

whose main difference from the usual semantic games for First Order Logic ([9]) is the following *uniformity condition*[3] over Player II's strategies:

– Let τ be a strategy for Player II in $H^M(\phi)$. Then τ is *uniform* if and only if for any two plays $\overline{p} = p_1 \ldots p_n$, $\overline{q} = q_1 \ldots q_{n'}$ of $H^M(\phi)$ such that
  1. Player II follows τ in both $\overline{p}$ and $\overline{q}$;
  2. p_n and $q_{n'}$ correspond to the same occurrence of the same atomic dependence formula $=(t_1 \ldots t_n)$;
  3. Player II is the active player in both p_n and $q_{n'}$;
  4. The current assignments in p_n and $q_{n'}$ are respectively s and s';
  5. The interpretations of the terms $t_1 \ldots t_{n-1}$ in s coincide with the interpretations of the same terms in s',

  the interpretation of t_n in s coincides with the interpretation of t_n in s'.

Then, as in the First Order case, one may define the truth conditions of IF-Logic sentences in terms of the existence of winning strategies:

Definition 2 (Truth in Dependence Logic). *Let M be a First Order model of signature Σ and let $\phi \in D$ be a sentence in the same signature. Then we say that ϕ is* true *in M, and we write $M \models \phi$, if and only if Player II has a winning strategy in $H^M(\phi)$.*

For some time, it was an open problem whether it was possible to develop a natural compositional semantics which coincided with the usual Game Theoretic Semantics for a logic of imperfect information: in particular, Hintikka stated in [2] that

> *... there is no realistic hope of formulating compositional truth-conditions for [IF-Logic], even though I have not given a strict impossibility proof to that effect.*

This conjecture was proved false in [4]: the elegant, compositional *trump semantics* introduced by Wilfrid Hodges for a variant of IF-Logic gives the same truth conditions for sentences than the Game Theoretic Semantics for the same logic, and furthermore it is easily adaptable to other logics of imperfect information. A complete definition of this semantics can be found in Hodges' above mentioned paper, and its adaptation to Dependence Logic is in [3]. Here we will only mention that this semantics is obtained by defining two inductive satisfaction relations $\models^+$ and $\models^-$ between First Order models, Dependence Logic formulas and *teams*, that is, sets of assignments[4].

[3] This condition is the reason why the $H^M(\phi)$ are games of imperfect information, whereas the corresponding games for First Order Logic are games of perfect information. However, it must be stressed that not all game semantics for Dependence Logic or any expressively equivalent logic need to be of imperfect information: for example, in [3] Väänänen developed a perfect information game semantics for Dependence Logic which is equivalent to the usual imperfect information one, and it is an easy exercise to adapt this semantics to the DF-Logic and IF-Logic cases.

[4] This notion of team was introduced by Väänänen; Hodges talks instead of sets of tuples of elements of the model.

In particular, given a formula ϕ one may identify the meaning of ϕ according to this semantics with the pair

$$\|\phi\|_M = (\{X : FV(X) = Dom(\phi), M \models^+_X \phi\}, \{X : Dom(X) = FV(\phi), M \models^-_X \phi\}).$$

If we identify each team X with the corresponding set of tuples, we can say (after Hodges) that the meaning of a Dependence Logic formula with k free variables is a *double k-suit*, that is, a pair $(\mathcal{A}, \mathcal{B})$ of sets of sets of k-tuples such that

- $\mathcal{A} \cap \mathcal{B} = \{\emptyset\}$;
- If $X \in \mathcal{A}$ then $X' \in \mathcal{A}$ for all $X' \subseteq X$;
- If $Y \in \mathcal{B}$ then $Y' \in \mathcal{B}$ for all $Y' \subseteq Y$.

The following result, proved by Väänänen in [3], is the adaptation to Dependence Logic of the corresponding result for IF-Logic proved by Hodges in [4]:

Theorem 1. *Let M be a model with signature Σ and let ϕ be an Dependence Logic sentence with the same signature. Then $M \models \phi$ according to the Game Theoretic Semantics if and only if $M \models^+_{\{\emptyset\}} \phi$ according to the above defined Team Semantics.*

Corollary 1. *Let M, Σ and ϕ be as above. Then $M \models \phi$ if and only if $\|\phi\|_M = (\{\emptyset, \{\emptyset\}\}, \{\emptyset\})$.*

Proof. Follows at once from the above theorem, from the fact that both "sides" of a double suit are downwards closed and from the fact that their intersection is precisely $\{\emptyset\}$.

Finally, it is worth mentioning that Hodges' semantics is also *fully abstract*:[5]

Proposition 1. *Let ψ and ψ' be two Dependence Logic formulas with $FV(\psi) = FV(\psi')$. Then the following are equivalent:*

1. *$\|\psi\|_M = \|\psi'\|_M$ for all models M in which ψ may be interpreted;*
2. *For every sentence ϕ, if ϕ' is obtained from ϕ by replacing an occurrence of ψ as a subformula of ϕ with an occurrence of ψ' then $\|\phi\|_M = \|\phi'\|_M$ for all models M in which ϕ may be interpreted.*

1.2 The Combinatorics of Imperfect Information

As summarized in the previous subsections, Wilfrid Hodges developed a compositional semantics for logics of imperfect information in which the meaning-carrying objects of such semantics are Double Suits, that is, pairs of downward-closed sets of sets of assignments which intersect only in the empty set of assignment.

[5] As for many of the previous results, the proof is in [4] and it refers to IF-Logic rather than to Dependence Logic. Adapting the proof to the latter formalism poses no difficulty whatsoever.

As Hodges showed in [10], the choice of these kinds of objects comes, in a very natural way, from a careful analysis of the Game Theoretic Semantics for IF-Logic; but is it possible to find an equivalent semantics whose meaning-carrying entities are simpler? In particular, is it possible to find such a semantics in which meanings are *sets of assignments*, as in the case of Tarski's semantics for First Order Logic?

In [1], a negative answer to this question was found, and the corresponding argument will now be briefly reported. In that paper, Cameron and Hodges introduced the concept of "adequate semantics" for IF-Logic, which can be easily adapted to Dependence Logic:

Definition 3 (Adequate semantics). *An* adequate semantics *for Dependence Logic is a function μ that associates to each pair (ϕ, M), where ϕ is a formula and M is a model whose signature includes that of ϕ, a value $\mu_M(\phi)$, and that furthermore satisfies the following two properties:*

1. *There exists a value TRUE such that, for all sentences ϕ and all models M, $\mu_M(\phi)$ = TRUE if and only if $M \models \phi$ (according to the Game Theoretic Semantics);*
2. *For any two formulas ϕ, ψ and for any sentence χ and any model M such that $\mu_M(\phi) = \mu_M(\psi)$, if χ' is obtained from χ by substituting an occurrence of ϕ in χ with one occurrence of ψ then*

$$\mu_M(\chi) = TRUE \Leftrightarrow \mu_M(\chi') = TRUE.$$

The first condition states that the semantics μ coincides with the game semantics on sentences, and the second one is a very weak notion of compositionality (which is easily verified to be implied by compositionality in the frameworks of both [11] and [12], the latter of whom can be seen as a descendant of that of [13]).

They also proved the following result:

Proposition 2. *Let $g(n)$ be the number of double* 1*-suits with domain over a model M with n elements. Then*

$$g(n) \in \Omega\left(2^{2^n/(\sqrt{\pi \lfloor n/2 \rfloor})}\right)$$

Cameron and Hodges then verified that there exist finite models in which every double 1-suit corresponds to the interpretation of a formula with one free variable, and hence that[6]

Proposition 3. *Let μ be an adequate semantics for Dependence Logic, let x be any variable, and let $n \in \mathbb{N}$. Then there exists a model A_n with n elements, such that*

$$|\{\mu_{A_n}(\phi(x)) : FV(\phi) = \{x\}\}| \geq g(n).$$

Furthermore, the signature of A_n contains only relations.

From this and from the previous proposition, they were able to conclude at once that, for any $k \in \mathbb{N}$, there exists no adequate semantics (and, as a consequence,

[6] Again, Cameron and Hodges' results refer to IF-Logic rather than to Dependence Logic, but it is easy to see that their arguments are still valid in the Dependence Logic case.

no compositional semantics) μ such that $\mu_M(\phi)$ is a set of k-tuples whenever $FV(\phi) = \{x\}$: indeed, the number of sets of k-tuples of assignments in a model with n elements is $2^{(n^k)}$, and there exists a $n_0 \in \mathbb{N}$ such that $g(n_0) > 2^{(n_0^k)}$. Then, since μ is adequate we must have that $|\{\mu_{A_{n_0}}(\phi(x)) : FV(\phi) = \{x\}\}| \geq g(n_0) > 2^{n_0^k}$, and this contradicts the hypothesis that μ interprets formulas with one free variables as k-tuples.

However, as Cameron and Hodges observe, this argument does not carry over if we let M range only over infinite structures: indeed, in Dependence Logic (or in IF-Logic) there only exist countably many classes of formulas modulo choice of predicate symbols[7], and therefore for every model A of cardinality $\kappa \geq \aleph_0$ there exist at most $\omega \cdot 2^\kappa = 2^\kappa$ distinct interpretations of IF-Logic formulas in A. Hence, there exists an injective function from the equivalence classes of formulas in A to 1-tuples of elements of A, and in conclusion there exists a semantics which encodes each such congruence class as a 1-tuple.

Cameron and Hodges then conjectured that there exists no reasonable way to turn this mapping into a semantics for IF-Logic:

> *Common sense suggests that there is no sensible semantics for [IF-Logic] on infinite structures A, using subsets of the domain of A as interpretations for formulas with one free variable. But we don't know a sensible theorem along these lines.*

What I will attempt to do in the rest of this work is to give a precise, natural definition of "sensible semantics" according to which Cameron and Hodges' conjecture may be turned into a formal proof: even though, by the cardinality argument described above, it is possible to find a compositional semantics for IF-Logic assigning sets of elements to formulas with one free variable, it will be proved that it is not possible for such a semantics to be also "sensible" according to this definition.

Furthermore, we will also verify that this property is satisfied by Hodges' Trump Semantics, by Tarski's Semantics for First Order Logic and by Kripke's semantics for Modal Logic: this, in addition to the naturalness (at least, according to the author's intuitions) of this condition, will go some way in suggesting that this is a property that we may wish to require any formal semantics to satisfy.

2 Sensible Semantics

2.1 Sensible Semantics of Imperfect Information

Two striking features of Definition 3. are that

1. The class $\mathcal{M}$ of all First Order models is not used in any way other than as an *index class* for the semantic relation: no matter what relation exists between two models M and N, no relation is imposed between the functions

[7] This is not the same of *countably many formulas*, of course, since the signature might contain uncountably many relation symbols.

μ_M and μ_N. Even if M and N were isomorphic, nothing could be said in principle about the relationship between $\mu_M(\phi)$ and $\mu_N(\phi)$!

2. The second part of the definition of adequate semantics does not describe a property of the semantics μ itself, but rather a property of the *synonymy modulo models* relation that it induces. This also holds for the notion of compositionality of [11], albeit not for that of [12]; in any case, in neither of these two formalisms morphisms between models are required to induce morphisms between the corresponding "meaning sets", and in particular isomorphic models may well correspond to non-isomorphic meaning sets.

These observations justify the following definition:

Definition 4 (Sensible Semantics). *Let L be a* partial algebra *representing the syntax of our logic*[8] *for some fixed signature*[9] *and let $\mathcal{M}$ be the category of the* models *of L for the same signature*[10]*. Then a* sensible semantics *for it is a triple $(\mathcal{S}, Me, \mu)$, where*

- *$\mathcal{S}$ is a subcategory of the category Set of all sets;*
- *Me is a functor from $\mathcal{M}$ to $\mathcal{S}$;*
- *For every $M \in \mathcal{M}$, μ_M is a function from L to $S_M = Me(M) \in \mathcal{S}$, called the* meaning set *for L in $\mathcal{S}$*

and such that

1. *For all $\phi, \psi, \chi \in L$ and for all $M \in \mathcal{M}$, if $\mu_M(\phi) = \mu_M(\psi)$ and χ' is obtained from χ by substituting an occurrence of ϕ as a subterm of χ with an occurrence of ψ, then $\chi' \in L$ and $\mu_M(\chi) = \mu_M(\chi')$;*
2. *If $f : M \to N$ is an isomorphism between two models $M, N \in \mathcal{M}$, then $\mu_N = \mu_M \circ Me(f)$ for all formulas $\phi \in L$.*

The first condition is, again, a weak variant of compositionality, plus a version of the *Husserl Property* of [11]: if two formulas have the same interpretation in a model M then the operation of substituting one for the other sends grammatical expressions into grammatical expressions with the same interpretation in M. One could strengthen this notion of compositionality after the fashion of [12], by imposing an algebraic structure over each set S_M with respect to the same signature of L and by requiring each μ_M to be an homomorphism between L and M, but as this is not necessary for the purpose of this work we will content ourselves with this simpler statement.

The second condition, instead, tells us something about the way in which isomorphisms between models induce isomorphisms between formula meanings, that is, that the diagram of Figure 1 commutes whenever f is an isomorphism: if

[8] That is, the objects of L are the well-formed formulas of our logic and the operations of L are its formation rules.

[9] If the notion of signature is applicable to the logic we are studying; otherwise, we implicitly assume that all models and formulas have the same empty signature.

[10] The choice of morphisms in $\mathcal{M}$ is supposed to be given, and to be part of our notion of model for the semantics which is being considered.

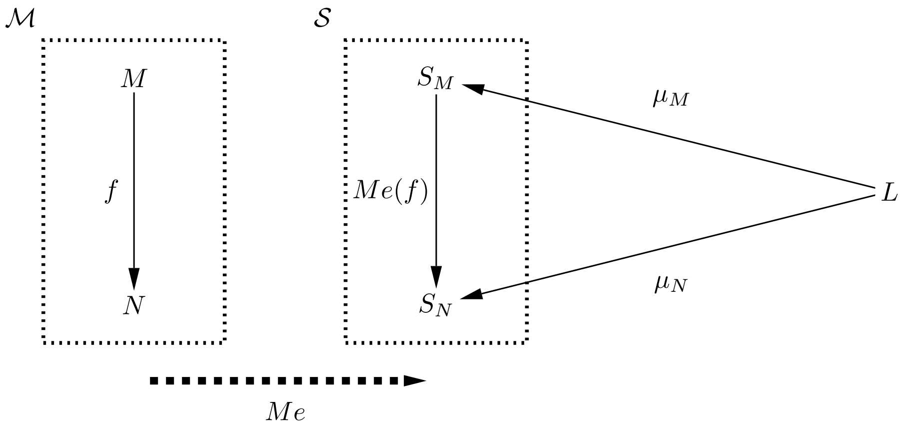


Fig. 1. Diagram representation of Condition 2 of Definition 4 (Sensible Semantics): if $f : M \to N$ is an isomorphism then $\mu_N(\phi) = Me(f)(\mu_N(\phi))$ for all formulas $\phi \in L$

M and N are isomorphic through f then the interpretation $\mu_N(\phi)$ of any formula ϕ in the model N can be obtained by taking the interpretation $\mu_M(\phi) \in S_M$ of ϕ in M and applying the "lifted isomorphism" $Me(f) : S_M \to S_N$.

Before applying this definition to the case of Dependence Logic, let us verify its naturality by checking that it applies to a couple of very well-known logics with their usual semantics, as well as to Dependence Logic with Trump Semantics:

Proposition 4. *Let FO be the language of First Order Logic (for some signature Σ which we presume fixed), and let $\mathcal{M}$ be the category of all First Order models for the same signature.*

Furthermore, for every $M \in \mathcal{M}$ let S_M be the disjoint union, for k ranging over $\mathbb{N}$, of all sets of k-tuples of elements of M[11] and let Me be such that $Me(M) = S_M$ for all $M \in \mathcal{M}$ and

$$Me(f)(H) = f_{\uparrow}(H) = \{(f(m_1) \dots f(m_k)) : (m_1 \dots m_k) \in H\} \tag{1}$$

for all $f : M \to N$ and all $H \in S_M$.

Now, let μ be the usual Tarski semantics, that is, for every model M and formula $\phi(x_1 \dots x_k)$ with $FV(\phi) = \{x_1 \dots x_k\}$ let

$$\mu_M(\phi(x_1 \dots x_k)) = \{(m_1 \dots m_k) \in M^K : M \models_{(x_1:m_1 \dots x_k:m_k)} \phi(x_1 \dots x_k)\}.$$

Then $(\mathcal{S}, Me, \mu)$ is a sensible semantics for the logic $(FO, \mathcal{M})$.

Proof. The first condition is an obvious consequence of the compositionality of Tarski's semantics: if $\Phi[\phi]$ is a well-formed formula, ϕ is equivalent to ψ in the model M and $FV(\phi) = FV(\psi)$ then $\Phi[\psi]$ is also a well-formed formula and it is equivalent to $\Phi[\phi]$ in M.

[11] In particular, this definition implies that S_M contains distinct "empty sets of k-tuples" for all $k \in \mathbb{N}$.

For the second one, it suffices to observe that if $f : M \to N$ is an isomorphism then

$$M \models_s \phi \Leftrightarrow N \models_{f \circ s} \phi \tag{2}$$

for all assignments s and all First Order formulas ϕ.

Mutatis mutandis, the same holds for Kripke's Semantics for Modal Logic:

Proposition 5. *Let ML be the language of modal logic and let $\mathcal{M}$ be the category of all Kripke models $M = (W, R, V)$, where W is the set of possible worlds, R is a binary relation over W and V is a* valutation function *from atomic propositions to subsets of W. Furthermore, for any $M = (W, R, V) \in \mathcal{M}$ let S_M be the powerset $\mathcal{P}(W)$ of W, and, for every $f : M \to N$, let $Me(f) : S_M \to S_N$ be such that*

$$Me(f)(X) = \{f(w) : w \in X\}$$

for all $X \subseteq W$.

Finally, let μ be Kripke's semantics choosing, for each model $M = (W, R, V)$ and each modal formula ϕ, the set $\mu_M(\phi) = \{w \in W : M \models_w \phi\}$: then $(\mathcal{S}, Me, \mu)$ is a sensible semantics for $(ML, \mathcal{M})$.

Proof. Again, the first part of the definition is an easy consequence of the compositionality of μ. For the second part, it suffices to observe that, if $f : M \to N$ is an isomorphism between Kripke models,

$$M \models_w \phi \Leftrightarrow N \models_{f(w)} \phi$$

for all w in the domain of M, as required.

Finally, Hodges' trump semantics for Dependence Logic, whose meaning sets are the disjoint unions over $k \in \mathbb{N}$ of the sets of all double k-suits, is also sensible: indeed, for all isomorphisms $f : M \to N$, all sets of k-tuples X and all formulas $\phi(x_1 \dots x_k)$, $M \models_X \phi(x_1 \dots x_k)$ if[12] and only if $N \models_{f_\uparrow(X)} \phi(x_1 \dots x_k)$, where $f_\uparrow$ is defined as in Equation 1.

Let us now get to the main result of this work. First, we need a simple lemma:

Lemma 1. *Let $(\mathcal{S}, Me, \mu)$ be a sensible semantics for $(D, \mathcal{M})$, where D is the language of Dependence Logic (seen as a partial algebra) and $\mathcal{M}$ is the category of all First Order models. Suppose, furthermore, that TRUE is a distinguished value such that $\mu_M(\phi) = TRUE$ if and only if $M \models \phi$ for all models M and sentences ϕ. Then μ is an adequate semantics for Dependence Logic.*

Proof. Obvious from Definition 3. and Definition 4.

[12] Here we are committing a slight abuse of notation, since X is a set of tuples rather than a set of assignments. The intended interpretation is that each tuple $(m_1 \dots m_k) \in X$ corresponds to the assignment s such that $s(x_1) = m_1, \dots, s(x_k) = m_k$.

Theorem 2. *Let $\mathcal{M}$ be the class of all infinite models for a fixed signature, let S_M be the set of all sets of k-tuples of elements of M (for all k), and for every $f : M \to N$ let $Me(f)$ be defined as*

$$Me(f)(X) = \{f_{\uparrow}(s) : s \in X\}.$$

for all sets of tuples $X \in S_M$.

Then, for every $k \in \mathbb{N}$, there exists no function μ such that

1. *For all models M and formulas $\phi(x)$ with only one free variable, $\mu_M(S)$ is a set of k-tuples;*
2. *$M \models \phi \Leftrightarrow \mu_M(\phi) = TRUE$ for all $M \in \mathcal{M}$, for all $\phi \in D$ and for some fixed value $TRUE$;*
3. *$(\mathcal{S}, Me, \mu)$ is a sensible semantics for Dependence Logic with respect to $\mathcal{M}$.*

Proof. Suppose that such a μ exists for some $k \in \mathbb{N}$: then, by Lemma 1, μ is an adequate semantics for Dependence Logic.

Let $g(n)$ be the number of double suits in a finite model M with n elements, let $h(n) = 2^{2(nk)^k}$, and let n_0 be the least number (whose existence follows from Proposition 2.) such that $g(n_0) > h(n_0)$. Furthermore, let A_{n_0} be the relational model with n_0 elements, defined as in Proposition 3., for which Cameron and Hodges proved that any compositional semantics for Dependence Logic must assign at least $g(n_0)$ distinct interpretations to formulas with exactly one free variable x.

Now, let the infinite model B_{n_0} be obtained by adding countably many new elements $\{b_i : i \in \mathbb{N}\}$ to A_{n_0}, by letting $R^{B_{n_0}} = R^{A_{n_0}}$ for all relations R in the signature of A_{n_0} and by introducing a new unary relation P with $P^{B_{n_0}} = Dom(A_{n_0})$.

It is then easy to see that, with respect to B_{n_0}, our semantics must assign at least $g(n_0)$ different meanings to formulas ϕ with $FV(\phi) = \{x\}$: indeed, if $\phi^{(P)}$ is the relativization of ϕ with respect to the predicate P we have that

$$\mu_{A_{n_0}}(\phi) = \mu_{A_{n_0}}(\psi) \Leftrightarrow \mu_{B_{n_0}}(\phi^{(P)}) = \mu_{B_{n_0}}(\psi^{(P)}),$$

and we already know that $|\{\mu_{A_{n_0}}(\phi) : FV(\phi) = \{x\}\}| = g(n_0)$.

Now, suppose that μ is sensible and $\mu_{B_{n_0}}(\phi)$ is a set of k-tuples for every formula $\phi(x)$: then, since every permutation $\pi : B_{n_0} \to B_{n_0}$ that pointwise fixes the element of A_{n_0} is an automorphism of B_{n_0}, we have that

$$Me(\pi)(\mu_{B_{n_0}}(\phi)) = \mu_{B_{n_0}}(\phi)$$

for all such π.

But then $|\mu_{B_{n_0}}(\phi) : FV(\phi) = \{x\}| \leq h(n_0)$, since there exist at most $2^{2(n_0k)^k}$ equivalence classes of tuples with respect to the relation

$$\overline{b} \equiv \overline{b}' \Leftrightarrow \exists f : B_{n_0} \to B_{n_0}, f \text{ automorphism, s.t. } f_{\uparrow}b = c.$$

Indeed, one may represent such an equivalence class by first specifying whether it contains any element of A_{n_0}, then listing without repetition all elements of

A_{n_0} occurring in $\overline{b}$, padding this into a list $\overline{m}$ to a length of k by repeating the last element, and finally encoding each item b_i of $\overline{b}$ as an integer t_i in $1 \dots k$ in such a way that

- If $b_i \in A_{n_0}$, $m_{t_i} = \overline{b}_i$ and $\overline{m}_{t_i-1} \neq \overline{m}_{t_i}$ whenever $t_i > 0$;
- If $b_i \notin A_{n_0}$, $\overline{m}_{t_i} = \overline{m}_{t_i-1}$ whenever $t_i > 0$;
- $t_i = t_j$ if and only if $b_i = b_j$.

In total, this requires $1 + k\log(n_0) + k\log(k)$ bits, and therefore there exist at most $2(n_0k)^k$ such equivalence classes; and since each $\mu_{B_{n_0}}(\phi)$ is an union of these equivalence classes, there are at most $2^{2(n_0k)^k}$ possible interpretations of formulas with one free variable.

But this contradicts the fact that $g(n_0) > h(n_0)$, and hence no such semantics exists.

References

1. Cameron, P., Hodges, W.: Some Combinatorics of Imperfect Information. The Journal of Symbolic Logic 66(2), 673–684 (2001)
2. Hintikka, J.: The Principles of Mathematics Revisited. Cambridge University Press, Cambridge (1996)
3. Väänänen, J.: Dependence Logic. Cambridge University Press, Cambridge (2007)
4. Hodges, W.: Compositional Semantics for a Language of Imperfect Information. Journal of the Interest Group in Pure and Applied Logics 5 (4), 539–563 (1997)
5. Henkin, L.: Some Remarks on Infinitely Long Formulas. In: Infinitistic Methods. Proc. Symposium on Foundations of Mathematics, pp. 167–183. Pergamon Press, Oxford (1961)
6. Hintikka, J., Sandu, G.: Informational independence as a semantic phenomenon. In: Fenstad, J., Frolov, I., Hilpinen, R. (eds.) Logic, Methodology and Philosophy of Science, pp. 571–589. Elsevier, Amsterdam (1989)
7. Janssen, T.: Independent Choices and the Interpretation of IF Logic. Journal of Logic, Language and Information 11, 367–387 (2002)
8. Caicedo, X., Dechesne, F., Jannsen, T.: Equivalence and Quantifier Rules for Logic with Imperfect Information. Logic Journal of the IGPL 17(1), 91–129 (2009)
9. Hintikka, J., Sandu, G.: Game-Theoretical Semantics. In: van Benthem, J., Meulen, A.T. (eds.) Handbook of Logic and Language, pp. 361–410. Elsevier, Amsterdam (1997)
10. Hodges, W.: Logics of imperfect information: why sets of assignments. In: van Benthem, J., Gabbay, D., Löwe, B. (eds.) Proceedings of 7th De Morgan Workshop 'Interactive Logic: Games and Social Software (2005)
11. Hodges, W.: Formal features of compositionality. Journal of Logic, Language, and Information 10(1), 7–28 (2001)
12. Janssen, T.: Compositionality. In: van Benthem, J., ter Meulen, A. (eds.) Handbook of Logic and Language, pp. 417–473. Elsevier, Amsterdam (1996)
13. Montague, R.: Universal grammar. Theoria 36, 373–398 (1970)

A Qualitative Approach to Uncertainty

Sujata Ghosh[1,⋆] and Fernando R. Velázquez-Quesada[2]

[1] Department of Artificial Intelligence
University of Groningen
sujata@ai.rug.nl
[2] Institute for Logic, Language and Computation
University of Amsterdam
F.R.VelazquezQuesada@uva.nl

Abstract. We focus on modelling dual epistemic attitudes (belief-disbelief, knowledge-ignorance, like-dislike) of an agent. This provides an interesting way to express different levels of uncertainties explicitly in the logical language. After introducing a *dual* modal framework, we discuss the different possibilities of an agent's attitude towards a proposition that can be expressed in this framework, and provide a preliminary look at the dynamics of the situation.

1 Introduction

Real life is not boolean. Many of the situations, events and concepts that we face in our everyday life cannot be classified in a simple binary hierarchy. When we wake up, the morning may not be dark with thunder clouds hovering, but it may not be sunny too (many-valuedness); we believe that our favorite team will win tonight's soccer match, but we still consider the possibility of a loss or even a tie (modality); the dice which is about to be cast will result in any of the six different outcomes (probabilistic); the director of the company seems to be young, but we are not sure (fuzziness). Thus, uncertainty arises in various forms and various connotations. What we will deal with here is not vagueness or impreciseness of concepts (which leads to fuzzy logic, many-valued logic; see [1]), but rather uncertainties in agent's attitudes (e.g. beliefs).

The main aim of this paper is to deal with negative attitudes of agents on a par with their positive attitudes; this will allow us to describe different levels of uncertainties in a more classical framework. Though there are previous works modelling these dual attitudes (e.g. [2,3]), the novelty of this paper lies in the fact that we are expressing different levels of agent uncertainty explicitly in the logical language.

We propose a bi-modal framework that allows us to express various kinds of attitudes toward a formula φ. Our work is based on the following observations. First, the modal operator $\Box$ allows us to express what is true in all worlds reachable by the accessibility relation. Such propositions, which are true in those

⋆ Acknowledges NWO grant # 600.065.120.08N201.

M. Banerjee and A. Seth (Eds.): ICLA 2011, LNAI 6521, pp. 90–104, 2011.

accessible worlds, may be false elsewhere. The worlds where the propositions are false should be considered, if we want to deal with attitudes like disbelief, dislike or ignorance of agents. Secondly, every interpretation of the accessibility relation (knowledge, belief, like) has its *dual* concept (ignorance, disbelief, dislike, respectively), which is usually not represented in the framework[1]. The reason that they are not expressed in the existing literature is that, usually it is assumed that these two concepts are complementary (that is, representable by negations). In general, this does not have to be the case: the fact that we dislike the rain does not imply that we like when it is not raining. On the other hand, suppose we believe that a horse will win a race if its rating is above certain number, and disbelieve it if its below some other number. "Disbelieving" will not equate to "not believing" in this case.

Let us consider a fair 100 ticket lottery. Though we do believe that exactly one of the tickets will win the lottery, we have doubts regarding the possibility of an individual ticket winning it. One can resolve this paradoxical situation by replacing the classical negation (believing that the ticket number 99 will not win) by the weaker notion of disbelief (disbelieving that the ticket number 99 will win). This applies to many practical situations as well. A crime has been committed and two of your very good friends are the prime suspects. It is really hard for you to believe that any one of them has committed the crime, yet the circumstantial evidence forces you to believe that either of them did it.

From the perspective of belief merging, consider k sources of information providing their opinions regarding a certain event p. Suppose that m of them state that p holds, and n of them state that p does not hold. Some of the sources may not have any opinion regarding p, but none of the sources are inconsistent in the sense that they do not simultaneously state that p holds and does not hold. So we have that $m + n \leq k$. The fraction m/k can be seen as the degree of certainty of the source that p holds and n/k that ¬p holds. Let $cr(p) \in [0, 1]$ denote the degree of certainty that p holds. We can think of threshold values t_1 and t_2 ($0 < t_1 \leq t_2 \leq 1$) for belief and disbelief, that is, p is believed if $t_2 \leq cr(p)$, and is disbelieved if $cr(p) < t_1$. In the remaining cases p is neither believed nor disbelieved. So, from the fact the p is not believed, we cannot say that p is disbelieved. Also, from the fact that p is disbelieved, it does not necessarily follow that $\neg p$ is believed. [2]

As exemplified above, considering the dual of positive notions of knowledge, belief and others are important in their own right. Moreover, this also opens the door for modeling different kinds of uncertain attitudes that arise, when the agent is placed in decision making scenarios.

Consider the following decision-making scenario. Suppose Alice has to elect a member of parliament from her constituency for the next five year term. Several candidates are in the fray, say, A, B, C, D, E. Generally, these situations are modelled by the notion of *preference*. But we feel that, while considering the agent's preference over candidates, the intricate feelings of uncertainties that

[1] Here we do not refer to the dual of the modal operator $\Box$, which is $\Diamond$.

[2] These examples are taken from [2,4].

the agent may have regarding the candidates get lost. Let I denote that I is a good candidate for the task at hand. Alice may be undecided about A and B, has strong positive opinion about C, negative opinion about D, and have not even heard of E. By introducing the dual framework (cf. Section 3), we will be able to model all the intricate levels of Alice's uncertainties.

2 Some Relevant Issues: A Prelude

In this section, we discuss different issues leading up to our proposal. We start by providing a systematic sketch about the attitudes expressible by a single modal operator. Then we give a brief overview of the relevant literature dealing with positive and negative sides of information, and finally, we provide a brief sketch about various approaches to uncertainty.

2.1 Attitudes Representable with a Single Modal Operator

Modal logic, one of the simplest intensional logics, allows us to express attitudes by means of the modal operator $\Box$. Formulas of its language are built from a set of atomic propositions $\mathtt{P}$ by closing it under negation $\neg$, conjunction $\wedge$ and the modal operator $\Box$. Formulas of the form $\Box\varphi$ can be read as *"the agent has a positive attitude towards φ"*, and $\Diamond\varphi$ is an abbreviation for $\neg\Box\neg\varphi$.

Such formulas are evaluated in Kripke models. Formulas of the form $\Box\varphi$ and $\Box\neg\varphi$ can be true or false, giving us four possible attitudes towards φ. These possibilities can be ordered according to the amount of information about φ each one provides, that is, the number of true formulas in the set $\{\Box\varphi, \Box\neg\varphi\}$. This yields the power set of a 2-element set ordered by inclusion:

$$\begin{tikzcd}
& \Box\varphi \wedge \Box\neg\varphi & \\
\neg\Box\varphi \wedge \Box\neg\varphi \arrow[ur] & & \Box\varphi \wedge \neg\Box\neg\varphi \arrow[ul] \\
& \neg\Box\varphi \wedge \neg\Box\neg\varphi \arrow[ul] \arrow[ur] &
\end{tikzcd}$$

The case with $\Box\varphi$ and $\Box\neg\varphi$ is undesirable under some interpretations (eg. knowledge) of $\Box$ since it represents having positive attitude towards both a formula and its negation (what we will call $\Box$-inconsistency). To avoid this case, it is usually asked for the accessibility relation in the Kripke model to be at least *serial* (in knowledge interpretations, the stronger *reflexivity* is required), and we are left with three possible attitudes which can also be ordered according to the attitude of the agent about φ: from positive towards $\neg\varphi$ to positive attitude about φ (see the diagram below).

$$\begin{tikzcd}
\Box\varphi \wedge \neg\Box\neg\varphi \\
\neg\Box\varphi \wedge \neg\Box\neg\varphi \arrow[u] \\
\neg\Box\varphi \wedge \Box\neg\varphi \arrow[u]
\end{tikzcd}$$

Two of these three $\Box$-consistent cases express certainty (towards φ or $\neg\varphi$), and we are left with just one attitude to express uncertainty. This is not enough to differentiate between say, "*not interested and having no opinion*" or "*interested but indecisive*" about φ.

2.2 Bipolar Representation of Information

Prevalent modal approaches that deal with information and attitudes mostly focus on representing only the positive information an agent has about a subject. But there is also a complementary view: just as we have positive information about a subject, we can also have (independent) negative information about it.

The idea of representing both positive and negative aspects of a subject is not new. There are approaches with such proposals in many areas, like decision theory [5], argumentation theory [6], and many others (see [7] for an overview). The concept of *bipolarity* is precisely about this: an explicit handling of the positive and negative aspects in information [8]. It is based on the fact that, when taking a decision or weighing some possibilities, we consider not only the positive aspects of the available options, but also the negative ones.

From this perspective, frameworks that consider only the positive aspect can be seen as special situations in which the positive and the negative information are mutually exclusive and mirror images of each other: I consider p as good if and only if I consider $\neg p$ as bad. But this does not need to be the case: we can imagine a situation in which, though p is good, its negation $\neg p$ is not necessarily bad, and the notion of bipolarity allows us to deal with such cases. These are precisely the kind of situations that we are interested in, and the framework presented in Section 3 allows us to deal with them.

2.3 Approaches for Modeling Uncertainty

Considering such dual frameworks for positive and negative information paves the way for an in-depth study of qualitative representation of uncertainty (cf. Section 3). We now give a brief overview of the highly active research area of uncertainty-modeling. The later half of the past century witnessed several proposals for modeling uncertainty, with a focus to formalize human (commonsense reasoning). To mention a few of the relevant approaches, we can refer to fuzzy set theory [9], possibility theory [10], rough set theory [11], probabilistic approaches [12], and Dempster-Shafer theory [13,14]. These comprise quantitative ways of dealing with uncertainties, but there have been some qualitative approaches based on ordered sets [12]. Different kinds of propositional and first-order frameworks have also been proposed describing different interpretations of uncertainty, e.g. probability logics [15], fuzzy logics [16], possibilistic logics, [17], rough logics [18], and many-valued logics [19].

Evidently, there are varied approaches to deal with uncertainties, but mathematical structures play a very important role in these formulations. The logical languages describing the uncertainty concept generally remain the same, only their interpretations change in the different theories. Some of these theories are more suitable for describing vague or uncertain concepts, while others are more appropriate for describing mental attitudes of agents. Our interests lie in describing diverse mental attitudes of agents *more explicitly in the logical language*. To achieve such a goal, the following section presents a special kind of dual framework, viz. a logical language talking about positive and negative attitudes of agents regarding events. This will give us a way to talk about uncertainties of agents in a bi-modal framework.

3 A Language for Positive and Negative Attitudes

We now introduce an extension of the classical modal language that allows us to express both positive and negative attitudes *explicitly*. After presenting the language, models, and semantic interpretation, we show how, even by imposing strong *consistency* requirements, we still can express more attitudes than the classical modal framework.

3.1 The Basic System

Definition 3.1. *Let* $\mathtt{P}$ *be a set of atomic propositions. Formulas* φ *of the language* $\mathcal{L}$ *are given by*

$$\varphi ::= p \mid \neg\varphi \mid \varphi \wedge \varphi \mid [+]\varphi \mid [-]\varphi$$

with $p \in \mathtt{P}$*. Formulas of the form* $[+]\varphi$ *(*$[-]\varphi$*) are read as "the agent has a positive (negative) attitude towards* φ*". The corresponding 'diamond' modalities are defined as usual.*

Having both the positive and the negative attitudes represented explicitly allow us to express combinations of them. For example, in a discussion about preferences, the notions can be read as *like* and *dislike*, and then we can express situations like *the agent is undecided whether she likes (that is, she likes and dislikes) rainy day* ($[+]rd \wedge [-]rd$). In a doxastic context the notions can be read as *belief* and *disbelief*, and we can express ideas like *the agent does not have any opinion (neither believes nor disbelieves) regarding whether it will rain tomorrow* ($\neg[+]rt \wedge \neg[-]rt$). We can even interpret them in terms of knowledge, and express combinations of *knowledge* and *ignorance*.

Definition 3.2 (Dual model). *Given a set of atomic propositions* $\mathtt{P}$*, a* dual model *is a tuple* $\mathcal{M} = \langle W, R^+, R^-, V\rangle$ *where* W *is a non-empty set of worlds,* R^+ *and* R^- *are binary relations on* W *and* $V : \mathtt{P} \to \wp(W)$ *is a valuation function. We denote by* $\boldsymbol{M}$ *the class of all semantic models.*

The difference between our system and an ordinary bi-modal framework relies on the interpretation of negative attitude formulas $[-]\varphi$.

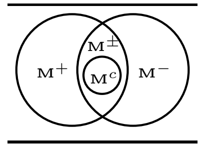
Definition 3.3. *Let* $\mathcal{M} = \langle W, R^+, R^-, V\rangle$ *be a dual semantic model and let* w *be a world in it. Atomic propositions, negation and conjunction are interpreted as usual. For the modalities, we have*

$$\begin{array}{lll} (\mathcal{M}, w) \models [+]\varphi & \textit{iff} & \textit{for all } w' \textit{ such that } R^+ww',\ (\mathcal{M}, w') \models \varphi \\ (\mathcal{M}, w) \models [-]\varphi & \textit{iff} & \textit{for all } w' \textit{ such that } R^-ww',\ (\mathcal{M}, w') \models \neg\varphi. \end{array}$$

Among the class of all dual semantic models, we distinguish four of them.

Definition 3.4. *A dual model $\mathcal{M}$ is in*

– *the class of* $[+]$-consistent *models* ($\mathbf{M}^+$) *iff* R^+ *is* serial*;*
– *the class of* $[-]$-consistent *models* ($\mathbf{M}^-$) *iff* R^- *is* serial*;*
– *the class of* $[\pm]$-consistent *models* ($\mathbf{M}^\pm$) *iff* R^+ *and* R^- *are* serial*;*
– *the class of* dual-consistent *models* ($\mathbf{M}^c$) *iff for every* $w \in W$ *we have* $R^+[w] \cap R^-[w] \neq \emptyset$.



It can be easily verified that in $\mathbf{M}^+$-models, the formula $[+](\varphi \wedge \neg\varphi)$ is not satisfiable, hence the name of the class. Similarly, $[-](\varphi \vee \neg\varphi)$ is not satisfiable in $\mathbf{M}^-$-models, and both formulas are not satisfiable in $\mathbf{M}^\pm$-models. The class $\mathbf{M}^c$ is more restrictive than $\mathbf{M}^\pm$ since, besides having both R^+ and R^- serial, their intersection should be non-empty. In models of such class, the two mentioned formulas are not satisfiable, and so is $[+]\varphi \wedge [-]\varphi$.

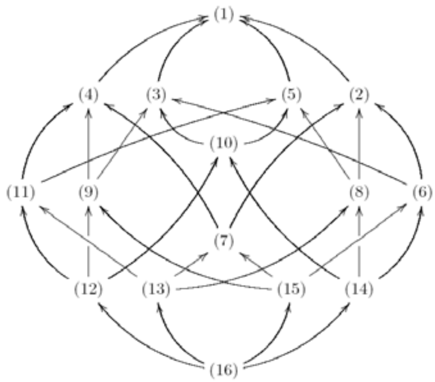
3.2 The New Attitudes

In a language with a single normal modality we can only express four attitudes with respect to any given formula φ, as discussed earlier. With our new modality $[-]$ we have sixteen combinations of truth-values for the formulas $[+]\varphi$, $[+]\neg\varphi$, $[-]\varphi$ and $[-]\neg\varphi$, all of them satisfiable in $\mathbf{M}$. To make a systematic analysis, we consider all the cases.

1:	$([+]\varphi \wedge [+]\neg\varphi) \wedge ([-]\varphi \wedge [-]\neg\varphi)$	2:	$(\neg[+]\varphi \wedge [+]\neg\varphi) \wedge ([-]\varphi \wedge [-]\neg\varphi)$
3:	$([+]\varphi \wedge \neg[+]\neg\varphi) \wedge ([-]\varphi \wedge [-]\neg\varphi)$	4:	$([+]\varphi \wedge [+]\neg\varphi) \wedge (\neg[-]\varphi \wedge [-]\neg\varphi)$
5:	$([+]\varphi \wedge [+]\neg\varphi) \wedge ([-]\varphi \wedge \neg[-]\neg\varphi)$	6:	$(\neg[+]\varphi \wedge \neg[+]\neg\varphi) \wedge ([-]\varphi \wedge [-]\neg\varphi)$
7:	$(\neg[+]\varphi \wedge [+]\neg\varphi) \wedge (\neg[-]\varphi \wedge [-]\neg\varphi)$	8:	$(\neg[+]\varphi \wedge [+]\neg\varphi) \wedge ([-]\varphi \wedge \neg[-]\neg\varphi)$
9:	$([+]\varphi \wedge \neg[+]\neg\varphi) \wedge (\neg[-]\varphi \wedge [-]\neg\varphi)$	10:	$([+]\varphi \wedge \neg[+]\neg\varphi) \wedge ([-]\varphi \wedge \neg[-]\neg\varphi)$
11:	$([+]\varphi \wedge [+]\neg\varphi) \wedge (\neg[-]\varphi \wedge \neg[-]\neg\varphi)$	12:	$([+]\varphi \wedge \neg[+]\neg\varphi) \wedge (\neg[-]\varphi \wedge \neg[-]\neg\varphi)$
13:	$(\neg[+]\varphi \wedge [+]\neg\varphi) \wedge (\neg[-]\varphi \wedge \neg[-]\neg\varphi)$	14:	$(\neg[+]\varphi \wedge \neg[+]\neg\varphi) \wedge ([-]\varphi \wedge \neg[-]\neg\varphi)$
15:	$(\neg[+]\varphi \wedge [+]\neg\varphi) \wedge (\neg[-]\varphi \wedge \neg[-]\neg\varphi)$	16:	$(\neg[+]\varphi \wedge \neg[+]\neg\varphi) \wedge (\neg[-]\varphi \wedge \neg[-]\neg\varphi)$

We explain these cases in terms of the mental attitudes they describe. Case 1 represents an agent's indecision about both $\neg\varphi$ and φ. Cases 2 and 3 correspond to the positive attitude towards $\neg\varphi$ and φ, respectively, whereas cases 4 and 5 give the corresponding negative attitudes. Case 6 corresponds to having no positive opinion whether φ holds, and case 11 to having no negative opinion. Case 7 corresponds to being undecided about $\neg\varphi$, case 10 about φ. Case 8 corresponds to positive about $\neg\varphi$, and negative about φ (strengthening of $\neg\varphi$), case 9 is just the opposite (strengthening of φ). Case 12 corresponds to having no negative attitude but only positive attitude towards φ, case 13 corresponds to having no negative attitude towards φ, and positive attitude towards $\neg\varphi$. Case 14 corresponds to having no positive attitude but only negative attitude towards φ, case 15 corresponds to no positive attitude towards φ, but negative attitude towards $\neg\varphi$. Case 16 corresponds to having no opinion whatsoever about φ.

We can order these 16 attitudes according to their informational content, just like we did with the four attitudes that can be expressed with a single normal modality (cf. Section 2.1). The order goes, again, from the case in which the four relevant formulas are false (case 16) to the case in which all of them are true (case 1); we get the power set of a 4-element set ordered by inclusion:



Of course, not all these cases are satisfiable in all classes of models:

- In $\mathbf{M}^+$-models, cases in $\{1, 4, 5, 11\}$ are *not* satisfiable.
- In $\mathbf{M}^-$-models, cases in $\{1, 2, 3, 6\}$ are *not* satisfiable.
- In $\mathbf{M}^\pm$-models, cases in $\{1, 2, 3, 4, 5, 6, 11\}$ are *not* satisfiable.
- In $\mathbf{M}^c$-models, *only* cases in $\{8, 9, 12, 13, 14, 15, 16\}$ are *satisfiable.*

In particular, the cases satisfiable in $\mathbf{M}^c$ can also be ordered according to the attitude towards a given formula, just like we did with the three $\Box$-consistent cases of the single normal modality (cf. Section 2.1). This time, the order goes from a completely negative attitude towards φ (case 8) to a completely positive attitude towards it (case 9), as the diagram on the right shows.

$$\begin{tikzcd}
 & 9 & \\
15 \arrow[ur] & & 12 \arrow[ul] \\
 & 16 \arrow[ul] \arrow[ur] & \\
14 \arrow[ur] & & 13 \arrow[ul] \\
 & 8 \arrow[ul] \arrow[ur] &
\end{tikzcd}$$

Moreover, considering the $\mathbf{M}$ models and the set of *uncertain* attitudes towards φ, viz. {1, 6, 7, 10, 11, 16}, we can have the following orderings of *indecision* and *no opinion*, in terms of increasing degrees of uncertainty. Note that these notions of uncertainty are independent of each other. The ordering for indecision and no opinion are as follows:

$$\begin{tikzcd}
 & 1 & \\
7 \arrow[ur] & & 10 \arrow[ul]
\end{tikzcd}
\qquad
\begin{tikzcd}
 & 16 & \\
6 \arrow[ur] & & 11 \arrow[ul]
\end{tikzcd}$$

indecision $\qquad$ $no - opinion$

We should note here that for all practical purposes, the above representations are too long to comprehend. We consider their reduced versions corresponding to the $\mathbf{M}^\pm$ models:

1: $\big([+]\varphi \wedge [+]\neg\varphi\big) \wedge \big([-]\varphi \wedge [-]\neg\varphi\big)$ 2: $\big(\neg[+]\varphi \wedge [+]\neg\varphi\big)$ 3: $\big([+]\varphi \wedge \neg[+]\neg\varphi\big)$
4: $\big(\neg[-]\varphi \wedge [-]\neg\varphi\big)$ 5: $\big([-]\varphi \wedge \neg[-]\neg\varphi\big)$ 6: $\big(\neg[+]\varphi \wedge \neg[+]\neg\varphi\big)$
7: $[+]\neg\varphi \wedge [-]\neg\varphi$ 8: $[+]\neg\varphi \wedge [-]\varphi$ 9: $[+]\varphi \wedge [-]\neg\varphi$

10: $[+]\varphi \wedge [-]\varphi$ 11: $(\neg[-]\varphi \wedge \neg[-]\neg\varphi)$ 12: $([+]\varphi \wedge \neg[+]\neg\varphi) \wedge (\neg[-]\varphi \wedge \neg[-]\neg\varphi)$
13: $(\neg[+]\varphi \wedge [+]\neg\varphi) \wedge (\neg[-]\varphi \wedge \neg[-]\neg\varphi)$ 14: $(\neg[+]\varphi \wedge \neg[+]\neg\varphi) \wedge ([-]\varphi \wedge \neg[-]\neg\varphi)$
15: $(\neg[+]\varphi \wedge [+]\neg\varphi) \wedge (\neg[-]\varphi \wedge \neg[-]\neg\varphi)$ 16: $(\neg[+]\varphi \wedge \neg[+]\neg\varphi) \wedge (\neg[-]\varphi \wedge \neg[-]\neg\varphi)$

We will have 15 different cases (case 1 being not satisfiable), preserving the corresponding attitudes they express. For our technical discussions we will consider all types of models and subsequently, 16 different cases.

As the readers can easily apprehend, though some of the cases are redundant, we have a plethora of notions of agent uncertainty like *indecision*, *no opinion*, and others. If we refer back to our example in the introduction, all the attitudes of Alice's mental state can now be expressed in the present framework. We consider the reduced versions. Alice is undecided about A and B: $([+]A \wedge [-]A) \wedge ([+]B \wedge [-]B)$, has strong positive opinion about C: $[+]C \wedge [-]\neg C$, negative opinion about D: $[-]D \wedge \neg[-]\neg D$, have not heard about E: $(\neg[+]E \wedge \neg[+]\neg E) \wedge (\neg[-]E \wedge \neg[-]\neg E)$.

3.3 The *K* System as a Particular Case

Classical modal logic assumes that having a positive attitude towards φ is the same as having a negative attitude towards $\neg\varphi$. A dual model in which the positive and the negative relations coincide satisfy this property.

Proposition 3.1. *Denote by $\boldsymbol{M}^{\Box}$ the class of dual models in which the positive and the negative relation are the same, that is, $R^+ = R^-$. in this class, the following formula is valid:*

$$[+]\varphi \leftrightarrow [-]\neg\varphi$$

Consider now the previous sixteen cases.

Proposition 3.2. *In the class $\boldsymbol{M}^{\Box}$, only cases 1, 8, 9 and 16 are satisfiable. In fact, by using the mentioned equivalence $[+]\varphi \leftrightarrow [-]\neg\varphi$, they become the four attitudes we can express with a single normal modality, that is,*

$$\begin{array}{llll} (1): & [+]\varphi \wedge [+]\neg\varphi & (9): & [+]\varphi \wedge \neg[+]\neg\varphi \\ (8): & \neg[+]\varphi \wedge [+]\neg\varphi & (16): & \neg[+]\varphi \wedge \neg[+]\neg\varphi \end{array}$$

Proposition 3.3. *The classical modal system K is a subsystem of the dual system obtained by using $\Box$ for $[+]$ and by asking for the positive and the negative relation to be the same.*

4 Profile of the *dual* Framework

4.1 Decidability

We now show that, in general, we can always decide whether a given formula φ is **M**-valid or not. This is the case because our system, just like standard modal logic, has the *strong finite model property*.

Fact 1. *Let φ be a formula of $\mathcal{L}$. If φ is satisfiable in an $\boldsymbol{M}$-model, then it is satisfiable in a finite $\boldsymbol{M}$-model whose size is at most $2^{|\varphi|}$, where $|\varphi|$ is the size of the set of sub-formulas of φ closed under single negation.*

Moreover, the logic L has the strong finite model property with respect to a recursive set of **M**-models. So we have a method for deciding whether a formula φ is satisfiable or not. All we have to do is to verify if it is satisfiable in some pointed model of size up to $|\varphi|$ and, since there are only finitely many such models, we will eventually get an answer. If we find one such model, then φ is satisfiable; if not, then it follows from Fact 1 that φ is not satisfiable.

But in our system negation behaves classically, so a formula is valid if and only if its negation is not satisfiable. Hence, we can decide whether a formula φ is valid by deciding whether $\neg\varphi$ is satisfiable.

4.2 Complexity

We now provide complexity results for model checking and satisfiability problem in the *dual* framework. First, we state the problems formally.

Definition 4.1. *The* model checking *problem is, given a formula φ in $\mathcal{L}$ and a pointed dual model $(\langle W, R^+, R^-, V\rangle, w)$, decide whether φ is true at $(\mathcal{M}, w)$ or not. The* satisfiability *problem is, given a formula φ in $\mathcal{L}$ and a class of models $\boldsymbol{C}$, decide whether φ is satisfiable in a model of the class $\boldsymbol{C}$.*

The next fact provides us with an efficient algorithm for checking whether a formula is true in a pointed model. For the case of the satisfiability problem, we have already discussed an algorithm that decides whether a given formula is satisfiable in an **M**-model. The following gives a more efficient way to do it.

Fact 2. *In a* dual *framework, the complexity of model checking is P, while the complexity of satisfiability in the class of all dual models ($\boldsymbol{M}$) is PSPACE.*

4.3 Axiom Systems

We can decide whether a given $\mathcal{L}$-formula is **M**-valid. Such formulas can also be syntactically characterized, as the following theorem shows.

Theorem 4.1. *The logic L, presented in the table below, is sound and complete for the language $\mathcal{L}$ with respect to models in $\boldsymbol{M}$.*

P	*All propositional tautologies*	**MP**	*If* $\vdash \varphi$ *and* $\vdash \varphi \to \psi$, *then* $\vdash \psi$
K+	$\vdash [+](\varphi \to \psi) \to ([+]\varphi \to [+]\psi)$	**K-**	$\vdash [-](\varphi \wedge \psi) \to ([-]\neg\varphi \to [-]\psi)$
Gen+	*If* $\vdash \varphi$, *then* $\vdash [+]\varphi$	**Gen-**	*If* $\vdash \neg\varphi$, *then* $\vdash [-]\varphi$

For the cases of $\mathbf{M}^+$ and $\mathbf{M}^-$, their validities are characterized by the logic L extended with the **D+** axiom $[+]\varphi \to \langle + \rangle\varphi$ (the logic L^+) and the **D-** axiom $[-]\varphi \to \langle - \rangle\varphi$ (the logic L^-), respectively. Naturally, validities for the class $\mathsf{L}^\pm$ are characterized by L extended with both **D+** and **D-** (the logic $\mathsf{L}^\pm$).

The case of the class $\mathbf{M}^c$ is different.

Theorem 4.2. *The logic* L^c*, extending* L *with the axiom* $[-]\varphi \rightarrow \neg[+]\varphi$*, is sound and complete for the language* $\mathcal{L}$ *with respect to* $\boldsymbol{M}^c$*-models.*

Finally, the case of validities in $\mathbf{M}^{\Box}$-models.

Theorem 4.3. *The logic* $\mathsf{L}^{\Box}$*, extending* L *with the axiom* $[+]\varphi \leftrightarrow [-]\neg\varphi$ *is sound and complete for the language* $\mathcal{L}$ *with respect to models in* $\boldsymbol{M}^{\Box}$*.*

5 Concrete Interpretations

We now consider a particular interpretation of the $[+]$ and the $[-]$ modalities: *belief* and *disbelief*. Some intuitive ways to relate them are: *"disbelieving φ"* is a stronger notion than *"not believing in φ"*, whereas, *"believing in $\neg\varphi$"* should imply *"disbelieving φ"*. In fact, the reading of the different cases of dual expressions that we have provided in Section 3.3 is motivated by our understanding of $[+]$ as belief and $[-]$ as disbelief.

Consideration of disbelief as a separate epistemic category came to the fore in the latter part of last decade [20,3]. Consideration of changing or revising disbeliefs as a process analogous to belief revision was taken up by [4]. Belief-disbelief pairs, i.e. simultaneous consideration of belief and disbelief sets, were also taken up [2]. A more recent proposal can be found in [21], where 'disbelieving φ' is modeled as 'considering $\neg\varphi$ to be plausible'.

5.1 Belief-Disbelief Logic

We now propose the model and axiom system of the belief-disbelief logic ($\mathbf{L}_{KD45}$).

Definition 5.1. *We denote by* $\mathbf{M}_{KD45}$ *the class of models in* $\boldsymbol{M}$ *for which the positive and negative relations, now denoted by* R^{B} *and* R^{D}*, are serial, reflexive and Euclidean. Their respective universal modalities are given by* B *and* D *(with* $\widehat{\mathrm{B}}$ *and* $\widehat{\mathrm{D}}$ *denoting the corresponding existential ones).*

Theorem 5.1. *The logic* $\mathbf{L}_{KD45}$ *given by the axiom system of Theorem 4.1 plus the axioms below is sound and complete for* $\mathcal{L}$ *with respect to* $\mathbf{M}_{KD45}$*.*

D+	$\vdash \mathrm{B}\,\varphi \rightarrow \widehat{\mathrm{B}}\varphi$	**D-**	$\vdash \mathrm{D}\,\varphi \rightarrow \widehat{\mathrm{D}}\varphi$
4+	$\vdash \mathrm{B}\,\varphi \rightarrow \mathrm{B}\,\mathrm{B}\,\varphi$	**4-**	$\vdash \mathrm{D}\,\varphi \rightarrow \mathrm{D}\,\neg\,\mathrm{D}\,\varphi$
5+	$\vdash \neg\,\mathrm{B}\,\varphi \rightarrow \mathrm{B}\,\neg\,\mathrm{B}\,\varphi$	**5-**	$\vdash \neg\,\mathrm{D}\,\varphi \rightarrow \mathrm{D}\,\mathrm{D}\,\varphi$

What we have now is a minimal logic of belief and disbelief. To make things more interesting and useful we should have inter-relations between the belief and disbelief modalities. The table below lists interesting axioms and their corresponding characterizing-criteria in the $\mathbf{M}_{KD45}$ class of models.

C	$\vdash \mathrm{D}\,\varphi \rightarrow \neg\,\mathrm{B}\,\varphi$	**Mc**	$\forall w \in W,\ R^{\mathrm{B}}[w] \cap R^{\mathrm{D}}[w] \neq \emptyset$
BD	$\vdash \mathrm{B}\,\neg\varphi \rightarrow \mathrm{D}\,\varphi$	**Mbd**	$R^{\mathrm{D}} \subseteq R^{\mathrm{B}}$
DB	$\vdash \mathrm{D}\,\neg\varphi \rightarrow \mathrm{B}\,\varphi$	**Mdb**	$R^{\mathrm{B}} \subseteq R^{\mathrm{D}}$
Intro1	$\vdash \mathrm{D}\,\varphi \rightarrow \mathrm{B}\,\mathrm{D}\,\varphi$	**MI-1**	$wR^{\mathrm{B}}w' \wedge wR^{\mathrm{D}}w'' \Rightarrow w'R^{\mathrm{D}}w''$
Intro2	$\vdash \mathrm{B}\,\varphi \rightarrow \mathrm{D}\,\neg\,\mathrm{B}\,\varphi$	**MI-2**	$wR^{\mathrm{D}}w' \wedge wR^{\mathrm{B}}w'' \Rightarrow w'R^{\mathrm{B}}w''$

5.2 Preference

There is a very close relationship between an agent's beliefs and her preferences which has been extensively discussed in [22]. In fact, both objective and subjective preferences over objects are described, along with preference over propositions. The dual framework of belief-disbelief can also provide a way to describe subjective preferences over propositions, taking into account agents' uncertainties as well. For example, consider the following notion of preference.

$$Pref(\varphi, \psi) : (O_3\varphi \vee O_9\varphi) \wedge (O_5\psi \vee O_6\psi \vee O_{10}\psi \vee O_{16}\psi),$$

where $O_i\chi$ represents the i-th possibility of the 16 different attitudes described earlier. The $Pref$ relation defined above is *transitive* but *neither reflexive nor linear*. This is a very weak notion of preference in the sense that it cannot distinguish between longer preference orders, e.g., orders of length more than 3. But, if we go back to the decision-making scenario described in the introduction, we can still deduce from the known facts that $Pref(C, X)$, for $X = A, B, D, E$. We leave a more detailed discussion on stronger notions of preference as well as their technical study for future work.

6 A Comparative Discussion

As mentioned in the introduction, there already have been past works modeling these dual attitudes. In this section we provide a comparative discussion with two such proposals which are very close in spirit to ours.

6.1 A Logic of Acceptance and Rejection

In [3], the authors present a nonmonotonic formalism AEL2 extending the framework of Moore's auto-epistemic logic [23] to deal with *uncertainty* of an agent. The underlying logical framework is the same as that of $\mathbf{L}_{KD45}$. In AEL2, accepted and rejected premises are separated to form a pair of sets of formulas (I_1, I_2). Then, the AEL2-extensions (T_1, T_2) are defined, where T_1 is expected to contain all the accepted formulas with respect to I_1 and T_2 to contain all the rejected formulas with respect to I_2.

Definition 6.1. *(T_1, T_2) is a stable* AEL2 *expansion of (I_1, I_2) if*

$$T_1 = Cn(I_1 \cup \{B\varphi : \varphi \in T_1\} \cup \{\neg B\varphi : \varphi \notin T_1\} \cup \{D\varphi : \varphi \in T_2\} \cup \{\neg D\varphi : \varphi \notin T_2\})$$
$$T_2 = Cn'(I_2 \cup \{\neg B\varphi : \varphi \in T_1\} \cup \{B\varphi : \varphi \notin T_1\} \cup \{\neg D\varphi : \varphi \in T_2\} \cup \{D\varphi : \varphi \notin T_2\}).$$

Here, Cn is the classical propositional consequence operator, and Cn' is the corresponding consequence operator for the propositional logic of contradictions.

Definition 6.2. *(T_1, T_2) is said to be a BD-dual extension of (I_1, I_2) if*

$$T_1 = \{\psi \mid I_1 \cup \{\neg B\varphi : \varphi \notin T_1\} \cup \{D\varphi : \varphi \in T_2\} \cup \{\neg D\varphi : \varphi \notin T_2\} \vdash_{\mathsf{L}_1} \psi\}$$
$$T_2 = \{\psi \mid I_2 \cup \{\neg B\varphi : \varphi \in T_1\} \cup \{B\varphi : \varphi \notin T_1\} \cup \{\neg D\varphi : \varphi \in T_2\} \vdash_{\mathsf{L}_2} \psi\}.$$

Here, L_1 is a fragment of $\mathbf{L}_{KD45}$ axiomatized by the propositional tautologies, MP, modal axioms **K+**, **4+**, and **5+**, and rule **Gen+** corresponding to the modal operator B, and L_2 is the logic axiomatized by the propositional contradictions, *Rej*, modal axioms **K-**, D_2, and D_3, and rule **Gen-** corresponding to the modal operator D, where D_2, D_3,and *Rej* are given below:

$$D_2 : D\varphi \rightarrow \neg DD\varphi \qquad \qquad Rej\!: \frac{\beta, \neg(\alpha \rightarrow \beta)}{\alpha}$$
$$D_3 : \neg D\varphi \rightarrow \neg D\neg D\varphi$$

It is a very well-known result that Moore's auto-epistemic logic corresponds to the non-monotonic modal logic weak $S5$, in particular the $K45$-logic (see [24] for a detailed discussion). Similarly, it can be proved that,

Proposition 6.1. *(T_1, T_2) is a consistent stable* AEL2 *expansion of (I_1, I_2) iff (T_1, T_2) is a BD-dual extension of (I_1, I_2).*

6.2 Some Logics of Belief and Disbelief

In [2], the authors present several logics for dealing with beliefs and disbeliefs from a syntactic perspective, together with providing a neighbourhood-like semantics for them. Given a set of atomic propositions P, denote by L_B the set of classical propositional logic formulas built from P, and define $L_D := \{\overline{\phi} \mid \phi \in L_B\}$. The underlying language L is then given by $L_B \cup L_D$. Suppose $\Gamma \subseteq L$. The agent *believes* every ϕ with $\phi \in \Gamma$ (denoted by Γ_B), and *disbelieves* every ϕ with $\overline{\phi} \in \Gamma$ (denoted by Γ_D). Based on specific closure properties of Γ_B and Γ_D, the authors define four different logics, but all of them can be interpreted on **WBD** models, tuples of the form $\langle M, \mathcal{N} \rangle$ where M is a set of propositional valuations and $\mathcal{N} \subseteq \wp(V)$ is a set of sets of propositional valuations (the particular logics are characterized by properties of the semantic model). Then,

$$\begin{array}{lll} \langle M, \mathcal{N} \rangle \models \phi & \text{iff} & \phi \text{ is true under every valuation in } M\text{, and} \\ \langle M, \mathcal{N} \rangle \models \overline{\phi} & \text{iff} & \neg\phi \text{ is true under every valuation of } N \textit{ for some } N \in \mathcal{N}. \end{array}$$

We now provide a semantic comparison between these logics and our framework. Consider a **WBD** model $\mathsf{M} = \langle M, \mathcal{N} \rangle$ in which $\mathcal{N}$ is finite, and denote by k its number of elements. We will build an extension of our dual models in which the domain consists of all the possible valuations for the given atomic propositions.

Definition 6.3. *Let P be a set of atomic propositions and let $\mathsf{M} = \langle M, \mathcal{N} \rangle$ be a **WBD** model based on them, with $\mathcal{N} = \{N_1, \ldots, N_k\}$ (i.e, $\mathcal{N}$ is finite). Denote by V the set of all propositional valuations over P, and denote by V_p the set of propositional valuations in V that make p true. The extended dual model $\mathcal{M}_{\mathsf{M}} = \langle W, R^{\mathrm{B}}, R^{\mathrm{B}}_s, R^{\mathrm{D}}, V \rangle$ has as domain the set of all valuations for P, that is, $W := \mathsf{V}$. Now select arbitrary $k+1$ worlds $\mathsf{w}, \mathsf{w}_1, \ldots, \mathsf{w}_k$ in W. Define $R^{\mathrm{B}}\mathsf{w}u$ iff $u \in M$. For each $i \in \{1, \ldots, k\}$, define $R^{\mathrm{D}}\mathsf{w}_\mathsf{i}u$ iff $u \in N_i$. Define $R^{\mathrm{B}}_s\mathsf{w}\mathsf{w}_\mathsf{i}$ for every $i \in \{1, \ldots, k\}$, and for every atomic proposition p, define $V(p) := \mathsf{V}_p$.*

For this special dual model, we use the modalities B, B_s and D for the relations R^{B}, R^{B}_s and R^{D}, respectively. Then, for every world $w \in W$,

$$(\mathcal{M}_{\mathsf{M}}, w) \models \mathrm{B}\,\varphi \quad \text{iff} \quad \text{for all } w' \text{ such that } R^{\mathrm{B}}ww',\ (\mathcal{M}_{\mathsf{M}}, w') \models \varphi$$
$$(\mathcal{M}_{\mathsf{M}}, w) \models \mathrm{B}_s\varphi \quad \text{iff} \quad \text{for all } w' \text{ such that } R^{\mathrm{B}}_s ww',\ (\mathcal{M}_{\mathsf{M}}, w') \models \varphi$$
$$(\mathcal{M}_{\mathsf{M}}, w) \models \mathrm{D}\,\varphi \quad \text{iff} \quad \text{for all } w' \text{ such that } R^{\mathrm{D}}ww',\ (\mathcal{M}_{\mathsf{M}}, w') \models \neg\varphi$$

Proposition 6.2. *Let $\mathsf{M} = \langle M, \mathcal{N}\rangle$ be a **WBD** with $\mathcal{N}$ finite. For every propositional formula γ, we have*

$$\mathsf{M} \models \gamma \ \textit{ iff } \ (\mathcal{M}_{\mathsf{M}}, \mathsf{w}) \models \mathrm{B}\,\gamma \qquad\qquad \mathsf{M} \models \overline{\gamma} \ \textit{ iff } \ (\mathcal{M}_{\mathsf{M}}, \mathsf{w}) \models \widehat{\mathrm{B}_s}\,\mathrm{D}\,\gamma$$

In other words, our formula $\mathrm{B}\,\gamma$ expresses the notion "the agent believes γ" of [2], and our formula $\widehat{\mathrm{B}_s}\,\mathrm{D}\,\gamma$ expresses their "the agent disbelieves γ".

In [25], the authors suggest combining universal and existential notions to describe *knowledge*. The above proposition puts us on similar track, but we leave the detailed study for future work.

7 Dynamics

The system proposed in Section 3 allows us to represent positive and negative attitudes by means of two modalities that allow us to build formulas of the form $[+]\varphi$ and $[-]\varphi$. While the first one is true at a world w iff φ is true in all the worlds R^+– reachable from w, the second is true at a world w iff φ is *false* in all the worlds R^-– reachable from w. We have also shown that, when the two relations R^+ and R^- are the same, we get the validity $[-]\varphi \leftrightarrow [+]\neg\varphi$ indicating that the agent has a negative attitude towards a formula iff she has a positive attitude towards its negation. This actually says that, when $R^+ = R^-$, the negative attitude collapses into classical negation, and therefore we get the classical K-system.

But from a more dynamic perspective, the case in which $R^+ = R^-$, that is, the K case can be thought of as not a particular case of the static system, but a possible result of some dynamic extension. In other words, the 'ideal' system K in which negative attitudes coincide with classical negation, can be seen not as the state of an ideal static agent, but as the possible final state of a non-ideal but dynamic one who can perform actions that make the two relations the same. This section looks briefly at possible results of such actions.

There are various ways to generate a new relation from two others, and in this case they represent the different policies through which the agent 'merges' her positive and negative attitudes. For example, she can be drastic in two different ways: give up her negative attitude ($R := R^+$) or give up the positive one ($R := R^-$). More reasonable are the policies that actually combine the two relations, like $R := R^+ \cup R^-$.

Definition 7.1 (Merging policies). *Let $\mathcal{M} = \langle W, R^+, R^-, V\rangle$ be a dual model. The relation R of a dual model $\langle W, R, V\rangle$ that results from the agent's merging of her positive and negative attitudes can be defined in several forms.*

- $R := R^+$ *(the* drastic positive *policy; new model denoted by* $\mathcal{M}_+$*).*
- $R := R^-$ *(the* drastic negative *policy; new model denoted by* $\mathcal{M}_-$*).*
- $R := R^+ \cup R^-$ *(the* liberal combining *policy; new model denoted by* $\mathcal{M}_\cup$*).*
- $R := \alpha(R^+, R^-)$*, where* $\alpha(R^+, R^-)$ *is a PDL-expression [26] based on* R^+ *and* R^- *(the* PDL *policy; new model denoted by* $\mathcal{M}_\alpha$*).*
- $R := R^+ \cap R^-$ *(the* skeptic combining *policy; new model denoted by* $\mathcal{M}_\cap$*).*
- $R := R^+ \setminus R^-$ *(new model denoted by* $\mathcal{M}_\pm$*).*
- $R := R^- \setminus R^+$ *(new model denoted by* $\mathcal{M}_\mp$*).*

For each policy $\circ$, we define a modality $[m_\circ]$ for building formulas of the form $\langle m_\circ \rangle \varphi$, read as "there is a way of merging attitudes with policy $\circ$ after which φ is the case". Their semantic interpretation is given by:

$$(\mathcal{M}, w) \models \langle m_\circ \rangle \varphi \quad \text{iff} \quad (\mathcal{M}_\circ, w) \models \varphi$$

Now, for an axiom system, we can provide reduction axioms for each policy. In each case, the relevant ones are those describing the way the new relations are created. Though they are the same, we will stick with $+$ and $-$ for representing the positive and negative relations after the operation.

The *drastic positive* policy:	$\langle m_+ \rangle \langle + \rangle \varphi \leftrightarrow \langle + \rangle \langle m_+ \rangle \varphi$ $\langle m_+ \rangle \langle - \rangle \varphi \leftrightarrow \langle + \rangle \langle m_+ \rangle \neg\varphi$
The *drastic negative* policy:	$\langle m_- \rangle \langle + \rangle \varphi \leftrightarrow \langle - \rangle \langle m_- \rangle \neg\varphi$ $\langle m_- \rangle \langle - \rangle \varphi \leftrightarrow \langle - \rangle \langle m_- \rangle \varphi$
The *liberal combining* policy:	$\langle m_\cup \rangle \langle + \rangle \varphi \leftrightarrow \langle + \rangle \langle m_\cup \rangle \varphi \vee \langle - \rangle \langle m_\cup \rangle \neg\varphi$ $\langle m_\cup \rangle \langle - \rangle \varphi \leftrightarrow \langle + \rangle \langle m_\cup \rangle \neg\varphi \vee \langle - \rangle \langle m_\cup \rangle \varphi$

For $*$-free *PDL* policies, reduction axioms for each particular $\alpha(R^+, R^-)$ can be obtained by following the technique introduced in [27]. To get sound and complete reduction axioms for the policies involving $\cap$ and $\setminus$, we may need to extend the language with nominals.

But we do not only need to look at actions that create a single relation in one shot. We can also look at procedures in which the single relation is a long-term result of small operations that merge the two of them in a step-wise form. We leave the detailed study on these dynamical aspects for future work.

Acknowledgements. The authors gratefully thank the anonymous ICLA 2011 reviewers for their interesting comments and suggestions which helped to improve the paper.

References

1. Hájek, P., Harmancová, D., Verbrugge, R.: A qualitative fuzzy possibilistic logic. International Journal of Approximate Reasoning 12(1), 1–19 (1995)
2. Chopra, S., Heidema, J., Meyer, T.A.: Some logics of belief and disbelief. In: Gedeon, T(T.) D., Fung, L.C.C. (eds.) AI 2003. LNCS (LNAI), vol. 2903, pp. 364–376. Springer, Heidelberg (2003)

3. Gomolinska, A.: On the logic of acceptance and rejection. Studia Logica 60(2), 233–251 (1998)
4. Gomolinska, A., Pearce, D.: Disbelief change. Electronic essays on the occasion of the fiftieth birthday of Peter Gardenfors (2001)
5. Tversky, A., Kahneman, D.: Advances in prospect theory: Cumulative representation of uncertainty. Journal of Risk and Uncertainty 5(4), 297–323 (1992)
6. Caminada, M.W.A., Gabbay, D.M.: A logical account of formal argumentation. Studia Logica 93(2-3), 109–145 (2009)
7. Dubois, D., Prade, H. (eds.): Special issue on Bipolarity (2008); International Journal of Intelligent Systems 23(8, 9, 10)
8. Dubois, D., Prade, H.: Foreword. International Journal of Intelligent Systems 23(8), 863–865 (2008)
9. Klir, G.J., Clair, U.S., Yuan, B.: Fuzzy set theory: foundations and applications. Prentice-Hall, Inc., Upper Saddle River (1997)
10. Dubois, D., Prade, H.: Possibility theory. Plenum Press, New-York (1988)
11. Pawlak, Z.: Rough Sets. Theoretical Aspects of Reasoning about Data. Springer, Formerly Kluwer Academic Publishers, Boston (1991)
12. Halpern, J.: Reasoning about Uncertainty. The MIT Press, Cambridge (2003)
13. Dempster, A.P.: A generalization of Bayesian inference. Journal of the Royal Statistical Society 30, 205–247 (1968)
14. Shafer, G.: A Mathematical Theory of Evidence. Princeton University Press, Princeton (1976)
15. Williamson, J.: Probability logic. In: Gabbay, D.M., Johnson, R., Ohlbach, H., Woods, J. (eds.) Handbook of the Logic of Inference and Argument, vol. 1, pp. 397–424. Elsevier, Amsterdam (2002)
16. Hájek, P.: Metamathematics of Fuzzy Logic. Kluwer Academic Publishers, Dordrecht (1998)
17. Dubois, D., Prade, H.: Possibilistic logic. In: Gabbay, D., Hogger, C., Robinson, J. (eds.) Handbook of Logic in Artificial Intelligence and Logic Programming, vol. 3. Claredon Press, Oxford (1994)
18. Pagliani, P., Chakraborty, M.: A Geometry of Approximation: Rough Set Theory Logic, Algebra and Topology of Conceptual Patterns. Springer, Heidelberg (2008)
19. Malinowski, G.: Many-Valued Logics. Oxford University Press, Oxford (1993)
20. Ghose, A., Goebel, R.: Belief states as default theories: Studies in non-prioritized belief change. In: ECAI, pp. 8–12 (1998)
21. de Jongh, D., Ghosh, S.: Comparing strengths of beliefs explicitly. In: Bilkova., M., et al. (eds.) MALLOW. CEUR Workshop Proceedings, CEUR-WS.org, vol. 494 (2009)
22. Liu, F.: Changing for the Better. Preference Dynamics and Agent Diversity. PhD thesis, ILLC, UvA, Amsterdam, The Netherlands (2008)
23. Moore, R.C.: Semantical considerations on nonmonotonic logic. Artificial Intelligence 25(1), 75–94 (1985)
24. Shvarts, G.F.: Autoepistemic modal logics. In: Parikh, R. (ed.) TARK, pp. 97–109. Morgan Kaufmann, San Francisco (1990)
25. van Benthem, J., Martínez, M.: The stories of logic and information. In: Adriaans, P., van Benthem, J. (eds.) Philosophy of Information, pp. 217–280. North-Holland, Amsterdam (2008)
26. Harel, D., Kozen, D., Tiuryn, J.: Dynamic Logic. MIT Press, Cambridge (2000)
27. van Benthem, J., Liu, F.: Dynamic logic of preference upgrade. Journal of Applied Non-Classical Logics 17(2), 157–182 (2007)

The Completion of the Emergence of Modern Logic from Boole's *The Mathematical Analysis of Logic* to Frege's *Begriffsschrift*

Priyedarshi Jetli

Visiting Faculty, Department of Philosophy,
University of Mumbai, Mumbai
pjetli@gmail.com

Abstract. Modern logic begins with Boole's *The Mathematical Analysis of Logic* when the algebra of logic was developed so that classical logic syllogisms were proven as algebraic equations and the turn from the logic of classes to propositional logic was suggested. The emergence was incomplete as Boole algebraised classical logic. Frege in *Begriffsschrift* replaced Aristotelian subject–predicate propositions by function and argument and displaced syllogisms with an axiomatic propositional calculus using conditionals, *modus ponens* and the law of substitution. Further Frege provided the breakthrough to lay down the groundwork for the development of quantified logic as well as the logic of relations. He achieved all of this through his innovative formal notations which have remained underrated. Frege hence completed the emergence of modern logic. Both Boole and Frege mathematised logic, but Frege's goal was to logicise mathematics. However the emergence of modern logic in Frege should be detached from his logicism.

Keywords: Boole, Frege, conditional, *modus ponens*, propositional calculus, quantifiers, function and argument, axioms.

1 Introduction

Can we pinpoint the moment when modern logic emerged? I propose that the emergence occurs in 1847 in Boole's *The Mathematical Analysis of Logic* [1] and in 1879 in Frege's *Begriffsschrift* [2, 3]. However this is to be complemented by the metalogic developed by Hilbert. Boole mathematised logic; whereas Frege raised logic to its highest pedestal by attempting to logicise mathematics and for Hilbert both logic and mathematics as a unified enterprise reached a second order level never developed before.

'Modern logic' may have various meanings such as (a) a fully developed first order propositional and predicate calculus, or (b) the development of higher order logics and metalogic including the relation of semantics to proof theory.[1] If 'modern' means latest, then (b) would be more appropriate. However, if 'modern' is taken in an

[1] I thank an anonymous referee for making this distinction clear and asking me to make clear in which sense I meant 'modern logic'.

M. Banerjee and A. Seth (Eds.): ICLA 2011, LNAI 6521, pp. 105–123, 2011.

historical sense, then (a) may be the more important meaning. 'Modern philosophy' is taken to begin with Descartes in the 17th century. In this historical sense 'Modern' is a break from medieval philosophy. 'Modern philosophy' includes the span of philosophers from Descartes to the 21st century. No crucial break from classical logic occurred in the 17th or 18th centuries, hence we mark the 19th century as when modern logic emerged. Just as we say that Modern philosophy begins with Descartes, I want to claim that modern logic begins with Boole and Frege. Hence, under 'modern logic' I include not only (a) and (b) but also the more recent developments such as modal logic, epistemic logic, deontic logic, quantum logic, multivalued logic, fuzzy logic and paraconsistent logic. However, I use 'emergence' as being restricted to (a). Just as the emergence of Modern philosophy in Descartes does not include Kant's critical philosophy, similarly, the emergence of modern logic does not include (b) and later developments. Nonetheless just as Descartes is a necessary antecedent for Kant's critical philosophy, so (a) is a necessary antecedent for (b). Second order and higher order logics would not have emerged without the initial emergence of first order axiomatic propositional and predicate logic.

When I pinpoint Boole (1815–1864) and Frege (1848–1925) as the founders of modern logic, by no means do I wish to marginalise the contributions made to the emergence of modern logic by great mathematicians and logicians such as DeMorgan (1806–1871), Lewis Carroll (1832–1898), MacColl (1837–1909), Peirce (1839–1914), Schröder (1841–1902), Peano (1858–1932) and others. Each of these may have made as great or greater contributions than Boole and Frege and their influence on Frege at least may also be of great significance as he may not have been able to develop the logic he did without them. Rather, I have pinpointed one work of each and I wish to consider these works in themselves, in as much as possible, independent of their authors, as this makes it very convenient to understand when and how modern logic emerged. Hence, even though Frege on the whole may not be sympathetic with Boole's algebraisation of logic, I have attempted to show how the *Begriffsschrift* complements *The Mathematical Analysis of Logic* in the emergence of modern logic. I am open to alternative or complementary accounts which may for example consider DeMorgan and Peirce as the founders of modern logic. I also make a heuristic distinction between the roots of modern logic, the emergence of modern logic and the development of modern logic. The roots may go as far back as Aristotle, but the emergence occurs in the 19th century and most of what happens in modern logic in the 20th century may fall under the development of modern logic. My concern in this paper is only with the emergence of modern logic. As I have dealt extensively with Boole's *The Mathematical Analysis of Logic* elsewhere [4], I will concentrate in this paper on Frege's *Begriffsschrift*.

The emergence of modern mathematics came in the golden age of mathematics from mid 18th century to early 19th century led by Euler (1707–1783), Lagrange (1736–1813), Laplace (1749–1827) and Gauss (1777–1855). Non-Euclidean geometry and abstract algebra also emerged in this period. The delay in the emergence of modern logic was because of certain historical developments in algebra. Non-Euclidean geometries had already raised the possibility of a geometry that did not deal solely with measurement. With the development of symbolical algebra it became possible to have a purely abstract algebra that did not deal with quantity.

Hence, Boole could develop the algebra of logic which was completely symbolic and completely devoid of content. However, despite laying the groundwork of algebraic logic as well as a propositional calculus, Boole stuck to the Aristotelian limitation of subject–predicate propositions. So his logic was only partially modern.

Frege put an end to subject–predicate propositions and syllogisms by considering propositions in terms of functions and arguments. So, he also mathematised logic, but his mathematisation was conceptual and this logic would serve all of mathematics. Frege hence displaced the portion of logic that had remained Aristotelian and modern logic emerged. Even though '*Begriffsschrift*' translates as 'conceptual content', it was his innovative notations that finalised the revolution in logic and a comprehensive propositional and predicate symbolic logic could then be developed in 1910 by Whitehead and Russell in *Principia Mathematica* (*PM*) [5].

2 The Beginning of Modern Logic in Boole's *The Mathematical Analysis of Logic*

Symbolic propositional and predicate calculus could not be developed because neither Aristotle nor any logician after Aristotle was able to mathematise logic. Leibniz anticipated the algebra of logic to be the art of combinations as Louis Couturat states:

> Leibniz had conceived the idea […] of all the operations of logic, […] was acquainted with the fundamental relations of the two copulas […] found the correct algebraic translation of the four classical propositions, […] discovered the principal laws of the logical calculus, […] he possessed almost all the principles of the logic of Boole and Schröder, and on certain points he was more advanced than Boole himself. (my translation) [6, pp. 385–6]

What Leibniz really needed was the development of symbolical algebra which occurred more than a century after his death. In 1830 George Peacock claimed that operations in symbolic algebra must be open to interpretations other than that in arithmetic:

> […] in framing the definitions of algebraical operations, […] we must necessarily omit every condition which is in any way connected with their specific value or representation: […] the definitions of some operations must regard the laws of their combination *only* […] in order that such operations may possess an invariable meaning and character, […] [7, pp. viii–x]

The primacy of combinations over what they combine is thereby established.

Boole developed a quantity free algebra of logic in *Mathematical Analysis of Logic*.

He began by laying down the foundations of the algebra of logic which is a logic of classes in which the three combinatory laws (1) $x(y + z) = xy + xz$ (distributive), (2) $xy = yx$ (commutative) and (3) $x^2 = x$ (index) (p. 15); when combined with the axiom that equivalent operations performed on equivalent subjects, produce equivalent results, constitute the axiomatic foundations for all of logic [1, p. 18]. First, Boole represents

Aristotelian categorical propositions as algebraic equations. Then, he captures valid syllogisms of classical logic by multiplying equations and eliminating y which represents the traditional middle term:

$$ay + b = 0$$
$$a'y + b' = 0$$

When y is eliminated this reduces to:

$$ab' - a'b = 0 \qquad [1, p. 32]$$

Boole then makes the crucial turn to propositional logic in his account of conditionals. First he presents conditionals in terms of classes as in syllogistic logic:

If A is B, then C is D,
But A is B, therefore, C is D.

But then he expresses it in terms of propositions without reference to classes:

If X is true, then Y is true,
But X is true, therefore, Y is true.

> […] Thus, what we have to consider is not objects and classes of objects, but the truths of Propositions, namely, of those elementary Propositions which are embodied in the terms of our hypothetical premises [1, pp. 48–9].

We can embalm page 48 as the long awaited turning point from classical to modern logic as a scheme to translate syllogisms into inferences involving conditionals is suggested and in the particular example, the rule of inference of *modus ponens* is given in its propositional conditional form as we know it today.

Using 0 for false and 1 for true Boole now comes up with the possibilities for truth tables [1, pp. 50–51] and goes on to define conjunction, disjunction (both exclusive and inclusive), and conditional truth functionally [1, pp. 52–4]. As truth values are algebraised, mathematics can provide important insights into logic. These equations can be used for understanding truth functionality in a way that may not be understood without mathematics. The equation for the exclusive disjunction 'Either x is true or y is true' is $x - 2xy + y = 1$, which is acquired from the second and third row of the truth table: $x(1 - y) + y(1 - x) = x - xy + y - xy = x - 2xy + y$, and this must be true, so $x - 2xy + y = 1$. Now, since $x^2 = x$, we get: $x^2 - 2xy + y^2 = 1$. Which reduces to $(x - y)^2 = 1$; $x - y = \pm 1$. When x is true having the value of 1, then y must be false having the value 0 and when x is false, having the value 0, then y is true, having the value 1 to satisfy the equation [1, p. 55]. Hence, we see from the inside of Boolean algebra how a simple algebraic operation, but without regard to quantity, as the rule $x^2 = x$ is not a rule of ordinary algebra, leads to a clear definition of a logical operation like exclusive disjunction.

Boole concludes: 'The general doctrine of elective symbols and all the more characteristic applications are quite independent of any quantitative origin' [1, p. 82]. Boole successfully developed the algebra of logic on the basis of symbolical algebra that divests itself of quantitative origin. However, Aristotelian logic was sustained as is clear by the title 'Aristotelian Logic and its Modern Extensions, Examined by the

Method of this Treatise' of the culminating chapter of the logic part of his major work *Laws of Thought* in 1854. [8, pp. 174–86]. Clearly then, modern logic had not yet emerged in 1847 or in 1854.

3 The Completion of the Emergence of Modern Logic in Frege's *Begriffsschrift*

Frege completed the emergence of modern logic: first, by his innovative notation for judgments where the content stroke represented the content of the judgment, he finally brought down the axe on subject–predicate propositions which Boole was unable to do; second, he introduced a formal notation for conditional statements which in turn led to the development of axiomatic logic as well as a rigorous proof technique using *modus ponens* that made Aristotelian syllogisms archaic; third, he introduced a perspicuous notation for the universal quantifier so that a predicate as well as propositional calculus could be developed; fourth, he imported from mathematics the notions of function and argument and placed them at the core of symbolic logic and there was no looking back. Frege made up for Leibniz's failure to develop modern logic due to a lack of formalisation of relations and modern logic finally emerged. By no means was Frege a lone ranger in the emergence and development of modern logic. Invaluable contributions, without which Frege would have been nowhere, were made by DeMorgan, Schröder, Peirce and others. Yet Frege perhaps put it all together better than anyone else. The master historians of logic, Kneale and Kneale, best capture Frege's contribution:

> Frege's *Begriffsschrift* is the first really comprehensive system of formal logic. Aristotle was interested chiefly in certain common varieties of general propositions. He did indeed formulate the principles of non-contradiction and excluded middle, which belong to a part of logic more fundamental than his theory of the syllogism; but he failed to recognize the need for a systematic account of primary logic. Such an account was supplied, at least in part, by Chrysippus; but neither he nor the medieval logicians who wrote about *consequentiae* succeeded in showing clearly the relation between primary and general logic. Leibniz and Boole, recognizing a parallelism between primary logic and certain propositions of general logic about attributes or classes, worked out in abstract fashion a calculus that seemed to cover both; but neither of these enlarged the traditional conception of logic to include the theory of relations. Working on some suggestions of De Morgan, Peirce explored this new field, and shortly after the publication of the *Begriffsschrift* he even produced independently a doctrine of functions with a notation adequate for expressing all the principles formulated by Frege; but he never reduced his thoughts to a system or set out a number of basic principles like those given in the last section. Frege's work, on the other hand, contains all the essentials of modern logic, and it is not unfair either to his predecessors or to his successors to say that 1879 is the most important date in the history of the subject. [9, pp. 510–11]

Others have also expressed highest praises for the *Begriffsschrift*. von Heijenoort says 'Modern logic began in 1879, the year in which Gottlob Frege (1848–1925) published his *Begriffsschrift*' [10, p. 242]. According to Haaparanta 'Still others argue that the beginning of modern logic was 1879, when Frege's *Begriffsschrift* appeared' [11, p. 5]. Christian Thiel states 'If Frege has been regarded as the founder of modern mathematical logic, this characterization refers to his creation of classical quantificational logic in his *Begriffsschrift* of 1879 without any predecessor' [12, p. 197].

I now proceed to capture Frege's contributions in the *Begriffsschrift*.

The Preface announces Frege's motivation as he believes that pure logic gives the most reliable proof, and this depends solely on those laws on which all knowledge rests. Aristotle felt that his greatest achievement in logic was the discovery of the laws of thought. Boole appropriately entitled his later book as *An Investigation into the Laws of Thought*. There is a remarkable structural affinity among Aristotle, Boole and Frege, yet they are the greatest revolutionaries in logic. Frege, as a philosopher, made explicit what Boole as a mathematician left only as implicit. Boole algebraised logic by importing symbolical algebra into logic but at the same time, he set up formal logic that could become the basis of algebra as well. In attempting to logicise arithmetic, that is, to make it bereft of facts, and thereby content, Frege wanted to express arithmetical sequences by representing the ordering of a sequence without bringing in intuition and the existing mathematical language made this a very difficult task. Hence, he created his own formula language, the central nerve of which is *conceptual content* (*begrifflichen inhalt*) [2, pp. iii–iv; 3, pp. 5–6][2]. This formula language is modelled after the formula language of arithmetic, yet it is the 'formula language for pure thought' including arithmetic. Frege next announces that argument and function replaces subject and predicate of traditional logic and this will stand the test of time [2, p. vii; 3, p. 7]. And indeed it has stood the test of time. The Preface ends:

> As I remarked at the beginning, arithmetic was the point of departure for the train of thought that led me to my ideography. And that is why I intend to apply it first of all to that science, attempting to provide a more detailed analysis of the concepts of arithmetic and a deeper foundation for its theorems. [2, p. viii; 3, p. 8]

This is a profound insight that arithmetic is begging for someone to build logic out of its language so that it (the new logic) can provide a sounder foundation for arithmetic itself.

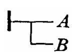
Part I is 'I. DEFINITION OF THE SYMBOLS'. In #1 Frege begins with the distinction between two types of signs; letters, like *a*, *b*, *c*, etc., that represent variability in meaning and symbols like +, -, √ , and 1, 2, 3, which have determinate meaning: '*I adopt this basic idea of distinguishing two types of signs*, which unfortunately is not strictly observed in the theory of magnitudes, *in order to apply it*

---

[2] Though my reading is of the English translation of the *Begriffsschrift* [3] and all except one of the quotations are from the English translation, nonetheless I also give the citation of page number from the German original first [2] so that the reader can refer to the original as well. Hence, the citations are given as here '[2, pp. iii–iv; 3, pp. 5–6]'. If only the German is referred to than [3] is left out.

in the more comprehensive domain of pure thought in general' [2, p. 1; 3, pp. 10–11]. Frege steals an important distinction from under the noses of mathematicians, which the mathematicians do not clearly see, and builds on it the new logic. #2 introduces '⊢—' for expressing judgments. The horizontal stroke is the content stroke representing the thought of the proposition and the vertical stroke is the judgment stroke [2, pp. 1–2; 3, pp. 11–12]. #3 begins: 'Eine Untersheidung von *Subject* und *Prüdicat* findet bei meiner Darstellung eines Urthiels *nicht statt*' [2, pp. 2–3]. {'The distinction between *subject* and *predicate* does *not occur* in my way of representing a judgment' [3, p. 12]}. This marks the death of Aristotelian logic and the emergence of modern logic. '⊢—A' does not represent a subject–predicate proposition such as (1) 'Archimedes perished at the capture of Syracuse' but it represents the *conceptual content* of it, which is equally captured by a distinct subject–predicate proposition such as (2) 'The capture of Syracuse led to the death of Archimedes' or (3) 'The violent death of Archimedes at the capture of Syracuse is a fact'. As in (3) all judgments may be thought of as having a propositional content as the subject and 'is a fact' as the common predicate that makes them true [2, pp. 2–3; 3, pp. 12–13].

In #4 many distinctions of classical logic are dissolved such as that between universal and particular judgments which is now the distinction between universal and particular content and not of categorically different propositions. Negation is an adjunct to the content so that negative judgments are not categorically different from positive ones. Boole represented the four Aristotelian categorical propositions as algebraic equations hence dissolving the categorical distinction between particular and universal on the one hand and negative and affirmative on the other, since as algebraic equations they are not categorically distinct. Frege builds on this accomplishment; since, from this point on, there is no need to deal with Aristotelian categorical propositions [2, pp. 4–5; 3, p. 13]. #5 introduces the notation for conditional judgments as:

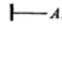


We symbolise this today as $B \supset A$. The horizontal lines to the left of A and B up to the middle vertical line are the content strokes of A and B respectively, and the horizontal stroke to the left of the top node of the vertical line is the content stroke of the meaning of the conditional regardless of the contents of A and B. First Frege gives four possibilities:

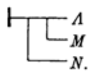
> (1) A is affirmed and B is affirmed; (2) A is affirmed and B is denied; (3) A is denied and B is affirmed; (4) A is denied and B is denied. [2, p. 5; 3, p. 13]

Then he defines the conditional as 'the judgment that *the third of these possibilities does not take place, but one of the three others does*' [2, p. 5; 3, p. 14]. This truth functional definition is the key that explains the axioms as laws of thought. Hugh MacColl had 'argued that the basic relation in logic is not class inclusion but implication between two propositions' [13, p. 373]. #6 develops *modus ponens* from the conditional as the only rule of inference:

(XX)::

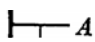


Where (XX) represents |—B and the '::' indicates that B would have to be formulated and put into the inference [2, p. 8; 3, p.16]. Boole's algebraised logic required disjunction and conjunction as the main connectives which are symbolised by addition and multiplication respectively. Boole also gave the example of *modus ponens* when he made the transition from Aristotelian syllogisms to modern propositional logic on page 48. However, Boole failed to develop or even set the ground for an axiomatic propositional calculus which is the task that Frege picks up. Frege truth functionally explains *modus ponens*; |—B⊃A, |—B, therefore |—A as 'Of the four cases enumerated above, the third is excluded by |—B⊃A and the second and fourth by |—B so that only the first remains' [2, pp. 8–9; 3, pp. 15–16], that is, that both A and B are true. Hence, the truth of A is affirmed by this inference. The more general rule of *modus ponens* is given as:

> For, the truth contained in some other kind of inference can be stated in one judgment, of the form : if *M* holds and if *N* holds, then Λ holds also, or, in signs,
>
> 
>
> From this judgment, together with |—N and |—M, there follow, as above, |—Λ. In this way an inference in accordance with any mode of inference can be reduced to our case. Since it is therefore possible to manage with a single mode of inference, it is a commandment of perspicuity to do so. [2, p. 9; 3, p. 17]

Since Frege, in the axiomatic development of logic, it has become almost a commandment to use *modus ponens* as the only rule of inference along with the rule of substitution. Frege makes conditionals foundational to the development of propositional logic with his notation of the conditional which is also used here to represent the rule of inference of *modus ponens*.

#7 introduces negation notationally as:



The small vertical stroke in the middle is the negation stroke and this represents 'not *A*' which means that the content of *A* does not take place. To the left of the negation stroke the horizontal line is the content stroke of the negation of *A* regardless of what *A* is, whereas the horizontal line to the right of the negation stroke is the content stroke of *A* [2, pp. 10–11; 3, pp. 17–18]. This seems to be a cumbersome way to represent negation. Why could Frege not simply have used '–A' or '~A' as we use today? The significance here has to do more with the philosophy of logic than with logic proper. With Frege's notation we can read negation as either 'not A is asserted' or 'the assertion of A is denied'. In the representation '~A' the literal reading when A is 'The Statue of Liberty is in Delhi' is 'it is not the case that the Statue of Liberty is in Delhi', since the negation sign is outside the sentence. However, with Frege's notation both 'The Statue of Liberty is not in Delhi' and 'it is not the case that the Statue of Liberty is in Delhi' are being expressed literally. If we read the negation as first the denial of A (as the content) and then the assertion of the denial; then the literal reading is that of 'The Statue of Liberty is not in Delhi'. However, if we read the negation stroke as the denial of the affirmation of A, then the literal reading is ‘It is not the case that the Statue of Liberty is in Delhi’. Frege most probably would go for the first reading which is also the ordinary language reading. Nonetheless, the possible ambiguity here is a virtue as we can have both the ordinary language reading and the formal reading of the negation stroke simultaneously.

In #8 Frege defines ‘$\vdash (A \equiv B)$’ as ‘the sign *A* and sign *B* have the same conceptual content’ [2, p. 15; 3, p. 21].

#9 introduces the notions of function and argument:

> *If in an expression, whose content need not be capable of becoming a judgment, a simple or a compound sign has one or more occurrences and if we regard the sign as replaceable in all or some of these occurrences by something else (but everywhere by the same thing), then we call the part that remains invariant in the expression a function, and the replaceable part the argument of the function.* [2, p. 16; 3, p. 22]

Whereas ‘the number 20’ is an independent idea, ‘every positive integer’ is an idea that depends on the context. Functions and arguments may be determinate or indeterminate, or more determinate and less determinate. Also, two different statements may be thought of as having the same function and two different arguments thought of as the same argument with two different functions. The definition of function and argument is not sufficient to claim that it can be implemented formally in logic. Hence, the symbolic representation in #10 is necessary to ground function and argument as foundational for modern logic: $\vdash \Psi(A, B)$, i.e., ‘*A* stands in the relation Ψ to *B*’ [2, p. 18; 3, p. 23]. The logic of relations can now be developed. Whereas Boole algebraised logical propositions he could not make the distinction between propositional logic and the logic of relations. Frege finally makes this distinction as a statement in the form of function and argument is about conceptual content or formal content and not about subjects and predicates.

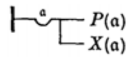
#11 ‘*Generality*’ introduces the universal quantifier:

$$\vdash\!\!\smile^{\mathfrak{a}}\!\!-\!\!-\, \Phi(\mathfrak{a})$$

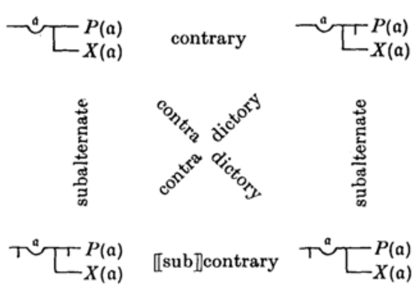
The horizontal stroke to the left of the concavity is the content stroke of the circumstances that whatever we put in for the argument $\Phi(a)$ holds; and the horizontal stroke to the right of the concavity is the content stroke of $\Phi(a)$ where we must imagine something definite for the argument *a* [2, p. 19; 3, p. 24]. Frege provides the notational representation of a bound universal quantifier just 12 lines below his formal representation of function and argument. ‘…the use of quantifiers to bound variables was one of the greatest intellectual inventions of the nineteenth century’ [9, p. 511]. This use was developed from DeMorgan and Schröder to Peirce.

Frege’s share of the emergence of modern logic begins on page 2 with the introduction of the notation that expresses the conceptual content and assertion in the judgment stroke and the declaration on pages 2 and 3 that his formalised representation of judgments no longer requires subject–predicate propositions; on page 16 with the definition of function and argument combined with the formal representation of functions within the judgment stroke on page 18; and on page 19 with the notational introduction of the universal quantifier. Complemented by Boole these pages of *Begriffsschrift* more or less complete the emergence of modern logic.

In #12 Frege uses his fundamental notation to express universal propositions:



i.e., ‘if something has the property X, then it also has the property P’ [2, p. 23; 3, p. 27]. Though this looks very different from the accepted notation today of ‘(x)(Fx⊃Gx)’ we can clearly see how the ‘a’ in the dip binds the conditional so that the notion of bound variable is concretised.[3] Is this the A statement of Aristotle? Boole had already suggested the transition from the propositions of terms to conditional propositions. Frege with his notation of conditional and the universal quantifier could hence represent A statements in their proper conditional form. This is how modern logic subsumes classical logic while eliminating subject–predicate propositions and syllogisms. And all of this can be done because of the notational representation of Frege here. Boole algebraised A propositions as: $x(1- y) = 0$, which means that either $x = 0$, or $y = 1$; that is, in ‘all humans are mortal’ either the antecedent ‘x is a human’ is false or the consequence ‘x is mortal’ is true. Frege replaces A propositions with: ‘if something, a, has the property of being human, then it also has the property of being mortal’, or ‘a is not mortal’ is denied. Frege is aware that in the Aristotelian A propositions the subject term must designate something existent so that ‘all unicorns are one-horned’ is false because unicorns don’t exist. In Fregean logic it is true as it is expressed in the above form, replacing P with Ω and X with O, where Ω stands for ‘one-horned’ and O stands for ‘unicorns’. It is true because the case of $O(a)$ holding, that is, being a unicorn, and $\Omega(a)$ not holding, that is, not being one-horned, does not occur precisely because there are no unicorns. What happens in the Boolean representation? Either ‘x is a unicorn’ is false or ‘x is one-horned’ is true. Since ‘x is a unicorn’ is never true, the A statement is always true. So Boole’s representation of A statements denies the Aristotelian prerequisite of existence, therefore it is not an A statement. Therefore Frege is correct in replacing classical A statements rather than representing them. Frege uses the Aristotelian methodological model of the square of opposition



[2, p. 24; 3, p. 28]

---

[3] I thank another anonymous referee who suggested that I incorporate a conceptual analysis of why and how Frege’s notations work. I have attempted a beginning of such a conceptual analysis in my comments on the conditional, negation, the universal quantifier and on *modus ponens* below. More detailed analyses of these will have to be undertaken in future in another paper.

but refuses to use the Aristotelian labels of A, E, I and O, because of their metaphysical commitment to existence. Hence we have here the desired emergence of modern predicate logic that transcends Aristotelian logic.

Part II is entitled 'REPRESENTATION AND DERIVATION OF SOME JUDGMENTS OF PURE THOUGHT'. Frege provides the axioms of first order propositional and predicate calculus and proves numerous theorems. In #13 Frege, like Aristotle, Leibniz and Boole, wants to find the minimum laws of thought on which all of logic can be built [2, pp. 25–26; 3, 28–29]. The method that follows pronounces this minimalist program. Most logic books today begin with a list of axioms and then, without questioning whether they are the correct axioms, derive theorems from these axioms. This is the classic model of Euclid's *Elements* [14], beginning with 23 definitions, 5 postulates (axioms) and 5 common notions and the propositions (theorems) of geometry are then one by one proved from these foundations. In successive theorems previously proved theorems are used as justifications for steps in the proofs. Frege preserves this structure of the Euclidean proving procedure. However, he does not provide all the axioms at the beginning. Rather the axioms are spread out throughout part II, each being stated when its need occurs. The axioms themselves emerge, so that we can pause and critically examine each axiom and convince ourselves that it really is a law of thought. Euclidean geometry could also have followed this Fregean method, as for example, the notorious postulate 5 is not used until proposition 29 [14, pp. 311–312]. It may have been better to state the postulate when it was needed after proposition 28. If Euclid had followed the Fregean method, then he himself would have wondered whether this postulate was correct when he stated it, as he may not have found it to be a law of thought but rather conjectural. Non-Euclidean geometry then would have begun with Euclid himself as the possibility of alternative postulates to postulate 5 or no postulate in place of it would have been entertained. Frege avoids this deficiency of the Euclidean system as in his system the grounds for alternative axiomatic systems to his own are provided since in the progressive emergence of axioms, any axiom may be replaced by alternative ones that then would lead to alternative theorems. If there are alternative axioms to any of Frege's axioms, then these would have to be tested to see whether they really are laws of thought. Frege scholars seem to have overlooked this Fregean insight on method which is not to be passed over as an idiosyncrasy.

#14 provides the first axiom {A1}[4]:

[2, p. 26; 3, p. 29]

$q \supset (p \supset q)$[5]. This is not an axiom of *PM*, but is prominent as it is listed as the first theorem:

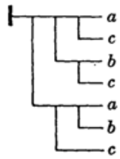
> The most important propositions proved in the present number are the following :
>
> *2.02. $\vdash : q . \supset . p \supset q$.

[4] All the labels for axioms, {A1} thorough {A9} are mine.

[5] Following each representation of an axiom in Fregean notation I immediately provide a symbolization of it in terms of notations that we are most familiar with.

> *I.e.* q implies that p implies q, *i.e.* a true proposition is implied by any proposition. [5, p. 103]

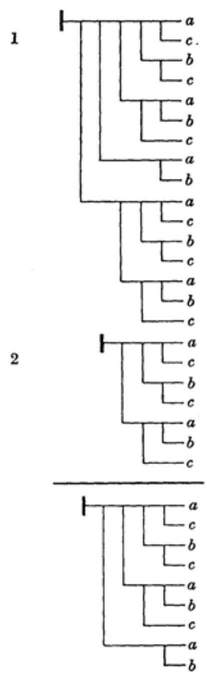
The reason makes it sound like a law of thought, then why did they not consider it as an axiom? Many logicians since *PM* nonetheless have followed Frege in making his first axiom, the first of their sets of axioms: Church [15, p. 72], Imai and Iséki [16, p. 19], Mendelson [17, p. 35]. Frege next gives the second axiom {A2}:



[2, p. 26; 3, p. 29]

$[p\supset(q\supset r)]\supset[(p\supset q)\supset(p\supset r)]$. This is second axiom for Church [15, p.72] and Mendelson [17, p. 35].

In #15 Frege proves the first theorem from {A1} and {A2}:



(3). [2, p. 30; 3, p. 32]

(3) says 'If b implies a, then any proposition c is such that "c implies that b implies a" implies that if "c implies b" then "c implies a"'. This may not be a law of thought, yet it is an immediate consequence of {A1} and {A2} which are laws of thought. Frege's proof employs *modus ponens* and the rule of substitution. From the Fregean picture we can clearly see that 1 is obtained by substitutions in {A1}, the substitution table being given on the previous page [2, 29; 3, p. 31]. 2 is simply {A2}. Since 2 is the bottom part of 1, that is the antecedent, hence we can deduce the top part, that is,

the consequent, which is formula (3), the first theorem. We can formulate this derivation as follows:

1. $\{[p\supset(q\supset r)]\supset[(p\supset q)\supset(p\supset r)]\}\supset\{(q\supset r)\supset([p\supset(q\supset r)]\supset[(p\supset q)\supset(p\supset r)])\}$ <substituting in {A1} '$q\supset r$' for 'p'; and '$[p\supset(q\supset r)]\supset[(p\supset q)\supset(p\supset r)]$' for '$q$'>.
2. $[p\supset(q\supset r)]\supset[(p\supset q)\supset(p\supset r)]$ <A2>

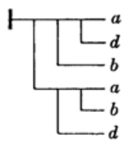
Therefore, 3. $(q\supset r)\supset([p\supset(q\supset r)]\supset[(p\supset q)\supset(p\supset r)])$ <1, 2, *modus ponens*>.

Frege's proof takes up much more space but it may actually be easier to follow as he provides a table for substitution as well. The diagram proof may be more easily understood by those whose right hemisphere is relatively more dominant, whereas the relatively more left hemisphered activity persons would see the three line proof as much more perspicuous and convenient and Frege's proof as cumbersome, difficult to type and wasteful. It is surprising that great logicians and mathematicians actually saw this as a fault in Frege. In 1918 C. I. Lewis stated:

> Besides the precision of notation and analysis, Frege's work is important as being the first in which the nature of rigorous demonstration is sufficiently understood. His proofs proceed almost exclusively by substitution for variables of values of those variables, and the substitution of defined equivalents. Frege's notation, it must be admitted is against him: it is almost diagrammatic, occupying unnecessary space and carrying the eye here and there in a way which militates against easy understanding. It is probably this forbidding character of his medium, combined with the unprecedented demands upon the reader's logical subtlety, which accounts for the neglect which his writings so long suffered. But for this, the revival of logistic proper might have taken place ten years earlier, and dated from Frege's *Grundlagen* rather than Peano's *Formulaire*. [18, p. 115]

After high praise of Frege for providing proofs that proceed from axioms and the rule of substitution alone, bringing in modern logic proving techniques as replacement of syllogistic proofs; Lewis then criticises Frege for his space occupying notations. Why does his notation and proof procedure 'militate against easy understanding'? Lewis does not explain. In 1914, Jourdain remarked: '[…] the using of FREGE'S symbolism as a calculus would be rather like using a three-legged stand-camera for what is called "snap-shot" photography [19, p. viii].' Why are snap-shots to be preferred? Rather, Frege's three-legged stand-camera photography is profound. The diagram of the full proof of (3) is an aesthetic marvel as it beautifully depicts how Frege moves naturally from conditionality to the first two axioms as laws of thought to the use of rule of substitution and *modus ponens* to derive the first theorem of his propositional calculus in formula (3). *Modus ponens* is appropriately often called the 'conditional elimination rule' ('$\supset$ elimination' or '$\rightarrow$ elimination'). When we look at the picture, we see the chunk of 2 as the bottom half of the chunk of 1 and eliminate it to end up with the top chunk of 1 as the conclusion, that is, formula (3). Modern logicians may not find this clarity of seeing very appealing and instead demand a 'metamathematical demonstration'[6]. Frege, being a philosopher, is closely connected to ordinary language and thought. Since *modus ponens* is not stated as an axiom but a rule of inference, we may ask whether it too is a law of thought. Frege might respond that it may not be an obvious law of thought as the axioms are, but it is a rule of inference that no one would deny, but everyone would accept, and that is why the visual mechanism used by him seems very convincing. The Fregean notationalised proof is the perfect complement to Boole's algebraisation of logic as there is a parallel here to Boole's representation of inferences as elimination in simultaneous equations. The two procedures are quite distinct, but there is a structural similarity in that the conclusion is being reached from the premises by using some type of elimination. That is why I use 'parallel' instead of 'similar'. Pages 29–30, where the proof for formula (3) is diagrammed should be added to the stock of pages where modern logic emerges. With axiomatic proof procedure the emergence is now approximately complete since Boole's algebra of logic was not an axiomatic system as both Peirce and Schröder realised [20, p. 59].

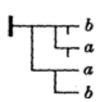
#16 provides the next axiom {A3}:



[2, p. 35; 3, 36]

$[p\supset(r\supset s)]\supset[\mathrm{r}\supset(p\supset s)]$. Is this a clear law of thought? Frege says that the antecedent, the bottom conditional in his picture, amounts to 'the case in which *a* is denied and *b* and *d* are affirmed does not take place', and the antecedent conditional 'means the same'. It means the same because of the commutativity of conjunction, that '*b* and *d*' is the same as '*d* and *b*', which is a law of thought. We know that '$p\supset(r\supset s)$' is the same as '$(p\&r)\supset s$' and Frege also knows that but he has not introduced conjunction as a logical connective. Hence, his reading in terms of a conjunction is a purely truth functional reading from the truth table. Why does Frege pick 'd' instead of 'c' here? Surely it makes no difference which letter one picks, though normally one would pick 'c' which is in sequence after 'a' and 'b'. My sense is that Frege wants to make this stand out as an axiom, and as the third axiom. The first axiom contained only 'a' and 'b', the second axiom contained 'a', 'b' and 'c', so let the third axiom contain 'a', 'b' and 'd' to make it stand out in comparison to the second axiom. In the rest of the section where formulas (9) through (27) are proven (without any more axioms appearing) only letters in sequence appear, which seems to confirm that Frege uses 'd' here to make the axiom stand out as an axiom. In *PM* {A3} appears as proposition *2.04. $\vdash:.p.\supset.q\supset r:\supset:q.\supset.p\supset r$ [5, p. 103]. However, this is an axiom neither for Church nor for Mendelson.

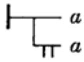
#17 provides the next axiom, {A4}:



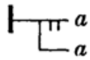
[2, p. 43; 3, p. 44]

[6] This point was again raised by the first anonymous referee.

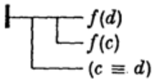
(p⊃q)⊃(~q⊃~p). In *PM* this is *2.16. $\vdash$:*p⊃q.⊃.~q⊃~p* [5, p. 107]. In #18 we get {A5}:

 [2, p. 44; 3, p. 45]

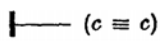
~~p⊃p. In *PM* this is *2.14. $\vdash$.~(~*p*)⊃*p* [5, p. 106]. In #19 we get {A6}:

 [2, p. 47; 3, p. 47]

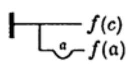
p⊃~~p. In *PM* this is *2.12. $\vdash$.*p*⊃~(~*p*) [5, p. 105]. In #20 we get {A7}:

 [2, 50; 3, p. 50]

(c≡d)⊃[*f*(*c*)⊃*f*(*d*)], i.e., whenever the content of *c* and *d* are the same then in any function we can substitute *d* for *c*. If all propositions can be expressed in terms of function and argument, then this also serves as the law of substitution which has been used all along. #21 provides {A8}:

 [2, 50; 3, 50]

This is obviously a law of thought as it is the Aristotelian law of identity, except here it is in terms of the self identity of the content of a proposition. #22 provides the final axiom {A9}:

 [2, p. 51; 3, p. 51]

(*x*)*f*(*x*)⊃*f*(*c*), which is universal instantiation. Is it a law of thought? Everyone would agree that if all humans are mortal then Socrates is mortal if he is a human. Whereas Boole demonstrated how the algebraisation of Aristotelian logic made a distinction between two types of invalid syllogisms that Aristotelian logic did not make [1, p. 41], Frege shows the reverse, that a distinction made between two types of valid syllogisms in Aristotelian logic, namely Felapton and Fesapo is not made here since their logical form is the same despite different placements of the subject and predicate in the first or major premise [2, pp. 51–2; 3, p. 52].

We come to the end of our trek of the emergence of modern logic in the *Begriffsschrift*.

Part III begins with a definition and derives theorems by applying the axioms to the definition, so that a theory of sequences is put forth which uses only axioms of logic that are pure laws of thought, the definition is also purely logical, and no intuition is involved which may be involved in the mathematicians' account of sequences. This is the beginning of the logicist program, that of reducing all of mathematics to logic. Most Frege scholars find a necessary link between Frege's logicism and his development of logic. Peter Sullivan states:

> Frege's major publications represent three stages in the project that occupied the core of his working life. Subsequently termed 'logicism', […] In *Begriffsschrift* (1879) Frege set out the system of logic without

> which rigorous demonstration of the logicist thesis could not be so much as attempted, and illustrated the power of his system by establishing general results about sequences, including a generalization of the principle of mathematical induction. [21, p. 660]

Peirce argued that though logic depends on mathematics, mathematics does not depend on logic [22, p. 96]. This is not inconsistent with Frege's logicism for both Boole and he have shown the dependence of logic on mathematics as modern logic would not have emerged without the emergence of symbolical algebra and without the mathematical notion of function and argument. Modern logic once it emerges can be used to recapture mathematical theories solely by using laws of thought and logical inferences. Mathematics however does not depend on this logic and mathematicians may develop new theories without framing them in terms of modern logic. The logicist's task then is to track these theories down and reduce them to pure logic. Peirce also made a subtle distinction between mathematics as the science which draws necessary conclusions and logic as the science of drawing necessary conclusions [22, p. 95]. To the mathematician it does not matter how the necessary conclusions are drawn so intuition could well play a role, whereas for a logician conclusions must be drawn by using laws of thought and purely logical inferences. Frege's contribution to the emergence of logic is therefore not necessarily tied to his logicism even though his motivation for developing it is logicism. I differ from most Frege scholars in that I believe that whatever Frege's motivation may have been, the first two parts of the *Begriffsschrift* are autonomous and self-sufficient and bring about the completion of the emergence of modern logic that began with Boole. I stress this since in the development of logic since Frege mathematical logic is often linked with logicism whereas this need not be the case.

One meaning of 'modern logic' as discussed earlier includes the development of higher order logics. Part III does undertake this to some extent as Jose Ferreiros states: 'Upon more careful reading it becomes clear that Frege's system is higher-order throughout, and that he actually deployed higher order tools (this is explicit in the theory of series in the last part of *Begriffsschrift*)' [23, p. 444]. Using my heuristic distinction mentioned at the beginning of the paper, I would not go as far as to say that higher order logics emerged with Frege, but rather the roots of higher order logics are found in Frege, with the emergence and development of higher order logics to come later.

4 Conclusion

Mathematics and philosophy are the proper parents of logic as both are second order disciplines dealing with pure form without content and unlike the sciences they are not directly about the world. Adamson states: 'The distinction of logic from the sciences, as dealing in the abstract with that which is concretely exemplified in each of them, […]' [24, p. 9]. Hence, the biggest names in the origins, emergence and development of formal logic and the roots and emergence of modern symbolic logic are of philosophers like Aristotle, Leibniz, Peirce, Frege, Russell and C. I. Lewis; and mathematicians like Boole, DeMorgan, Schröder and Hilbert. I have pinpointed the emergence of modern logic from Boole's *Mathematical Analysis of Logic* to Frege's

Begriffsschrift. The emergence of modern logic begins with Booole's book where logic was algebraised, and particularly on page 48 where the transition from the Aristotelian logic of classes to propositional calculus is strongly suggested. In Frege's book we find the near completion of the emergence with his innovative notation of the judgment and content strokes that allows us to replace subject–predicate propositions of classical logic [2, pp. 1–3]; of providing a notational representation of the conditional along with a truth functional definition to be read out with the picture of the notation [2, p. 5]; of replacing subject–predicate propositions by function and argument [2, p. 16, 18]; of the symbolic representation of the universal quantifier (p. 19); and the technique of proving theorems by using axioms, the sole inference rule of *modus ponens*, and the rule of substitution [2, p. 30]. The completion of the emergence of modern logic came with the development of metamathematics and metalogic in which the soundness, consistency and completeness of axiomatic propositional as well as predicate logic can be proven.

I have argued that the emergence of modern logic in Boole is partial as he ends up algebraising Aristotelian syllogisms. Frege lays the grounds for the completion of the emergence as the classical pillar of subject–predicate propositions is replaced by functions and arguments, and that of syllogisms is replaced by axioms and axiomatic proofs. By making Aristotelian logic archaic and replacing it with symbolic axiomatic logic Frege completed the task that Boole began and for all intents and purposes from 1847 to 1879 modern logic finally emerged. I say 'finally' because whereas modern science and modern philosophy are usually marked as having emerged in the 17th century and modern mathematics in the 18th century, modern logic did not emerge until the 19th century. Whereas Boole algebraised logic borrowing from mathematics, specifically from symbolical algebra, the idea of combinations without regard to content; Frege borrowed from mathematics the notions of function and argument, but his real intention was to logicise arithmetic once modern logic itself had emerged. Some would argue that the completion did not take place until the beginning of the 20th century until Russell and Whitehead's *Principia Mathematica* and Hilbert's metalogic. I would label these as 'the development of modern logic' rather than the 'emergence of modern logic'. However, I will not push this distinction here within the scope of this paper and will concede to those who want to include these latter contributions as part and parcel of the emergence of modern logic.

Frege's contribution can best be assessed by highlighting an important distinction as stated by Jourdain:

> […] the distinction pointed out by LEIBNIZ between a *calculus ratiocinator* and a […] *lingua characteristica*. […] The objects of a complete logical symbolism are: firstly, […] providing an *ideography*, in which the signs represent ideas and the relations between them *directly* […], and secondly, […], from given premises, we can, in this ideography, draw all the logical conclusions which they imply by means of rules of transformation of formulas analogous to those of algebra,—in fact, in which we can replace reasoning by the almost mechanical process of calculation. This second requirement is the requirement of a *calculus ratiocinator*. It is essential that the ideography should be complete […] {and} concise. The merits of such an ideography are obvious: rigor of reasoning is ensured by the calculus character; we are sure of not

> introducing unintentionally any premise; and we can see exactly on what propositions any demonstration depends. We can […] characterize the dual development of the theory of symbolic logic during the last sixty years as follows: The *calculus ratiocinator* aspect of symbolic logic was developed by BOOLE, DE MORGAN, JEVONS, VENN, C. S. PEIRCE, SCHRODER, Mrs. LADD FRANKLIN and others; the *lingua characteristica* aspect was developed by FREGE, PEANO and RUSSELL. […] FREGE has remarked that his own symbolism is meant to be a *calculus ratiocinator* as well as a *lingua characteristica* […] [19, pp. vii–viii).

Lingua characteristica provides an ideography in which the signs represent concepts and the relations among concepts and *calculus ratiocinator* is the rigorous proofs from axioms in this ideography. Frege, according to the findings of this paper, is right on the mark in claiming that his symbolism is both a *calculus ratiocinator* and a *lingua characteristica* because not only do his notations capture the relations among concepts but they themselves are conceptual inventions, and the understanding of the axioms as laws of thought as well as the proving techniques are totally dependent on the notations and formulations using these notations themselves, hence Frege also provides a *calculus ratiocinator*. The emergence of modern logic is incomplete in Boole who provides a rigorous *calculus ratiocinator* by using algebra to capture logic, but fails to create a new ideography to relate concepts, but more or less accepts the old ideography of Aristotle which is not a universal *lingua characteristica*. Frege creates this new ideography and also provides a rigorous *calculus ratiocinator* for his new *lingua characteristica* which is more universal than Aristotle's, hence satisfying Leibniz's dream, as it can be used not only for all of mathematics, but for science and philosophy as well. I began the discussion of Frege with how he stole an important distinction from under the noses of mathematicians. I end here with a distinction that I steal from the mathematician Jourdain to elevate Frege to the rank of the greatest logician, whereas Frege was mostly ignored by mathematicians for about three decades after the publication of the *Begriffsschrift*.

References

1. Boole, G.: The mathematical analysis of logic: being an essay towards a calculus of deductive reasoning. Macmillan, Barclay and Macmillan, Cambridge (1847)
2. Frege, G.: Begriffsschrift, Eine der arithmetischen nachgebildete Formelsprache des reinen Denkens. Verlag von Louis Nebert, Halle (1879)
3. Frege, G.: Begriffsschrift, a formula language, modeled upon that of arithmetic, for pure thought (translated by Stefan Bauer-Mengelberg). In: van Heijenoort, J. (ed.) From Frege to Gödel: a source book in mathematical logic, 1879–1931, pp. 1–82 (1879). Excel Press, New York (1967)
4. Jetli, P.: The emergence of modern logic. Paper presented at the Third Indian School of Logic and its Applications, University of Hyderabad, India (January 18, 2010), `http://ali.cmi.ac.in/isla2010/slides/jetli-lec.pdf`
5. Whitehead, A.N., Russell, B.: Principia Mathematica, vol. I. Cambridge University Press, Cambridge (1910)

6. Couturat, L.: La logique de Leibniz d'après des documents inédits. Felix Alcan, Paris (1901)
7. Peacock, G.: A treatise on algebra. J. and J. J. Deighton, Cambridge (1830)
8. Boole, G.: An investigation of the laws of thought on which are founded the mathematical theories of logic and probabilities. Walton and Maberly, London (1854)
9. Kneale, W., Kneale, M.: The development of logic. Clarendon Press, Oxford (1962)
10. van Heijenoort, J.: Historical development of modern logic. The Review of Modern Logic 2(3), 242–255 (1992)
11. Haaparanta, L.: Introduction. In: Haaparanta, L. (ed.) The development of modern logic, pp. 3–10. Oxford University Press, Oxford (2009)
12. Thiel, C.: Gottlob Frege and the interplay between logic and mathematics. In: Haaparanta, L. (ed.) The development of modern logic, pp. 196–202. Oxford University Press, Oxford (2009)
13. Cavaliere, F.: L'opera di Hugh MacColl alle origini delle logiche non-classiche. The Review of Modern Logic 6(4), 373–402 (1995)
14. Euclid: The thirteen books of Euclid's Elements. Translated by Sir Thomas L. Heath, vol. I. Dover, New York (1956)
15. Church, A.: Introduction to mathematical logic. Princeton University Press, Princeton (1956)
16. Imai, Y., Iséki, K.: On axiom systems of propositional calculi. Proceedings of the Japan Academy 42(1), 19–22 (1966)
17. Mendelson, E.: Introduction to mathematical logic, 4th edn, 1997. Chapman & Hall, New York (1964)
18. Lewis, C.I.: A survey of symbolic logic. University of California Press, Berkeley (1918)
19. Jourdain, P.E.B.: Preface. In: Couterat, L. (ed.) The algebra of logic (translated by Lydia Gilingham Robinson), pp. iii–x. The Open Court Publishing Company, Chicago (1914)
20. Green, J.: The algebra of logic: what Boole really started. The Review of Modern Logic 4(1), 48–62 (1994)
21. Sullivan, P.M.: Frege's logic. In: Gabbay, D.M., Woods, J. (eds.) Handbook of the history of logic. The rise of modern logic: from Leibniz to Frege, vol. 3, pp. 659–750. Elsevier, London (2004)
22. Houser, N.: Peirce's logic today: (a report on the logic program of the Peirce Sesquicentennial Congress). The Review of Modern Logic 1(1), 92–101 (1990)
23. Ferreiros, J.: The road to modern logic: an interpretation. The Bulletin of Symbolic Logic 7(4), 441–484 (2001)
24. Adamson, R.: A brief history of logic. William Blackwood and Sons, London (1911)

A Modal Logic for Multiple-Source Tolerance Approximation Spaces

Md. Aquil Khan[1,⋆] and Minghui Ma[2]

[1] Department of Mathematics and Statistics,
Indian Institute of Technology,
Kanpur 208016, India
mdaquilkhan@gmail.com
[2] Institute of logic and Intelligence, Chongqing
Department of Philosophy, Tsinghua University, Beijing
mmh.thu@gmail.com

Abstract. Notions of lower and upper approximations are proposed for multiple-source tolerance approximation spaces which consist of a number of tolerance relations over the same domain. A modal logic is proposed for reasoning about the defined notions of approximations. A sound and complete deductive system for the logic is presented. Decidability is also proved.

1 Introduction

Pawlak's rough set theory [15] deals with the approximations of sets using the indiscernibility relations. Pawlak considers the indiscernibility relations to be equivalence, but later it is observed that transitivity is not an obvious property of indiscernibility relations. Thus, the notion of a tolerance approximation space comes into the picture where the indiscernibility relation is taken to be a tolerance, that is, reflexive and symmetric relation (cf. [17,12,7]). The lower and upper approximations in a tolerance approximation space (W, R) are defined in a natural way as follows. Let $R(x) := \{y \in W : (x, y) \in R\}$. Then for $X \subseteq W$, $\underline{X}_R := \{x \in W : R(x) \subseteq X\}$ and $\overline{X}_R := \{x \in W : R(x) \cap X \neq \emptyset\}$. The elements belonging to $\underline{X}_R$, $W \setminus \overline{X}_R$ and $\overline{X}_R \setminus \underline{X}_R$ are respectively called *positive, negative and boundary element* of X.

In the current article, our interest is on the multiple-source extension of the tolerance approximation spaces. Thus, the structures of the form $(W, \{R_i\}_{i \in N})$ are considered, where N is a proper initial segment of the set $\mathbb{N}$ of positive integers representing the set of sources, and each R_i is a tolerance relation on W. Such a structure will be called a *multiple-source tolerance approximation space* ($MTAS$). For each $i \in N$, R_i represents the knowledge base of the i^{th} source of the system. Thus each source perceives the same domain differently

⋆ The first author acknowledges the support of ILLC, Amsterdam, where he was visiting scholar with an EMECW grant financed by the European Commission when the paper was written.

M. Banerjee and A. Seth (Eds.): ICLA 2011, LNAI 6521, pp. 124–136, 2011.

depending on what information she has about the domain. In literature one can find works on the multiple-source situation in rough set theory [13,16]. But our aim is to propose suitable notions of approximations of sets for such a situation. A natural requirement for such a notion should be that it takes into account the information provided by each source of the system. Thus, we may have notions based on either of the following principles.

1. It takes into account views of the sources of the system regarding the membership of the objects. That is, we take into account whether the source considers the object positive, negative or boundary element of the set. So, in that case, we use the approximations $\underline{X}_R$ and $\overline{X}_R$ corresponding to knowledge base R of each source of the system.
2. It takes into account only the information that sources have about the objects, but it does not consider sources' view regarding the membership. That is, we only use the knowledge base R of the sources.

In [8,9,11,10], rough set is studied in multiple-source situation and the notions of approximations are proposed based on the first principle. Although, the indiscernibility relations are taken to be equivalence, results of these articles can obviously be reformulated for tolerance relations as well. In this article, we propose notions of approximations based on the second principle. Furthermore, a modal logic is presented where one can express properties of these approximations.

The remainder of this article is organized as follows. In Sect. 2, we propose some notions of approximations of sets and their properties are explored. Then we focus on a possible logic where one can express these notions. In Sect. 3, we propose such a logic. Coalgebraic perspective of the proposed logic is given in Sect. 3.3. A deductive system is proposed in Sect. 4 and the corresponding soundness and completeness theorems are proved. Finally, decidability of the logic is obtained in Sect. 5. Section 6 concludes the article.

2 Notion of Approximations in Multiple-Source Scenario

The notion of $MTAS$ is defined in Sect. 1 to capture the multiple-source situation. But the notions of approximations that we are going to propose now, is not directly based on it. In fact, we shall consider a particular type of tolerance space [2] and fuzzy approximation space [3]. Let $\mathbb{Q}$ denote the set of rational numbers.

Definition 1. *A* $[0,1] \cap \mathbb{Q}$-fuzzy approximation space *is a tuple* $\mathfrak{S} := (W, \sigma)$*, where* W *is a non-empty set of objects and* $\sigma : W \times W \to [0,1] \cap \mathbb{Q}$*.*

A $[0,1] \cap \mathbb{Q}$*-fuzzy approximation space is called* $[0,1] \cap \mathbb{Q}$-tolerance space *if it satisfies the following additional conditions: for all* $x, y \in W$*,*

– *(reflexivity)* $\sigma(x,x) = 1$*.*
– *(symmetry)* $\sigma(x,y) = \sigma(y,x)$*.*

Let us see how the $[0,1] \cap \mathbb{Q}$-fuzzy approximation spaces are related with the $MTAS$s. Given a $MTAS$ $\mathfrak{F} := (W, \{R_i\}_{i \in N})$, we obtain a $[0,1] \cap \mathbb{Q}$-tolerance space $\mathfrak{G}_{\mathfrak{F}} := (W, \sigma_{\mathfrak{F}})$, where $\sigma_{\mathfrak{F}}(x,y) := \frac{|Ag_{\mathfrak{F}}(x,y)|}{|N|}$. $Ag_{\mathfrak{F}}(x,y)$ denotes the set $\{i \in N : (x,y) \in R_i\}$ consisting of all the sources of $\mathfrak{F}$ which considers x and y indiscernible. Let us call it *standard* $[0,1] \cap \mathbb{Q}$*-tolerance space generated by* $\mathfrak{F}$ following Vakarelov [18]. Is every $[0,1] \cap \mathbb{Q}$-tolerance space standard? Answer is no, but we have the following. For a function $f : X \to Y$, we use the notation $f[X]$ to denote the set $\{f(x) : x \in X\}$.

Proposition 1. *A* $[0,1] \cap \mathbb{Q}$*-tolerance space* $\mathfrak{G} := (W, \sigma)$ *is standard if and only if there exists a positive integer* n *such that* $\sigma[W \times W] \subseteq \{\frac{r}{n} : 0 \leq r \leq n\}$.

Proof. One direction is obvious. In fact, if $\mathfrak{G}$ is generated by $\mathfrak{F} := (W, \{R_i\}_{i \in N})$, then $|N|$ is the desired choice of n. For the converse part, consider the $MTAS$ $\mathfrak{F} := (W, \{R_i\}_{1 \leq i \leq n})$, where

$(x,y) \in R_i$ if and only if $i \leq n \times \sigma(x,y)$.

One can verify that $\mathfrak{G}_{\mathfrak{F}} = \mathfrak{G}$. □

Corollary 1. *A* $[0,1] \cap \mathbb{Q}$*-tolerance space* (W, σ) *with finite* $\sigma[W \times W]$ *is standard.*

From the definition of standard $[0,1] \cap \mathbb{Q}$-tolerance space, it is clear that the function σ of the standard $[0,1] \cap \mathbb{Q}$-tolerance space (W, σ) generated by a $MTAS$ $(W, \{R_i\}_{i \in N})$ gives the possibilities of indiscernibility of two objects *keeping in view the information provided by the sources of the system.* In fact, $\sigma(x,y)$ is the ratio of the number of sources which considers x and y indiscernible to the total number of sources of the system.

Next, we define the notions of approximations for $[0,1] \cap \mathbb{Q}$-fuzzy approximation spaces. It is based on the following idea. Suppose we are given a number of relations instead of just one, representing the knowledge base of the sources. Now, for a given threshold $\lambda \in (0,1] \cap \mathbb{Q}$, we consider a new indiscernibility relation R_λ defined as follows.

$(x,y) \in R_\lambda$ if and only if the ratio of the number of sources which considers x and y indiscernible with the total number of sources of the system exceeds λ.

Now, using this relation, one can define the approximations of the sets in the usual way. Thus, we have the following. For a $[0,1] \cap \mathbb{Q}$-fuzzy approximation space $\mathfrak{G} := (W, \sigma)$, $x \in W$ and $\lambda \in (0,1] \cap \mathbb{Q}$, we will write $x \uparrow^\lambda$ to denote the set $\{y \in W : \sigma(x,y) \geq \lambda\}$.

Definition 2. *The lower approximation* $L_{\mathfrak{G}}$ *and upper approximation* $U_{\mathfrak{G}}$ *of* X *of degree* $\lambda \in (0,1] \cap \mathbb{Q}$ *are defined as follows:*

$$L_{\mathfrak{G}}(X, \lambda) := \{x \in U : x \uparrow^\lambda \subseteq X\},$$
$$U_{\mathfrak{G}}(X, \lambda) := \{x \in U : x \uparrow^\lambda \cap X \neq \emptyset\}.$$

If there is no confusion, we shall omit $\mathfrak{G}$ as the subscript in the above definition. The tolerance approximation space $\mathfrak{F} := (W, R)$ generates the $[0,1]\cap\mathbb{Q}$-tolerance space $\mathfrak{G}_{\mathfrak{F}} := (W, \sigma_{\mathfrak{F}})$, where $\sigma_{\mathfrak{F}}$ is the characteristic function of the relation R. Moreover, for any $\lambda \in (0,1] \cap \mathbb{Q}$, we obtain $x \uparrow^{\lambda} = R(x)$. Thus, we have $\underline{X}_R = L(X, \lambda)$ and $\overline{X}_R = U(X, \lambda)$.

Let us recall the notion of fuzzy approximation space (W, R), $R : W \times W \to [0, 1]$, considered in [3]. The fuzzy upper approximation of a crisp set X is defined as follows:

$$\overline{R}X(x) := \sup_{y \in X} |R(x, y)|.$$

Note that $[0,1] \cap \mathbb{Q}$-fuzzy approximation spaces are also fuzzy approximation spaces. Moreover, for a standard $[0,1]\cap\mathbb{Q}$-tolerance space (W, σ) and $\lambda \in (0,1]\cap \mathbb{Q}$, we obtain

$$x \in U(X, \lambda) \text{ if and only if } \overline{\sigma}X(x) \geq \lambda.$$

Example 1. Let us consider a $MTAS$ $\mathfrak{F} := (W, \{R_i\}_{i\in N})$, where

- $N := \{1, 2, 3\}$ and $U := \{O_1, O_2, \ldots, O_5\}$,
- each R_i is an equivalence relation such that
  - $U|R_1 := \{\{O_1, O_2\}, \{O_4\}, \{O_3\}, \{O_5\}\}$,
  - $U|R_2 := \{\{O_1, O_4\}, \{O_2, O_3\}, \{O_5\}\}$,
  - $U|R_3 := \{\{O_2\}, \{O_1, O_4\}, \{O_3, O_5\}\}$.

Then, we obtain the $[0,1] \cap \mathbb{Q}$-tolerance space $\mathfrak{G}_{\mathfrak{F}} := (W, \sigma_{\mathfrak{F}})$, where

- $\sigma_{\mathfrak{F}}(O_1, O_3) = \sigma_{\mathfrak{F}}(O_1, O_5) = \sigma_{\mathfrak{F}}(O_2, O_4) = \sigma_{\mathfrak{F}}(O_2, O_5) = \sigma_{\mathfrak{F}}(O_3, O_4) = \sigma_{\mathfrak{F}}(O_4, O_5) = 0$,
- $\sigma_{\mathfrak{F}}(O_1, O_2) = \sigma_{\mathfrak{F}}(O_2, O_3) = \sigma_{\mathfrak{F}}(O_3, O_5) = \frac{1}{3}$,
- $\sigma_{\mathfrak{F}}(O_1, O_4) = \frac{2}{3}$.

Suppose it is decided that two objects will be considered indiscernible if it is so for at least half of the sources of the system. In that case, we need to take $\lambda := \frac{1}{2}$. Moreover, we obtain the corresponding lower and upper approximations of the set $X := \{O_2, O_3, O_4\}$ as

$$L(X, \tfrac{1}{2}) = \{O_2, O_3\} \text{ and } U(X, \tfrac{1}{2}) = \{O_1, O_2, O_3, O_4\}.$$

Similarly, if we want to consider two objects discernible if any source of the system can distinguish them, then we need to take $\lambda := 1$. In that case we obtain $L(X, 1) = U(X, 1) = \{O_2, O_3, O_4\}$.

Observe that O_4 is in the lower approximation of X with respect to the knowledge base of source 1, but not in the lower approximation $L(X, \lambda)$ when $\lambda := \frac{1}{2}$. Moreover, O_4 moves to the lower approximation when we relax the restriction and consider $\lambda := 1$.

We end this section with the following proposition listing a few properties of the notions of approximations. We write X^c to denote the set theoretic complement of the set X.

Proposition 2

1. *The following holds in all* $[0,1]\cap\mathbb{Q}$*-fuzzy approximation spaces.*
   (a) $L(X,\lambda) = (U(X^c,\lambda))^c$.
   (b) $L(X\cap Y,\lambda) = L(X,\lambda)\cap L(Y,\lambda)$.
   (c) $L(X,\lambda)\cup L(Y,\lambda) \subseteq L(X\cup Y,\lambda)$.
   (d) *For* $X\subseteq Y$, $L(X,\lambda)\subseteq L(Y,\lambda)$.
   (e) $L(W,\lambda) = W$, *where* W *is the domain of the* $[0,1]\cap\mathbb{Q}$*-fuzzy approximation space.*
   (f) $L(X,\lambda_1)\subseteq L(X,\lambda_2)$ *for* $\lambda_1\leq\lambda_2$.
2. *The following holds in all* $[0,1]\cap\mathbb{Q}$*-tolerance spaces.*
   (a) $L(X,\lambda)\subseteq X$.
   (b) $X\subseteq L(U(X,\lambda),\lambda)$.
3. *In the standard* $[0,1]\cap\mathbb{Q}$*-tolerance space* $\mathfrak{G}_{\mathfrak{F}}$ *generated by a MTAS* $\mathfrak{F} := (W,\{R_i\}_{i\in N})$, *we have* $L(X,1) = \underline{X}_{R_N}$, *where* $R_N := \bigcap_{i\in N} R_i$.

Item (1a) shows that $U(X,\lambda)$ is the dual of $L(X,\lambda)$.

3 Logic for $[0,1]\cap\mathbb{Q}$-Fuzzy Approximation Space

In this section, we shall propose a logic for reasoning about the notions of approximations based on $[0,1]\cap\mathbb{Q}$-fuzzy approximation spaces proposed in Sect. 2. The language $\mathbb{L}$ of the logic consists of a set *Prop* of propositional variables and a set of unary modalities $\{\Diamond_\lambda : \lambda\in(0,1]\cap\mathbb{Q}\}$. Using the Boolean logical connectives $\neg$ (negation) and $\wedge$ (conjunction), well-formed formulae (wffs) of $\mathbb{L}$ are then defined recursively as

$$\alpha := \bot \mid \top \mid p\in Prop \mid \neg\alpha \mid \alpha\wedge\beta \mid \Diamond_\lambda\alpha,$$

where $\bot$ and $\top$ are the logical constants for false and true respectively. The connectives $\rightarrow$, $\leftrightarrow$, $\vee$ and $\Box_\lambda$ are defined in the usual way. We will use $\mathbb{L}$ also to denote the set of all wffs.

3.1 Semantics

The semantics of $\mathbb{L}$, as desired, is based on $[0,1]\cap\mathbb{Q}$-fuzzy approximation spaces. Thus, a model of $\mathbb{L}$ is a tuple $\mathfrak{M} := (\mathfrak{G}, V)$, where $\mathfrak{G} := (W,\sigma)$ is a $[0,1]\cap\mathbb{Q}$-fuzzy approximation space and $V : Prop \rightarrow \wp(W)$ is a valuation function.

Definition 3. *The satisfiability of a wff* α *in a model* $\mathfrak{M}$ *at* $x\in W$, *denoted as* $\mathfrak{M},x\models\alpha$, *is defined inductively:*

– *For each propositional variable* p, $\mathfrak{M},x\models p$, *if and only if* $x\in V(p)$;
– *The standard definitions for the Boolean cases;*
– $\mathfrak{M},x\models\Diamond_\lambda\alpha$ *if and only if there exists* y *with* $\sigma(x,y)\geq\lambda$ *and* $\mathfrak{M},y\models\alpha$.

For any wff α and model $\mathfrak{M}$, let $[\![\alpha]\!]_{\mathfrak{M}} := \{x \in W : \mathfrak{M}, x \models \alpha\}$. α is *valid in* $\mathfrak{M}$, denoted $\mathfrak{M} \models \alpha$, if and only if $[\![\alpha]\!]_{\mathfrak{M}} = W$. α is said to be valid in a $[0,1] \cap \mathbb{Q}$-fuzzy approximation space $\mathfrak{G}$, if and only if $\mathfrak{M} \models \alpha$ for all models $\mathfrak{M}$ based on $\mathfrak{G}$.

α is *valid* in a given class $\mathcal{C}$ of $[0,1] \cap \mathbb{Q}$-fuzzy approximation spaces if α is valid in all $[0,1] \cap \mathbb{Q}$-fuzzy approximation spaces of the class $\mathcal{C}$. Thus, α is valid in the class of all $[0,1] \cap \mathbb{Q}$-tolerance spaces, denoted as $\models_T \alpha$, if α is valid in all $[0,1] \cap \mathbb{Q}$-tolerance spaces.

A wff α is said to be *satisfiable* in a model $\mathfrak{M}$ if $[\![\alpha]\!]_{\mathfrak{M}} \neq \emptyset$.

Remark 1. The above formal language and semantics appear to be similar to that of standard graded model logic (GML) proposed by Kit Fine [6] and explored by some logicians in 1980s (e.g. [5,1]). But there are differences. In GML, we have modal operators $\Diamond_n$ for each non-negative integer n. On the other hand, $\mathbb{L}$ contains modal operators $\Diamond_\lambda$ for each $\lambda \in (0,1] \cap \mathbb{Q}$. Moreover, the satisfiability of a GML wff $\Diamond_n \alpha$ at an object (world) x asks for at least n objects accessible from x where α is true. This is obviously very different from the semantics given in Definition 3.

Remark 2. The modal operators $\Diamond_\lambda$ are very closely related with the modal operators $\Diamond_\lambda^c$ of [4]. The satisfiability condition for the operator $\Diamond_\lambda^c$ can be given as:

$$\mathfrak{M}, x \models \Diamond_\lambda^c \alpha \text{ if and only if } \sup\{\sigma(x,y) : \mathfrak{M}, y \models \alpha\} \geq \lambda.$$

In general, as observed in [4], the modal operators $\Diamond_\lambda$ and $\Diamond_\lambda^c$ are not semantically equivalent, but we have the following relationship:

- $\mathfrak{M}, x \models \Diamond_\lambda^c \alpha$ implies $\mathfrak{M}, x \models \Diamond_\lambda \alpha$.
- $\mathfrak{M}, x \models \Diamond_{\lambda_1} \alpha$ implies $\mathfrak{M}, x \models \Diamond_{\lambda_2}^c \alpha$ for $\lambda_2 < \lambda_1$.

Moreover, satisfiability conditions for $\Diamond_\lambda$ and $\Diamond_\lambda^c$ coincide in a model based on a $[0,1] \cap \mathbb{Q}$-fuzzy approximation space (W, σ) with finite $\sigma[W \times W]$.

The following illustrates how the modalities are used to express lower/upper approximations.

Proposition 3. *Given a model* $\mathfrak{M}$ *and a wff* α*, we have*

- $[\![\Box_\lambda \alpha]\!]_{\mathfrak{M}} = L([\![\alpha]\!]_{\mathfrak{M}}, \lambda)$;
- $[\![\Diamond_\lambda \alpha]\!]_{\mathfrak{M}} = U([\![\alpha]\!]_{\mathfrak{M}}, \lambda)$.

Notation 1. *In the remainder of the article, for any wff* α*,* $\Diamond_0 \alpha$ *will denote* $\top$.

Proposition 2 leads us to the following result.

Proposition 4

1. $\Box_\lambda(\alpha \wedge \beta) \leftrightarrow \Box_\lambda \alpha \wedge \Box_\lambda \beta$ *is valid in all* $[0,1] \cap \mathbb{Q}$*-fuzzy approximation spaces.*
2. *For* $\lambda_1 \leq \lambda_2$*,* $\Box_{\lambda_1} \alpha \to \Box_{\lambda_2} \alpha$ *is valid in all* $[0,1] \cap \mathbb{Q}$*-fuzzy approximation spaces.*
3. $\models_T \Box_\lambda \alpha \to \alpha$.
4. $\models_T \alpha \to \Box_\lambda \Diamond_\lambda \alpha$.

3.2 Relationship with Basic Modal Logic

We assume that the modal language and language $\mathbb{L}$ are based on the same set of propositional variables, and adopt the following notations and definitions.

- Λ: the set of modal wffs with $\triangle$ as the modal operator for possibility;
- $\mathbb{K}$: the class of all Kripke frames;
- $\mathbb{K}_T$: the class of all KTB-Kripke frames (i.e. Kripke frame (W, R), where R is tolerance relation);

A logic L_1 is *embeddable* into a logic L_2, denoted as $L_1 \rightharpoonup L_2$, provided there is a translation $\star$ of wffs of L_1 into L_2, such that $\alpha \in L_1$ if and only if $\alpha^\star \in L_2$ for any wff α of L_1. We use the denotation $L_1 \rightleftharpoons L_2$ when $L_1 \rightharpoonup L_2$ and $L_2 \rightharpoonup L_1$.

Let $\mathbb{K}^M$ ($\mathbb{K}_T^M$) be the class of all $[0,1]\cap\mathbb{Q}$-fuzzy approximation spaces ($[0,1]\cap\mathbb{Q}$-tolerance spaces) (W, σ) satisfying the additional condition $\sigma[W\times W] \subseteq \{0,1\}$. Moreover, we denote by $\mathbb{L}(\mathbb{K}^M)$ and $\mathbb{L}(\mathbb{K}_T^M)$, the logics containing all the wffs valid in all $[0,1]\cap\mathbb{Q}$-fuzzy approximation spaces belonging to $\mathbb{K}^M$ and $\mathbb{K}_T^M$ respectively.

Consider the mapping $\eta : \mathbb{K} \to \mathbb{K}^M$ which maps a Kripke frame (W, R) to (W, σ_R), where σ_R is the characteristic function of R, that is, $\sigma_R(x, y) = 1$ if and only if $(x, y) \in R$. Note that $\eta(\mathbb{K}_T) \subseteq \mathbb{K}_T^M$. Moreover, it is not difficult to see that $\eta : \mathbb{K} \to \mathbb{K}^M$ and $\eta|_{\mathbb{K}_T} : \mathbb{K}_T \to \mathbb{K}_T^M$ are bijections.

Let us consider the translations $* : \Lambda \to \mathbb{L}$ and $+ : \mathbb{L} \to \Lambda$, which fixes propositional variables and $(\triangle\, \alpha)^* := \Diamond_1\alpha^*$, $(\Diamond_\lambda\alpha)^+ := \triangle\, \alpha^+$.

Proposition 5. *Let $\mathfrak{F} := (W, R)$ be a Kripke frame and $V : Prop \to \wp(W)$. Then for all $x \in W$, $\alpha \in \Lambda$ and $\beta \in \mathbb{L}$, we have,*

1. *$(\mathfrak{F}, V), x \models \alpha$ if and only if $(\eta(\mathfrak{F}), V), x \models \alpha^*$;*
2. *$(\mathfrak{F}, V), x \models \beta^+$ if and only if $(\eta(\mathfrak{F}), V), x \models \beta$.*

Proof. The proof is by simple induction on the complexity of α and β. □

Using the fact that $\eta : \mathbb{K} \to \mathbb{K}^M$ and $\eta|_{\mathbb{K}_T} : \mathbb{K}_T \to \mathbb{K}_T^M$ are bijections, we obtain the following as a direct consequence of Proposition 5.

Proposition 6. *Let $\alpha \in \Lambda$ and $\beta \in \mathbb{L}$. Consider the normal modal logic K and KTB.*

1. *$\alpha \in K$ if and only if $\alpha^* \in \mathbb{L}(\mathbb{K}^M)$.*
2. *$\alpha \in KTB$ if and only if $\alpha^* \in \mathbb{L}(\mathbb{K}_T^M)$.*
3. *$\beta^+ \in K$ if and only if $\beta \in \mathbb{L}(\mathbb{K}^M)$.*
4. *$\beta^+ \in KTB$ if and only if $\beta \in \mathbb{L}(\mathbb{K}_T^M)$.*

Thus, the translations $*$ and $+$ gives us,

Proposition 7

1. *$K \rightleftharpoons \mathbb{L}(\mathbb{K}^M)$.*
2. *$KTB \rightleftharpoons \mathbb{L}(\mathbb{K}_T^M)$.*

Proposition 7 shows that the semantics proposed in this article is actually a generalization of that of basic modal logic.

3.3 Coalgebraic Perspective

In this section, we shall use some standard notations related with the general theory of coalgebras. We refer to [19,14] for details. We will use X_ω^Y to denote the set of all functions $g : Y \to X$ with finite $g[Y]$.

Recall that for a set functor T, a T-coalgebra is a tuple (A, ρ), where A is a set and $\rho : A \to TA$ is a function. Let us consider the functor Ω which maps a set X to $([0,1] \cap \mathbb{Q})_\omega^X$ and a function $h : X \to Y$ to the function $\Omega h : ([0,1] \cap \mathbb{Q})_\omega^X \to ([0,1] \cap \mathbb{Q})_\omega^Y$, where for $g \in [0,1] \cap \mathbb{Q})_\omega^X$ and $y \in Y$,

$$(\Omega h(g))(y) := \begin{cases} \max g[h^{-1}(y)], & \text{if } h^{-1}(y) \neq \emptyset; \\ 0, & \text{otherwise.} \end{cases}$$

Note that given an Ω-coalgebra $\mathcal{C} := (W, \rho)$, we obtain a $[0,1] \cap \mathbb{Q}$-fuzzy approximation space $\mathfrak{G}_\mathcal{C} := (W, \sigma_\rho)$, where $\sigma_\rho(x, y) := (\rho(x))(y)$.

Once we observed that the Ω-coalgebras are instances of $[0,1] \cap \mathbb{Q}$-fuzzy approximation spaces, one would like to see whether the semantics proposed in Sect. 3.1 can be seen as a coalgebraic semantics. In fact, it is the case.

Let us associate with each modal operator $\Box_\lambda$, a set-indexed family of functions $\{[\Box_\lambda]_X\})_{X \in Set}$, where $[\Box_\lambda]_X : \wp(X) \to \wp(\Omega X)$ is defined as

$$[\Box_\lambda]_X(A) := \{g \in ([0,1] \cap \mathbb{Q})_\omega^X : \{x \in X : g(x) \geq \lambda\} \subseteq A\},$$

for $A \subseteq X$. One can verify that $\{[\Box_\lambda]_X\})_{X \in Set}$ is a predicate lifting for the functor Ω [14].

Now, we define the satisfiability of a $\mathbb{L}$ wff α in a Ω-coalgebra $\mathcal{C} := (W, \rho)$ under the valuation $V : Prop \to \wp(W)$ at $x \in W$, denoted as $\mathcal{C}, V, x \models^c \alpha$, inductively as follows.

- Standard definitions for the propositional and Boolean cases.
- $\mathcal{C}, V, x \models^c \Box_\lambda \alpha$ if and only if $\rho(x) \in [\Box_\lambda]_W(V_\mathcal{C}(\alpha))$, where $V_\mathcal{C}(\alpha) := \{y \in W : \mathcal{C}, V, y \models^c \alpha\}$.

One can obtain the following proposition relating the two semantics of the language $\mathbb{L}$.

Proposition 8. *Consider an Ω-coalgebra $\mathcal{C} := (W, \rho)$, a valuation $V : Prop \to \wp(W)$ and $[0,1] \cap \mathbb{Q}$-fuzzy approximation space $\mathfrak{G}_\mathcal{C}$. Then for all $x \in W$ and all wffs α, we have*

$$(\mathcal{C}, V), x \models \alpha \text{ if and only if } \mathfrak{G}_\mathcal{C}, V, x \models^c \alpha.$$

4 Axiomatization and Completeness

In this section we present an axiomatic system Λ and prove the corresponding soundness and completeness theorems.

Axiom schema:

1. All instances of propositional tautologies.
2. $\Diamond_{\lambda_1} \alpha \to \Diamond_{\lambda_2} \alpha$ for $\lambda_2 < \lambda_1$.

3. $\Diamond_\lambda(\alpha \vee \beta) \leftrightarrow \Diamond_\lambda \alpha \vee \Diamond_\lambda \beta$.
4. $\Diamond_\lambda \bot \leftrightarrow \bot$.
5. $\alpha \rightarrow \Diamond_1 \alpha$.
6. $\alpha \rightarrow \Box_\lambda \Diamond_\lambda \alpha$.

Rules of inference:

(MON) From $\alpha \rightarrow \beta$ infer $\Diamond_\lambda \alpha \rightarrow \Diamond_\lambda \beta$.
(GEN) From α infer $\Box_\lambda \alpha$.
(MP) From α and $\alpha \rightarrow \beta$ infer β.

The notion of theoremhood is defined in the usual way. We will write $\vdash \alpha$ to denote that α is a theorem of the above deductive system.

Remark 3. The deductive system given above is different from that of GML given in [6,5]. The key observation is that axiom 3, which says that the modalities $\Diamond_\lambda$ distributes over disjunction, is not sound in GML. On the other hand, this axiom is valid in all $[0,1] \cap \mathbb{Q}$-fuzzy approximation spaces. Moreover, the deductive systems given in [6,5] use the derived connectives $\Diamond!_n \alpha := \Diamond_n \alpha \wedge \neg \Diamond_{n+1} \alpha$, where n is a positive integer. But such derived connectives do not make sense in the current frame work.

It is not difficult to obtain the following soundness theorem.

Theorem 1 (Soundness). *If* $\vdash \alpha$*, then* α *is valid in all (standard)* $[0,1] \cap \mathbb{Q}$*-tolerance spaces.*

Now, we prove the completeness theorem. Following the standard technique, we obtain,

Proposition 9. *Every consistent set of wffs has a maximally consistent extension.*

Given a wff α, I_α denotes the set $\{\lambda \in (0,1] \cap \mathbb{Q} : \Diamond_\lambda \text{ occurs in } \alpha\} \cup \{0,1\}$. Note that I_α is finite. Let $\mathbb{L}(I_\alpha)$ be the set of all wffs which involves modalities only from the set $\{\Diamond_\lambda : \lambda \in I_\alpha\}$. Note that $\alpha \in \mathbb{L}(I_\alpha)$.
For maximal consistent sets Γ and Γ', let us consider the set

$$\Delta_\alpha(\Gamma, \Gamma') := \bigcup_{\beta \in \mathbb{L}(I_\alpha) \cap \Gamma'} \{\lambda_1 \in I_\alpha : \Diamond_{\lambda_1} \beta \in \Gamma \ \& \ \neg \Diamond_{\lambda_2} \beta \in \Gamma \text{ for all } \lambda_2 \in I_\alpha \text{ with } \lambda_2 > \lambda_1\}.$$

We note the following fact about $\Delta_\alpha(\Gamma, \Gamma')$.

Proposition 10. $\Delta_\alpha(\Gamma, \Gamma')$ *is non-empty.*

Proof. Consider the set $S := \bigcup_{\beta \in \mathbb{L}(I_\alpha) \cap \Gamma'} \{\lambda \in I_\alpha : \lambda > 0 \ \& \ \Diamond_\lambda \beta \in \Gamma\}$. If S is empty, then we obtain $0 \in \Delta_\alpha(\Gamma, \Gamma')$. Otherwise, the largest integer belonging to S will also belong to $\Delta_\alpha(\Gamma, \Gamma')$. $\Box$

Let us now describe the *canonical model* $\mathfrak{M}^\Lambda_\alpha := (\mathfrak{G}^\Lambda_\alpha, V^\Lambda)$, $\mathfrak{G}^\Lambda_\alpha := (W^\Lambda, \sigma^\Lambda_\alpha)$, required for the proof of the completeness theorem.

Definition 4 (Canonical model)

$W^{\Lambda} = \{\Gamma : \Gamma$ *is a maximally consistent set*$\}$;
$\sigma^{\Lambda}_{\alpha} : W^{\Lambda} \times W^{\Lambda} \to [0,1] \cap \mathbb{Q}$ *is such that* $\sigma^{\Lambda}_{\alpha}(\Gamma, \Gamma') := \min \Delta_{\alpha}(\Gamma, \Gamma')$;
$V^{\Lambda} : Prop \to W^{\Lambda}$ *is such that* $V^{\Lambda}(p) = \{\Gamma \in W^{\Lambda} : p \in \Gamma\}$.

Observe that due to Proposition 10, $\sigma^{\Lambda}_{\alpha}$ is well-defined.

In order to obtain the existence lemma, we will require the following.

Lemma 1. *Let* Γ *and* Γ' *be maximal consistent sets and* $\lambda \in I_{\alpha}$. *Then,*

$$\sigma^{\Lambda}_{\alpha}(\Gamma, \Gamma') \geq \lambda \text{ if and only if } \Diamond_{\lambda}\beta \in \Gamma \text{ for all } \beta \in \Gamma' \cap \mathbb{L}(I_{\alpha}).$$

Proof. The result follows easily when $\lambda = 0$. So, let us consider the case when $\lambda > 0$. First suppose $\Diamond_{\lambda}\beta \in \Gamma$ for all $\beta \in \Gamma' \cap \mathbb{L}(I_{\alpha})$. If possible, let $\sigma^{\Lambda}_{\alpha}(\Gamma, \Gamma') = \lambda' < \lambda$. Then, we must have a $\beta \in \Gamma' \cap \mathbb{L}(I_{\alpha})$ such that (i) $\Diamond_{\lambda'}\beta \in \Gamma$ and (ii) $\Diamond_{\lambda^*}\beta \notin \Gamma$ for all $\lambda^* \in I_{\alpha}$ with $\lambda^* > \lambda'$. But this will give us $\Diamond_{\lambda}\beta \notin \Gamma$, a contradiction.

Conversely, suppose $\sigma^{\Lambda}_{\alpha}(\Gamma, \Gamma') \geq \lambda$. If possible, let $\beta \in \Gamma' \cap \mathbb{L}(I_{\alpha})$ be such that $\Diamond_{\lambda}\beta \notin \Gamma$. Let $k := \min\{\lambda' \in I_{\alpha} : \Diamond_{\lambda'}\beta \notin \Gamma\}$. Then $k \leq \lambda$ and $k \neq 0$ as $\Diamond_{k}\beta \notin \Gamma$. If $k = \min(I_{\alpha} \setminus \{0\})$, then we obtain $0 \in \Delta_{\alpha}(\Gamma, \Gamma')$ (using axiom 2) and hence $\sigma^{\Lambda}_{\alpha}(\Gamma, \Gamma') = 0$, a contradiction. So, we assume $k \neq \min(I_{\alpha} \setminus \{0\})$. Choose the largest $\lambda_1 \in I_{\alpha}$ such that $\lambda_1 < k$. Then $\Diamond_{\lambda_1}\beta \in \Gamma$ and $\Diamond_{\lambda_2}\beta \notin \Gamma$ for all $\lambda_2 > \lambda_1$ and $\lambda_2 \in I_{\alpha}$. Thus, we obtain $\lambda_1 \in \Delta_{\alpha}(\Gamma, \Gamma')$. Therefore, $\lambda_1 \geq \min \Delta_{\alpha}(\Gamma, \Gamma') \geq \lambda \geq k$. This contradicts that $\lambda_1 < k$. $\square$

Lemma 2 (Existence Lemma). *Let* Γ *be a maximal consistent set and* $\lambda \in I_{\alpha} \setminus \{0\}$. *Suppose* $\beta \in \mathbb{L}(I_{\alpha})$ *is such that* $\Diamond_{\lambda}\beta \in \Gamma$. *Then, there exists a maximal consistent set* Γ' *such that* $\sigma^{\Lambda}_{\alpha}(\Gamma, \Gamma') \geq \lambda$ *and* $\beta \in \Gamma'$.

Proof. Follows from Proposition 9, Lemma 1 and the fact that the set $\{\beta\} \cup \{\gamma \in \mathbb{L}(I_{\alpha}) : \Box_{\lambda}\gamma \in \Gamma\}$ is consistent. $\square$

Lemma 3 (Truth Lemma). *For any wff* $\beta \in \mathbb{L}(I_{\alpha})$ *and* $\Gamma \in W^{\Lambda}$,

$$\beta \in \Gamma \text{ if and only if } \mathfrak{M}^{\Lambda}_{\alpha}, \Gamma \models \beta.$$

We also note the following fact about the canonical $[0,1] \cap \mathbb{Q}$-fuzzy approximation space $\mathfrak{G}^{\Lambda}_{\alpha}$.

Proposition 11. $\mathfrak{G}^{\Lambda}_{\alpha}$ *is a standard* $[0,1] \cap \mathbb{Q}$*-tolerance space.*

Proof. It is not difficult to verify that $\mathfrak{G}^{\Lambda}_{\alpha}$ is a $[0,1] \cap \mathbb{Q}$-tolerance space using Lemma 1. Also note that $\sigma^{\Lambda}_{\alpha}[W^{\Lambda} \times W^{\Lambda}]$ is a finite set containing 1. Thus, using Corollary 1, we obtain $\mathfrak{G}^{\Lambda}_{\alpha}$ as a standard $[0,1] \cap \mathbb{Q}$-tolerance space. $\square$

As a direct consequence of Truth Lemma, we obtain the following completeness theorem.

Theorem 2 (Completeness). *If* β *is valid in the class of all standard* $[0,1] \cap \mathbb{Q}$*-tolerance spaces, then* $\vdash \beta$.

Proof. If possible, let $\nvdash \beta$. Then, $\neg\beta$ is consistent and hence we obtain $\Gamma \in W^{\Lambda}$ containing $\neg\beta$. Now consider the canonical model $\mathfrak{M}^{\Lambda}_{\beta}$. As $\neg\beta \in \mathbb{L}(I_{\beta}) \cap \Gamma$, by Truth Lemma, we obtain $\mathfrak{M}^{\Lambda}_{\beta}, \Gamma \models \neg\beta$. But this contradicts that β is valid in all standard tolerance $[0,1] \cap \mathbb{Q}$-fuzzy approximation spaces. Thus we obtain the desired result. □

Using soundness and completeness theorems, we also obtain the following proposition showing that the logics for the $[0,1] \cap \mathbb{Q}$-tolerance spaces and standard $[0,1] \cap \mathbb{Q}$-tolerance spaces are same.

Proposition 12. *A wff α is valid in the class of all $[0,1] \cap \mathbb{Q}$-tolerance spaces if and only if it is valid in the class of all standard $[0,1] \cap \mathbb{Q}$-tolerance spaces.*

We end this section with the following observation. Let Λ_1 be the deductive system consisting of the axioms 1-4, and the inference rules (MON), (GEN) and (MP). Following the technique given above, one can show the soundness and completeness of Λ_1 with respect to the class of $[0,1] \cap \mathbb{Q}$-fuzzy approximation spaces.

5 Finite Model Property and Decidability

In this section, our aim is to prove the following result.

Theorem 3. *Given a wff α, we can decide whether there exists a model $\mathfrak{M}$ based on a standard $[0,1] \cap \mathbb{Q}$-tolerance space $\mathfrak{G}$ such that α is satisfiable in $\mathfrak{M}$.*

Observe that due to Proposition 12, this will also give us the decidability with respect to the class of all $[0,1] \cap \mathbb{Q}$-tolerance spaces.

In order to prove Theorem 3, we shall show that the proposed logic has the finite model property. For this, we shall use the technique similar to filtration.

Let Σ denote a finite sub-wff closed set of wffs. Let $\mathfrak{M} := (\mathfrak{G}, V)$ be a model based on a standard tolerance $[0,1] \cap \mathbb{Q}$-fuzzy approximation space $\mathfrak{G} := (W, \sigma)$. Note that $\sigma[W \times W]$ is finite.

We define an equivalence relation $\equiv_{\Sigma}$ on W as follows:

$$x \equiv_{\Sigma} y, \text{ if and only if for all } \beta \in \Sigma,\ \mathfrak{M}, x \models \beta \text{ if and only if } \mathfrak{M}, y \models \beta.$$

Definition 5 (Filtration model). *Given a model $\mathfrak{M} = (W, \sigma, V)$ and Σ as above, we define a model $\mathfrak{M}^f = (W^f, \sigma^f, V^f)$, where*

- *$W^f := \{[x] : x \in W\}$, $[x]$ is the equivalence class of x with respect to the equivalence relation $\equiv_{\Sigma}$;*
- *$\sigma^f([x],[y]) := \max\{\sigma(x', y') : x' \in [x],\ y' \in [y]\}$.*
- *$V^f(p) := \{[x] : x \in V(p)\}$.*

We note the following facts.

Proposition 13

1. *$\sigma^f[W^f \times W^f] \subseteq \sigma[W \times W]$.*
2. *$\mathfrak{M}^f$ is a standard $[0,1] \cap \mathbb{Q}$-tolerance space.*
3. *W^f contains at most $2^{|\Sigma|}$ elements.*

Proposition 14 (Filtration Theorem). *For all wffs $\beta \in \Sigma$ and all $x \in W$,*

$$\mathfrak{M}, x \models \beta \text{ if and only if } \mathfrak{M}^f, [x] \models \beta.$$

Let us write $Sub(\alpha)$ to denote the set of all sub-wffs of α. Also recall that I_α denotes the finite set $\{\lambda \in (0,1] \cap \mathbb{Q} : \Diamond_\lambda \text{ occurs in } \alpha\} \cup \{0,1\}$.

Theorem 4. *If a wff α is satisfiable in a standard $[0,1] \cap \mathbb{Q}$-tolerance space, then it is also satisfiable in a standard $[0,1] \cap \mathbb{Q}$-tolerance space $\mathfrak{S}^* := (W^*, \sigma^*)$ with $|W^*| \leq 2^{|Sub(\alpha)|}$ and $\sigma^*(W^* \times W^*) \subseteq I_\alpha$.*

Proof. If α is satisfiable in a standard $[0,1] \cap \mathbb{Q}$-tolerance space, then using soundness theorem, we obtain $\nvdash \neg\alpha$. This implies that α is satisfiable in the canonical model $\mathfrak{M}_\alpha^\Lambda$. Now taking $\Sigma := Sub(\alpha)$ and using Propositions 13 and 14, we obtain the desired result. □

Theorem 3 follows from Theorem 4.

6 Conclusions

Notions of $[0,1] \cap \mathbb{Q}$-fuzzy approximation space and multiple-source tolerance approximation space are considered and relationship between them is determined. Notions of lower and upper approximations for these structures are defined. A modal logic is proposed which can express these notions of approximations. It is observed that although the syntax and semantics of the proposed logic appears to be similar to that of the graded modal logic, but there are differences. Questions of axiomatization and decidability of the logic are also addressed.

In this paper, we have considered indiscernibility relations to be tolerance. A natural question would be about the extension of the current work to the case where indiscernibility relations are equivalence. The situation will not remain so simple in that case. In fact, one needs to come up with a suitable condition on the $[0,1] \cap \mathbb{Q}$-fuzzy approximation spaces which will give us the counterpart of Proposition 1 when the tolerance relation is replaced with the equivalence relation.

References

1. De Caro, F.: Graded modalities, II (canonical models). Studia Logica 47(1), 1–10 (1988)
2. Doherty, P., Łukaszewicz, W., Szałas, A.: Tolerance spaces and approximative representational structures. In: Günter, A., Kruse, R., Neumann, B. (eds.) KI 2003. LNCS (LNAI), vol. 2821, pp. 475–489. Springer, Heidelberg (2003)
3. Dubois, D., Prade, H.: Rough fuzzy sets and fuzzy rough sets. International Journal of General Systems 17, 191–200 (1990)
4. Esteva, F., Garcia, P., Godo, L.: A modal account of similarity-based reasoning. International Journal of Approximate Reasoning 16, 235–260 (1997)
5. Fattorosi-Barnaba, M., De Caro, F.: Graded modalities, I. Studia Logica 44, 197–221 (1985)

6. Fine, K.: In so many possible worlds. Notre Dame Journal of Formal Logic 13(4), 516–520 (1972)
7. Järinen, J.: Approximations and rough sets based on tolerances. In: Ziarko, W.P., Yao, Y. (eds.) RSCTC 2000. LNCS (LNAI), vol. 2005, pp. 182–189. Springer, Heidelberg (2001)
8. Khan, M.A., Banerjee, M.: Formal reasoning with rough sets in multiple-source approximation systems. International Journal of Approximate Reasoning 49(2), 466–477 (2008)
9. Khan, M.A., Banerjee, M.: Multiple-source approximation systems: membership functions and indiscernibility. In: Wang, G., Li, T., Grzymala-Busse, J.W., Miao, D., Skowron, A., Yao, Y. (eds.) RSKT 2008. LNCS (LNAI), vol. 5009, pp. 80–87. Springer, Heidelberg (2008)
10. Khan, M.A., Banerjee, M.: A preference-based multiple-source rough set model. In: Szczuka, M., Kryszkiewicz, M., Ramanna, S., Jensen, R., Hu, Q. (eds.) RSCTC 2010. LNCS, vol. 6086, pp. 247–256. Springer, Heidelberg (2010)
11. Khan, M.A., Banerjee, M.: A study of multiple-source approximation systems. In: Peters, J.F., Skowron, A., Słowiński, R., Lingras, P., Miao, D., Tsumoto, S. (eds.) Transactions on Rough Sets XII. LNCS, vol. 6190, pp. 46–75. Springer, Heidelberg (2010)
12. Komorowski, J., Pawlak, Z., Polkowski, L., Skowron, A.: Rough sets: a tutorial. In: Pal, S.K., Skowron, A. (eds.) Rough Fuzzy Hybridization: A New Trend in Decision-Making, pp. 3–98. Springer, Singapore (1999)
13. Orłowska, E., Pawlak, Z.: Expressive power of knowledge representation systems. International Journal of Man-Machine Studies 20(5), 485–500 (1984)
14. Pattinson, D.: Coalgebraic logics and applications. In: Fourth International Joint Conference on Automated Reasoning (2008), http://2008.IJCAR.org
15. Pawlak, Z.: Rough sets. International Journal of Computer and Information Science 11(5), 341–356 (1982)
16. Rauszer, C.M.: Rough logic for multiagent systems. In: Masuch, M., Polos, L. (eds.) Logic at Work 1992. LNCS (LNAI), vol. 808, pp. 161–181. Springer, Heidelberg (1994)
17. Skowron, A., Stepaniuk, J.: Tolerance approximation spaces. Fundamenta Informaticae 27, 245–253 (1996)
18. Vakarelov, D.: Abstract characterization of someknowledge representation systems and the logic NIL ofnondeterministic information. In: Jorrand, P., Sgurev, V. (eds.) Artificial Intelligence II, pp. 255–260. North-Holland, Amsterdam (1987)
19. Venema, Y.: Algebras and coalgebras. In: van Benthem, J., Blackburn, P., Wolter, F. (eds.) Handbook of Modal Logic, pp. 331–426. Elsevier, Amsterdam (2006)

A Note on Nathanial's Invariance Principle in Polyadic Inductive Logic

Jeff B. Paris and Alena Vencovská*

School of Mathematics
University of Manchester
Manchester M13 9PL, UK
jeff.paris@manchester.ac.uk,
alena.vencovska@manchester.ac.uk

Abstract. As one direction of a Representation Theorem for probability functions satisfying Nathanial's Invariance Principle in Polyadic Inductive Logic we exhibit a family of basic probability functions satisfying this principle. We conjecture that every probability function satisfying this principle can be approximated arbitrarily closely by a convex combination of these basic solutions.

Keywords: Nathanial's Invariance Principle, Symmetry, Polyadic Inductive Logic, Probability Logic, Logical Probability, Rationality.

Introduction

The motivation for this contribution comes from polyadic inductive logic which is now a well studied area that could be viewed as a continuation of Carnap's program from the mid 20th century to formalize probabilistic reasoning and spell out its implications for induction. Carnap (see for example [1], [2], [3], [4], [10]), and others investigated mainly the case involving unary properties whilst the more modern study also involves relations, binary or of higher arities.

Our framework consists of a language L with constants $a_1, a_2, \ldots$ and predicates $\{R_1, \ldots, R_q\}$ (with arities $r_1, \ldots, r_q$ respectively), no function symbols or equality.

A *probability function on* L is a function w from the set SL of sentences of L into $[0,1] \subset \mathbb{R}$ satisfying that for $\theta, \phi, \exists x\, \psi(x) \in SL$:

(P1) If $\vdash \theta$ then $w(\theta) = 1$.

(P2) If $\vdash \neg(\theta \wedge \phi)$ then $w(\theta \vee \phi) = w(\theta) + w(\phi)$.

(P3) $w(\exists x\, \psi(x)) = \lim_{n\to\infty} w(\bigvee_{i=1}^{n} \psi(a_i))$.

* Supported by a UK Engineering and Physical Sciences Research Council (EPSRC) Research Associateship.

M. Banerjee and A. Seth (Eds.): ICLA 2011, LNAI 6521, pp. 137–146, 2011.

By a theorem of Gaifman [6], we know that such a probability function is determined[1] by its values on the *state descriptions* of L, that is the sentences of the form $\Theta(b_1, b_2, \dots, b_n)$,

$$\Theta(b_1, b_2, \dots, b_n) \;=\; \bigwedge_{d=1}^{q} \;\bigwedge_{i_1, i_2, \dots, i_{r_d} \in \{1, \dots, n\}} \pm R_d(b_{i_1}, b_{i_2}, \dots, b_{i_{r_d}}), \qquad (1)$$

where $\pm R$ stands for R or $\neg R$ respectively and $b_1, b_2, \dots, b_n$ stand for some distinct choices from the constants a_i. Notice that the state descriptions $\Theta(b_1, b_2, \dots, b_n)$ for $b_1, b_2, \dots, b_n$ are disjoint and exhaustive.

The problem in its basic form is to determine which probability functions are good candidates for rational inference: given some information on probabilities of some sentences, what probabilities should a rational agent adopt for other sentences? It can be argued (see [5] for a discussion of "The Received View" (TVR)) that via conditional probability the problem reduces to justifying a choice of a probability function which is good for inference on the basis of no information and this is the problem we deal with. (Carnap's conclusion for the unary case was that these priors should be from what is now called *Carnap's Continuum of Inductive Methods.*)

Many principles have been proposed to direct the quest for such a probability distribution and narrow the possibilities. The most commonly accepted one is the

Constant Exchangeability Principle, Ex
If $\theta(x_1, \dots, x_n)$ is a formula of L which does not mention any constants then

$$w(\theta(b_1, \dots, b_n)) = w(\theta(b'_1, \dots, b'_n) \qquad (2)$$

(where the $b_1, \dots, b_n$ and $b'_1, \dots, b'_n$ are some distinct choices from the a_i). We remark that an equivalent formulation of Ex can be obtained by requiring that (2) holds merely for all *state formulae*

$$\Theta(x_1, \dots, x_n) = \bigwedge_{s=1}^{q} \;\bigwedge_{i_1, i_2, \dots, i_{r_s} \in \{1, \dots, n\}} \pm R_s(x_{i_1}, \dots, x_{i_{r_s}}).$$

Another well-studied principle useful in the polyadic context is the Principle of Spectrum Exchangeability (see for example [9]). To formulate it, we need some

[1] Conversely, any function f from the set of all state descriptions to nonnegative real numbers satisfying $f(\emptyset) = 1$ (where by convention $\emptyset$ is the state description for no individuals understood to be a tautology) and

$$\sum_{\substack{\Phi(b_1, \dots, b_n, b_{n+1}) \vdash \Theta(b_1, \dots, b_n) \\ \Phi \text{ is a state description}}} f(\Phi(b_1, \dots, b_n, b_{n+1})) = f(\Theta(b_1, \dots, b_n))$$

whenever $b_1, \dots, b_n, b_{n+1}$ are distinct constants and $\Theta(b_1, \dots, b_n)$ is a state description, extends uniquely to a probability function.

notation. Given a state description $\Theta(b_1, b_2, \ldots, b_n)$ and $i, j \in \{1, \ldots, n\}$, we say that b_i, b_j are *indistinguishable* mod Θ, written $b_i \sim_\Theta b_j$, if $\Theta(b_1, \ldots, b_n) \,\&\, b_i = b_j$ is consistent[2]. Clearly $\sim_\Theta$ is an equivalence relation.

We define the *spectrum* of Θ, denoted $\mathcal{S}(\Theta)$, to be the multiset of sizes of the (non-empty) equivalence classes with respect to $\sim_\Theta$. By convention we list any spectrum in a decreasing order.

We will use a simple example of L having a single binary predicate R to illustrate our concepts throughout the paper. In this situation a state description for $b_1, \ldots, b_n$ is a conjunction $\bigwedge_{i,j \in \{1,\ldots,n\}} \pm R(b_i, b_j)$ and can be represented by a $n \times n$ matrix with entries 0 and 1, where 1 or 0 respectively as the ij-th entry means that the state description implies $R(b_i, b_j)$ or $\neg R(b_i, b_j)$ respectively. If a state description $\Theta(b_1, b_2, b_3, b_4)$ is represented by

$$\begin{matrix} 1 & 1 & 1 & 1 \\ 1 & 0 & 1 & 0 \\ 1 & 1 & 1 & 1 \\ 1 & 0 & 1 & 0 \end{matrix}$$

then $b_1 \sim_\Theta b_3$ and $b_2 \sim_\Theta b_4$, and $\mathcal{S}(\Theta) = \{2, 2\}$.

The Spectrum Exchangeability Principle, Sx

If $\Theta(b_1, \ldots, b_n), \Phi(b'_1, \ldots, b'_n)$ are state description and $\mathcal{S}(\Theta) = \mathcal{S}(\Phi)$ then

$$w(\Theta(b_1, \ldots, b_n)) = w(\Phi(b'_1, \ldots, b'_n))$$

(where again the $b_1, \ldots, b_n$ and $b'_1, \ldots, b'_n$ are some distinct choices from the a_i).

Note that Sx clearly implies Ex. In the case that all the predicates are unary (call them $P_1, \ldots, P_q$ to emphasize that we consider a special case) state descriptions are

$$\bigwedge_{i=1}^{n} \alpha_{h_i}(b_i)$$

where the $\alpha_h(x)$, $h = 1, \ldots, 2^q$ are the *atoms* of L, that is formulae of the form

$$\pm P_1(x) \wedge \pm P_2(x) \wedge \ldots \wedge \pm P_q(x).$$

Sx thus reduces to *Atom Exchangeability*, Ax, asserting that

$$w\left(\bigwedge_{i=1}^{n} \alpha_{h_i}(b_i)\right)$$

depends only on the multiset of $|\{i \,|\, h_i = j\}|$ for $j = 1, \ldots, 2^q$.

---

[2] Equivalently, for any R_d $(d = 1, \ldots, q)$ and any $t_1, \ldots, t_{r_d} \in \{1, \ldots, n\}$ (not necessarily distinct), the sentence $R_d(b_{t_1}, b_{t_2}, \ldots, b_{t_{r_d}})$ appears positively as a conjunct in $\Theta(b_1, b_2, \ldots, b_n)$ if and only if $R_i(b_{s_1}, b_{s_2}, \ldots, b_{s_{r_d}})$ also appears positively as a conjunct in $\Theta(b_1, b_2, \ldots, b_n)$ where $\langle s_1, \ldots, s_{r_d}\rangle$ is the result of replacing any number of occurrences of i in $\langle t_1, \ldots, t_{r_d}\rangle$ by j or vice-versa.

As a natural generalization of Ax, Sx inherits some respectability since Ax (also called *attribute symmetry*) was endorsed by Carnap and other researchers in the field (for a discussion see for example [13]), and it is satisfied by the Carnap Continuum functions. [However, there are criticisms that can and have been raised against Ax, in particular its denial of analogical support, as observed already by Carnap himself in [4].]

The principle of Spectrum Exchangeability, likewise, has pleasing consequences. It became considerably easier to understand the situation after some de Finetti style representation theorems were proved (see [7], [8], [12]), which state that under some fairly natural additional conditions, probability functions satisfying Sx are continuous averages of some canonical relatively simple probability functions satisfying Sx.

However, it may be the case that Spectrum Exchangeability is in some senses too strong a generalization of Ax. A recent effort at clarifying precisely what could be meant when principles are advocated on grounds of symmetry (involving considering automorphisms of the underlying structures) sheds some light on this: whilst Ax has a very good and clear justification along these lines, Sx has so far remained unaccounted for, and a principle somewhat more delicate than Sx appears to arise much more naturally from these considerations. A detailed account of the motivation for this new principle, the so called *Nathanial's Invariance Principle*, is given in [11]. It derives from the requirement that a rational probability function should be invariant under automorphisms of the underlying structure, in particular that state descriptions that can be mapped to each other via such an automorphism should receive the same probability. A criterion for this to hold, and which in turn begets Nathanial's Invariance Principle, is the relation of *similarity* between state descriptions, a relation which we now describe.

For σ a surjection from the set of (distinct) constants $\{b_1, \ldots, b_s\}$ onto the set of (distinct) constants $\{b'_1 \ldots, b'_t\}$, denoted $\sigma : \{b_1, \ldots, b_s\} \twoheadrightarrow \{b'_1, \ldots, b'_t\}$, and $\Phi(b'_1, \ldots, b'_t)$ a state description, there is a unique state description $\Theta(b_1, \ldots, b_s)$ such that

$$\Theta(\sigma(b_1), \ldots, \sigma(b_s)) \equiv \Phi(b'_1, \ldots, b'_t)$$

($\equiv$ stands for logical equivalence). We denote this state description Θ by

$$(\Phi(b'_1, \ldots, b'_t))_\sigma(b_1, \ldots, b_s)$$

or Φ_σ if the constants are clear from the context.

Thus Φ_σ as above is a state description for $b_1, \ldots, b_s$, and all the b_i mapped by σ to the same b'_j are indistinguishable in Φ_σ from each other. Their role in Φ_σ is like that of b'_j in Φ. Φ and Φ_σ in a sense share the same structure except that in Φ_σ some constants have further 'clones'.

Continuing our example with a single binary relation, let Φ be the state description for b'_1, b'_2 represented by the matrix

$$\begin{matrix} 1 & 1 \\ 1 & 0 \end{matrix}$$

If $\sigma : \{b_1, b_2, b_3, b_4\} \twoheadrightarrow \{b'_1, b'_2\}$ maps b_1 and b_3 to b'_1, and b_2 and b_4 to b'_2 then Φ_σ is the state description Θ for b_1, b_2, b_3, b_4 (which we have considered before) represented by the matrix

$$\begin{matrix} 1 & 1 & 1 & 1 \\ 1 & 0 & 1 & 0 \\ 1 & 1 & 1 & 1 \\ 1 & 0 & 1 & 0 \end{matrix}$$

Now let $\Phi(b_1, \ldots, b_n)$ be a state description and $t_1, \ldots, t_k$ some numbers from $\{1, \ldots, n\}$. We define $\Phi(b_1, \ldots, b_n)[b_{t_1}, \ldots, b_{t_k}]$ to be the restriction of $\Phi(b_1, \ldots, b_n)$ to $\{b_{t_1}, \ldots, b_{t_k}\}$, that is, the conjunction of those $\pm R_d(i_1, \ldots, i_{r_d})$ implied by $\Phi(b_1, \ldots, b_n)$ with all $i_1, \ldots, i_{r_d}$ from amongst the $t_1, \ldots, t_k$. We may omit $(b_1, \ldots, b_n)$ from $\Phi(b_1, \ldots, b_n)[b_{t_1}, \ldots, b_{t_k}]$ and write simply $\Phi[b_{t_1}, \ldots, b_{t_k}]$ if it is clear in the context that Φ is a state description for $b_1, \ldots, b_n$.

Using our example again, if $\Theta(b_1, b_2, b_3, b_4)$ is as above, then $\Theta[b_1, b_3]$ and $\Theta[b_1, b_2]$ respectively are represented by

$$\begin{matrix} 1 & 1 \\ 1 & 1 \end{matrix} \qquad\qquad \begin{matrix} 1 & 1 \\ 1 & 0 \end{matrix}$$

respectively. We say that the state descriptions $\Theta(b_1, \ldots, b_n)$ and $\Phi(b_1, \ldots, b_n)$ are *similar* if for each $t_1, \ldots, t_m$ (distinct) and $s_1, \ldots, s_k$ (also distinct) from $\{1, \ldots, n\}$ and $\sigma : \{s_1, \ldots, s_k\} \twoheadrightarrow \{t_1, \ldots, t_m\}$ we have

$$\Theta[b_{s_1}, \ldots, b_{s_k}] \equiv (\Theta[b_{t_1}, \ldots, b_{t_m}])_\sigma \iff \Phi[b_{s_1}, \ldots, b_{s_k}] \equiv (\Phi[b_{t_1}, \ldots, b_{t_m}])_\sigma$$

In particular, with $m = k$ and $\sigma(s_i) = t_i$ for each $i \in \{1, \ldots, k\}$ we obtain that

$$\Theta[b_{s_1}, \ldots, b_{s_k}] \equiv (\Theta[b_{t_1}, \ldots, b_{t_k}])(b_{s_1}/b_{t_1}, \ldots, b_{s_k}/b_{t_k})$$
$$\iff \Phi[b_{s_1}, \ldots, b_{s_k}] \equiv (\Phi[b_{t_1}, \ldots, b_{t_k}])(b_{s_1}/b_{t_1}, \ldots, b_{s_k}/b_{t_k})$$

where $(\Theta[b_{s_1}, \ldots b_{s_k}])(b_{t_1}/b_{s_1}, \ldots, b_{t_k}/b_{s_k})$ results from $\Theta[b_{s_1}, \ldots b_{s_k}]$ by replacing each b_{s_i} by b_{t_i}, $i \in \{1, \ldots, k\}$ and similarly for Φ. This property also holds when $t_1, \ldots, t_k$ and $s_1, \ldots, s_k$ are not necessarily distinct.

Also, if $t_1, \ldots, t_k, t_{k+1}$ are from $\{1, \ldots, n\}$ and $b_{t_k}, b_{t_{k+1}}$ are indistinguishable in $\Theta[b_{t_1}, \ldots, b_{t_k}, b_{t_k+1}]$ then they must also be indistinguishable in $\Phi[b_{t_1}, \ldots, b_{t_k}, b_{t_{k+1}}]$ because $b_{t_k}, b_{t_{k+1}}$ being indistinguishable in $\Theta[b_{t_1}, \ldots, b_{t_k}, b_{t_{k+1}}]$ is equivalent to

$$\Theta[b_{t_1}, \ldots, b_{t_k}, b_{t_{k+1}}] \equiv (\Theta[b_{t_1}, \ldots, b_{t_k}])_\sigma$$

where $\sigma : \{b_{t_1}, \ldots, b_{t_{k+1}}\} \twoheadrightarrow \{b_{t_1}, \ldots, b_{t_k}\}$ maps $b_{t_{k+1}}$ to b_{t_k} and b_{t_i} to itself for $i \leq k$.

For example, considering state descriptions $\Theta_1, \Theta_2, \Theta_3$ for b_1, b_2, b_3 represented by

$$\begin{matrix} 1\,1\,1 \\ 1\,0\,1 \\ 1\,0\,0 \end{matrix} \qquad\qquad \begin{matrix} 1\,1\,1 \\ 1\,0\,0 \\ 1\,1\,0 \end{matrix} \qquad\qquad \begin{matrix} 1\,1\,1 \\ 1\,0\,1 \\ 1\,1\,0 \end{matrix}$$

respectively. With some effort Θ_1 and Θ_2 can be seen to be similar, but Θ_3 is not similar to them since $\Theta_3[b_2, b_3]$ is logically equivalent to $\Theta_3[b_2, b_3](b_3/b_2, b_2/b_3)$ which is not the case for Θ_1 and Θ_2.

Considerations such as the above lead us to propose the following new principle:

Nathanial's Invariance Principle, NIP. If the state descriptions $\Theta(b_1, \ldots, b_n)$ and $\Phi(b_1, \ldots, b_n)$ are similar then

$$w(\Theta(b_1, \ldots, b_n)) = w(\Phi(b_1, \ldots, b_n)).$$

Given how useful the representation theorems proved to be when studying Sx we would naturally wish to have such powerful tools at our disposal in the case of NIP. This contribution defines what appear to be the right canonical probability functions for this case, and includes proofs that they do satisfy NIP and Ex, and also the Principle of Language Invariance defined later. The technically more difficult conjectured result about probability distributions satisfying NIP and Ex being in some sense averages of these canonical probability distributions is deferred to a forthcoming paper.

The NIP Representation Theorem

For each $k \geq 1$ let $\mathbb{E}_k$ be the set of equivalence relations $\equiv_k$ on $\{1, 2, \ldots\}$.

Let $\mathbb{E} \subseteq \mathbb{E}_1 \times \mathbb{E}_2 \times \mathbb{E}_3 \times \ldots$ contain those sequences of equivalence relations

$$\bar{E} = \langle \equiv_1, \equiv_2, \equiv_3, \ldots \rangle$$

such that the following condition hold:

If $\langle c_1, \ldots, c_k \rangle \equiv_k \langle d_1, \ldots, d_k \rangle$ then for $s_1, \ldots, s_m \in \{1, \ldots, k\}$ (not necessarily distinct)

$$\langle c_{s_1}, \ldots, c_{s_m} \rangle \equiv_m \langle d_{s_1}, \ldots, d_{s_m} \rangle;$$

Note that if $\bar{E} = \langle \equiv_1, \equiv_2, \equiv_3, \ldots \rangle$ is in $\mathbb{E}$ then the following must hold:

- If $\langle c_1, \ldots, c_k \rangle \equiv_k \langle d_1, \ldots, d_k \rangle$ and σ is a permutation of $\{1, 2, \ldots, k\}$ then
$$\langle c_{\sigma(1)}, \ldots, c_{\sigma(k)} \rangle \equiv_k \langle d_{\sigma(1)} \ldots, d_{\sigma(k)} \rangle,$$
- If $\langle c_1, \ldots, c_k \rangle \equiv_k \langle d_1, \ldots, d_k \rangle$ and $d_s = d_t$ for some $s < t$ then
$$\langle c_1, \ldots, c_s \ldots, c_t, \ldots, c_k \rangle \equiv_k \langle c_1, \ldots, c_s, \ldots, c_s, \ldots, c_k \rangle$$

($\boldsymbol{c}$ is equivalent to $\boldsymbol{c}$ with c_t replaced by c_s).

Let $\mathbb{B}_0$ be the set of sequences of real numbers $\langle p_1, p_2, \ldots \rangle$ such that $0 \leq p_i \leq 1$ for all i, $\Sigma_{i=1}^{\infty} p_i = 1$, and $p_1 \geq p_2 \geq p_3 \geq \ldots$.

The probability functions $u_L^{\bar{p},\bar{E}}$. For $\bar{p} \in \mathbb{B}_0$, $\langle \equiv_1, \equiv_2, \equiv_3, \ldots \rangle = \bar{E} \in \mathbb{E}$ and $\Theta(b_1, \ldots, b_n)$ a state description we define $u_L^{\bar{p},\bar{E}}(\Theta)$ to be the probability that Θ is arrived at by the process described below.

We take an urn containing balls with colours $1, 2, \ldots$ in proportions $p_1, p_2, \ldots$.

We choose a sequence $\langle c_1, \ldots, c_n \rangle$ so that each c_i is chosen independently to be k with probability p_k. We define a binary relation $\sim_m^{\bar{c}}$ on $\{b_1, \ldots, b_n\}^m$ for each m as follows:

$$\langle b_{i_1}, \ldots, b_{i_m} \rangle \sim_m^{\bar{c}} \langle b_{j_1}, \ldots, b_{j_m} \rangle \Longleftrightarrow \langle c_{i_1}, \ldots, c_{i_m} \rangle \equiv_m \langle c_{j_1}, \ldots, c_{j_m} \rangle.$$

$\sim_m^{\bar{c}}$ is clearly an equivalence relation.

For each $d = 1, \ldots, q$ and each equivalence class A of $\sim_{r_d}^{\bar{c}}$ we choose one of

$$\bigwedge_{\langle b_{i_1}, \ldots, b_{i_{r_d}} \rangle \in A} R_d(b_{i_1}, \ldots, b_{i_{r_d}}), \qquad \bigwedge_{\langle b_{i_1}, \ldots, b_{i_{r_d}} \rangle \in A} \neg R_d(b_{i_1}, \ldots, b_{i_{r_d}})$$

(each with probability $\frac{1}{2}$). Given the sequence $\langle c_1, \ldots, c_n \rangle$, there are 2^g possible state descriptions which we can thus obtain by taking the conjunction of these choices, where g is the sum for $d = 1, \ldots q$ of the number of equivalence classes associated with R_d in $\{b_1, \ldots, b_n\}^{r_d}$ with respect to the equivalence $\sim_{r_d}^{\bar{c}}$, and each of these state descriptions is obtained with probability $\frac{1}{2^g}$. $u_L^{\bar{p},\bar{E}}(\Theta(b_1, \ldots, b_n))$ is the sum of the probabilities of choosing $\langle c_1, \ldots, c_n \rangle$ and a state description respecting the $\sim_{r_d}^{\bar{c}}$ as above which equals $\Theta(b_1, \ldots, b_n)$.

Note that due to the condition on $\bar{E}$, for a sequence $\langle c_1, \ldots, c_n \rangle$ to contribute a non-zero factor to $u_L^{\bar{p},\bar{E}}(\Theta(b_1, \ldots, b_n))$ it must be the case that for any natural number $m \geq 1$, if for some $i_1, \ldots, i_m, j_1, \ldots, j_m$ from $\{1, \ldots, n\}$

$$\langle c_{i_1}, \ldots, c_{i_m} \rangle \equiv_m \langle c_{j_1}, \ldots, c_{j_m} \rangle \tag{3}$$

then $b_{i_1}, \ldots, b_{i_m}$ and $b_{j_1}, \ldots, b_{j_m}$ 'behave' the same as each other within $\Theta(b_1, \ldots, b_n)$. I.e. for any $d \in \{1, \ldots, q\}$ and any $k_1, \ldots, k_{r_d} \in \{1, \ldots, m\}$, $R_d(b_{i_{k_1}}, \ldots, b_{i_{k_{r_d}}})$ appears positively as a conjunct in $\Theta(b_1, \ldots, b_n)$ if and only if $R_d(b_{j_{k_1}}, \ldots, b_{j_{k_{r_d}}})$ also appears positively as a conjunct in $\Theta(b_1, \ldots, b_n)$, which is perhaps better expressed as

$$\Theta[b_{i_1}, \ldots b_{i_m}] \equiv (\Theta[b_{j_1}, \ldots b_{j_m}])(b_{i_1}/b_{j_1}, \ldots, b_{i_m}/b_{j_m}).$$

The same conclusion follows from

$$\langle b_{i_1}, \ldots, b_{i_m} \rangle \sim_m^{\bar{c}} \langle b_{j_1}, \ldots, b_{j_m} \rangle$$

in place of (3).

Another important point to notice is that if $\langle c_1, \ldots, c_n \rangle$ contributes a non-zero factor to $u_L^{\bar{p},\bar{E}}(\Theta(b_1, \ldots, b_n))$ and $c_i = c_j$ for some $i, j \in \{1, \ldots, n\}$ then b_i and b_j are indistinguishable in $\Theta(b_1, \ldots, b_n)$. This is true since if $d \in \{1, \ldots, q\}$ and $\langle t_1, \ldots, t_{r_d} \rangle$ is an r_d-tuple of numbers from $\{1, \ldots, n\}$ and $\langle s_1, \ldots, s_{r_d} \rangle$ obtains from $\langle t_1, \ldots, t_{r_d} \rangle$ by replacing any number of occurrences of i by j or vice-versa then $\langle b_{t_1}, \ldots, b_{t_{r_d}} \rangle \sim^{\bar{c}}_{r_d} \langle b_{s_1}, \ldots, b_{s_{r_d}} \rangle$.

Finally we remark that in the special case when $\bar{E} = \langle Id_1, Id_2, \ldots, \rangle$ (with Id_k being the finest equivalence on $\mathbb{E}_k$, that is, equality), $u_L^{\bar{p},\bar{E}}$ are the same as the previously studied canonical functions satisfying Sx, $u_L^{\bar{p}}$ (with $p_0 = 0$), see [7], [8], [12].

Theorem 1. *$u_L^{\bar{p},\bar{E}}$ defined above for state descriptions determines a probability function satisfying Ex and NIP.*

Proof. The process described above yields a state description for $b_1, \ldots, b_n$ with probability 1 so the values that $u_L^{\bar{p},\bar{E}}$ gives to state descriptions for $b_1, \ldots, b_n$ sum to 1 and are non-negative. The value given to a particular state description for $b_1, \ldots, b_n$ is the sum of the values given to all state descriptions for $b_1, \ldots, b_n, b_{n+1}$ which extend it. To see this, notice that for $\bar{c} = \langle c_1, \ldots, c_n \rangle$ and $\bar{c}^+ = \langle c_1, \ldots, c_n, c_{n+1} \rangle$, $\sim^{\bar{c}^+}_{r_d}$ differs from $\sim^{\bar{c}}_{r_d}$ by having possibly some additional classes (h_d of them) containing r_d-tuples featuring b_{n+1}, and by extending some old classes by adding some r_d-tuples featuring b_{n+1} to them. If $\bar{c} = \langle c_1, \ldots, c_n \rangle$ contributes a non-zero factor (which must be equal to $\left(\prod_{j=1}^n p_{c_j}\right) \frac{1}{2^g}$, where g is the sum for $d = 1, \ldots q$ of the number of equivalence classes in $\{b_1, \ldots, b_n\}^{r_d}$ with respect to the equivalence $\sim^{\bar{c}}_{r_d}$) to $u_L^{\bar{p},\bar{E}}(\Theta(b_1, \ldots, b_n))$ then there are $2^{\sum_{d=1}^q h_d}$ extensions of Θ that $\bar{c}^+ = \langle c_1, \ldots, c_n, c_{n+1} \rangle$ contributes to, and the factor contributed to each of them is

$$p_{c_{n+1}} \times \left(\prod_{j=1}^n p_{c_j} \right) \frac{1}{2^g} \frac{1}{2^{\sum_{d=1}^q h_d}}.$$

Hence the sum of all factors contributing to extensions of Θ is $u_L^{\bar{p},\bar{E}}(\Theta)$. These observations allow us to conclude that $u_L^{\bar{p},\bar{E}}$ determines a probability function. Ex holds since it makes no difference in the process which particular constants are involved. To show NIP, suppose that $\Theta(b_1, \ldots, b_n) \sim \Phi(b_1, \ldots, b_n)$ and let $c_1, \ldots, c_n$ be a sequence of colours. If non-zero, then as above the contributions of this sequence to $u_L^{\bar{p},\bar{E}}(\Theta)$, $u_L^{\bar{p},\bar{E}}(\Phi)$ must be the same[3]. Hence the contributions could only differ if one was zero and the other one non-zero, so assume the contribution of $\langle c_1, \ldots, c_n \rangle$ to $u_L^{\bar{p},\bar{E}}(\Theta)$ is 0, whilst its contribution to $u_L^{\bar{p},\bar{E}}(\Phi)$ is non-zero.

The contribution of $\langle c_1, \ldots, c_n \rangle$ to $u_L^{\bar{p},\bar{E}}(\Theta)$ being 0 means that there is some $d \in \{1, \ldots, q\}$ and $i_1, \ldots, i_{r_d}, j_1, \ldots, j_{r_d}$ such that $\langle b_{i_1}, \ldots, b_{i_r} \rangle \sim^{\bar{c}}_{r_d} \langle b_{j_1}, \ldots, b_{j_{r_d}} \rangle$ with $\Theta \models R_d(b_{i_1}, \ldots, b_{i_{r_d}})$ and $\Theta \models \neg R_d(b_{j_1}, \ldots, b_{j_{r_d}})$. However, $\bar{c}$ contributing a non-zero factor to $u_L^{\bar{p},\bar{E}}(\Phi)$ also means that

$$\Phi[b_{i_1}, \ldots b_{i_{r_d}}] \equiv (\Phi[b_{j_1}, \ldots b_{j_{r_d}}])(b_{i_1}/b_{j_1}, \ldots, b_{i_{r_d}}/b_{j_{r_d}})$$

[3] Namely $\left(\prod_{j=1}^n p_{c_j}\right) \frac{1}{2^g}$, where g is the sum for $d = 1, \ldots q$ of the number of equivalence classes in $\{b_1, \ldots, b_n\}^{r_d}$ with respect to the equivalence $\sim^{\bar{c}}_{r_d}$.

which by similarity implies

$$\Theta[b_{i_1},\dots b_{i_{r_d}}] \equiv (\Theta[b_{j_1},\dots b_{j_{r_d}}])(b_{i_1}/b_{j_1},\dots,b_{i_{r_d}}/b_{j_{r_d}}),$$

contradiction.

A principle which proved to be remarkably powerful in combination with Sx is that of Language Invariance, see in particular [7]. The motivation behind this principle is that whilst we may at any one time wish to consider only a certain fixed collection of predicates and work with a probability function on this fixed language, it should be possible to add other predicates and have probability functions for such richer languages so that when considering sentences from the original language we still get the original probabilities.

Language Invariance The probability function w on L satisfies *Language Invariance* if there exists a class of probability functions $w_{\mathcal{L}}$ for each finite predicate language[4] $\mathcal{L}$ such that whenever $\mathcal{L}'$ is a sublanguage of $\mathcal{L}$ then w restricted to $S\mathcal{L}'$ equals $w_{\mathcal{L}'}$ and $w_L = w$.

In this case we shall describe the $w_{\mathcal{L}}$ as a *language invariant family* containing w. If all the $w_{\mathcal{L}}$ satisfy NIP then we refer to them as *language invariant family with NIP.*

We conclude this paper with a proof that for a fixed $\bar{p} \in \mathbb{B}_0$ and $\bar{E} \in \mathbb{E}$ the functions $u_L^{\bar{p},\bar{E}}$ in fact provide such a family.

Theorem 2. *Let $\bar{p} \in \mathbb{B}_0$ and $\bar{E} \in \mathbb{E}$. Then the probability functions $u_L^{\bar{p},\bar{E}}$ form a language invariant family with NIP.*

Proof. Let $\Theta(b_1,\dots,b_n)$ be a state description of L and let $L' = L \cup \{R\}$ where R is a new, k-ary predicate. Assume that $\overline{c} = c_1,\dots,c_n$ contributes $\left(\prod_{i=1}^n p_{c_i}\right)\frac{1}{2^g}$ to $u_L^{\bar{p},\bar{E}}(\Theta(b_1,\dots,b_n))$. Let h be the number of equivalence classes, with respect to $\sim_k^{\bar{c}}$, in $\{b_1,\dots,b_n\}^k$. Each state description in L' that extends Θ and is contributed to in $u_{L'}^{\bar{p},\bar{E}}(\Theta(b_1,\dots,b_n))$ is determined by h choices, one for each of these equivalence classes, and the contribution from $c_1,\dots,c_n$ to such a state description is $\left(\prod_{i=1}^n p_{c_i}\right)\frac{1}{2^{g+h}}$. But since there are 2^h of these extensions, the overall contribution to Θ is the same using L' as it was using L.

References

1. Carnap, R.: Logical Foundations of Probability. University of Chicago Press, Chicago. Routledge & Kegan Paul Ltd. (1950)
2. Carnap, R.: The continuum of inductive methods. University of Chicago Press, Chicago (1952)
3. Carnap, R., Jeffrey, R.C. (eds.): Studies in inductive logic and probability, vol. I. University of California Press, Berkeley (1971)

[4] As before, with constants a_i and no equality nor function symbols.

4. Carnap, R.: A basic system of inductive logic. In: Jeffrey, R.C. (ed.) Studies in Inductive Logic and Probability, vol. II, pp. 7–155. University of California Press, Berkeley (1980)
5. Fitelson, B.: Inductive Logic, http://fitelson.org/il.pdf
6. Gaifman, H.: Concerning measures on first order calculi. Israel Journal of Mathematics 2, 1–18 (1964)
7. Landes, J., Paris, J.B., Vencovská, A.: A Characterization of the Language Invariant Families satisfying Spectrum Exchangeability in Polyadic Inductive Logic. Annals of Pure and Applied Logic 161, 800–811 (2010)
8. Landes, J., Paris, J.B., Vencovská, A.: Representation Theorems for Probability Functions satisfying Spectrum Exchangeability in Inductive Logic. International Journal of Approximate Reasoning 51, 35–55 (2009)
9. Landes, J., Paris, J.B., Vencovská, A.: Some aspects of Polyadic Inductive Logic. Studia Logica 90, 3–16 (2008)
10. Paris, J.B.: Pure Inductive Logic. In: Pettigrew, R., Horsten, L. (eds.) The Continuum Companion to Philosophical Logic. Continuum International Publishing Group, London (to appear)
11. Paris, J.B., Vencovská, A.: Symmetry Principles in Polyadic Inductive Logic. Submitted to the Journal of Logic, Language and Information
12. Paris, J.B., Vencovská, A.: From unary to binary inductive logic. To appear in the Proceedings of the Second Indian Conference on Logic and its Applications, Mumbai (2007)
13. Zabel, S.L.: Carnap and the Logic of Induction. In: Handbook of the History of Logic Logic, vol. 10. Elsevier, Amsterdam (2009)

First-Order Inquisitive Pair Logic

Katsuhiko Sano

Department of Humanistic Informatics, Graduate School of Letters
Kyoto University / JSPS
katsuhiko.sano@gmail.com

Abstract. We introduce two different calculi for a first-order extension of inquisitive pair semantics (Groenendijk 2008): Hilbert-style calculus and Tree-sequent calculus. These are first-order generalizations of (Mascarenhas 2009) and (Sano 2009), respectively. First, we show the strong completeness of our Hilbert-style calculus via canonical models. Second, we establish the completeness and soundness of our Tree-sequent calculus. As a corollary of the results, we semantically establish that our Tree-sequent calculus enjoys a cut-elimination theorem.

1 Introduction

Groenendijk [1] first introduced the *inquisitive pair semantics* for a language of propositional logic to capture both classical and inquisitive meanings of a sentence. For example, the classical meaning of $\mathbf{p} \vee \mathbf{q}$ is $|\mathbf{p} \vee \mathbf{q}|$ and the inquisitive meaning of it is $\{ |\mathbf{p}|, |\mathbf{q}| \}$, where $|A|$ is the set of all truth functions making A true. In the first logical study for inquisitive pair semantics [2], Mascarenhas revealed that the corresponding *inquisitive pair logic* is an axiomatic extension of intuitionistic logic (however, it is not closed under uniform substitutions) and established the completeness of it. Independently, following the idea of [3], the author gave a complete and cut-free Gentzen-style sequent calculus for inquisitive pair logic [4]. After these studies, Ciardelli and Roelofsen [5] generalized inquisitive pair semantics within the propositional level and revealed that their *generalized inquisitive logic* has various beautiful logical properties.

Disjunction $\vee$ allows us to formalize an English sentence containing 'or'. However, in order to handle the sentences containing quantifications as well as 'which', 'who', etc., we need a first-order extension of inquisitive semantics. Ciardelli [6] studied how to give a recursive definition of inquisitive meaning in a first-order setting. As far as the author knows, however, there is no complete axiomatization of first-order inquisitive logic, though there was a preliminary study toward this direction [7, Ch.6]. This paper contributes to this point. In this paper, we focus on a first-order extension of the original *inquisitive pair semantics* and give two different complete calculi for a *first-order inquisitive pair logic*: Hilbert-style calculus and Gentzen-style sequent calculus. We can regard these as first-order generalizations of [2] and [4], respectively.

There are various ways of considering first-order extensions of intuitionistic logic for Kripke semantics: e.g. by expanding the domain or keeping it constant. Following [7, Ch.6], this paper also concerns the constant-domain semantics, which means that we adopt **CD**: $\forall x. (A \vee B(x)) \to (A \vee \forall x. B(x))$ (x is not free in A) as our logical axiom. In the first part of this paper, we establish the correspondence between the first-order

M. Banerjee and A. Seth (Eds.): ICLA 2011, LNAI 6521, pp. 147–161, 2011.

inquisitive models and a specific class of constant-domain Kripke models (Theorem 1). After introducing the Hilbert-style axiomatization of first-order inquisitive pair logic, we use the correspondence above and the canonical model method [8, Ch.7.2] to establish the strong completeness (Corollary 1). In the second part, we first extend the sequent calculus of [4] to cover the quantifiers (**CD** gives us the simpler rule, cf. [3,9]), and then, we establish the completeness (Theorem 3) and soundness (Theorem 5) of our Tree-sequent calculus. By combining these with the results of the first part, we can semantically establish the cut-elimination theorem of our sequent calculus.

In the propositional level, the generalized inquisitive logic is a 'limit' of a hierarchy of inquisitive logics [7, Ch.6], one of which is the inquisitive pair logic. Therefore, based on this study, the author hopes that we could also 'approximate' a generalized first-order inquisitive logic by considering the corresponding first-order hierarchy.

2 Inquisitive Semantics and Constant-Domain Kripke Semantics

2.1 Inquisitive Pair Semantics

Our syntax $\mathcal{L}$ consists of a countable set $\mathsf{VAR} = \{ x_i \mid i \in \omega \}$ of variables, a countable set $\{ c_i \mid i \in \omega \}$ of constant symbols, a countable set of predicate symbols P, the propositional connectives: $\bot$, $\neg$, $\rightarrow$, $\wedge$, $\vee$, the quantifiers: $\forall$, $\exists$, and the parentheses: (,). t is a *term* if t is a variable or a constant symbol. Then, the *formulas* of $\mathcal{L}$ are defined as usual. We use Γ and Δ, etc. to denote a (possibly infinite) set of formulas. For a finite Γ, $\bigwedge \Gamma$ (or, $\bigvee \Gamma$) is defined as the conjunction (or, disjunction) of all formulas of Γ, if Γ is non-empty; otherwise $\top$ (or, $\bot$, respectively). $A[t/x]$ denotes the result of the simultaneous substitution of t for all free occurrences of x in A.

An (*first-order*) *inquisitive model* $\mathfrak{M}$ consists of a non-empty set W, a non-empty set D, and a valuation V satisfying $c^V \in D$ and $P_w^V \subseteq D^n$ ($w \in W$), where n is the arity of P[1]. Given any $W \neq \emptyset$, we say that $s \subseteq W$ is *pairwise* if $\#s \leq 2$ and $s \neq \emptyset$. Given any inquisitive model $\mathfrak{M} = \langle W, R, D \rangle$, any pairwise $s \subseteq W$, any *assignment* $g : \mathsf{VAR} \rightarrow D$, and any formula A, the satisfaction relation $s, g \models_{\mathfrak{M}} A$ is defined by:

$$
\begin{array}{lll}
s, g \models_{\mathfrak{M}} P(t_1, \ldots, t_n) & \text{iff} & \langle \overline{g}(t_1), \ldots, \overline{g}(t_n) \rangle \in P_w^V \text{ for any } w \in s; \\
s, g \models_{\mathfrak{M}} \bot & & \text{Never ;} \\
s, g \models_{\mathfrak{M}} \neg A & \text{iff} & \text{for any pairwise } s' \subseteq s\text{: } s', g \not\models_{\mathfrak{M}} A; \\
s, g \models_{\mathfrak{M}} A \wedge B & \text{iff} & s, g \models_{\mathfrak{M}} A \text{ and } s, g \models_{\mathfrak{M}} B; \\
s, g \models_{\mathfrak{M}} A \vee B & \text{iff} & s, g \models_{\mathfrak{M}} A \text{ or } s, g \models_{\mathfrak{M}} B; \\
s, g \models_{\mathfrak{M}} A \rightarrow B & \text{iff} & \text{for any pairwise } s' \subseteq s\text{: } s', g \models_{\mathfrak{M}} A \text{ implies } s', g \models_{\mathfrak{M}} B; \\
s, g \models_{\mathfrak{M}} \forall x. A & \text{iff} & \text{for any } a \in D\text{: } s, g(x|a) \models_{\mathfrak{M}} A; \\
s, g \models_{\mathfrak{M}} \exists x. A & \text{iff} & \text{for some } a \in D\text{: } s, g(x|a) \models_{\mathfrak{M}} A,
\end{array}
$$

where $\overline{g}(t) := g(x)$ (if $t \equiv x$); c^V (if $t \equiv c$), and $g(x|a)$ is the *x-variant* of g such that $g(x|a)(x) = a$. We usually drop the subscript $\mathfrak{M}$ from $\models_{\mathfrak{M}}$, if it is clear from the context.

Given any $\mathfrak{M} = \langle W, D, V \rangle$, A is *valid in* $\mathfrak{M}$ (notation: $\models_{\mathfrak{M}} A$) if for any pairwise $s \subseteq W$ and for any $g : \mathsf{VAR} \rightarrow D$, $s, g \models_{\mathfrak{M}} A$. Let M be a class of inquisitive models. $\Gamma \models_{\mathsf{M}} A$ means that, for any $\mathfrak{M} \in \mathsf{M}$, any assignment g and any pairwise s, if $s, g \models_{\mathfrak{M}} B$ for all

[1] For a propositional variable $\mathbf{p}$ (i.e. 0-ary predicate symbol), we define $\mathbf{p}_w^V \in \{ \mathbf{true}, \mathbf{false} \}$.

$B \in \Gamma$ then $s, g \models_{\mathfrak{M}} A$. We say that A is *valid in* M (notation: $\mathsf{M} \Vdash A$) if $\emptyset \models_{\mathsf{M}} A$. Define $\mathsf{M}_{\mathrm{all}}$ as the class of *all* inquisitive models.

In [6] and [7, Ch.6], the following special class of inquisitive models are considered: Let us fix $D \neq \emptyset$ and fix a mapping $\mathscr{I}\colon \{\, c_i \mid i \in \omega \,\} \to D$, i.e., an *interpretation* on D of all the constant symbols. Let $W_{(D,\mathscr{I})}$ be the collection of all first-order classical structures for $\mathcal{L}$ such that the universe of $\mathfrak{A}$ is D and, $c^{\mathfrak{A}} = \mathscr{I}(c)$ for any $\mathfrak{A} \in W_{(D,\mathscr{I})}$. Define the valuation V of inquisitive model by: $c^V := c^{\mathfrak{A}}$ for some fixed $\mathfrak{A}$ and $P^V_{\mathfrak{A}} = P^{\mathfrak{A}}$. Then, $\langle\, W_{(D,\mathscr{I})}, D, V \,\rangle$ is an inquisitive model. Let us define that an *intended inquisitive model* is such a tuple $\langle\, W_{(D,\mathscr{I})}, D, V \,\rangle$ for some D and $\mathscr{I}$. Fix an assignment g. Remark that we can rewrite the satisfaction clause for atoms as follows: $s, g \models P(t_1, \ldots, t_n)$ iff $\mathfrak{A} \models P(t_1, \ldots, t_n)[g]$ for any $\mathfrak{A} \in s$, where $\mathfrak{A} \models A[g]$ means the ordinary *classical* satisfaction relation.

Definition 1. $\mathsf{M}_{\mathrm{int}} = \{\, \langle\, W_{(D,\mathscr{I})}, D, V \,\rangle \mid D \neq \emptyset$ *and* $\mathscr{I} : \{\, c_i \mid i \in \omega \,\} \to D \,\}$.

So, $\mathsf{M}_{\mathrm{int}}$ is the class of all intended inquisitive models. We will show that there is no difference between $\mathsf{M}_{\mathrm{all}}$ and $\mathsf{M}_{\mathrm{int}}$ with respect to the logical consequence (Theorem 1).

Let us explain why this paper studies first-order inquisitive pair semantics: While inquisitive pair semantics shows a peculiar logical-phenomena in calculating the inquisitive meaning of $\mathbf{p} \vee \mathbf{q} \vee \mathbf{r}$ (i.e. all the *possibilities* (defined below) for $\mathbf{p} \vee \mathbf{q} \vee \mathbf{r}$) in the propositional level, it still forms a good starting point to investigate *first-order inquisitive logic*, i.e, all valid formulas on $\mathsf{M}_{\mathrm{int}}$ in first-order inquisitive semantics [7, Ch.6] by Ciardelli. In what follows in this subsection, let us pay attention only to $\mathsf{M}_{\mathrm{int}}$. Before explaining the detail above, we would like to introduce some terminology. Define that $s \subseteq W_{(D,\mathscr{I})}$ is *n-tuplewise* if $1 \leq \#s \leq n$. '2-tuplewise' is the same notion as 'pairwise'. If we replace 'pairwise' with 'n-tuplewise' or 'non-empty' in the satisfaction clauses above, then we obtain *first-order inquisitive n-tuple semantics* or *first-order inquisitive semantics* [7, Ch.6] by Ciardelli[2], respectively.

Consider the propositional counterpart of our inquisitive pair semantics and define that a *possibility* for a propositional formula A is a $\supseteq$-maximal element s such that $s \models A$ (cf. [1]). Denote all the possibilities for A by $[A]$. Then, $[\mathbf{p} \vee \mathbf{q}] = \{\, |\mathbf{p}|, |\mathbf{q}| \,\}$ holds, where $|A|$ is all the truth functions making A true. Ciardelli, however, showed that $[\mathbf{p} \vee \mathbf{q} \vee \mathbf{r}] \neq \{\, |\mathbf{p}|, |\mathbf{q}|, |\mathbf{r}| \,\}$ in inquisitive pair semantics [7, Ch.5]). Inquisitive 3-tuplewise semantics can fix this defeat for $\mathbf{p} \vee \mathbf{q} \vee \mathbf{r}$. However, in order to avoid such peculiar phenomena for any formula containing $\vee$, we should drop the cardinality restriction of the upper bound of $\#s$ in the satisfaction clauses above. Such a consideration leads us to (propositional) inquisitive semantics by Ciardelli and Roelofsen [5].

Let InqQL_n (or, InqQL) be all the valid formulas on $\mathsf{M}_{\mathrm{int}}$ in first-order inquisitive n-tuplewise semantics (or, first-order inquisitive semantics, respectively). Let InqL_n and InqL be their propositional counterparts. Then, $\bigcap_{2 \leq n} \mathsf{InqL}_n = \mathsf{InqL}$ holds [7, Corollary 4.1.6.], and so, InqL_2 forms a starting point of approximating InqL. When we move to the first-order level, we do not know whether $\bigcap_{2 \leq n} \mathsf{InqQL}_n = \mathsf{InqQL}$ in this stage. However, it is obvious that $\bigcap_{2 \leq n} \mathsf{InqQL}_n \subseteq \mathsf{InqQL}$. Therefore, first-order inquisitive pair semantics still forms a good starting point to investigate InqQL.

[2] Ciardelli also observed that the restriction $\#s \leq 2$ gives us the equivalent semantics to the original inquisitive pair semantics by Groenendijk (see [7, Ch.5, pp.55-6]). In this sense, we still call our semantics '(first-order) inquisitive *pair* semantics'.

2.2 Constant-Domain Kripke Semantics

If we extend the first-order intuitionistic logic **IQL** with the axiom **CD** in Table 1 below, then we can obtain the following simpler Kripke semantics [8, Ch.3.4]. A *constant-domain Kripke model* (in short: *cd-model*) is a tuple $\langle W, \leq, D, V \rangle$, where $W \neq \emptyset$, $\leq$ on W is a pre-order, $D \neq \emptyset$, and V is a valuation satisfying $c^V \in D$, $P^V_w \subseteq D^n$, and $P^V_w \subseteq P^V_v$ if $w \leq v$ (the *hereditary condition*). Given any cd-model $\langle W, \leq, D, V \rangle$, any $g : \mathsf{VAR} \to D$, $w \in W$, and any A of $\mathcal{L}$, the satisfaction relation $\Vdash$ is defined by:

$$
\begin{array}{lll}
\mathfrak{M}, w, g \Vdash P(t_1, \ldots, t_n) & \text{iff} & \langle \overline{g}(t_1), \ldots, \overline{g}(t_n) \rangle \in P^V_w; \\
\mathfrak{M}, w, g \Vdash \bot & & \text{Never}; \\
\mathfrak{M}, w, g \Vdash \neg A & \text{iff} & \text{for any } w' \geq w\text{: } \mathfrak{M}, w', g \nVdash A; \\
\mathfrak{M}, w, g \Vdash A \wedge B & \text{iff} & \mathfrak{M}, w, g \Vdash A \text{ and } \mathfrak{M}, w, g \Vdash B; \\
\mathfrak{M}, w, g \Vdash A \vee B & \text{iff} & \mathfrak{M}, w, g \Vdash A \text{ or } \mathfrak{M}, w, g \Vdash B; \\
\mathfrak{M}, w, g \Vdash A \to B & \text{iff} & \text{for any } w' \geq w\text{: } w', g \Vdash A \text{ implies } w', g \Vdash B; \\
\mathfrak{M}, w, g \Vdash \forall x. A & \text{iff} & \text{for any } a \in D\text{: } \mathfrak{M}, w, g(x|a) \Vdash A; \\
\mathfrak{M}, w, g \Vdash \exists x. A & \text{iff} & \text{for some } a \in D\text{: } \mathfrak{M}, w, g(x|a) \Vdash A.
\end{array}
$$

Given any cd-model $\mathfrak{M} = \langle W, \leq, D, V \rangle$, A is *valid in* $\mathfrak{M}$ (notation: $\mathfrak{M} \Vdash A$) if for any $w \in W$ and for any $g : \mathsf{VAR} \to D$, $\mathfrak{M}, w, g \Vdash A$. By the following procedure, we can

Table 1. All Additional Axioms for First-Order Inquisitive Pair Logic

CD	$\forall x. (A \vee B(x)) \to (A \vee \forall x. B(x))$, where x is not free in A.
H2	$A \vee (A \to B \vee \neg B)$
W2	$(A \to B) \vee (B \to A) \vee ((A \to \neg B) \wedge (B \to \neg A))$
ADN	$\neg\neg P(t_1, \ldots, t_n) \to P(t_1, \ldots, t_n)$ for any atomic $P(t_1, \ldots, t_n)$

regard any inquisitive model $\mathfrak{M} = \langle W, D, V \rangle$ as a cd-model $\langle W', \leq, D', V' \rangle$ for first-order intuitionistic logic with the axiom **CD**. Put $W' := \{ s \subseteq W \mid s \text{ is pairwise} \}$. Define a pre-order $\leq$ on W' by $s \leq t$ iff $t \subseteq s$. Define $D' := D$. As for the valuation V', we define $c^{V'} = c^V$ and $\langle d_1, \ldots, d_n \rangle \in P^{V'}_s$ iff $\langle d_1, \ldots, d_n \rangle \in P^V_w$ for any $w \in s$ (s: pairwise). It is easy to see that V satisfies the hereditary condition. Then, we can show that $s, g \models_{\mathfrak{M}} A$ iff $\langle W', \leq, D', V' \rangle, s, g \Vdash A$, for any pairwise $s \subseteq W$ and any A. This observation allows us to say that all theorems of first-order intuitionistic logic as well as **CD** are valid in any inquisitive model.

Moreover, we can specify the class of cd-models corresponding to M_{all} as Mascarenhas [2] did for the propositional language. $\langle W', \leq, D' \rangle$ satisfies:

($h2$) the maximum length of $\leq$-chains is 2 (or, it is of depth ≤ 2, simply);
($w2$) each state can have no more than two distinct successors.

These observations tells us that both **H2** and **W2** in Table 1 are valid on any inquisitive model $\langle W, D, V \rangle$ by ($h2$) and ($w2$), respectively (see [2, Theorem 35]). There is one more feature of the above $\langle W', \leq, D', V' \rangle$:

Definition 2. $\mathfrak{M} = \langle W, \leq, D, V \rangle$ *has the* intersection property *if, for any* $w \in W$, $P^V_w = \bigcap \left\{ P^V_v \mid w \leq v \textit{ and } v \textit{ is an endpoint} \right\}$.

This feature validates the axiom **ADN** in Table 1 on any inquisitive model:

Proposition 1. *Let $\mathfrak{M} = \langle W, \leq, D, V \rangle$ be a Kripke model such that $\{ v \mid w \leq v \}$ is finite* ($w \in W$) *and $\mathfrak{M}$ satisfies the intersection property. Then,* **ADN** *is valid in $\mathfrak{M}$.*

Proof. Fix any $w \in W$ and any assignment g. Assume $\mathfrak{M}, w, g \Vdash \neg\neg P(t_1, \ldots, t_n)$. We show $\mathfrak{M}, w, g \Vdash P(t_1, \ldots, t_n)$. By assumption, for any $v \geq w$, we can find some $u \geq v$ such that $\mathfrak{M}, u, g \Vdash P(t_1, \ldots, t_n)$. Since $\{ w' \mid w \leq w' \}$ is finite, we can find $u^* \geq w$ such that u^* is an endpoint. Then, $\mathfrak{M}, u^*, g \Vdash P(t_1, \ldots, t_n)$. By the intersection property, we can conclude that $\mathfrak{M}, w, g \Vdash P(t_1, \ldots, t_n)$, as desired. □

Clearly, the above $\langle W', \leq, D', V' \rangle$ has the intersection property. Under ($h2$) and ($w2$), $\{ v \mid w \leq v \}$ is always finite ($w \in W$). Therefore, **ADN** is valid in M_{all}.

Definition 3. *Let VI be the class of all cd-models satisfying* ($w2$), ($h2$) *and the intersection property.*

$\Gamma \Vdash_{\mathsf{VI}} A$ means that for any $\mathfrak{M} \in \mathsf{VI}$, any assignment g and any state w in $\mathfrak{M}$, if $\mathfrak{M}, w, g \models B$ for all $B \in \Gamma$ then $\mathfrak{M}, w, g \models A$. We denote $\emptyset \Vdash_{\mathsf{VI}} A$ by $\mathsf{VI} \Vdash A$. The following is a generalization of [2, Theorem 36] to this setting.

Theorem 1. $\Gamma \models_{\mathsf{M}_{\text{all}}} A$ iff $\Gamma \models_{\mathsf{M}_{\text{int}}} A$ iff $\Gamma \Vdash_{\mathsf{VI}} A$.

Proof. $\Gamma \Vdash_{\mathsf{VI}} A \implies \Gamma \models_{\mathsf{M}_{\text{all}}} A$ is clear from the above argument. By definition, $\Gamma \models_{\mathsf{M}_{\text{all}}} A \implies \Gamma \models_{\mathsf{M}_{\text{int}}} A$. So, it suffices to show $\Gamma \models_{\mathsf{M}_{\text{int}}} A \implies \Gamma \Vdash_{\mathsf{VI}} A$. We establish the contrapositive implication. Assume $\Gamma \nVdash_{\mathsf{VI}} A$, i.e., there exists some cd-model $\mathfrak{M} \in \mathsf{VI}$, some w in $\mathfrak{M}$ and some g such that $\mathfrak{M}, w, g \Vdash B$ ($B \in \Gamma$) and $\mathfrak{M}, w, g \nVdash A$. Take the point-generated submodel $\mathfrak{M}_w$ by w of $\mathfrak{M}$. It is easy to see that $\mathfrak{M}, w, g \Vdash C$ iff $\mathfrak{M}_w, w, g \Vdash C$ for any formula C. Thus, $\mathfrak{M}_w, w, g \Vdash B$ ($B \in \Gamma$) and $\mathfrak{M}_w, w, g \nVdash A$. Since ($w2$), ($h2$) (and the intersection property) still hold in $\mathfrak{M}_w$, we can state that $\mathfrak{M}_w$ has one of the following shapes: (i) one point reflexive model; (ii) 'I'-shape; (iii) 'V'-shape. Write $\mathfrak{M}_w := \langle W, \leq, D, V \rangle$. First, consider the case (i). Define an interpretation $\mathscr{I}$ on D of constants by $\mathscr{I}(c) = c^V$. Define a first-order classical structure $\mathfrak{A}$ by: $|\mathfrak{A}| = D$, $c^{\mathfrak{A}} = \mathscr{I}(c)$, and $P^{\mathfrak{A}} = P^V_w$. Then, we can establish that $\mathfrak{M}_w, w, g \Vdash C$ iff $\{ \mathfrak{A} \}, g \models C$ for any formula C. Therefore, we have found $\mathfrak{A} \in W_{(D,\mathscr{I})}$ such that $\{ \mathfrak{A} \}, g \models B$ ($B \in \Gamma$) and $\{ \mathfrak{A} \}, g \not\models A$, i.e., $\Gamma \not\models_{\mathsf{M}_{\text{int}}} A$, as required. Second, consider the case (ii). We can put $W = \{ w, v \}$. By the intersection property, however, P^V_v are the same as P^V_w. So, we can reduce this case to the case (i). Third, let us consider (iii). Put $W = \{ w, v, u \}$. We regard v and u as the 'leaves' of the 'V'-shape tree with the root w. Similarly to (i), define an interpretation $\mathscr{I}$ on D of constants by $\mathscr{I}(c) = c^V$. In this case, however, we need to define two first-order classical structures $\mathfrak{A}$ and $\mathfrak{B}$ by: $|\mathfrak{A}| = |\mathfrak{B}| = D$, $c^{\mathfrak{A}} = c^{\mathfrak{B}} = \mathscr{I}(c)$, and $P^{\mathfrak{A}} = P^V_v$ and $P^{\mathfrak{B}} = P^V_u$. By induction, we can show that $\mathfrak{M}_w, w, g \Vdash C$ iff $\{ \mathfrak{A}, \mathfrak{B} \}, g \models C$ for any C. By the similar argument to (i), we can conclude that $\Gamma \not\models_{\mathsf{M}_{\text{int}}} A$. □

By this correspondence, we can easily show the following propositions (cf. [4]).

Proposition 2. *Let $s \subseteq W$ be pairwise and $w, v \in W$ distinct.* (i) *If $s, g \models A$ and $s' \subseteq s$ is pairwise, then $s', g \models A$;* (ii) $\{ w, v \}, g \models \neg A$ iff $\{ w \}, g \not\models A$ *and* $\{ v \}, g \not\models A$; (iii) $\{ w \}, g \models \neg A$ iff $\{ w \}, g \not\models A$; (iv) $\{ w \}, g \models A \to B$ iff $\{ w \}, g \models A$ *implies* $\{ w \}, g \models B$.

Let $\mathsf{M}_k := \{ \langle W, D, V \rangle \mid \#W = k \}$ ($k = 1$ or 2) and $\mathsf{M}_{\geq 2} := \{ \langle W, D, V \rangle \mid \#W \geq 2 \}$.

Proposition 3. (i) *Assume that* $\#W \geq 2$. *Then, A is valid in an inquisitive model* $\mathfrak{M}$ iff $s, g \models A$ *for any pairwise s with* $\#s = 2$ *and any g in* $\mathfrak{M}$. (ii) $\mathsf{M}_1 \models A$ iff *A is classically valid.* (iii) *If* $\mathsf{M}_{\geq 2} \models A$, *then A is classically valid.* (iv) $\mathsf{M}_{\text{all}} \models A$ iff $s, g \models_{\langle W,D,V \rangle} A$ *for any pairwise* $s \subseteq W$ *with* $\#s = 2$, *any g, and any* $\langle W, D, V \rangle \in \mathsf{M}_{\geq 2}$.

3 A Complete Hilbert-Style Calculus for Inquisitive Pair Logic

Definition 4. *Define* $\mathbf{QLV}^+$ *is* **IQL** *extended with all the axioms in Table 1.*

The reader can find the axiomatization of the first-order intuitionistic logic **IQL** in [10]. Define $\Gamma \vdash A$ if $\vdash \bigwedge \Gamma' \to A$ for some finite $\Gamma' \subseteq \Gamma$. If $\Gamma = \emptyset$, we write $\mathbf{QLV}^+ \vdash A$ but we usually drop '$\mathbf{QLV}^+$' from it and write $\vdash A$, when no confusion arises. In order to show the completeness of $\mathbf{QLV}^+$, we adopt the known canonical model method as in [8]. We, however, include the detailed outline to make this section self-contained.

Remark 1. We have two different axiomatizations of the set InqL_2 of all valid propositional formulas in inquisitive pair semantics. One proposed by Mascarenhas is the propositional intuitionistic logic IL extended with **W2**, **H2**, and atomic double negations ($\neg\neg\mathbf{p} \to \mathbf{p}$ for any atom $\mathbf{p}$). Another one proposed by Ciardelli and Roelofsen is IL extended with Kreisel-Putnam axiom **KP**: $(\neg A \to B \lor C) \to (\neg A \to B) \lor (\neg A \to C)$ and **H2**, and atomic double negations. And, if we drop **H2** from Ciardelli and Roelofsen's axiomatization, then we obtain the axiomatization of InqL, i.e., all valid propositional formulas in (generalized) inquisitive semantics. However, if we consider the first-order extension with **CD** of these logics, strong completeness of **IQL** extended with **CD** and **KP** for constant-domain Kripke semantics seems an open problem (p.c. by Valentin Shehtman and Silvio Ghilardi). Therefore, we choose Mascarenhas' axiomatization as a basis of our first-order inquisitive pair logic $\mathbf{QLV}^+$.

Let us expand our language $\mathcal{L}$ with a countable set $\{ \mathbf{c}_i \mid i \in \omega \}$ of *new* constant symbols. Let $\mathcal{L}^+$ be this expanded language of $\mathcal{L}$. We say that $\langle \Gamma; \Delta \rangle$ of $\mathcal{L}^+$ is *consistent* if $\nvdash \bigvee \Gamma_1 \to \bigwedge \Delta_1$ for any finite $\Gamma_1 \subseteq \Gamma$ and any finite $\Delta_1 \subseteq \Delta$. $\langle \Gamma; \Delta \rangle$ of $\mathcal{L}^+$ is *maximal* if $A \in \Gamma$ or $A \in \Delta$ for any formula A. $\langle \Gamma; \Delta \rangle$ of $\mathcal{L}^+$ is $\exists\forall$-*maximally consistent* if it is consistent and maximal and satisfies the following: ($L\exists$-**property**): For any formula of the form $\exists x. A$, if $\exists x. A \in \Gamma$, then $A[\mathbf{c}/x] \in \Gamma$ for some $\mathbf{c}$, and ($R\forall$-**property**): For any formula of the form $\forall x. A$, if $\forall x. A \in \Delta$, then $A[\mathbf{c}/x] \in \Delta$ for some $\mathbf{c}$. By consistency and maximality, it is obvious that $\Delta = \Gamma^c$, the complement of Γ[3]. So, if $\langle \Gamma; \Delta \rangle$ is $\exists\forall$-maximally consistent, then we usually say that Γ is an $\exists\forall$-*MCS*.

Lemma 1. (i) *If* $\langle \Gamma \cup \{ \exists x. A \}; \Delta \rangle$ *is consistent, then* $\langle \Gamma \cup \{ \exists x. A, A[\mathbf{c}/x] \}; \Delta \rangle$ *is consistent, where* $\mathbf{c}$ *is fresh in* $\langle \Gamma \cup \{ \exists x. A \}; \Delta \rangle$. (ii) *If* $\langle \Gamma; \Delta \cup \{ \forall x. A \} \rangle$ *is consistent, then* $\langle \Gamma; \Delta \cup \{ \forall x. A, A[\mathbf{c}/x] \} \rangle$ *is consistent, where* $\mathbf{c}$ *is fresh in* $\langle \Gamma; \Delta \cup \{ \forall x. A \} \rangle$. (iii) *If* $\langle \Gamma; \Delta \rangle$ *is consistent, then either* $\langle \Gamma \cup \{ A \}; \Delta \rangle$ *or* $\langle \Gamma; \Delta \cup \{ A \} \rangle$ *is consistent.*

Proof. We only establish (ii), since we need **CD** here. Suppose for contradiction that there exists some $\Gamma' \subseteq \Gamma$ and some $\Delta' \subseteq \Delta$ such that $\vdash \bigwedge \Gamma' \to \bigvee \Delta' \lor \forall x. A \lor A[\mathbf{c}/x]$. Fix some fresh y in $\langle \Gamma; \Delta \cup \{ \forall x. A \} \rangle$. It is clear that $(A[y/x])[\mathbf{c}/y] \equiv A[\mathbf{c}/x]$. Since y

[3] Remark that we can easily derive from the consistency of $\langle \Gamma; \Delta \rangle$ that $\Gamma \cap \Delta = \emptyset$.

and $\mathbf{c}$ are fresh, we obtain: $\vdash \bigwedge \Gamma' \to \forall y. (\bigvee \Delta' \vee \forall x. A \vee A[y/x])$. We deduce from **CD** that $\vdash \bigwedge \Gamma' \to (\bigvee \Delta' \vee \forall x. A)$ (remark that $\forall x. A$ and $\forall y. (A[y/x])$ are equivalent), which gives us the desired contradiction. □

Lemma 2. *If $\langle \Gamma; \Delta \rangle$ of $\mathcal{L}$ is consistent, then there exists $\langle \Gamma^+; \Delta^+ \rangle$ of $\mathcal{L}^+$ such that $\Gamma \subseteq \Gamma^+$, $\Delta \subseteq \Delta^+$, and Γ^+ is an $\exists\forall$-MCS.*

Proof. Let us enumerate all the formulas of $\mathcal{L}^+$ as $(F_n)_{n \in \omega}$. Recall that all the new constant symbols $\{ \mathbf{c}_i \mid i \in \omega \}$ are indexed by $i \in \omega$. In what follows, we define a sequence $(\langle \Gamma_n; \Delta_n \rangle)_{n \in \omega}$ such that each $\langle \Gamma_n; \Delta_n \rangle$ is consistent, and obtain $\langle \Gamma^+; \Delta^+ \rangle := \langle \bigcup_{n \in \omega} \Gamma_n; \bigcup_{n \in \omega} \Delta_n \rangle$ as its limit. (Basis) Put $\Gamma_0 := \Gamma$ and $\Delta_0 := \Delta$. (Inductive Step) Suppose that we have defined a consistent $\langle \Gamma_n; \Delta_n \rangle$. We subdivide our argument into the following three cases: (a) $F_n \equiv \exists x. A$ and $\langle \Gamma_n \cup \{ F_n \}; \Delta_n \rangle$ is consistent; (b) $F_n \equiv \forall x. A$ and $\langle \Gamma_n; \Delta_n \cup \{ F_n \} \rangle$ is consistent; (c) Otherwise. First, we show the case (c). Since either $\langle \Gamma_n \cup \{ F_n \}; \Delta_n \rangle$ or $\langle \Gamma_n; \Delta_n \cup \{ F_n \} \rangle$ is consistent by Lemma 1 (iii), choose a consistent pair and define it as $\langle \Gamma_{n+1}, \Delta_{n+1} \rangle$. As for the case (a), let us choose a fresh $\mathbf{c}$ in $\langle \Gamma_n \cup \{ F_n \}; \Delta_n \rangle$ by Lemma 1 (i) and define $\langle \Gamma_{n+1}, \Delta_{n+1} \rangle := \langle \Gamma_n \cup \{ \exists x. A, A[\mathbf{c}/x] \}; \Delta_n \rangle$. As for the case (b) (similarly to (a)), let us choose a fresh $\mathbf{c}$ in $\langle \Gamma_n; \Delta_n \cup \{ F_n \} \rangle$ by Lemma 1 (ii) and define $\langle \Gamma_{n+1}, \Delta_{n+1} \rangle := \langle \Gamma_n; \Delta_n \cup \{ \forall x. A, A[\mathbf{c}/x] \} \rangle$.

Finally, it is easy to see that $\langle \bigcup_{n \in \omega} \Gamma_n; \bigcup_{n \in \omega} \Delta_n \rangle$ is $\exists\forall$-maximally consistent. □

Γ is *ω-closed* if, for any formula of the form $\forall x. A$ in $\mathcal{L}^+$, if $\Gamma \vdash A[\mathbf{c}/x]$ for all constants $\mathbf{c}$ then $\Gamma \vdash \forall x. A$. $\langle \Gamma; \Delta \rangle$ is *ω-closed-finite-consistent* (in short, ωfc) if Γ is ω-closed and Δ is finite and $\langle \Gamma; \Delta \rangle$ is consistent. We can easily show the following:

Lemma 3. *If Γ is an $\exists\forall$-MCS, then Γ is ω-closed.*

Lemma 4. (i) *If Γ is ω-closed, then $\Gamma \cup \{ A \}$ is also ω-closed.* (ii) *If $\langle \Gamma \cup \{ \exists x. A \}; \Delta \rangle$ is ωfc, then there exists some $\mathbf{c}$ such that $\langle \Gamma \cup \{ \exists x. A, A[\mathbf{c}/x] \}; \Delta \rangle$ is consistent.* (iii) *If $\langle \Gamma; \Delta \cup \{ \forall x. A \} \rangle$ is ωfc there exists some $\mathbf{c}$ such that $\langle \Gamma; \Delta \cup \{ \forall x. A, A[\mathbf{c}/x] \} \rangle$ is consistent.*

Proof. We only establish (iii), since we need **CD** here. Suppose that $\langle \Gamma; \Delta \cup \{ \forall x. A \} \rangle$ is ωfc. Assume for contradiction that $\langle \Gamma; \Delta \cup \{ \forall x. A, A[\mathbf{c}/x] \} \rangle$ is inconsistent for all constant symbol $\mathbf{c}$. By finiteness of Δ, we can assume w.l.o.g. that x does not occur in Δ (otherwise, it suffices to rename the bounded variable). Then, for all constant $\mathbf{c}$, we have $\Gamma \vdash \bigvee \Delta \vee (\forall x. A) \vee A[\mathbf{c}/x]$, i.e., $\Gamma \vdash (\bigvee \Delta \vee (\forall x. A) \vee A)[\mathbf{c}/x]$. Since Γ is ω-closed, $\Gamma \vdash \forall x. (\bigvee \Delta \vee (\forall x. A) \vee A)$. By **CD**, $\vdash \forall x. (\bigvee \Delta \vee (\forall x. A) \vee A) \to \bigvee \Delta \vee (\forall x. A)$. Therefore, we get $\Gamma \vdash \bigvee \Delta \vee (\forall x. A)$, which contradicts the consistency of $\langle \Gamma; \Delta \cup \{ \forall x. A \} \rangle$. □

Lemma 5. *If $\langle \Gamma; \Delta \rangle$ of $\mathcal{L}^+$ is ωfc, then there exists $\langle \Gamma^+; \Delta^+ \rangle$ of $\mathcal{L}^+$ such that $\Gamma \subseteq \Gamma^+$, $\Delta \subseteq \Delta^+$, and Γ^+ is an $\exists\forall$-MCS.*

Proof. The proof is similar to the proof of Lemma 2. We, however, need to care about the fact that $\langle \Gamma; \Delta \rangle$ is ωfc. Fix any enumeration $(F_n)_{n \in \omega}$ of all the formulas of $\mathcal{L}^+$. In what follows, we only describe the difference from the proof of Lemma 2. Below, we define a sequence $(\langle \Gamma_n; \Delta_n \rangle)_{n \in \omega}$ such that *each $\langle \Gamma_n; \Delta_n \rangle$ is ωfc*, and obtain $\langle \Gamma^+; \Delta^+ \rangle := \langle \bigcup_{n \in \omega} \Gamma_n; \bigcup_{n \in \omega} \Delta_n \rangle$. The basis step is the same as before. As for the inductive step, suppose that we have defined an ωfc $\langle \Gamma_n; \Delta_n \rangle$. We subdivide our argument into the cases (a), (b), and (c) in the same way as in the proof of Lemma 2. The definition of $\langle \Gamma_{n+1}; \Delta_{n+1} \rangle$ for each case is exactly the same as before. However, we need to check that we can find some constant **c** in both the cases (a) and (b) (the most important point is: there is no need for **c** to be *fresh*) and that $\langle \Gamma_{n+1}; \Delta_{n+1} \rangle$ is also ωfc. We can ensure these points by Lemma 4. □

Lemma 6. *Let Γ be an $\exists\forall$-MCS. Then:* (i) $A \wedge B \in \Gamma$ iff ($A \in \Gamma$ *and* $B \in \Gamma$), (ii) $A \vee B \in \Gamma$ iff ($A \in \Gamma$ *or* $B \in \Gamma$), (iii) $\forall x. A \in \Gamma$ iff $A[t/x] \in \Gamma$ *for any term t,* (iv) $\exists x. A \in \Gamma$ iff $A[t/x] \in \Gamma$ *for some term t,* (v) *If* $A \to B \in \Gamma$ *and* $A \in \Gamma$, *then* $B \in \Gamma$, (vi) ($\neg A \in \Gamma$ *and* $A \in \Gamma$) *fails.*

Proof. Suppose $\langle \Gamma; \Delta \rangle$ is $\exists\forall$-maximally consistent. We only show (iii). By $\vdash \forall x. A \to A[t/x]$, we can establish the left-to-right direction. As for the right-to-left direction, assume $\forall x. A \notin \Gamma$. By maximality, $\forall x. A \in \Delta$. By $R\forall$-property, $A[\mathbf{c}/x] \in \Delta$ for some constant **c**. So, there exists a term t such that $A[t/x] \notin \Gamma$ by the consistency. □

Definition 5. *The* canonical model for $\mathbf{QLV}^+$ $\mathfrak{M} = \langle W, \leq, D, V \rangle$ *is defined by:* (i) $W = \{ \Gamma \mid \Gamma$ *is an* $\exists\forall$*-MCS* $\}$[4], (ii) $\Gamma \leq \Pi$ iff $\Gamma \subseteq \Pi$, (iii) $D = \{ t \mid t$ *is a term of* $\mathcal{L}^+ \}$, (vi) $c^V = c$ *for any constant symbol c in* $\mathcal{L}^+$, (v) $\langle t_1, \ldots, t_n \rangle \in P^V_\Gamma$ iff $P(t_1, \ldots, t_n) \in \Gamma$.

Lemma 7 (Truth Lemma). *Let* $\mathfrak{M} = \langle W, \leq, D, V \rangle$ *be the canonical model for* $\mathbf{QLV}^+$. *Define the canonical assignment g by* $g(x) = x$. *Then,* $\mathfrak{M}, \Gamma, g \Vdash A$ iff $A \in \Gamma$.

Proof. By induction on A. First, let us remark that $\overline{g}(t) = t$ for any term t of $\mathcal{L}^+$. By Lemma 6 and the definition of the canonical model, we can easily establish the cases where $A \equiv P(t_1, \cdots, t_n)$, $B \vee C$, $B \wedge C$, $\exists x. B$ or $\forall x. B$ (if $A \equiv \exists x. B$ or $\forall x. B$, we need to use: $\mathfrak{M}, \Gamma, g(x|t) \Vdash A$ iff $\mathfrak{M}, \Gamma, g \Vdash A[t/x]$). So, let us only show the case where $A \equiv B \to C$. In order to establish the left-to-right direction, assume $B \to C \notin \Gamma$. By maximality, $B \to C \in \Delta$, where $\Delta = \Gamma^c$. By consistency of $\langle \Gamma; \Delta \rangle$, $\langle \Gamma \cup \{B\}; \{C\} \rangle$ is consistent. By Lemma 3 and Lemma 4 (i), $\langle \Gamma \cup \{B\}; \{C\} \rangle$ is ωfc. It follows from Lemma 5 that there exists some $\langle \Gamma^+; \Delta^+ \rangle$ such that Γ^+ is an $\exists\forall$-MCS and $\Gamma \cup \{B\} \subseteq \Gamma^+$ and $C \in \Delta^+$ (i.e., $C \notin \Gamma^+$ by the consistency). By the induction hypothesis, we obtain: $\mathfrak{M}, \Gamma, g \Vdash B$ and $\mathfrak{M}, \Gamma, g \nVdash C$. Since $\Gamma \subseteq \Gamma^+$, we conclude that $\mathfrak{M}, \Gamma, g \nVdash B \to C$. Finally, let us show the right-to-left direction. Assume $\mathfrak{M}, \Gamma, g \nVdash B \to C$, i.e., there exists some $\exists\forall$-MCS Γ' such that $\mathfrak{M}, \Gamma', g \Vdash B$ and $\mathfrak{M}, \Gamma', g \nVdash C$. By the induction hypothesis, we obtain: $B \in \Gamma'$ and $C \notin \Gamma'$. It follows from Lemma 6 (v) that $B \to C \notin \Gamma'$. □

Lemma 8. *Let* $\mathfrak{M} = \langle W, \leq, D, V \rangle$ *be the canonical model for* $\mathbf{QLV}^+$. *Then,* (i) $\mathfrak{M}$ *satisfies* ($h2$), (ii) $\mathfrak{M}$ *satisfies* ($w2$), (iii) $\mathfrak{M}$ *has the intersection property.*

Proof. We can show (i) and (ii) in the same way as in the propositional case [2, Theorem 35] (for (i), the reader can also refer to [8, Lemma 7.3.3 (1)]). So, we only show (iii). Let Γ be an $\exists\forall$-MCS. It suffices to show that: $P(t_1, \ldots, t_n) \in \Gamma$ iff $P(t_1, \ldots, t_n) \in$

[4] Remark that any MCS Γ is a $\mathbf{QLV}^+$-theory. This is shown as follows: Given any MCS Γ, assume that $\varphi \in \Gamma$ and $\varphi \vdash \psi$. Suppose for contradiction that $\psi \notin \Gamma$. By maximality, $\psi \in \Delta$. By consistency, we get $\nvdash \varphi \to \psi$, which contradicts $\varphi \vdash \psi$.

$\bigcap\{\Gamma' \mid \Gamma \subseteq \Gamma' \text{ and } \Gamma' \text{ is an endpoint}\}$ (remark that ($w2$) and ($h2$) assure us that, for any Γ in $\mathfrak{M}$, there exists some endpoint $\Gamma' \supseteq \Gamma$). We can easily show the left-to-right direction. So, let us establish the right-to-left direction. Assume that $P(t_1, \ldots, t_n) \in \Gamma'$ for any $\Gamma' \supseteq \Gamma$ such that Γ' is an endpoint. By ($w2$) and ($h2$), we can state that, for any state $\Pi \supseteq \Gamma$, there exists an endpoint $\Theta \supseteq \Pi$. Thus, we deduce from Truth Lemma that $\mathfrak{M}, \Gamma, g \Vdash \neg\neg P(t_1, \ldots, t_n)$, i.e., $\neg\neg P(t_1, \ldots, t_n) \in \Gamma$. Since $\vdash \neg\neg P(t_1, \cdots, t_n) \to P(t_1, \ldots, t_n)$, we can conclude that $P(t_1, \ldots, t_n) \in \Gamma$. □

Theorem 2. $\Gamma \Vdash_{\mathsf{VI}} A$ iff $\Gamma \vdash A$.

Proof. We can easily show that $\Gamma \vdash A$ implies $\Gamma \Vdash_{\mathsf{VI}} A$. So, let us establish the left-to-right direction. We show the contrapositive implication. Assume $\Gamma \nvdash A$ (remark that Γ might be infinite). Then, $\langle \Gamma, A \rangle$ is consistent. By Lemma 2, there exists some $\langle \Gamma^+; \Delta^+ \rangle$ such that $\Gamma \subseteq \Gamma^+$, $A \in \Delta^+$, and Γ^+ is an $\exists\forall$-MCS. By consistency of $\langle \Gamma^+; \Delta^+ \rangle$, $A \notin \Gamma^+$. It follows from Truth Lemma that $\mathfrak{M}, \Gamma^+, g \Vdash B$ ($B \in \Gamma$) and $\mathfrak{M}, \Gamma^+, g \nVdash A$. By Lemma 8, $\Gamma \nVdash_{\mathsf{VI}} A$, as desired. □

Corollary 1. *The following are all equivalent:* (i) $\Gamma \Vdash_{\mathsf{VI}} A$; (ii) $\Gamma \models_{\mathsf{M}_{\text{all}}} A$; (iii) $\Gamma \models_{\mathsf{M}_{\text{int}}} A$; (iv) $\Gamma \vdash A$.

Proof. Theorem 1 gives us the equivalence among (i), (ii), and (iii). Theorem 2 ensures the equivalence between (i) and (iv). □

4 Tree-Sequent Calculus for First-Order Inquisitive Pair Logic

In this section, we first introduce a tree-sequent calculus for $\mathsf{InqQL}_2 = \{A \mid \mathsf{M}_{\text{all}} \models A\}$, as a special form of Labelled Deductive Systems [11].

Let $\mathcal{T} = \langle \{0, 1, 2\}, \leq \rangle$ be the tree equipped with the order $\leq := \{\langle 0, 1 \rangle, \langle 0, 2 \rangle\} \cup \{\langle x, x \rangle \mid x \in \{0, 1, 2\}\}$. A *label* is an element of $\{0, 1, 2\}$. We use letters α, β, etc. for labels. A *labelled formula* is a pair $\alpha : A$, where α is a label and A is a formula of the language $\mathcal{L}$. In what follows in this paper, we use Γ, Δ, etc. to denote a set of *labelled formulas*. A *tree-sequent* is an expression $\Gamma \Rightarrow \Delta$ where Γ and Δ are finite sets of labelled formulas.

Now, let us introduce the tree-sequent calculus TInqQL_2 for first-order inquisitive pair logic InqQL_2. This system defines inference schemes which allow us to manipulate tree-sequents. The axioms of TInqQL_2 are of the following forms:

$$\alpha : A, \Gamma \Rightarrow \Delta, \alpha : A \ \text{(Ax)} \qquad \alpha : \bot, \Gamma \Rightarrow \Delta \ (\bot\text{L}).$$

The inference rules of TInqQL_2 are the following:

$$\frac{0 : P(t_1, \ldots, t_n), \Gamma \Rightarrow \Delta}{1 : P(t_1, \ldots, t_n), 2 : P(t_1, \ldots, t_n), \Gamma \Rightarrow \Delta} \ \text{(Atom L)} \qquad \frac{1 : A, 2 : A, \Gamma \Rightarrow \Delta}{0 : A, \Gamma \Rightarrow \Delta} \ \text{(Move)}$$

$$\frac{\alpha : A, \alpha : B, \Gamma \Rightarrow \Delta}{\alpha : A \wedge B, \Gamma \Rightarrow \Delta} \ (\wedge\text{L}) \qquad \frac{\Gamma \Rightarrow \Delta, \alpha : A \quad \Gamma \Rightarrow \Delta, \alpha : B}{\Gamma \Rightarrow \Delta, \alpha : A \wedge B} \ (\wedge\text{R})$$

$$\frac{\alpha : A, \Gamma \Rightarrow \Delta \quad \alpha : B, \Gamma \Rightarrow \Delta}{\alpha : A \vee B, \Gamma \Rightarrow \Delta} \ (\vee\text{L}) \qquad \frac{\Gamma \Rightarrow \Delta, \alpha : A, \alpha : B}{\Gamma \Rightarrow \Delta, \alpha : A \vee B} \ (\vee\text{R})$$

$$\frac{\Gamma \Rightarrow \Delta, \alpha : A}{\alpha : \neg A, \Gamma \Rightarrow \Delta}\ (\neg\mathrm{L}) \qquad \frac{\alpha : A, \Gamma \Rightarrow \Delta}{\Gamma \Rightarrow \Delta, \alpha : \neg A}\ (\neg\mathrm{R}_{1,2}) \text{ where } \alpha \neq 0$$

$$\frac{1 : A, \Gamma \Rightarrow \Delta \quad 2 : A, \Gamma \Rightarrow \Delta}{\Gamma \Rightarrow \Delta, 0 : \neg A}\ (\neg\mathrm{R}_0)$$

$$\frac{\Gamma \Rightarrow \Delta, \alpha : A \quad \alpha : B, \Gamma \Rightarrow \Delta}{\alpha : A \to B, \Gamma \Rightarrow \Delta}\ (\to \mathrm{L}) \qquad \frac{\alpha : A, \Gamma \Rightarrow \Delta, \alpha : B}{\Gamma \Rightarrow \Delta, \alpha : A \to B}\ (\to \mathrm{R}_{1,2}) \text{ where } \alpha \neq 0$$

$$\frac{0 : A, \Gamma \Rightarrow \Delta, 0 : B \quad 1 : A, \Gamma \Rightarrow \Delta, 1 : B \quad 2 : A, \Gamma \Rightarrow \Delta, 2 : B}{\Gamma \Rightarrow \Delta, 0 : A \to B}\ (\to \mathrm{R}_0)$$

$$\frac{\alpha : A[t/x], \Gamma \Rightarrow \Delta}{\alpha : \forall x. A, \Gamma \Rightarrow \Delta}\ (\forall L) \qquad \frac{\Gamma \Rightarrow \Delta, \alpha : A[z/x]}{\Gamma \Rightarrow \Delta, \alpha : \forall x. A}\ (\forall R)\dagger$$

$$\frac{\alpha : A[z/x], \Gamma \Rightarrow \Delta}{\alpha : \exists x. A, \Gamma \Rightarrow \Delta}\ (\exists L)\dagger \qquad \frac{\Gamma \Rightarrow \Delta, \alpha : A[t/x]}{\alpha : \Gamma \Rightarrow \Delta, \exists x. A}\ (\exists R)$$

$$\frac{\Gamma \Rightarrow \Delta, \alpha : A \quad \alpha : A, \Gamma \Rightarrow \Delta}{\Gamma \Rightarrow \Delta}\ (\mathrm{Cut})$$

where † means the *eigenvariable condition*: z does not occur in the conclusion. The tree-sequent calculus **cutfreeTInqQL**$_2$ is obtained by dropping (Cut) from **TInqQL**$_2$. Whenever a tree-sequent $\Gamma \Rightarrow \Delta$ is provable in **TInqQL**$_2$ (or, in **cutfreeTInqQL**$_2$), we write **TInqQL**$_2 \vdash \Gamma \Rightarrow \Delta$ (or, **cutfreeTInqQL**$_2 \vdash \Gamma \Rightarrow \Delta$, respectively).

4.1 Completeness of Tree-Sequent Calculus

In this subsection, we show that the tree-sequent calculus **cutfreeTInqQL**$_2$ is sufficient to prove all formulas that are valid in $\mathsf{M}_{\mathrm{all}}$.

In the following, Γ, Δ are possibly infinite in the expression $\Gamma \Rightarrow \Delta$ of a tree-sequent. In the case where Γ, Δ are all finite, the tree-sequent $\Gamma \Rightarrow \Delta$ said to be *finite*. A (possibly infinite) tree-sequent $\Gamma \Rightarrow \Delta$ is *provable* in **cutfreeTInqQL**$_2$, if **cutfreeTInqQL**$_2 \vdash \Gamma' \Rightarrow \Delta'$ for some finite tree-sequent $\Gamma' \Rightarrow \Delta'$ such that $\Gamma' \subseteq \Gamma$ and $\Delta' \subseteq \Delta$. In what follows, we extend our notation **cutfreeTInqQL**$_2 \vdash \Gamma \Rightarrow \Delta$ to cover any possibly infinite tree-sequent in the sense explained above.

Definition 6. *A tree-sequent $\Gamma \Rightarrow \Delta$ is* saturated *if it satisfies the following:*

(consistency) (i) *If $\alpha : A \in \Gamma$, then $\alpha : A \notin \Delta$,* (ii) $\alpha : \bot \notin \Gamma$.
(persistence condition) *If $0 : A \in \Gamma$, then $1 : A \in \Gamma$ and $2 : A \in \Gamma$.*
(atom l) *If $1 : P(t_1, \ldots, t_n) \in \Gamma$ and $2 : P(t_1, \ldots, t_n) \in \Gamma$, then $0 : P(t_1, \ldots, t_n) \in \Gamma$.*
(∧l) *If $\alpha : A \wedge B \in \Gamma$, then $\alpha : A \in \Gamma$ and $\alpha : B \in \Gamma$.*
(∧r) *If $\alpha : A \wedge B \in \Delta$, then $\alpha : A \in \Delta$ or $\alpha : B \in \Delta$.*
(∨l) *If $\alpha : A \vee B \in \Gamma$, then $\alpha : A \in \Gamma$ or $\alpha : B \in \Gamma$.*
(∨r) *If $\alpha : A \vee B \in \Delta$, then $\alpha : A \in \Delta$ and $\alpha : B \in \Delta$.*
(¬l) *If $\alpha : \neg A \in \Gamma$, then $\alpha : A \in \Delta$.*
(¬r$_{1,2}$) *If $\alpha : \neg A \in \Delta$ and $\alpha \neq 0$, then $\alpha : A \in \Gamma$.*
(¬r$_0$) *If $0 : \neg A \in \Delta$, then $1 : A \in \Gamma$ or $2 : A \in \Gamma$.*
(→l) *If $\alpha : A \to B \in \Gamma$, then $\alpha : A \in \Delta$ or $\alpha : B \in \Gamma$.*
(→r$_{1,2}$) *If $\alpha : A \to B \in \Delta$ and $\alpha \neq 0$, then $\alpha : A \in \Gamma$ and $\alpha : B \in \Delta$.*
(→r$_0$) *If $0 : A \to B \in \Delta$, then ($\alpha : A \in \Gamma$ and $\alpha : B \in \Delta$) for some $\alpha \in \{0, 1, 2\}$.*
(∀l) *If $\alpha : \forall x. A \in \Gamma$, then $\alpha : A[t/x] \in \Gamma$ for any term t.*
(∀r) *If $\alpha : \forall x. A \in \Delta$, then $\alpha : A[z/x] \in \Delta$ for some variable z.*
(∃l) *If $\alpha : \exists x. A \in \Gamma$, then $\alpha : A[z/x] \in \Gamma$ for some variable z.*
(∃r) *If $\alpha : \exists x. A \in \Delta$, then $\alpha : A[t/x] \in \Delta$ for any term t.*

Lemma 9. *If a finite tree-sequent $\Gamma \Rightarrow \Delta$ is not provable in* $\mathbf{cutfreeT}\mathsf{InqQL}_2$*, then there exists a saturated tree-sequent $\Gamma^+ \Rightarrow \Delta^+$ such that $\Gamma \subseteq \Gamma^+$ and $\Delta \subseteq \Delta^+$ and $\Gamma^+ \Rightarrow \Delta^+$ is not provable in* $\mathbf{cutfreeT}\mathsf{InqQL}_2$.

The proof of this lemma can be found in Appendix A. Each node α of a tree-sequent $\Gamma \Rightarrow \Delta$ is associated with a sequent $\Gamma_\alpha \Rightarrow \Delta_\alpha$ where Γ_α (or, Δ_α) is the set of formulas such that $\alpha : A \in \Gamma$ (or, $\alpha : A \in \Delta$, respectively). We define a translation of tree-sequents into formulas of $\mathcal{L}$. In the following definition, tree-sequents are all finite. Let $\Gamma \Rightarrow \Delta$ be a tree-sequent and $\mathbf{s}$, $\mathbf{t}$ be fresh *propositional variables* in $\Gamma \Rightarrow \Delta$. The *formulaic translation* $[\![\Gamma \Rightarrow \Delta]\!]$ is defined as (note that the following formulaic translation depends on the choice of $\mathbf{s}$ and $\mathbf{t}$):

$$[\![\Gamma \Rightarrow \Delta]\!] \equiv \bigwedge \Gamma_0 \to ((\mathbf{s} \vee \mathbf{t}) \vee \bigvee \Delta_0 \vee [\![\Gamma \Rightarrow \Delta]\!]_1 \vee [\![\Gamma \Rightarrow \Delta]\!]_2) \text{ where:}$$
$$[\![\Gamma \Rightarrow \Delta]\!]_1 \equiv \mathbf{s} \wedge \bigwedge \Gamma_1 \to \mathbf{t} \vee \bigvee \Delta_1; \qquad [\![\Gamma \Rightarrow \Delta]\!]_2 \equiv \mathbf{t} \wedge \bigwedge \Gamma_2 \to \mathbf{s} \vee \bigvee \Delta_2.$$

An idea behind fresh $\mathbf{s}$ and $\mathbf{t}$ is to *name* three pairwise subsets (corresponding to 0, 1, 2 in our fixed tree) in an inquisitive model. Recall that $\mathsf{M}_{\mathrm{int}}$ is the class of all *intended* inquisitive models.

Theorem 3. *If $\mathsf{M}_{\mathrm{int}} \models [\![\Gamma \Rightarrow \Delta]\!]$, then $\mathbf{cutfreeT}\mathsf{InqQL}_2 \vdash \Gamma \Rightarrow \Delta$. Therefore, if $\mathsf{M}_{\mathrm{all}} \models [\![\Gamma \Rightarrow \Delta]\!]$, then $\mathbf{cutfreeT}\mathsf{InqQL}_2 \vdash \Gamma \Rightarrow \Delta$.*

Proof. It suffices to establish the first part. We show the contrapositive implication of it. Assume that $\Gamma \Rightarrow \Delta$ is unprovable in $\mathbf{cutfreeT}\mathsf{InqQL}_2$. Then, by Lemma 9, there exists some saturated tree-sequent $\Gamma^+ \Rightarrow \Delta^+$ such that $0 : A \in \Delta^+$ and $\mathbf{cutfreeT}\mathsf{InqQL}_2 \nvdash \Gamma^+ \Rightarrow \Delta^+$. Define $D = \{\, t \,|\, t \text{ is a term of } \mathcal{L} \,\}$. We define an interpretation $\mathscr{I}$ of constant symbols on D by $\mathscr{I}(c) := c$ and an assignment g by $g(x) = x$. Let us define the following two first-order classical structure $\mathfrak{A}_1$ and $\mathfrak{A}_2$: $|\mathfrak{A}_1| = |\mathfrak{A}_2| = D$, $c^{\mathfrak{A}_1} = c^{\mathfrak{A}_2} = \mathscr{I}(c)$, $P^{\mathfrak{A}_1} = \{\, \langle t_1, \ldots, t_n \rangle \,|\, 1 : P(t_1, \ldots, t_n) \in \Gamma^+ \,\}$, $P^{\mathfrak{A}_2} = \{\, \langle t_1, \ldots, t_n \rangle \,|\, 2 : P(t_1, \ldots, t_n) \in \Gamma^+ \,\}$. Now we show by induction on X of $\mathcal{L}$ that:

- (i) If $0 : X \in \Gamma^+$ then $\{\, \mathfrak{A}_1, \mathfrak{A}_2 \,\}, g \models X$; (ii) If $0 : X \in \Delta^+$ then $\{\, \mathfrak{A}_1, \mathfrak{A}_2 \,\}, g \not\models X$.
- (iii) If $\alpha : X \in \Gamma^+$ and $\alpha \neq 0$, then $\{\, \mathfrak{A}_\alpha \,\}, g \models X$; (iv) If $\alpha : X \in \Delta^+$ and $\alpha \neq 0$, then $\{\, \mathfrak{A}_\alpha \,\}, g \not\models X$.

Here we consider only the cases where X is of the form $P(t_1, \ldots, t_n)$ and of the form $\forall\, x.\, B$ (for the cases X is of the form $\neg B$ and of the form $B \to C$, the reader can find an essential argument in the proof of [4, Theorem 1]).

(The case where X is of the form $P(t_1, \ldots, t_n)$) We only show the cases (i) and (ii). (i) Suppose that $0 : P(t_1, \ldots, t_n) \in \Gamma^+$. Since $\Gamma^+ \Rightarrow \Delta^+$ is saturated, $1 : P(t_1, \ldots, t_n), 2 : P(t_1, \ldots, t_n) \in \Gamma^+$ by (**persistence condition**). So, $\langle t_1, \ldots, t_n \rangle \in P^{\mathfrak{A}_1}$ and $\langle t_1, \ldots, t_n \rangle \in P^{\mathfrak{A}_2}$. Since $\overline{g}(t) = t$, we can deduce that $\{\, \mathfrak{A}_1, \mathfrak{A}_2 \,\}, g \models P(t_1, \ldots, t_n)$. (ii) Suppose that $0 : P(t_1, \ldots, t_n) \in \Delta^+$. Since $\mathbf{cutfreeT}\mathsf{InqQL}_2 \nvdash \Gamma^+ \Rightarrow \Delta^+$ and $\Gamma^+ \Rightarrow \Delta^+$ is saturated, $0 : P(t_1, \ldots, t_n) \notin \Gamma^+$ by (**consistency**). $0 : P(t_1, \ldots, t_n) \notin \Gamma^+$ means that $1 : P(t_1, \ldots, t_n) \notin \Gamma^+$ or $2 : P(t_1, \ldots, t_n) \notin \Gamma^+$ by (**atomI**). So, $\langle t_1, \ldots, t_n \rangle \notin P^{\mathfrak{A}_1}$ or $\langle t_1, \ldots, t_n \rangle \notin P^{\mathfrak{A}_2}$. Therefore, by $\overline{g}(t) = t$, $\{\, \mathfrak{A}_1, \mathfrak{A}_2 \,\}, g \models P(t_1, \ldots, t_n)$, as desired.

(The case where X is of the form $\forall\, x.\, B$) We only show the cases (i) and (ii). (i) Suppose that $0 : \forall\, x.\, B \in \Gamma^+$. Since $\Gamma^+ \Rightarrow \Delta^+$ is saturated, $0 : B[t/x] \in \Gamma^+$ for any term t by ($\forall$**l**). By the induction hypothesis, we have: for any term t, $\{\mathfrak{A}_1, \mathfrak{A}_2\}, g \models B[t/x]$, i.e., $\{\mathfrak{A}_1, \mathfrak{A}_2\}, g(x|t) \models B$. Therefore, $\{\mathfrak{A}_1, \mathfrak{A}_2\}, g \models \forall x. B$. (ii) Suppose that $0 : \forall x. B \in \Delta^+$. Since $\Gamma^+ \Rightarrow \Delta^+$ is saturated, $0 : B[z/x] \in \Delta^+$ for any some variable z by ($\forall$**r**). By the induction hypothesis, we have: for some variable z, $\{\mathfrak{A}_1, \mathfrak{A}_2\}, g \models B[z/x]$, i.e., $\{\mathfrak{A}_1, \mathfrak{A}_2\}, g(x|z) \models B$. Therefore, $\{\mathfrak{A}_1, \mathfrak{A}_2\}, g \not\models \forall x. B$.

Let us choose fresh **s** and **t** in $\Gamma^+ \Rightarrow \Delta^+$ for $[\![\Gamma \Rightarrow \Delta]\!]$ and expand our model above so that **s** is true only under $\mathfrak{A}_1$ and **t** is true only under $\mathfrak{A}_2$. Then, we can conclude that $[\![\Gamma \Rightarrow \Delta]\!]$ is not valid in $\mathsf{M}_{\mathrm{int}}$ by construction of our model and (i) - (iv) above. □

4.2 Cut-Elimination Theorem and Soundness of Tree-Sequent Calculus

In this subsection, we establish that TInqQL_2 (i.e., **cutfree**TInqQL_2 with (Cut)) enjoys a cut-elimination theorem and that it is sound with respect to the class $\mathsf{M}_{\mathrm{all}}$ of all inquisitive models.

Lemma 10. *If* $\mathsf{TInqQL}_2 \vdash \Gamma \Rightarrow \Delta$*, then* $\mathsf{M}_{\mathrm{all}} \models [\![\Gamma \Rightarrow \Delta]\!]$.

The proof of this lemma can be found in Appendix B.

Theorem 4. *If* $\mathsf{TInqQL}_2 \vdash \Gamma \Rightarrow \Delta$*, then* **cutfree**$\mathsf{TInqQL}_2 \vdash \Gamma \Rightarrow \Delta$.

Proof. It follows from Lemma 10 and Theorem 3. □

In order to establish the soundness through our formulaic translation with fresh variables, we need to show the following, which lets us use the fresh propositional variables **s** and **t** to *name* three pairwise subsets (corresponding to 0, 1, 2 in our fixed tree) in an inquisitive model.

Lemma 11. *If* $\mathsf{M}_{\mathrm{all}} \models (\mathbf{s} \vee \mathbf{t}) \vee A \vee (\mathbf{s} \to \mathbf{t}) \vee (\mathbf{t} \to \mathbf{s})$*, then* $\mathsf{M}_{\mathrm{all}} \models A$*, where* **s** *and* **t** *are fresh in* A.

Proof. Assume $\mathsf{M}_{\mathrm{all}} \not\models A$. By Proposition 3 (iv), there exists some inquisitive model $\mathfrak{M} = \langle W, D, V \rangle$, some $w, v \in W$ and some assignment g such that $w \neq v$ and $\#W \geq 2$ and $\{w, v\}, g \not\models_{\mathfrak{M}} A$. Let V' be the same valuation as V except that **s** is true only at w and **t** is true only at v under V'. Write $\mathfrak{M}' = \langle W, D, V' \rangle$. Then, $s, g \models_{\mathfrak{M}} B$ iff $s, g \models_{\mathfrak{M}'} B$, for any $s \subseteq \{w, v\}$ and any subformula B of A. Thus, $\{w, v\}, g \not\models_{\mathfrak{M}'} A$. By definition of V', $\{w, v\}, g \not\models_{\mathfrak{M}'} (\mathbf{s} \vee \mathbf{t}) \vee A \vee (\mathbf{s} \to \mathbf{t}) \vee (\mathbf{t} \to \mathbf{s})$, as required. □

Theorem 5. *If* $\mathsf{TInqQL}_2 \vdash\Rightarrow 0 : A$*, then* $\mathsf{M}_{\mathrm{all}} \models A$.

Proof. By Lemma 10, $[\![\Rightarrow 0 : A]\!]$ is valid in $\mathsf{M}_{\mathrm{all}}$, i.e., $(\mathbf{s} \vee \mathbf{t}) \vee A \vee (\mathbf{s} \to \mathbf{t}) \vee (\mathbf{t} \to \mathbf{s})$ is valid in $\mathsf{M}_{\mathrm{all}}$. It follows from Lemma 11 that A is valid in $\mathsf{M}_{\mathrm{all}}$. □

5 Conclusion

Corollary 2. *All of the following are equivalent:* (i) **cutfree**$\mathsf{TInqQL}_2 \vdash \Rightarrow 0 : A$; (ii) $\mathsf{TInqQL}_2 \vdash \Rightarrow 0 : A$; (iii) $\mathsf{M}_{\mathrm{all}} \models A$; (iv) $\mathsf{M}_{\mathrm{int}} \models A$; (v) $\mathsf{VI} \Vdash A$; (vi) $\mathbf{QLV}^+ \vdash A$.

Proof. By Corollary 1, we establish the equivalence among (iii), (iv), (v) and (vi) (put $\Gamma = \emptyset$). By Theorem 3, (iii) $\Rightarrow$ (i). Trivially, (i) $\Rightarrow$ (ii). By Theorem 5, (ii) $\Rightarrow$ (iii). □

Our proof process for Corollary 2 also reveals that TInqQL_2 corresponds to $\mathbf{QLV}^+$ extended with the following non-standard proof rule: From $(\mathbf{s} \vee \mathbf{t}) \vee A \vee (\mathbf{s} \to \mathbf{t}) \vee (\mathbf{t} \to \mathbf{s})$, we may infer A, where $\mathbf{s}$ and $\mathbf{t}$ are fresh propositional variables in A. One of the main causes of such logical phenomena consists in the fact that we use the fixed tree $\mathcal{T}$, unlike the previous studies [12,9] which employ 'growing' tree-sequents. Therefore, this study also contributes to witness a logical connection between labelled deductive systems with a *fixed set of labels* and Hilbert-style axiomatizations with *non-standard proof-rules*[5,6].

References

1. Groenendijk, J.: Inquisitive semantics: Two possibilities for disjunction. Prepublication Series PP-2008-26, ILLC (2008)
2. Mascarenhas, S.: Inquisitive semantics and logic. Master's thesis, Institute for Logic, Language and Computation, University of Amsterdam (2009)
3. Kashima, R.: Sequent calculi of non-classical logics - Proofs of completeness theorems by sequent calculi. In: Proceedings of Mathematical Society of Japan Annual Colloquium of Foundations of Mathematics, pp. 49–67 (1999) (in Japanese)
4. Sano, K.: Sound and complete tree-sequent calculus for inquisitive logic. In: Ono, H., Kanazawa, M., de Queiroz, R. (eds.) WoLLIC 2009. LNCS (LNAI), vol. 5514, pp. 365–378. Springer, Heidelberg (2009)
5. Ciardelli, I., Roelofsen, F.: Inquisitive logic. To appear in Journal of Philosophical Logic, http://sites.google.com/site/inquisitivesemantics/
6. 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)
7. Ciardelli, I.: Inquisitive semantics and intermediate logics. Master's thesis, Institute for Logic, Language and Computation, University of Amsterdam (2009)
8. Gabbay, D.M., Shehtman, V., Skvortsov, D.P.: Quantification in Nonclassical Logic. Studies in logic and the foundations of mathematics, vol. 153. Elsevier, Amsterdam (2009)
9. Ishigaki, R., Kikuchi, K.: Tree-sequent methods for subintuitionistic predicate logics. In: Olivetti, N. (ed.) TABLEAUX 2007. LNCS (LNAI), vol. 4548, pp. 149–164. Springer, Heidelberg (2007)
10. Gabbay, D.M.: Semantical investigations in Heyting's intuitionistic logic. Synthese Library, vol. 148. D. Reidel Pub. Co., Dordrecht (1981)

[5] We can reduce the completeness of TInqQL_2 for M_{all} to the completeness of $\mathbf{QLV}^+$ for M_{all} as follows: Suppose $\mathsf{M}_{\text{all}} \models A$. By the completeness of $\mathbf{QLV}^+$ for M_{all}, $\mathbf{QLV}^+ \vdash A$. Then, we can deduce by induction on the derivation for A that $\mathsf{TInqQL}_2 \vdash \Rightarrow 0 : A$. This argument, however, does not give us a cut-elimination theorem of TInqQL_2.

[6] I would like to thank Ryo Kashima for giving me a clear understanding of strong Kripke completeness proof of first-order intuitionistic logic with the axiom **CD**. I also would like to thank Dick de Jongh, Valentin Shehtman, Tadeusz Litak, Silvio Ghilardi and Floris Roelofsen for their discussion with me and/or comments to this study. Finally, I would like to thank the anonymous referees for their very helpful comments and suggestions. All errors, however, are mine.

11. Gabbay, D.M.: Labelled Deductive Systems. Oxford Logic Guides, vol. 33. Clarendon Press/Oxford Science Publications, Oxford (1996)
12. Hasuo, I., Kashima, R.: Kripke completeness of first-order constructive logics with strong negation. Logic Journal of the IGPL 11(6), 615–646 (2003)

A A Proof of Lemma 9

Proof. The idea of this proof is essentially the same as in the proof of [4, Lemma 1]. The difference is: we need to care about $\forall$ and $\exists$. So, we basically concentrate on stating the difference from the proof of [4, Lemma 1] below. Suppose that a finite tree-sequent $\Gamma \Rightarrow \Delta$ is not provable in **cutfreeT**InqQL_2. In the following, we construct a sequence $(\Gamma^i \Rightarrow \Delta^i)_{i\in\omega}$ of finite tree-sequents and obtain $\Gamma^+ \Rightarrow \Delta^+$ as the union of them.

Let $(\alpha_i : F_i)_{i>1}$ be an enumeration of all labelled formulas such that each formula of $\mathcal{L}$ appears infinitely many times. We also enumerate all variables as $(x_i)_{i\in\omega}$ and all terms as $(t_i)_{i\in\omega}$. From now on, we construct $(\Gamma^i \Rightarrow \Delta^i)_{i\in\omega}$ such that **cutfreeT**$\mathsf{InqQL}_2 \nvdash \Gamma^i \Rightarrow \Delta^i$. (Basis) Let $\Gamma^0 \Rightarrow \Delta^0 \equiv \Gamma \Rightarrow \Delta$. By assumption, **cutfreeT**$\mathsf{InqQL}_2 \nvdash \Gamma^0 \Rightarrow \Delta^0$. (Inductive step) Suppose that we have already defined $\Gamma^{k-1} \Rightarrow \Delta^{k-1}$ such that **cutfreeT**$\mathsf{InqQL}_2 \nvdash \Gamma^{k-1} \Rightarrow \Delta^{k-1}$. In this k-th step, we define $\Gamma^k \Rightarrow \Delta^k$ so that unprovability of the tree-sequent is preserved. The operations executed in the k-th step are as follows: First, for any $0 : A \in \Gamma^k$, we add $1 : A$ and $2 : A$ to Γ^{k-1}. Unprovability is preserved because of the rule (Move). We denote the result of this step by $(\Gamma^{k-1})' \Rightarrow \Delta^{k-1}$. Second, according to the form of $\alpha_k : F_k$, one of the following operation is executed:

(1) The case where $F_k \equiv P(t_1, \dots, t_n)$ and $\alpha_k \neq 0$ and $\alpha_k : F_k \in (\Gamma^{k-1})'$. Define:

$$\Gamma^k \Rightarrow \Delta^k \equiv \begin{cases} 0 : P(t_1, \dots, t_n), (\Gamma^{k-1})' \Rightarrow \Delta^{k-1} & \text{if } (3-\alpha_k) : P(t_1, \dots, t_n) \in (\Gamma^{k-1})'; \\ (\Gamma^{k-1})' \Rightarrow \Delta^{k-1} & \text{o.w.} \end{cases}$$

Unprovability is preserved because of (Atom L).
(2) The case where $F_k \equiv A \wedge B$ and $\alpha_k : F_k \in (\Gamma^{k-1})'$. See [4].
(3) The case where $F_k \equiv A \wedge B$ and $\alpha_k : F_k \in \Delta^{k-1}$. See [4].
(4) The case where $F_k \equiv A \vee B$ and $\alpha_k : F_k \in (\Gamma^{k-1})'$. Similar to (3).
(5) The case where $F_k \equiv A \vee B$ and $\alpha_k : F_k \in \Delta^{k-1}$. Similar to (2).
(6) The case where $F_k \equiv \neg A$ and $\alpha_k : F_k \in (\Gamma^{k-1})'$. See [4].
(7) The case where $F_k \equiv \neg A$ and $\alpha_k : F_k \in \Delta^{k-1}$. See [4].
(8) The case where $F_k \equiv A \to B$ and $\alpha_k : F_k \in (\Gamma^{k-1})'$. See [4].
(9) The case where $F_k \equiv A \to B$ and $\alpha_k : F_k \in \Delta^{k-1}$. See [4].
(10) The case where $F_k \equiv \forall x. A$ and $\alpha_k : F_k \in (\Gamma^{k-1})'$. Define $\Gamma^k \Rightarrow \Delta^k \equiv \alpha_k : A[t_0/x], \dots, \alpha_k : A[t_{k-1}/x], (\Gamma^{k-1})' \Rightarrow \Delta_k$. Unprovability is preserved because of ($\forall$L).
(11) The case where $F_k \equiv \forall x. A$ and $\alpha_k : F_k \in \Delta^{k-1}$. Take a fresh variable z, and define $\Gamma^k \Rightarrow \Delta^k \equiv (\Gamma^{k-1})' \Rightarrow \Delta_k, \alpha_k : A[z/x]$. Unprovability is preserved because of ($\forall$R).
(12) The case where $F_k \equiv \exists x. A$ and $\alpha_k : F_k \in (\Gamma^{k-1})'$. Similar to (11).
(13) The case where $F_k \equiv \exists x. A$ and $\alpha_k : F_k \in \Delta^{k-1}$. Similar to (10).
(14) Otherwise. It suffices to define $\Gamma^k \Rightarrow \Delta^k \equiv (\Gamma^{k-1})' \Rightarrow \Delta^{k-1}$.

Now let $\Gamma^+ \Rightarrow \Delta^+$ be $(\bigcup_{i\in\omega} \Gamma^i) \Rightarrow (\bigcup_{i\in\omega} \Delta^i)$. It is easy to verify that the tree-sequent $\Gamma^+ \Rightarrow \Delta^+$ is saturated. $\square$

B A Proof of Lemma 10

By induction on the derivation of $\Gamma \Rightarrow \Delta$ in TInqQL_2. First, let us choose some fresh propositional variables **s**, **t** not occurring in the derivation. We assume that all formulaic translations in this proof depend on **s** and **t**. All cases in our induction immediately follow from the following Lemmas 12 and 13. We can easily establish Lemma 12 by definition of $[\![\Gamma \Rightarrow \Delta]\!]$.

Lemma 12. *If* $\mathsf{M}_{\text{all}} \models [\![\Gamma \Rightarrow \Delta]\!]_\alpha$ *for some* $\alpha \in \{1, 2\}$, *then* $\mathsf{M}_{\text{all}} \models [\![\Gamma \Rightarrow \Delta]\!]$.

Lemma 13. *The following formulas are valid in* M_{all}.

(**ax**) $A \wedge C \to A \vee D$.

($\bot$**left**) $\bot \wedge C \to D$.

(**atom left**) $X_1 \to X_2$, *where:*
$X_1 \equiv P(\overrightarrow{t_i}) \to (S \vee T) \vee D \vee (S \wedge E \to T \vee F) \vee (T \wedge G \to S \vee H)$;
$X_2 \equiv (S \vee T) \vee D \vee (P(\overrightarrow{t_i}) \wedge S \wedge E \to T \vee F) \vee (P(\overrightarrow{t_i}) \wedge T \wedge G \to S \vee H)$.

(**move**) $((E \wedge A \to F) \vee (G \wedge A \to H)) \to (A \to (E \to F) \vee (G \to H))$.

($\wedge$**right**) $(C \to D \vee A) \wedge (C \to D \vee B) \to (C \to (D \vee (A \wedge B)))$.

($\vee$**left**) $(A \wedge C \to D) \wedge (B \wedge C \to D) \to (((A \vee B) \wedge C) \to D)$.

($\neg$**left**) $(C \to D \vee A) \to (\neg A \wedge C \to D)$.

($\neg$**right**$_{1,2}$) $(C \wedge A \to D) \to (C \to D \vee \neg A)$.

($\neg$**right**$_0$) $X_3 \wedge X_4 \to X_5$, *where:*
$X_3 \equiv (S \vee T) \vee D \vee (S \wedge E \wedge A \to F \vee T) \vee (T \wedge G \to S \vee H)$;
$X_4 \equiv (S \vee T) \vee D \vee (S \wedge E \to F \vee T) \vee (T \wedge G \wedge A \to S \vee H)$;
$X_5 \equiv (S \vee T) \vee \neg A \vee D \vee (S \wedge E \to F \vee T) \vee (T \wedge G \to S \vee H)$.

($\to$ **left**) $(C \to D \vee A) \wedge (C \wedge B \to D) \to (C \wedge (A \to B) \to D)$.

($\to$ **right**$_{1,2}$) $(C \wedge A \to D \vee B) \to (C \to (D \vee (A \to B)))$.

($\to$ **right**$_0$) $(X_6 \wedge X_7 \wedge X_8) \to X_9$, *where:*
$X_6 \equiv A \to ((S \vee T) \vee D \vee B \vee (S \wedge E \to T \vee F) \vee (T \wedge G \to S \vee H))$;
$X_7 \equiv (S \vee T) \vee D \vee (S \wedge E \wedge A \to T \vee F) \vee (T \wedge G \to S \vee H)$;
$X_8 \equiv (S \vee T) \vee D \vee (S \wedge E \to T \vee F) \vee (T \wedge G \wedge A \to S \vee H)$;
$X_9 \equiv (S \vee T) \vee (A \to B) \vee D \vee (S \wedge E \to T \vee F) \vee (T \wedge G \to S \vee H)$.

($\forall$**left**) $(C \wedge A[t/x] \to D) \to (C \wedge \forall x.A \to D)$.

($\forall$**right**) $(C \to D \vee A[z/x]) \to (C \to D \vee \forall x.A)$, *where z is fresh in C, D and $\forall x.A$.*

($\exists$**left**) $(C \wedge A[z/x] \to D) \to (C \wedge \exists x.A \to D)$, *where z is fresh in C, D and $\exists x.A$.*

($\exists$**right**) $(C \to D \vee A[t/x]) \to (C \to D \vee \exists x.A)$.

(**cut**) $(C \to D \vee A) \wedge (C \wedge A \to D) \to (C \to D)$.

Proof. Formulas except (**atom left**), ($\neg$ **right**$_0$) and ($\to$ **right**$_0$) are all theorems of first-order intuitionistic logic with **CD** (we need **CD** for ($\forall$ right)). Therefore, they are all valid in M_{all}. So, it suffices to check (**atom left**), ($\neg$ **right**$_0$) and ($\to$ **right**$_0$). The essential arguments for these are the same as in the propositional case [4, p.373, Lemma 3]. $\square$

Ultrafilter Extensions of Models*

Denis I. Saveliev

Moscow State University, Russia

Dedication
To V.A. Uspensky on the 80th anniversary of his birth

Abstract. We show that any model $\mathfrak{A}$ can be extended, in a canonical way, to a model $\beta\mathfrak{A}$ consisting of ultrafilters over it. The extension procedure preserves homomorphisms: any homomorphism of $\mathfrak{A}$ into $\mathfrak{B}$ extends to a continuous homomorphism of $\beta\mathfrak{A}$ into $\beta\mathfrak{B}$. Moreover, if a model $\mathfrak{B}$ carries a compact Hausdorff topology which is (in a certain sense) compatible, then any homomorphism of $\mathfrak{A}$ into $\mathfrak{B}$ extends to a continuous homomorphism of $\beta\mathfrak{A}$ into $\mathfrak{B}$. This is also true for embeddings instead of homomorphisms.

We present a result in general model theory. We show that any model can be extended, in a canonical way, to the model (of the same language) consisting of ultrafilters over it such that the extended model inherits the universality property of the largest compactification.

Recall standard facts concerning topology of ultrafilters. The set βX of ultrafilters over a set X carries a natural topology generated by elementary (cl)open sets of form

$$\tilde{S} = \{u \in \beta X : S \in u\}$$

for all $S \subseteq X$. The space βX is compact Hausdorff, extremally disconnected (the closure of any open set is open), and in fact, the Stone–Čech (and also Wallman) compactification of the discrete space X, i.e. its *largest* compactification. This means that X is dense in βX (one lets $X \subseteq \beta X$ by identifying each $x \in X$ with the principal ultrafilter $\hat{x}$), and any continuous mapping h of X into any compact space Y can be uniquely extended to a continuous mapping $\tilde{h}$ of βX into Y. There is a one-to-one correspondence between filters over X and closed subsets of βX (a filter D corresponds to $\{u \in \beta X : D \subseteq u\}$ while a closed $C \subseteq \beta X$ corresponds to $\bigcap C$); in fact, the compactness of βX is equivalent to the claim that $\{u \in \beta X : D \subseteq u\}$ is nonempty for each D and thus unprovable in ZF alone (see [5]).

We show that if $F, \ldots, P, \ldots$ are operations and relations on X, there is a canonical way to extend them to operations and relations $\tilde{F}, \ldots, \tilde{P}, \ldots$ on βX, thus extending the model $\mathfrak{A} = (X, F, \ldots, P, \ldots)$ to the model $\beta\mathfrak{A} = (\beta X, \tilde{F}, \ldots, \tilde{P}, \ldots)$. We show that the extension procedure preserves homomorphisms: if h is a homomorphism of $\mathfrak{A}$ into $\mathfrak{B}$, then $\tilde{h}$ is a homomorphism of $\beta\mathfrak{A}$ into $\beta\mathfrak{B}$. Moreover, if $\mathfrak{B}$ carries a compact Hausdorff topology which is compatible in a certain sense,

* Partially supported by an INFTY grant of ESF.

M. Banerjee and A. Seth (Eds.): ICLA 2011, LNAI 6521, pp. 162–177, 2011.

and h is a homomorphism of $\mathfrak{A}$ into $\mathfrak{B}$, then $\tilde{h}$ is a homomorphism of $\beta\mathfrak{A}$ into $\mathfrak{B}$; thus extended models inherit the universality property of the Stone–Čech (or Wallman) compactification. We note also that both facts remain true if one replaces homomorphisms by embeddings or some other relationships between models.

The construction, although it looks old and should be known, did not appear before, except one very particular case when models are semigroups [1].[1] The reader can also consult on semigroups of ultrafilters and their applications in various areas (number theory, algebra, dynamics, ergodic theory) in [2]; an analogous technique for non-associative groupoids and some infinitary generalizations are discussed in [3].

Definition of extensions. The first main theorem

Here we define the extensions of models by ultrafilters. Then we establish our first main result showing that the extension procedure preserves homomorphisms.

To extend a model $\mathfrak{A} = (X, F, \ldots, P, \ldots)$, we extend operations $F, \ldots$ on X, i.e. mappings of Cartesian products of X into X itself, and relations $P, \ldots$ on X, i.e. subsets of such products. Let us provide a slightly more general definition involving n-ary mappings of $X_1 \times \ldots \times X_n$ into Y, and n-ary relations that are subsets of $X_1 \times \ldots \times X_n$. We shall use it e.g. when we shall show that any mapping h of a certain type between models extends to $\tilde{h}$ of the same type.

Definition 1. Given an n-ary mapping $F : X_1 \times \ldots \times X_n \to Y$, let $\tilde{F} : \beta X_1 \times \ldots \times \beta X_n \to \beta Y$ be *defined* as follows:

$$\tilde{F}(u_1, \ldots, u_n) =$$
$$\big\{S \subseteq Y : \{x_1 \in X_1 : \ldots \{x_n \in X_n : F(x_1, \ldots, x_n) \in S\} \in u_n \ldots\} \in u_1\big\}$$

for every $u_1 \in \beta X_1, \ldots, u_n \in \beta X_n$.

Lemma 2. *For all* $z_1 \in X_1$ *and* $u_2 \in \beta X_2, \ldots, u_n \in \beta X_n$,

$$\begin{aligned} &\big\{x_1 : \big\{x_2 : \ldots \{x_n : F(x_1, x_2, \ldots, x_n) \in S\} \in u_n \ldots\big\} \in u_2\big\} \in \hat{z}_1 \\ &\textit{iff} \quad \big\{x_2 : \ldots \{x_n : F(z_1, x_2, \ldots, x_n) \in S\} \in u_n \ldots\big\} \in u_2 \,. \end{aligned}$$

Proof. Clear.

Proposition 3. *If* $F : X_1 \times \ldots \times X_n \to Y$*, then* $\tilde{F} : \beta X_1 \times \ldots \times \beta X_n \to \beta Y$*. Moreover, the restriction of* $\tilde{F}$ *on* $\mathrm{dom}\,(F)$ *is* F*.*

Proof. By definition, $\mathrm{dom}\,(\tilde{F}) = \beta X_1 \times \ldots \times \beta X_n$, and a standard argument shows that values of $\tilde{F}$ are ultrafilters. It follows from Lemma 2 that for all $z_1 \in X_1, \ldots, z_n \in X_n$,

$$\big\{x_1 : \ldots \{x_n : F(x_1, \ldots, x_n) \in S\} \in \hat{z}_n \ldots\big\} \in \hat{z}_1 \;\leftrightarrow\; F(z_1, \ldots, z_n) \in S,$$

and therefore,

$$\tilde{F}(\hat{z}_1, \ldots, \hat{z}_n) = \hat{y} \quad \text{whenever} \quad F(z_1, \ldots, z_n) = y,$$

and thus $\tilde{F}$ extends F up to identification of x and $\hat{x}$.

[1] See also Remark at the end of the paper.

Let us discuss the construction.

First, in the unary case, an $F : X \to Y$ extends to $\tilde{F} : \beta X \to \beta Y$ by

$$\tilde{F}(u) = \big\{S \subseteq Y : \{x \in X : F(x) \in S\} \in u\big\}.$$

This gives the standard unique continuous extension of F. Indeed, it is easy to see that $\tilde{F}$ is continuous, and continuous extensions agreeing on a dense subset coincide.

Next, consider the binary case. $F : X_1 \times X_2 \to Y$ extends to $\tilde{F} : \beta X_1 \times \beta X_2 \to \beta Y$ by

$$\tilde{F}(u_1, u_2) = \big\{S \subseteq Y : \{x_1 \in X_1 : \{x_2 \in X_2 : F(x_1, x_2) \in S\} \in u_2\} \in u_1\big\}.$$

This can be considered as the extension fulfilled in two steps: first one extends left translations, then right ones. In the extended F, all right translations are continuous; in other words, the groupoid $(\beta X, \tilde{F})$ is *right topological*. Moreover, all left translations by *principal* ultrafilters are continuous, and such an extension is unique.

The extensions of mappings of arbitrary arity have analogous topological properties: If $F : X_1 \times \ldots \times X_n \to Y$ and $1 \leq i \leq n$, then for every $x_1 \in X_1, \ldots, x_{i-1} \in X_{i-1}$ and $u_{i+1} \in \beta X_{i+1}, \ldots, u_n \in \beta X_n$, the mapping

$$u \mapsto \tilde{F}(\hat{x}_1, \ldots, \hat{x}_{i-1}, u, u_{i+1}, \ldots, u_n)$$

of βX_i into βY is continuous, moreover, $\tilde{F}$ is a unique such extension of F. A proof of this fact will be done in the next section (Lemma 13).

Definition 4. Given $P \subseteq X_1 \times \ldots \times X_n$, let $\tilde{P}$ be *defined* as follows:

$$\langle u_1, \ldots, u_n\rangle \in \tilde{P} \quad \text{iff}$$
$$\big\{x_1 \in X_1 : \ldots \{x_n \in X_n : \langle x_1, \ldots, x_n\rangle \in P\} \in u_n \ldots\big\} \in u_1$$

for every $u_1 \in \beta X_1, \ldots, u_n \in \beta X_n$.

Proposition 5. *If $P \subseteq X_1 \times \ldots \times X_n$, then $\tilde{P} \subseteq \beta X_1 \times \ldots \times \beta X_n$. Moreover, $\tilde{P} \cap (X_1 \times \ldots \times X_n)$ is P.*

Proof. By Lemma 2.

Let us discuss the construction.

If P is a unary relation on X, $P \subseteq X$, one has

$$u \in \tilde{P} \quad \text{iff} \quad P \in u.$$

(The definition involves n-tuples; a 1-tuple $\langle x\rangle$ is just x.) Thus $\tilde{P}$ is an elementary open set of βX; the extensions of all unary relations on X form the standard open basis of the topology of βX. As we noted, the $\tilde{P}$ are in fact clopen.

If P is a binary relation, $P \subseteq X_1 \times X_2$, one has

$$\langle u_1, u_2\rangle \in \tilde{P} \quad \text{iff} \quad \big\{x_1 \in X_1 : \{x_2 \in X_2 : \langle x_1, x_2\rangle \in P\} \in u_2\big\} \in u_1.$$

There is an easier way to say the same. Let $\langle\,\rangle\tilde{}$ denote the extension of the pairing function $\langle\,\rangle$ (cf. Definition 11.1 in Hindman–Strauss' book, there $\langle\,\rangle\tilde{}$ is denoted by $\otimes$ and referred as a "tensor product"; another name that is used is a "Fubini product"). Then

$$\langle u_1, u_2\rangle \in \tilde{P} \quad \text{iff} \quad P \in \langle u_1, u_2\rangle\tilde{}.$$

This formula displays a similarity to the formula with unary P explicitly.

As for topological properties of extended binary relations, it is easy to see that for any $x_1 \in X_1$ and $u_2 \in \beta X_2$, the set $\{u_1 \in \beta X_1 : \langle u_1, u_2\rangle \in \tilde{P}\}$ is clopen in βX_1, and the set $\{u_2 \in \beta X_2 : \langle \hat{x}_1, u_2\rangle \in \tilde{P}\}$ is clopen in βX_2.

Likewise, if $\langle\,\rangle\tilde{}$ denotes the extension of taking n-tuples, one gets the following redefinition:

Proposition 6. *Let $P \subseteq X_1 \times \ldots \times X_n$. Then for all $u_1 \in \beta X_1, \ldots, u_n \in \beta X_n$,*

$$\langle u_1, \ldots, u_n\rangle \in \tilde{P} \quad \textit{iff} \quad P \in \langle u_1, \ldots, u_n\rangle\tilde{}\,.$$

Proof. Clear.

The extensions of relations of arbitrary arity have analogous topological properties: If $P \subseteq X_1 \times \ldots \times X_n$ and $1 \leq i \leq n$, then for every $x_1 \in X_1, \ldots, x_{i-1} \in X_{i-1}$ and $u_{i+1} \in \beta X_{i+1}, \ldots, u_n \in \beta X_n$, the subset

$$\{u \in \beta X_i : \langle \hat{x}_1, \ldots, \hat{x}_{i-1}, u, u_{i+1}, \ldots, u_n\rangle \in \tilde{P}\}$$

of βX_i is clopen. A proof of this fact is also postponed to the next section (Lemma 17).

Remark. It is worth to note that, strictly speaking, the symbol $\tilde{}$ carries an ambiguity (although the context usually leaves no doubts). First, given a relation P, one gets distinct extensions $\tilde{P}$ depends on its implicit arity. Say, let $P \subseteq X \times X$. If P is regarded as a binary relation on X, then $\tilde{P}$ is a binary relation on βX, while if P is considered as a unary relation on $X \times X$, then $\tilde{P}$ is a unary relation on $\beta(X \times X)$. Similarly for extensions of mappings. Second, the same object can have distinct extensions when regarded as a function or as a relation. Say, let P be a binary relation that is a function, and let F_P denote this unary function. If F_P is an injection, then $\tilde{P}$ and $\tilde{F}_P$ do not coincide: $\tilde{P} = P$, while $\tilde{F}_P \neq F_P$ whenever $\beta X \neq X$.

This case near characterizes relations coinciding with their extensions. Let us say that a relation P is *almost injective* iff for any i and all fixed $x_1, \ldots, x_{i-1}$, $x_{i+1}, \ldots, x_n$, the set

$$P_{x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_n} = \{x_i : \langle x_1, \ldots, x_n\rangle \in P\}$$

is finite. Note that a unary relation is almost injective iff it is finite. Then it can be shown that $\tilde{P} = P$ iff P is almost injective. The 'only if' part assumes that any infinite set carries a non-principal ultrafilter, which is weaker than the compactness of βX but still unprovable in ZF. The result can be restated in ZF alone if we redefine almost injective relations by replacing "is finite" with "carries no ultrafilter" (in ZFC the definitions coincide, while in any model without ultrafilters all relations are almost injective).

Lemma 7. *Let $h_1 : X_1 \to Y_1, \ldots, h_n : X_n \to Y_n$, and $G : Y_1 \times \ldots \times Y_n \to Z$. For all $S \subseteq Z$ and $u_1 \in \beta X_1, \ldots, u_n \in \beta X_n$, the following are equivalent:*

$$S \in \tilde{G}(\tilde{h}_1(u_1), \ldots, \tilde{h}_n(u_n))\,,$$

$$\big\{y_1 \in Y_1 : \ldots \{y_n \in Y_n : G(y_1, \ldots, y_n) \in S\} \in \tilde{h}_n(u_n) \ldots\big\} \in \tilde{h}_1(u_1)\,,$$

$$\big\{x_1 \in X_1 : \ldots \{x_n \in X_n : G(h_1(x_1), \ldots, h_n(x_n)) \in S\} \in u_n \ldots\big\} \in u_1\,.$$

Proof. The first and the second formulas are equivalent by definition of $\tilde{F}$.

That the second and the third formulas are equivalent can be proved by a straightforward induction on n. First one gets

$$\begin{aligned} &\big\{y_1 : \ldots \{y_n : G(y_1, \ldots, y_n) \in S\} \in \tilde{h}_1(u_n) \ldots\big\} \in \tilde{h}_1(u_1) \\ \text{iff}\quad &\big\{x_1 : h_1(x_1) \in \{y_1 : \ldots \{y_n : G(y_1, \ldots, y_n) \in S\} \in \tilde{h}_n(u_n) \ldots\}\big\} \in u_1 \\ \text{iff}\quad &\big\{x_1 : \{y_2 : \ldots \{y_n : G(h_1(x_1), y_2, \ldots, y_n) \in S\} \in \tilde{h}_n(u_n) \ldots\} \in \tilde{h}_2(u_2)\big\} \in u_1\,. \end{aligned}$$

Then similarly

$$\begin{aligned} &\big\{y_2 : \ldots \{y_n : G(h_1(x_1), y_2, \ldots, y_n) \in S\} \in \tilde{h}_n(u_n) \ldots\big\} \in \tilde{h}_2(u_2) \\ \text{iff}\quad &\big\{x_2 : \ldots \{y_n : G(h_1(x_1), h_2(x_2), \ldots, y_n) \in S\} \in \tilde{h}_n(u_n) \ldots\big\} \in u_2\,, \end{aligned}$$

etc. After n steps we obtain the required equivalence.

Corollary 8. *The following are equivalent:*

$$\begin{gathered} \langle \tilde{h}_1(u_1), \ldots, \tilde{h}_n(u_n)\rangle\tilde{} \in \tilde{P}, \\ P \in \langle \tilde{h}_1(u_1), \ldots, \tilde{h}_n(u_n)\rangle\tilde{}, \\ \{x_1 : \ldots \{x_n : \langle x_1, \ldots, x_n\rangle \in P\} \in \tilde{h}_n(u_n) \ldots\} \in \tilde{h}_1(u_1), \\ \{x_1 : \ldots \{x_n : \langle h_1(x_1), \ldots, h_n(x_n)\rangle \in P\} \in u_n \ldots\} \in u_1. \end{gathered}$$

Proof. The first and the second formulas are equivalent by Proposition 6, while the second and two last formulas are equivalent by Lemma 7 with $\langle\,\rangle$ as G.

Definition 9. Given a model $\mathfrak{A} = (X, F, \ldots, P, \ldots)$, let $\beta\mathfrak{A}$ denote the extended model $(\beta X, \tilde{F}, \ldots, \tilde{P}, \ldots)$.

As a corollary of Lemma 7, we get that continuous extensions of homomorphisms are homomorphisms.

Theorem 10 (The First Main Theorem). *Let $\mathfrak{A}$ and $\mathfrak{B}$ be two models. If h is a homomorphism of $\mathfrak{A}$ into $\mathfrak{B}$, then $\tilde{h}$ is a homomorphism of $\beta\mathfrak{A}$ into $\beta\mathfrak{B}$.*

Proof. Let $\mathfrak{A} = (X, F, \ldots, P, \ldots)$ and $\mathfrak{B} = (Y, G, \ldots, Q, \ldots)$.

Operations. As h is a homomorphism of (X, F) into (Y, G), we have for all $x_1, \ldots, x_n \in X$,

$$h(F(x_1, \ldots, x_n)) = G(h(x_1), \ldots, h(x_n)).$$

Then by Lemma 7, for all $u_1, \dots, u_n \in \beta X$,

$$\begin{aligned} &\tilde{h}(\tilde{F}(u_1, \dots, u_n)) \\ &= \big\{S : \{x_1 : \dots \{x_n : h(F(x_1, \dots, x_n)) \in S\} \in u_n \dots\} \in u_1\big\} \\ &= \big\{S : \{x_1 : \dots \{x_n : G(h(x_1), \dots, h(x_n)) \in S\} \in u_n \dots\} \in u_1\big\} \\ &= \tilde{G}(\tilde{h}(u_1), \dots, \tilde{h}(u_n)), \end{aligned}$$

thus $\tilde{h}$ is a homomorphism of $(\beta X, \tilde{F})$ into $(\beta Y, \tilde{G})$.

Relations. As h is a homomorphism of (X, P) into (Y, Q), we have for all $x_1, \dots, x_n \in X$,

$$\langle x_1, \dots, x_n \rangle \in P \quad \text{implies} \quad \langle h(x_1), \dots, h(x_n) \rangle \in Q.$$

We must verify that for all $u_1, \dots, u_n \in \beta X$,

$$\langle u_1, \dots, u_n \rangle \in \tilde{P} \quad \text{implies} \quad \langle \tilde{h}(u_1), \dots, \tilde{h}(u_n) \rangle \in \tilde{Q},$$

thus $\big\{x_1 : \dots \{x_n : \langle x_1, \dots, x_n \rangle \in P\} \in u_n\} \dots\big\} \in u_1$ implies

$$\big\{x_1 : \dots \{x_n : \langle x_1, \dots, x_n \rangle \in Q\} \in \tilde{h}(u_n) \dots\big\} \in \tilde{h}(u_1).$$

By Corollary 8, the latter formula is equivalent to

$$\big\{x_1 : \dots \{x_n : \langle h(x_1), \dots, h(x_n) \rangle \in Q\} \in u_n \dots\big\} \in u_1.$$

That h is a homomorphism means just $P \subseteq \{\langle x_1, \dots, x_n \rangle : \langle h(x_1), \dots, h(x_n) \rangle \in Q\}$. Therefore, the implication holds since $u_1, \dots, u_n$ are filters, thus $\tilde{h}$ is a homomorphism of $(\beta X, \tilde{P})$ into $(\beta Y, \tilde{Q})$.

Topological properties of extensions. The second main theorem

Here we describe specific topological structure of our extensions. Then we establish our second main result showing that the extensions are universal in the class of models carrying a topology with similar properties.

We start from an explicit description of extensions of (unary) mappings to arbitrary compact Hausdorff spaces.

Definition 11. If $F : X \to Y$ where Y is a compact Hausdorff topological space, let $\tilde{F} : \beta X \to Y$ be *defined* as follows:

$$\tilde{F}(u) = v \quad \text{iff} \quad \{v\} = \bigcap_{A \in u} \mathrm{cl}_Y(F\text{``}A).$$

It is routine to check that the intersection consists of a single point, so the definition is correct, and that $\tilde{F}$ is a continuous extension of F, unique since Y is Hausdorff.

If the compact space is βY, the ultrafilter $\tilde{F}(u)$ can be rewritten in a form closer to that we known already.

Lemma 12. *If $F : X \to \beta Y$, then*

$$\tilde{F}(u) = \big\{S \subseteq Y : \{x \in X : F(x) \in \tilde{S}\} \in u\big\}.$$

Proof. It easily follows from the definition that

$$\tilde{F}(u) = \{S \subseteq Y : (\forall A \in u)\ (\exists x \in A)\ \ F(x) \in \tilde{S}\}.$$

It remains to verify

$$(\forall A \in u)\ (\exists x \in A)\ \ F(x) \in \tilde{S} \quad \text{iff} \quad \{x \in X : F(x) \in \tilde{S}\} \in u.$$

'If' uses the fact that u is a filter, while 'only if' uses that u is ultra.

In particular, if $F : X \to Y \subseteq \beta Y$ with Y discrete, then $\tilde{F}$ in the sense of the first definition coincide with $\tilde{F}$ in the sense of this new definition, thus witnessing we do not abuse notation.

Lemma 13. *Let $F : X_1 \times \ldots \times X_n \to Y$. For each i, $1 \le i \le n$, and for every $x_1 \in X_1, \ldots,\ x_{i-1} \in X_{i-1}$ and $u_{i+1} \in \beta X_{i+1}, \ldots, u_n \in \beta X_n$, the mapping $\tilde{F}_{x_1,\ldots,x_{i-1},u_{i+1},\ldots,u_n}$ of βX_i into βY defined by*

$$u \mapsto \tilde{F}(x_1, \ldots, x_{i-1}, u, u_{i+1}, \ldots, u_n)$$

is continuous. Moreover, $\tilde{F}$ is the only such extension of F.

Proof. We shall show that $\tilde{F}$ can be constructed by fixing successively all but one arguments and extending resulting unary functions. First we describe the construction and verify that the constructed extension has the required continuity properties. Then we verify that it coincides with $\tilde{F}$.

Step 1. Fix all but the last arguments: $x_1 \in X_1, \ldots, x_{n-1} \in X_{n-1}$, and put

$$f_{x_1,\ldots,x_{n-1}}(x) = F(x_1, \ldots, x_{n-1}, x).$$

Thus $f_{x_1,\ldots,x_{n-1}} : X_n \to Y$. We extend it to $\tilde{f}_{x_1,\ldots,x_{n-1}} : \beta X_n \to \beta Y$ and put

$$F_1(x_1, \ldots, x_{n-1}, u) = \tilde{f}_{x_1,\ldots,x_{n-1}}(u).$$

Thus $F_1 : X_1 \times \ldots \times X_{n-1} \times \beta X_n \to \beta Y$. It is obvious from the construction that F_1 is continuous in its last argument (since then it coincides with $\tilde{f}_{x_1,\ldots,x_{n-1}}$). And it is continuous in any other of its arguments (since then its domain is discrete).

Step 2. Fix all but the $(n-1)$th arguments: $x_1 \in X_1, \ldots, x_{n-2} \in X_{n-2}$, $u_n \in \beta X_n$, and put

$$f_{x_1,\ldots,x_{n-2},u_n}(x) = F_1(x_1, \ldots, x_{n-2}, x, u_n).$$

Thus $f_{x_1,\ldots,x_{n-2},u_n} : X_{n-1} \to \beta Y$. We extend it to $\tilde{f}_{x_1,\ldots,x_{n-2},u_n} : \beta X_{n-1} \to \beta Y$ and put

$$F_2(x_1, \ldots, x_{n-2}, u, u_n) = \tilde{f}_{x_1,\ldots,x_{n-2},u_n}(u).$$

Thus $F_2 : X_1 \times \ldots \times \beta X_{n-2} \times \beta X_n \to \beta Y$. The mapping F_2 is continuous in its $(n-1)$th argument (since then it coincides with $\tilde{f}_{x_1,\ldots,x_{n-2},u_n}$). Moreover, it is continuous in its nth argument whenever the fixed $(n-1)$th argument is in X_{n-1} (since then it coincides with F_1).

Arguing so, after $n-1$ steps we get $F_{n-1} : X_1 \times \beta X_2 \times \ldots \times \beta X_n \to \beta Y$, which is continuous in its ith argument whenever any jth fixed argument is in X_j, for all i, $1 \le i \le n$, and all $j < i$.

Step n. Fix all but the first arguments: $u_2 \in \beta X_2, \ldots, u_n \in \beta X_n$, and put

$$f_{u_2,\ldots,u_n}(x) = F_{n-1}(x, u_2, \ldots, u_n).$$

Thus $f_{u_2,\ldots,u_n} : X_1 \to \beta Y$. We extend it to $\tilde{f}_{u_2,\ldots,u_n} : \beta X_1 \to \beta Y$ and put

$$F_n(u, u_2, \ldots, u_n) = \tilde{f}_{u_2,\ldots,u_n}(u).$$

Thus $F_n : \beta X_1 \times \ldots \times \beta X_n \to \beta Y$. The mapping F_n is continuous in its first argument (since then it coincides with $\tilde{f}_{u_2,\ldots,u_n}$). Moreover, it is continuous in its ith argument whenever any jth fixed argument is in X_j, for all i, $1 \le i \le n$, and all $j < i$.

The uniqueness of such an extension follows from the uniquiness of continuous extensions of unary mappings by induction.

It remains to verify that F_n coincides with $\tilde{F}$. We have:

$$\begin{aligned} &F_1(x_1, \ldots, x_{n-1}, u_n) = \tilde{f}_{x_1,\ldots,x_{n-1}}(u_n) \\ &= \big\{S : \{x : f_{x_1,\ldots,x_{n-1}}(x) \in S\} \in u_n\big\} \\ &= \tilde{F}(\hat{x}_1, \ldots, \hat{x}_{n-1}, u_n). \end{aligned}$$

Then

$$\begin{aligned} &F_2(x_1, \ldots, x_{n-2}, u_{n-1}, u_n) = \tilde{f}_{x_1,\ldots,x_{n-2},u_n}(u_{n-1}) \\ &= \big\{S : \{x_{n-1} : f_{x_1,\ldots,x_{n-2},u_n}(x_{n-1}) \in \tilde{S}\} \in u_{n-1}\big\} \\ &= \big\{S : \{x_{n-1} : F_1(x_1, \ldots, x_{n-2}, x_{n-1}, u_n) \in \tilde{S}\} \in u_{n-1}\big\} \\ &= \big\{S : \{x_{n-1} : \tilde{f}_{x_1,\ldots,x_{n-1}}(u_n) \in \tilde{S}\} \in u_{n-1}\big\} \\ &= \big\{S : \{x_{n-1} : \{T : \{x : f_{x_1,\ldots,x_{n-1}}(x) \in T\} \in u_n\} \in \tilde{S}\} \in u_{n-1}\big\} \\ &= \big\{S : \{x_{n-1} : S \in \{T : \{x : f_{x_1,\ldots,x_{n-1}}(x) \in T\} \in u_n\}\} \in u_{n-1}\big\} \\ &= \big\{S : \{x_{n-1} : \{x_n : f_{x_1,\ldots,x_{n-1}}(x_n) \in S\} \in u_n\} \in u_{n-1}\big\} \\ &= \tilde{F}(\hat{x}_1, \ldots, \hat{x}_{n-2}, u_{n-1}, u_n). \end{aligned}$$

Likewise we get $F_n(u_1, \ldots, u_n) = \tilde{F}(u_1, \ldots, u_n)$, as required.

Remark. This description of continuity of extended mappings cannot be improved. If some of $u_1, \ldots, u_{i-1}$ is non-principal, then the mapping $\tilde{F}_{u_1,\ldots,u_{i-1},u_{i+1},\ldots,u_n}$ of βX_i into βY defined by

$$u \mapsto \tilde{F}(u_1, \ldots, u_{i-1}, u, u_{i+1}, \ldots, u_n)$$

is not necessarily continuous. E.g. let F be a usual (binary) addition of natural numbers; then the mapping $u \mapsto u_1 \tilde{+} u$ is discontinuous. Also for fixed only $x_1 \in X_1, \ldots, x_{i-1} \in X_{i-1}$, the $(n-i+1)$-ary mapping $\tilde{F}_{x_1,\ldots,x_{i-1}}$ of $\beta X_i \times \ldots \times \beta X_n$ into βY defined by

$$\langle u_i, \ldots, u_n \rangle \mapsto \tilde{F}(x_1, \ldots, x_{i-1}, u_i, \ldots, u_n)$$

is not necessarily continuous. E.g. let $F(x_1, x_2, x_3) = x_2 + x_3$ and use the previous observation.

To name shortly the established topological property of $\tilde{F}$, let us introduce a terminology.

Definition 14. Let $X_1, \ldots, X_n, Y$ be topological spaces, and let $C_1 \subseteq X_1, \ldots, C_n \subseteq X_n$. We shall say that an n-ary function $F : X_1 \times \ldots \times X_n \to Y$ is *right continuous* w.r.t. $C_1, \ldots, C_n$ iff for each i, $1 \leq i \leq n$, and every $c_1 \in C_1, \ldots, c_{i-1} \in C_{i-1}$ and $x_{i+1} \in X_{i+1}, \ldots, x_n \in X_n$, the mapping

$$x \mapsto F(c_1, \ldots, c_{i-1}, x, x_{i+1}, \ldots, x_n)$$

of X_i into Y is continuous. If all the C_i coincide with, say C, we shall say that F is right continuous w.r.t. C.

In particular, F is right continuous w.r.t. the empty set iff for any $x_2 \in X_2, \ldots, x_n \in X_n$, the mapping

$$x \mapsto F(x, x_2, \ldots, x_n)$$

of X_1 into Y is continuous. Clearly, a unary F is right continuous iff it is continuous. If the operation is binary, the right continuity w.r.t. the empty set means that all right translations are continuous, and usually referred as "right continuity", see e.g. [2]. If F is right continuous w.r.t. the whole $X_1, \ldots, X_n$, it is called *separately continuous*.

The following proposition notes obvious properties of compositions of right continuous functions.

Proposition 15. (i) *Let $F : X_1 \times \ldots \times X_n \to Y$ be right continuous w.r.t. $C_1, \ldots, C_n$, and let $g : Y \to Z$ be continuous. Then $H : X_1 \times \ldots \times X_n \to Z$ defined by*

$$H(x_1, \ldots, x_n) = g(F(x_1, \ldots, x_n))$$

is right continuous w.r.t. $C_1, \ldots, C_n$.
(ii) *Let all $f_1 : X_1 \to Y_1, \ldots, f_n : X_n \to Y_n$ be continuous, and let $G : Y_1 \times \ldots \times Y_n \to Z$ be right continuous w.r.t. $D_1, \ldots, D_n$. Then $H : X_1 \times \ldots \times X_n \to Z$ defined by*

$$H(x_1, \ldots, x_n) = F(h_1(x_1), \ldots, h_n(x_n))$$

is right continuous w.r.t. $f_1^{-1} D_1, \ldots, f_n^{-1} D_n$. □

Proof. Clear.

Definition 16. We shall say that an algebra is *right topological* with C a *topological center* iff all its operations are strongly right continuous w.r.t. C.

In this terms, Lemma 13 states that for any algebra $\mathfrak{A} = (X, F, \ldots)$, its extension $\beta\mathfrak{A} = (\beta X, \tilde{F}, \ldots)$ is right topological with X a topological center.

Lemma 17. *Let $P \subseteq X_1 \times \ldots \times X_n$. For every i, $1 \leq i \leq n$, and for any $x_1 \in X_1, \ldots, x_{i-1} \in X_{i-1}$ and $u_{i+1} \in \beta X_{i+1}, \ldots, u_n \in \beta X_n$, the subset*

$$\tilde{P}_{x_1,\ldots,x_{i-1},u_{i+1},\ldots,u_n} = \{u \in \beta X_i : \langle \hat{x}_1, \ldots, \hat{x}_{i-1}, u, u_{i+1}, \ldots, u_n \rangle \in \tilde{P}\}$$

of βX_i is clopen.

Proof. Let

$$f_{x_1,\ldots,x_{i-1},u_{i+1},\ldots,u_n}(u) = \langle \hat{x}_1, \ldots, \hat{x}_{i-1}, u, u_{i+1}, \ldots, u_n \rangle\tilde{}.$$

The mapping $f_{x_1,\ldots,x_{i-1},u_{i+1},\ldots,u_n}$ of βX_i into $\beta(X_1 \times \ldots \times X_n)$ is continuous by the previous lemma. Hence

$$\begin{aligned}
\tilde{P}_{x_1,\ldots,x_{i-1},u_{i+1},\ldots,u_n} &= \{u \in \beta X_i : \langle \hat{x}_1, \ldots, \hat{x}_{i-1}, u, u_{i+1}, \ldots, u_n \rangle \in \tilde{P}\} \\
&= \{u \in \beta X_i : P \in \langle \hat{x}_1, \ldots, \hat{x}_{i-1}, u, u_{i+1}, \ldots, u_n \rangle\tilde{}\} \\
&= \{u \in \beta X_i : P \in f_{x_1,\ldots,x_{i-1},u_{i+1},\ldots,u_n}(u)\} \\
&= \{u \in \beta X_i : f_{x_1,\ldots,x_{i-1},u_{i+1},\ldots,u_n}(u) \in \tilde{Q}\}
\end{aligned}$$

where Q is P considered as a unary relation on $X_1 \times \ldots \times X_n$, thus $\tilde{Q}$ is a unary relation on $\beta(X_1 \times \ldots \times X_n)$. Since Q is clopen, so is its preimage $\tilde{P}_{x_1,\ldots,x_{i-1},u_{i+1},\ldots,u_n}$ under the continuous mapping $f_{x_1,\ldots,x_{i-1},u_{i+1},\ldots,u_n}$.

To name shortly the established topological property of $\tilde{P}$, let us introduce a terminology.

Definition 18. Let $X_1, \ldots, X_n$ be topological spaces, and let $C_1 \subseteq X_1, \ldots, C_n \subseteq X_n$. We shall say that an n-ary relation $P \subseteq X_1 \times \ldots \times X_n$ is *right open* w.r.t. $C_1, \ldots, C_n$ iff for each i, $1 \leq i \leq n$, and every $c_1 \in C_1, \ldots, c_{i-1} \in C_{i-1}$ and $x_{i+1} \in X_{i+1}, \ldots, x_n \in X_n$, the subset

$$P_{c_1,\ldots,c_{i-1},x_{i+1},\ldots,x_n} = \{x \in X_i : \langle c_1, \ldots, c_{i-1}, x, x_{i+1}, \ldots, x_n \rangle \in P\}$$

of X_i is open. That a relation is *right closed* (or *right clopen*, etc.) is *defined* likewise.

In particular, P is right open w.r.t. the empty set iff for every $x_2 \in X_2, \ldots, x_n \in X_n$, the subset

$$P_{x_2,\ldots,x_n} = \{x \in X_1 : \langle x, x_2, \ldots, x_n \rangle \in P\}$$

of X_1 is open. Clearly, a unary P is right open iff it is open. Likewise for right closed (right clopen, etc.) relations.

The following proposition notes an obvious interplay of right open (right closed, right clopen, etc.) relations and right continuous functions.

Proposition 19. (i) *Let* $F : X_1 \times \ldots \times X_n \to Y$ *be right continuous w.r.t.* $C_1, \ldots, C_n$*, and let* $Q \subseteq Y$ *be open. Then*

$$P = \{\langle x_1, \ldots, x_n \rangle \in X_1 \times \ldots \times X_n : F(x_1, \ldots, x_n) \in Q\}$$

is right open w.r.t. $C_1, \ldots, C_n$.
(ii) *Let all* $F_1 : X_1 \to Y_1, \ldots, F_n : X_n \to Y_n$ *be continuous, and let* $Q \subseteq Y_1 \times \ldots \times Y_n$ *be right open w.r.t.* $D_1, \ldots, D_n$*. Then*

$$P = \{\langle x_1, \ldots, x_n \rangle \in X_1 \times \ldots \times X_n : \langle F_1(x_1), \ldots, F_n(x_n) \rangle \in Q\}$$

is right open w.r.t. $F_1^{-1} D_1, \ldots, F_n^{-1} D_n$.
Both clauses also hold for right closed (right clopen, etc.) relations.

Proof. Clear.

Definition 20. Let $\mathfrak{A} = (X, F, \ldots, P, \ldots)$ be a model equipped with a topology, and $C \subseteq X$. We shall say that $\mathfrak{A}$ is *right open*, and C is its *topological center* iff all its operations are right continuous w.r.t. C and all its relations are right open w.r.t. C. Likewise for *right closed* (*right clopen*, etc.) models.

Note that if the model is an algebra (i.e. does not have relations), each of these properties means that the algebra is right topological with C a topological center.
In this terms, two last lemmas state the following.

Corollary 21. *For any model* $\mathfrak{A}$*, its extension* $\beta\mathfrak{A}$ *is right clopen with* $\mathfrak{A}$ *a topological center.*

Proof. Lemmas 13 and 17.

The following theorem concerns rather arbitrary right open and right closed models with dense topological centers than ultrafilter extensions.

Theorem 22. *Let* $\mathfrak{A}$ *be a right open model,* $\mathfrak{B}$ *a Hausdorff right closed model, and* $\mathfrak{C} \subseteq \mathfrak{A}$ *a dense submodel and a topological center of* $\mathfrak{A}$*. Let* h *be a continuous mapping of* $\mathfrak{A}$ *into* $\mathfrak{B}$ *such that*

(i) $h \restriction \mathfrak{C}$ *is a homomorphism, and*
(ii) $h``\,\mathfrak{C}$ *is a topological center of* $\mathfrak{B}$.

Then h *is a homomorphism of* $\mathfrak{A}$ *into* $\mathfrak{B}$.

Proof. Let $\mathfrak{A} = (X, F, \ldots, P, \ldots)$ and $\mathfrak{B} = (Y, G, \ldots, Q, \ldots)$.
Operations. We argue by induction on arity of F (and G).
Step 1. Fix $c_1, \ldots, c_{n-1} \in C$ and put for all $x \in X$ and $y \in Y$,

$$f_{c_1,\ldots,c_{n-1}}(x) = F(c_1, \ldots, c_{n-1}, x),$$
$$g_{h(c_1),\ldots,h(c_{n-1})}(y) = G(h(c_1), \ldots, h(c_{n-1}), y).$$

The functions $f_{c_1,\ldots,c_{n-1}}$ and $g_{h(c_1),\ldots,h(c_{n-1})}$ are continuous (since $c_1,\ldots,c_{n-1}$ are in C, C is a topological center of $\mathfrak{A}$, and $h``C$ is a topological center of $\mathfrak{B}$). Therefore the functions $h \circ f_{c_1,\ldots,c_{n-1}}$ and $g_{h(c_1),\ldots,h(c_{n-1})} \circ h$ (both of X to Y) are continuous too (as compositions of continuous functions). Moreover, they agree on the dense subset C of X (since $\mathfrak{C}$ is a subalgebra and $h \restriction C$ is a homomorphism), i.e. for all $c \in C$,

$$h(f_{c_1,\ldots,c_{n-1}}(c)) = g_{h(c_1),\ldots,h(c_{n-1})}(h(c)).$$

Hence (as Y is Hausdorff) they coincide, i.e. for all $x \in X$,

$$h(f_{c_1,\ldots,c_{n-1}}(x)) = g_{h(c_1),\ldots,h(c_{n-1})}(h(x)).$$

Thus we proved that for all $c_1,\ldots,c_{n-1} \in C$ and $x_n \in X$,

$$h(F(c_1,\ldots,c_{n-1},x_n)) = G(h(c_1),\ldots,h(c_{n-1}),h(x_n)).$$

Step 2. Fix $c_1,\ldots,c_{n-2} \in C$ and $x_n \in X$, and put for all $x \in X$ and $y \in Y$,

$$f_{c_1,\ldots,c_{n-2},x_n}(x) = F(c_1,\ldots,c_{n-2},x,x_n),$$
$$g_{h(c_1),\ldots,h(c_{n-2}),h(x_n)}(y) = G(h(c_1),\ldots,h(c_{n-2}),y,h(x_n)).$$

Again, the functions $f_{c_1,\ldots,c_{n-2},x_n}$ and $g_{h(c_1),\ldots,h(c_{n-2}),h(x_n)}$ are continuous (since $c_1,\ldots,c_{n-2}$ are in C, C is a topological center of $\mathfrak{A}$, and $h``C$ is a topological center of $\mathfrak{B}$). Therefore the compositions $h \circ f_{c_1,\ldots,c_{n-2},x_n}$ and $g_{h(c_1),\ldots,h(c_{n-2}),h(x_n)} \circ h$ (both of X to Y) are continuous too. Moreover, they agree on the dense subset C of X (by Step 1), i.e. for all $c \in C$,

$$h(f_{c_1,\ldots,c_{n-2},x_n}(c)) = g_{h(c_1),\ldots,h(c_{n-2}),h(x_n)}(h(c)).$$

Hence they coincide, i.e. for all $x \in X$,

$$h(f_{c_1,\ldots,c_{n-2},x_n}(x)) = g_{h(c_1),\ldots,h(c_{n-2}),h(x_n)}(h(x)).$$

Thus we proved that for all $c_1,\ldots,c_{n-2} \in C$ and $x_{n-1},x_n \in X$,

$$h(F(c_1,\ldots,c_{n-2},x_{n-1},x_n)) = G(h(c_1),\ldots,h(c_{n-2}),h(x_{n-1}),h(x_n)).$$

After n steps, we get $h(F(x_1,\ldots,x_n)) = G(h(x_1),\ldots,h(x_n))$ for all $x_1,\ldots,x_n \in X$, thus showing that h is a homomorphism of (X,F) into (Y,G), as required.

Relations. Assuming $\langle x_1,\ldots,x_n\rangle \in P$, we shall show $\langle h(x_1),\ldots,h(x_n)\rangle \in Q$ by induction on n.

Step 1. First we suppose $c_1,\ldots,c_{n-1} \in C$. Pick arbitrary neighborhood V of $h(x_n)$. Since h is continuous, there exists a neighborhood U of x_n such that $h``U \subseteq V$. The set $U \cap P_{c_1,\ldots,c_{n-1}}$ is open ($P_{c_1,\ldots,c_{n-1}}$ is open as $c_1,\ldots,c_{n-1}$ are in the topological center C) and nonempty (x_n belongs to it), and so there is $c \in C \cap U \cap P_{c_1,\ldots,c_{n-1}}$ (since C is dense). Therefore, we have $\langle c_1,\ldots,c_{n-1},c\rangle \in P$, and so $\langle h(c_1),\ldots,h(c_{n-1}),h(c)\rangle \in Q$ (since $h \restriction C$ is a homomorphism).

So we see that any neighborhood of $h(x_n)$ has a point y with $\langle h(c_1), \ldots, h(c_{n-1}), y\rangle \in Q$. Since the set

$$Q_{h(c_1),\ldots,h(c_{n-1})} = \{y : \langle h(c_1), \ldots, h(c_{n-1}), y\rangle \in Q\}$$

is closed (as $h(c_1), \ldots, h(c_{n-1})$ are in the topological center $h``C$), it has the point $h(x_n)$. Thus we proved that whenever $c_1, \ldots, c_{n-1} \in C$ and $\langle c_1, \ldots, c_{n-1}, x_n\rangle \in P$, then

$$\langle h(c_1), \ldots, h(c_{n-1}), h(x_n)\rangle \in Q.$$

Step 2. Now we suppose $c_1, \ldots, c_{n-2} \in C$ and $x_n \in X$. Pick arbitrary neighborhood V of $h(x_{n-1})$. Since h is continuous, there exists a neighborhood U of x_{n-1} such that $h``U \subseteq V$. Again, the set $U \cap P_{c_1,\ldots,c_{n-2},x_n}$ is open and nonempty, so there is $c \in C \cap U \cap P_{c_1,\ldots,c_{n-2},x_n}$. Hence, $\langle c_1, \ldots, c_{n-2}, c, x_n\rangle \in P$, and so $\langle h(c_1), \ldots, h(c_{n-2}), h(c), h(x_n)\rangle \in Q$ (by Step 1).

So any neighborhood of $h(x_{n-1})$ has a point y with $\langle h(c_1), \ldots, h(c_{n-2}), y, h(x_n)\rangle \in Q$. Since the set

$$Q_{h(c_1),\ldots,h(c_{n-2}),h(x_n)} = \{y : \langle h(c_1), \ldots, h(c_{n-2}), y, h(x_n)\rangle \in Q\}$$

is closed, it has the point $h(x_{n-1})$. Thus we proved that whenever $c_1, \ldots, c_{n-2} \in C$ and $\langle c_1, \ldots, c_{n-2}, x_{n-1}, x_n\rangle \in P$, then

$$\langle h(c_1), \ldots, h(c_{n-2}), h(x_{n-1}), h(x_n)\rangle \in Q.$$

After n steps, we conclude that whenever $\langle x_1, \ldots, x_n\rangle \in P$, then $\langle h(x_1), \ldots, h(x_n)\rangle \in Q$, thus h is a homomorphism of (X, P) into (Y, Q), as required.

The following theorem states the universal property of $\mathfrak{A}$ completely analogous to that of the Stone–Čech (or Wallman) compactification.

Theorem 23 (The Second Main Theorem). *Let $\mathfrak{A}$ and $\mathfrak{B}$ be two models, and let $\mathfrak{B}$ be compact Hausdorff right closed. Let h be a homomorphism of $\mathfrak{A}$ into $\mathfrak{B}$ such that $h``\mathfrak{A}$ is a topological center of $\mathfrak{B}$. Then $\tilde{h}$ is a homomorphism of $\beta\mathfrak{A}$ into $\mathfrak{B}$.*

Proof. By Corollary 21 and Theorem 22.

Note that the First Main Theorem (Theorem 10) follows from this one.

Generalizations

Here we note that the results establishing universality of ultrafilter extensions of models w.r.t. homomorphisms remain true if one replaces homomorphisms by more general relationships between models.

The concepts of homotopy and isotopy are customarily used for groupoids, especially, in quasigroup theory. Let us give a general definition of homotopy and isotopy between arbitrary models.

Definition 24. Let F and G be n-ary operations on X and Y resp. Mappings $h, h_1, \ldots, h_n$ of X into Y form a *homotopy* of (X, F) into (Y, G) iff

$$h(F(x_1, \ldots, x_n)) = G(h_1(x_1), \ldots, h_n(x_n))$$

for all $x_1, \ldots, x_n \in X$. The homotopy is an *isotopy* iff all the $h, h_1, \ldots, h_n$ are bijective.

Definition 25. Let P and Q be n-ary relations on X and Y resp. Mappings $h_1, \ldots, h_n$ of X into Y are a *homotopy* of (X, P) into (Y, Q) iff

$$P(x_1, \ldots, x_n) \quad \text{implies} \quad Q(h_1(x_1), \ldots, h_n(x_n))$$

for all $x_1, \ldots, x_n \in X$. The homotopy is an *isotopy* iff all the $h_1, \ldots, h_n$ are bijective and

$$P(x_1, \ldots, x_n) \quad \text{iff} \quad Q(h_1(x_1), \ldots, h_n(x_n)).$$

Note that when all the $h, h_1, \ldots, h_n$ coincide, then the homotopy is an homomorphism (and the isotopy is an isomorphism). In particular, homotopies of unary relations are homomorphisms.

If $\mathfrak{A}$ and $\mathfrak{B}$ have more than one operation or relation, there are various ways to define homotopies (and isotopies) between them, the weakest of which is as follows.

Definition 26. A family H of mappings of X into Y form a *homotopy* of $\mathfrak{A}$ into $\mathfrak{B}$ iff for any m-ary operation F in $\mathfrak{A}$ there are mappings $h, h_1, \ldots, h_m$ in H forming a homotopy of (X, F) into (Y, G) with the corresponding operation G in $\mathfrak{B}$, and for any n-ary relation P in $\mathfrak{A}$ there are mappings $h_1, \ldots, h_n$ in H forming a homotopy of (X, P) into (Y, Q) with the corresponding relation Q in $\mathfrak{B}$. The homotopy H is an *isotopy* iff all mappings in H are bijective.

Obviously, a homotopy H is a homomorphism iff $|H| = 1$. In general, the size of H can be regarded as a degree of its dissimilarity to a homomorphism.

Proposition 27. *Let $F : X \to Y$.*

(i) *If F is surjective, then so is $\tilde{F}$.*
(ii) *If F is injective, then so is $\tilde{F}$. Moreover, $(\tilde{F})^{-1} = (F^{-1})^{\sim}$.*
(iii) *If F is bijective, then $\tilde{F}$ is a homeomorphism of βX onto βY.*

Proof. (i) We must show that for any $v \in \beta Y$ there is $u \in \beta X$ such that $\tilde{F}(u) = v$, i.e.

$$S \in v \quad \text{iff} \quad \{x : F(x) \in S\} \in u$$

for all $S \subseteq Y$. Given v, let

$$D = \big\{\{x : F(x) \in S\} : S \in v\big\}.$$

D has the finite intersection property: Given $S', S'' \in v$, we have $\{x : F(x) \in S'\} \cap \{x : F(x) \in S''\} = \{x : F(x) \in S' \cap S''\}$, so this set is in D (since $S' \cap S''$ is in v).

Let u be any ultrafilter that extends D. Then u is as required: The 'only if' part holds by definition of u. To verify the 'if' part, notice that if $S \notin v$ then $Y \setminus S \in v$, and so $\{x : F(x) \in Y \setminus S\} \in u$, whence it follows $\{x : F(x) \in S\} \notin u$ (as preimages of disjoint sets are disjoint).

(ii) We must show that if $u', u'' \in \beta X$ are distinct, then so are $\tilde{F}(u'), \tilde{F}(u'') \in \beta Y$, i.e. there is $T \in \tilde{F}(u') \setminus \tilde{F}(u'')$, and thus $\{x : F(x) \in T\} \in u' \setminus u''$. As $u' \neq u''$, there is $S \in u' \setminus u''$. Since F is injective, we have $\{x : F(x) \in F``S\} = S$, so we can put $T = F``S$.

The equality $(\tilde{F})^{-1} = (F^{-1})\tilde{}$ follows immediately.

(iii) This follows from (i) and (ii).

Remark. Clause (i) uses the assumption that any filter extends to an ultrafilter, which is, as we mentioned above, equivalent to the compactness of βX.

By using Lemma 7 and Proposition 27 and modifying arguments of the proofs of our main results, one gets the following generalization (we leave details for the reader).

Theorem 28. *Both Main Theorems (Theorems 10 and 23), as well as Theorem 22, remain true by replacing homomorphisms with homotopies, isotopies, and embeddings.* □

Question. Characterize relationships between models such that both theorems remain true by replacing homomorphisms with these relationships.

Another interesting question is about theories of extended models.

Question. Characterize formulas that are preserved under β.

In [4] we answer the question for the case when the formulas are identities and the models are groupoids.

Finally, let us mention that certain types of ultrafilters form submodels of extended models. In particular, so are κ-complete ultrafilters, for any given κ. A proof generalizes the proof in [5] given for groupoids.

Remark. When I prepared the paper for publishing, I recognized V. Goranko's unpublished manuscript *Filter and Ultrafilter Extensions of Structures: Universal-algebraic Aspects*, 2007 (the author said me the first version was written ten years before). Goranko extends models by arbitrary filters. However his filter extension of operations does not work for ultrafilters, so he defines this case separately, in the same way that in Definition 1 here. His extension of relations differs from that given in Definition 4. Goranko proved a theorem analogous to the First Main Theorem, both theorems coincide for ultrafilter extension of operations. He asks whether there is another ultrafilter extension of operations that would be "better". Perhaps the Second Main Theorem can be considered as the negative answer.

References

1. Berglund, J., Junghenn, H., Milnes, P.: Analysis on semigroups. Wiley, N.Y. (1989)
2. Hindman, N., Strauss, D.: Algebra in the Stone–Čech compactification. W. de Gruyter, Berlin (1998)
3. Saveliev, D.I.: Groupoids of ultrafilters (2008) (manuscript)
4. Saveliev, D.I.: Identities stable under ultrafilter extensions (in progress)
5. Saveliev, D.I.: On ultrafilters without the axiom of choice (in progress)

Logic in the Community

Jeremy Seligman[1], Fenrong Liu[2], and Patrick Girard[3]

[1] University of Auckland
j.seligman@auckland.ac.nz
[2] Tsinghua University
fenrong@tsinghua.edu.cn
[3] University of Auckland
p.girard@auckland.ac.nz

1 Reasoning about Social Relations

Communities consist of individuals bounds together by social relationships and roles. Within communities, individuals reason about each other's beliefs, knowledge and preferences. Knowledge, belief, preferences and even the social relationships are constantly changing, and yet our ability to keep track of these changes is an important part of what it means to belong to a community. In the past 50 years, our patterns of reasoning about knowledge, beliefs and preferences have been extensively studied by logicians (cf. notably, [12], [19], [1], [6], [10], [18], and [17].), but the way in which we are influenced by social relationships has received little attention. The country and culture in which we are born, our families, friends, partners or work colleagues all play a part in the formation, rejection and modification of our attitudes. One might update one's beliefs about the impact of human activities on climate change after reading a scientific report, become vegetarian after moving to California, decide to change one's appearance because of peer pressure, vote for a candidate one doesn't like personally for the sake of one's department, or argue in a court of law for the innocence of someone one believes to be guilty. From the perspective of individual rationality, such changes are difficult to understand, but they are not arbitrary and are governed by norms that we internalise as readily as the rules of logic. It is the logic of these internalised norms of social behaviour, a social conception of rationality, that we intend to investigate from the standpoint of logic.

This paper lays out the problems we wish to address, with a view to promoting the logic of community as an interesting area of research in applied philosophical logic. As a small test case, we will provide some technical details for the first of the following examples. But the main aim of our paper is to describe what we take to be a coherent and fruitful topic for future research, some of which is already under way.

1.1 Facebook Friends

Perhaps the simplest example of a social structure is that of online communities such as Facebook (www.facebook.com). Cutting out bells and whistles, the structure is just that of a symmetric relation of 'friendship'. The relation is

M. Banerjee and A. Seth (Eds.): ICLA 2011, LNAI 6521, pp. 178–188, 2011.

not necessarily transitive and is arguably irreflexive. Yet even with this simple structure, we get an interesting model of communities. Define an agent's 'community' to be the set of her friends (together with the agent herself, if friendship is not reflexive). In this way, we get a picture of the social world as a collection of overlapping communities. Even this is a little more subtle and flexible than the naive view of communities as isolated groups: nations, schools, families, etc. Moreover, we also get a simple model of the interaction between social relations and propositional attitudes, specifically knowledge. Privacy protocols ensure that information can be restricted to be viewable only to one's friends (in theory at least!). This can be implemented (in logic) with an announcement operator – details will be given in Section 2 below. The social network itself is also subject to change: one can add or remove friends, so altering both social and epistemic relations.

1.2 Distributed Knowledge

> I have a friend in Minsk, who has a friend in Pinsk, whose friend in Omsk, has friend in Tomsk, with friend in Akmolinsk. His friend in Alexandrovsk has friend in Petropavlovsk, whose friend somehow is solving now the problem in Dnepropetrovsk. (Tom Lehrer, Lobachevsky)

Within any community, knowledge propagates via social relations. Actual transmission of knowledge depends on communication but as a first approximation, one can reason about who knows what on the basis of social relations. If you tell your colleague about some important secret at work, the chances are that his or her lover will also know. The same is true of academic networks, as immortalised by Tom Lehrer. A network of 'friends' gives rise to degrees of accessibility of information, which can be captured by the sequence of propositions

0. I know p
1. I have a friend who knows p
2. I have a friend who has a friend who knows p

∞ I am connected by friendship to someone who knows that p

The inference from p being accessible to an agent to that agent knowing p is defeasible, but there is no need to build in such defeasibility to a logic of distributed knowledge. Instead, one can reason directly about accessible information, defining an operator 'I have access to the information that p' or even the more fine-grained 'I have access in n steps to the information that p' for proposition n in the above list.

1.3 Social Information Flow

The mechanism for distributing knowledge is communication, which is a hugely complex matter in reality. Nonetheless, one can model the transmission of information within a simple social network on the assumption that announcements are made to friends. This gives an analogous sequence of propositions:

0. I tell a friend that p
1. My friend tells her friends that p
2. My friends' friends' tell their friends that p

∞ I told your secret to only one person, but now everyone knows!

By modelling this using announcement operators restricted to friends, we get an elementary logic of gossip.

1.4 Deference to Expert Opinion

Whereas friendship is a symmetric relation, the way in which our attitudes are shaped by society is typically asymmetric. Different people have different access to information and different capacities to absorb, process and transmit it. Our reasoning about knowledge in the community often takes these asymmetries into account. A clear example is the way in which we defer to expert opinion on matters that require specific training, ability or experience. In a court of law, or in policy making committees, the testimony of experts carries more weight than the opinion of ordinary folk. Even in our daily lives, when we seek council from older or more experienced members of our communities, we do it with an attitude of deference. If the opinion of the expert is in line with my own, then I may feel more confident in my attitudes but there is little practical consequence. The interesting case is when there is some difference.

Suppose I initially believe that $\sim p$ but that I consult Prof. X, who is an expert on matters to do with p. Prof. X is of the opinion that p. In reasoning about community belief, it may be sufficient to leave it at this point. There is a conflict, but Prof. X's opinion is that of an expert. Thus we have

1. I believe $\sim p$
2. Concerning p, I defer to the opinion of Prof X
3. Prof X believes p

This is consistent, but if I do nothing with Prof. X's advice, then I have wasted my time in consulting him, at possibly also the large sum of money paid as his fee. One action I might take is to change my belief to $\sim p$. A great deal has been written on the subject of belief revision and much of it applies to the analysis of this situation. But another possibility is for me to defer to Prof. X in the sense that I act in accordance with his belief, taking it to be a safer guide to action, but retaining a private conviction in $\sim p$. This is consistent but requires further analysis of the relation between belief, desire and action, all of which may also have a social dimension. Further possibilities are opened when considering the behaviour of groups. We, as a society, may agree to consult a panel of scientific experts when formulating policy regarding climate change. Given the expert opinion that significantly unpleasant consequences will result if the rate of carbon emissions is not drastically reduced (p for short), we may revise the group belief accordingly, admitting that some members of the group – perhaps most of them – retain their belief that $\sim p$. For all this to be modelled in a logical system, we must have a mechanism for belief aggregation that is sensitive to social relations within the group.

1.5 Peer Pressure

Deference to experts requires the addition of asymmetric social relationships to our model. But even within the symmetric friendship model, there is an asymmetry between myself and others. Suppose that I have a magnificent, well-developed and well-groomed handlebar moustache. I like it very much but most (perhaps all) of my friends think it is ridiculous and are even somewhat embarrassed to be seen with me in public. Our preferences are clear. Other things being equal, I prefer to have the moustache than not to have it, but my friends prefer the opposite. We can then define 'peer pressure' as adopting a deferential attitude to one's friends. Although the friendship relation is symmetric, it is important that it is also irreflexive: I am not my own friend. There are two further differences with the case of deference to expert opinion. First, the attitude involved is preference rather than belief. And second, one is deferential not to an individual but to a group. We therefore have to employ techniques of preference aggregation. Interestingly, the group is indexically determined. When bowing to peer pressure, I am deferential to the aggregated preference of *my* friends.

1.6 Community Norms

A somewhat similar scenario occurs whenever individual preferences are contrary to community norms, such as paying one's taxes. Everyone (let's assume) prefers not to pay tax but also prefers that everyone else in the community does pay tax. The Golden Rule of many ethical systems tells us what we ought to do in such situations but duty and preference may diverge. Again we have the logical structure of peer pressure but with an added asymmetry. In this case, everyone's preferences are the same *de re*; they only differ *de se*. Fleshing this out a bit, letting $T(x)$ stand for the predicate 'x pays his/her taxes', we can say that everyone agrees to the following indexical proposition

I prefer $\sim T(I)$ and $T(x)$ for all $x \neq I$

The socially acceptable resolution of this problem is for everyone to adopt a deferential attitude to the group's aggregated preference, which if we assume that a majority of $n-1$ to 1 is sufficient for suitably large n, results in everyone paying their taxes - even if they retain a private preference not to.

1.7 Mutual Subordination

Our final example of an interesting puzzle concerning logic in communities commonly arises in more intimate settings. There is a young couple, a boy and a girl, desperately in love and yet lacking a little in self-assurance. They have just moved in together. All is well except for one small problem about their sleeping arrangements. Both are used to sleeping on the right side of the bed and they both prefer this strongly. Yet they also both prefer to sacrifice their personal preferences in favour of the other - such is the power of love. Communication is obviously the answer to this problem but they are faced with the paradox of Mutual Subordination:

1. He prefers to sleep on the right.
2. He knows that she also prefers to sleep on the right.
3. He is deferential to her preferences and so revises his preference to that of sleeping on the left.
4. She does the same.
5. Now they both prefer to sleep on the left.

The last two examples have a game theoretical flavour and similar game theoretical scenarios have been extensively studied.[1] Our analysis, however, stresses the interplay between individuals and their communities, and how each is affected in attitude attribution, something which is hardly captured in a utility oriented calculus. For instance, there is a distinction between *my* preferences, *my friends'* preferences and *our* aggregated preferences. In a lot of cases, *my* preferences may not correspond to those of my friends, nor to *our* aggregated preferences. When *I* am saying that x is preferable, I might be reporting my own preference, that of my friends, or that of our community. I might in the same day say that "x is preferable" and "x is not preferable" without contradicting myself, as I might be reporting preferences in the name of my community on one occasion and my own on another occasion. We think that rationality for community has to accommodate this if it is to make sense at all, and the logic of the next section is devised with this purpose in mind.

2 Facebook Logic

In the remainder of the paper, we develop an epistemic logic of communities. This logic emphasises the multi-faceted attitude analysis of the above examples with a two-dimensional approach, one dimension standing for each agent's epistemic possibilities, the second for each agent's community (one's friends). As a starting point for this new paradigm of research, we sketch an approach to modelling the first of the applications mentioned above: that of Facebook Friends.

Define a **social network** $\langle A, \asymp \rangle$ to consists of a set A of *agents* and a binary relation $\asymp$ of *friendship* between agents that is irreflexive and symmetric. In the simplest case, we will only be interested in one propositional attitude: knowledge. For this, we adopt a minor variant of the standard definition from epistemic logic (e.g. [6]). An **epistemic model** $\langle W, A, \sim, V \rangle$ consists of a set W of *epistemic alternatives*, a set A of *agents*, a partial equivalence relation $\sim_a$ on W for each agent a in A, and a propositional valuation function V, assigning a subset of $W \times A$ to each propositional variable.

There are two main differences from the standard definition. First the relation $\sim_a$, which is interpreted by the relation between epistemic alternatives of being indistinguishable by a, is a partial equivalence relation. This means that it is symmetric and transitive but not necessarily reflexive. We do not insist on reflexivity because we allow for the possibility that some epistemic alternatives have

[1] One can even trace back such analyses to traditional community wisdom, for instance in Indian culture, the so-called *tragedy of the commons* and *Birbal story* (cf. [14]).

been ruled out by some but not all of the agents; this will be important when we consider the dynamics of announcements. Second, propositional variables (and formulas more generally) are interpreted as expressing *indexical* propositions, represented as subsets of $W \times A$ instead of subsets of W.

Now, combining the two ideas we define an **epistemic social network model** M to consists of a social network model $\langle A, \asymp_w \rangle$ for each w in W and an epistemic model $\langle W, A, \sim, V \rangle$. A social network model is linked to each epistemic alternative so that we can represent an agent's ignorance about the structure of the social network. We use indexical modal operators K and F, read as 'I know that' and 'all my friends' with a semantics in which satisfaction is relative to both an epistemic alternative w and an agent a. The salient clauses are:

$$
\begin{array}{lll}
M, w, a \models p & \text{iff} & (w, a) \in V(p) \\
M, w, a \models K\varphi & \text{iff} & M, v, a \models \varphi \text{ for every } v \sim_a w \\
M, w, a \models F\varphi & \text{iff} & M, w, b \models \varphi \text{ for every } b \asymp_w a
\end{array}
$$

A simple example illustrates the difference between the alternations of modalities K and F. Let p be the proposition 'I am in danger'. Then

KFp : I know that all my friends are in danger
FKp : Each of my friends knows that s/he is in danger

We define the existential duals as usual: $\langle K \rangle = {\sim}K{\sim}$, $\langle F \rangle = {\sim}F{\sim}$.

2.1 Distributed Knowledge

The basic scenario of Distributed Knowledge, as discussed above, can be represented as follows:

${\sim}(Kp \vee K{\sim}p)\ \&\ \langle F \rangle (Kp \vee K{\sim}p)$
I don't know whether p, but I have a friend who does.
Kp, $\langle F \rangle Kp$, $\langle F \rangle \langle F \rangle Kp$, etc.
I know p, I have a friend who knows p, I have a friend who has a friend who knows p, etc.
$\langle F^* \rangle Kp$
I am connected by friendship to someone who knows that p

The latter requires a new operator, F^*, which can be introduced (following *PDL* in [11]) as the modality of the transitive closure of the friendship relation.

2.2 Talking about Friends

To talk about your friends, you need to give them *names*. We therefore introduce a syntactic category of nominals and extend the valuation function V to apply to nominals as well as propositional variables (for further details of hybrid logic, we refer to [3]). We will assume that names are 'rigid designators' in the epistemic sense, i.e., that every agent knows who is whom. So for each nominal n we insist that there is an agent $\underline{n} \in A$ such that for all $a \in A$ and $w \in W$, $(w, a) \in V(n)$ iff $a\underline{n}$.

Now we can say

$\langle F \rangle n$: n is my friend

Also borrowed from hybrid logic is an operator $@_n$ for shifting the evaluation to the agent named n. This enables us to say

$@_n Kp$: n knows that p

Finally, another hybrid logic device: a way of *indexically referring* to the current agent. This is provided by the operator $\downarrow x$ which names the current agent 'x'. This enables us to express some nice interactions between friendship and knowledge:

$\downarrow x \langle F \rangle K @_n \langle F \rangle x$: I have a friend who knows that n is friends with me.

To capture the semantics of $\downarrow x$ we need the help of an assignment function g assigning agents to variables. Variables are of the same syntactic category as nominals and so we also write $\underline{x}$ for $g(x)$. With the help of assignment functions, we get the following satisfaction conditions:

$$\begin{array}{lll} M, g, w, a \models x & \text{iff} & g(x) = a \\ M, g, w, a \models @_n\varphi & \text{iff} & M, g, w, \underline{n} \models \varphi \\ M, g, w, a \models \downarrow x\varphi & \text{iff} & M, g^x_a, w, a \models \varphi \end{array}$$

where, as usual, g^x_a is defined by $g^x_a(y) = a$ if $x = y$ and $g(y)$ otherwise.

2.3 Indexical Public Announcements

In dynamic epistemic logic, the result of publicly announcing that p is given by eliminating epistemic alternatives in which p is not true.[2]. The operator $[!\varphi]$ for 'after announcing φ' is defined by

$$M, w \models [!\varphi]\psi \quad \text{iff} \quad \text{if } M, w \models \varphi \text{ then } M_\varphi, w \models \psi$$

where M_φ is the result of restricting M to the set of epistemic alternatives v such that $M, v \models \varphi$. The logic is pleasingly simple, thanks to the following (now well-known) reduction axioms:

$$\begin{array}{l} [!\varphi]p \;\equiv\; (\varphi \supset p) \\ [!\varphi]{\sim}\psi \;\equiv\; (\varphi \supset {\sim}[!\varphi]\psi) \\ [!\varphi](\psi_1 \,\&\, \psi_2) \;\equiv\; ([!\varphi]\psi_1 \,\&\, [!\varphi]\psi_2) \\ [!\varphi]K_a\psi \;\equiv\; (\varphi \supset K_a[!\varphi]\psi) \end{array}$$

With these axioms, a completeness result for the base epistemic logic can be lifted to its dynamic extension. A crucial feature of the operator is the restriction to announcements that are true. Without this, the model M_φ would not contain w and the satisfaction condition would be rendered meaningless.

[2] Public announcement logic was introduced in [15].

To interpret public announcement in indexical epistemic models, we give the obvious definition:

$$M, w, a \models [!\varphi]\psi \quad \text{iff} \quad \text{if } M, w, a \models \varphi \text{ then } M_{a,\varphi}w, a \models \psi$$

where $M_{a,\varphi}$ is the restriction of M to those epistemic alternatives v such that $M, v, a \models \varphi$. But there is a problem: public announcements cannot be reduced when we add the hybrid shifting operator $@_n$. The equivalence

$$[!\varphi]@_n\psi \;\equiv\; (\varphi \supset @_n[!\varphi]\psi)$$

is not in general valid. If φ is a non-indexical proposition then the equivalence holds: in fact we have the simpler equivalence $[!\varphi]@_n\psi \;\equiv\; @_n[!\varphi]\psi$. Since the truth of a non-indexical φ does not depend on the agent, it does not matter which agent announces it. But when the truth of φ is indexical, varying by agent, then the equivalence breaks down.

Suppose, for example that a but not $\underline{n}$ is in danger. Then evaluating at a, the following two propositions are not equivalent:

1. $[!p]@_np$
   After I announce that I am in danger, n is in danger.
2. $(p \supset @_n[!p]p)$
   If I am in danger then after n announces that he is in danger, he is in danger.

Proposition 1 can easily be falsified; there is no implication from a's being in danger to $\underline{n}$'s being in danger. But Proposition 2 is true: if $\underline{n}$ is not in danger then he cannot announce that he is and so the consequent of the conditional is trivially true. Moreover, there is no way of avoiding the problem. To do so, we would need an announcement by $\underline{n}$ that is equivalent to a's announcement, but for indexical announcements this is impossible.

Our solution is to introduce a new operator $[n!\varphi]$ for 'after n announces φ', with satisfaction conditions

$$M, g, w, a \models [n!\varphi]\psi \quad \text{iff} \quad \text{if } M, g, w, \underline{n}, \models \varphi \text{ then } M_{\underline{n},\varphi}, g, w, a \models \psi$$

This has the advantage of admitting reduction equivalences as follows:

$$\begin{aligned}
&[n!\varphi]p \;\equiv\; (@_n\varphi \supset p)\\
&[n!\varphi]{\sim}\psi \;\equiv\; (@_n\varphi \supset {\sim}[n!\varphi]\psi)\\
&[n!\varphi](\psi_1 \,\&\, \psi_2) \;\equiv\; ([n!\varphi]\psi_1 \,\&\, [n!\varphi]\psi_2)\\
&[n!\varphi]@_m\psi \;\equiv\; @_m[n!\varphi]\psi\\
&[n!\varphi] \downarrow x\psi \equiv\; \downarrow x[n!\varphi]\psi\\
&[n!\varphi]F\psi \;\equiv\; F[n!\varphi]\psi\\
&[n!\varphi]K\psi \;\equiv\; K[n!\varphi]\psi
\end{aligned}$$

(with a change of bound variables in the line for $\downarrow x$, if necessary.)

The new operator also allows us to recover reduction for the indexical notion of public announcement via the equivalence

$$[!\varphi]\psi \equiv\; \downarrow x[x!\varphi]\psi$$

in which x is a new variable.

2.4 Talking to Friends

Of greater interest to logic in the community is the possibility of making announcements only to one's friends. Here we adopt a simplistic approach, noting some of its limitations and a direction for further research.

We define an operator $[F!\varphi]$ for 'after I announce φ to my friends' by

$$M, g, w, a \models [F!\varphi]\psi \quad \text{iff} \quad \text{if } M, g, w, a \models \varphi \text{ then } M', g, w, a \models \psi$$

where $M' = \langle W, A, F, \sim', V\rangle$ has the same set W of epistemic alternatives as M but has an indistinguishability relation $\sim'$ defined as follows:

$$\begin{array}{l} \text{if } b \asymp a \text{ then} \\ \qquad u \sim'_b v \quad \text{iff} \quad u \sim_b v \text{ and } M, g, u, a \models \varphi \text{ and } M, g, v, a \models \varphi \\ \text{otherwise } \sim'_b = \sim_b \end{array}$$

In other words, the epistemic indistinguishability relation of agents that are not friends with a remains unchanged, but that of a's friends is changed so as to remove links between alternatives that are incompatible with a's announcement.

For example, suppose that $\underline{n}$ is in danger (p) and announces this to her friends. After the announcement, all of $\underline{n}$'s friends will know $@_n p$ that $\underline{n}$ is in danger. So the formula

$$@_n[F!p]FK@_n p$$

is valid.

Scenarios of this kind are somewhat similar to what is called *private announcement* in [4]. Our way of handling the announcement here is to take them to be *soft information* (see detailed discussions on soft information vs. hard information in [16]). We think that the approach of product update with event model in [4] can be adapted to this context, too.

3 Prospects

The sketch of Facebook Logic is only a beginning. Even within this simple model of communities there is much to investigate. We hope that this case study has shown the readers where we are heading: our goal is to use recent developments in dynamic logics of knowledge, belief and preference to model the subtleties of the communication and relationship between agents in communities. Going back to the topics outlined in Section 1, there are immediate directions we would like to explore. Due to limitations of space, we finish with a few preliminary remarks on how to proceed.

1. **Preference and belief.** To model preference and belief, we can introduce two orderings to the model, one for preference relation, one for plausibility relation, between alternatives. From there, we can consider changes in preference and beliefs within communities, again extending the existing framework on preference change and belief revision, e.g., [16], [13], [5] and [9].

2. **Dominance.** To model *asymmetric* social relations, we can replace friendship relation in the model with a new preorder S (for 'is subordinate to') between agents, and investigate, for instance, what the paradox of mutual subordination in Section 1.7 means to us within communities.
3. **Aggregation.** As agents are modelled explicitly, we can easily add groups of agents by imposing an algebra on the set of A, such as a semilattice $\sqcup$ whose atoms are interpreted as individuals. Different aggregation procedures can be defined in terms of the structure of the social network, see studies in this line [2], [7] and [8].

References

1. Alchourrón, C., Gärdenfors, P., Makinson, D.: On the logic of theory change: Partial meet contraction and revision functions. Journal of Symbolic Logic 50, 510–530 (1985)
2. Andréka, H., Ryan, M., Schobbens, P.-Y.: Operators and laws for combining preferential relations. Journal of Logic and Computation 12, 12–53 (2002)
3. Areces, C., ten Cate, B.: Hybrid logics. In: van Benthem, J., Blackburn, P., Wolter, F. (eds.) Handbook of modal logic. Elsevier, Amsterdam (2006)
4. Baltag, A., Moss, L.S., Solecki, S.: The logic of public announcements, common knowledge and private suspicions. In: TARK 1998 (1998)
5. Baltag, A., Smets, S.: A qualitative theory of dynamic interactive belief revision. In: van der Hoek, W., Bonanno, G., Wooldridge, M. (eds.) Logic and the Foundations of Game and Decision Theory. Texts in Logic and Games, vol. 3. Amsterdam University Press (2008)
6. Fagin, R., Halpern, J.Y., Moses, Y., Vardi, M.Y.: Reasoning about Knowledge. The MIT Press, Cambridge (1995)
7. Girard, P.: Modal Logics for Belief and Preference Change. PhD thesis, Stanford University (2008)
8. Girard, P., Seligman, J.: An analytic logic of aggregation. In: Ramanujam, R., Sarukkai, S. (eds.) Logic and Its Applications. LNCS (LNAI), vol. 5378, pp. 146–161. Springer, Heidelberg (2009)
9. Grune-Yanoff, T., Hansson, S.O.: Preference Change: Approaches from Philosophy, Economics and Psychology. In: Theory and Decision Library. Springer, Heidelberg (2009)
10. Hansson, S.O.: Preference logic. In: Gabbay, D., Guenthner, F. (eds.) Handbook of Philosophical Logic, ch. 4, vol. 4, pp. 319–393. Kluwer, Dordrecht (2001)
11. Harel, D., Kozen, D., Tiuryn, J.: Dynamic Logic. The MIT Press, Cambridge (2000)
12. Hintikka, J.: Knowledge and Belief. Cornell University Press, Ithaca (1962)
13. Liu, F.: Changing for the Better: Preference Dynamics and Agent Diversity. PhD thesis, ILLC, University of Amsterdam (2008)
14. Parikh, R.: Knowledge, games and tales from the east. In: Ramanujam, R., Sarukkai, S. (eds.) Logic and Its Applications. LNCS (LNAI), vol. 5378, pp. 65–76. Springer, Heidelberg (2009)
15. Plaza, J.A.: Logics of public announcements. In: Proceedings of the 4th International Symposium on Methodologies for Intelligent Systems (1989)

16. van Benthem, J.: Dynamic logic for belief revision. Journal of Applied Non-classical Logic 17(2), 129–155 (2007)
17. van Benthem, J.: Logical dynamics of information and interaction. Cambridge University Press, Cambridge (to appear, 2010)
18. van Ditmarsch, H., van der Hoek, W., Kooi, B.: Dynamic Epistemic Logic. Springer, Berlin (2007)
19. von Wright, G.H.: The Logic of Preference, Edinburgh (1963)

Reasoning about Protocol Change and Knowledge

Yanjing Wang

Department of Philosophy, Peking University, Beijing 100871, China

Abstract. In social interactions, protocols govern our behaviour and assign meaning to actions. In this paper, we investigate the dynamics of protocols and their epistemic effects. We develop two logics, inspired by Propositional Dynamic Logic (PDL) and Public Announcement Logic (PAL), for reasoning about protocol change and knowledge updates. We show that these two logics can be translated back to the standard PDL and PAL respectively.

1 Introduction

Protocols are the rules that govern the actions of humans or machines. They have two major functions in our everyday life. First of all, protocols regulate behaviours and thus let us (or machines) know what to do or what not to do. For example, when you are driving a car you are also driving according to various traffic protocols; in case an accident happens legal protocols are called into play; while you are sending emails or SMS messages to a friend to complain about the bad luck, communication protocols on computers are running to make sure the messages are delivered. Second, protocols assign meaning to actions. For example, we are educated to be polite by following social protocols such as *shaking hands*. The handshaking action itself does not mean anything, it is the conventional protocol: "**if** you want to say hello formally and politely **then** shake hands" which lets this simple action carry some extra information. We can also create a new meaning for an action by setting a new protocol. For example, by popularizing the slogan "Love her, take her to Häagen-Dazs!" in China, the ice cream company Häagen-Dazs managed to let many young Chinese couples believe that buying an ice cream shows their love, no matter what love actually means. Because of the existence of such protocols we save our civilization from chaos and make it more meaningful everyday. Without doubt, protocols rule the world.

Already from the Häagen-Dazs example, we can see that it is important to understand how we can "install" new protocols to people. More generally, we are interested in the changes of protocols and their epistemic effects. Here we give some more examples of dynamics of protocols in social interactions. Imagine that you were told to close the door and on your way to do it you are told again not to close it. Now what to do? As another example, a well-trained spokesman may

M. Banerjee and A. Seth (Eds.): ICLA 2011, LNAI 6521, pp. 189–203, 2011.

respond to a yes-no question (viewed as a protocol that says: answer 'Yes!' **or** answer 'No!') by inserting yet another protocol: "before answering your question, tell me what you meant by ϕ." Here is a more interesting example that involves meaning changing: The Chinese are *non-confrontational* in the sense that they will not overtly say "No.", instead they say "I will think about it." or "We will see.". For a western businessman, "We will see.", according to the standard interpretation, means it is still possible. However, if he is updated with the Chinese protocol: say "We will see." **if** you want to say "No.", then he should understand this is just another way of saying "No.". Clearly the difference in protocols is sometimes the reason behind many conflicts and misunderstandings. Such phenomena invite a formal investigation on the question "*how* to change protocols and *update* your knowledge?". This paper reports an attempt in addressing this very question.

Our approach is inspired by two well-known logics: Propositional Dynamic Logic (PDL [7]) and Public Announcement Logic (PAL [20,9]). First of all, as we have seen above, the protocols usually have program-like structures, which suggests a formalization of protocols by regular expressions as in PDL. In the pioneering work of [10] and [6], where protocols involving knowledge were studied, protocols were indeed treated as simple programs in the form of $?\phi \cdot a$ (**if** ϕ holds **then** do a). Here we also need such tests $?\phi$ together with regular expressions to encode how protocols assign meaning to actions, e.g., $?p_{love} \cdot a_{buy}$ for our Häagen-Dazs example. On the other hand, the simplest protocol changing action might be a *public protocol announcement* and it is useful, e.g., a public announcement of what the Chinese means by using "We will see." could solve many problems in advance. Differing from the public announcements $!\phi$ in PAL, an announcement of a protocol may not have truth values, instead it changes the set of possible (sequences of) actions in addition to the inherent restrictions of actions according to the model. Putting the protocol announcements $[!\pi]$ (π is a regular expression) together with program modalities $[\pi]$ as in PDL, we may express that "Although b is possible according to the current protocol, after the announcement of the new protocol $a \cdot b$, we can not execute b as the first event any more" by the formula $\langle b\rangle\top \wedge [!(a \cdot b)]\neg\langle b\rangle\top$[1]. The more interesting case is when knowledge (expressed by $K_i\phi$) comes in. For instance, in our we-will-see story we would like $[a_{\text{will-see}}]K_i p_{no}$ to be not valid while $[!(?p_{no} \cdot a_{\text{will-see}})][a_{\text{will-see}}]K_i p_{no}$ to be valid. We will define the formal semantics for such an enriched epistemic language in this paper.

The contributions of this paper can be summarized as follows:

– We introduce various protocol announcement operators to PDL-based logics.
– The new logics can be used to reason about: 1. the executable actions according to the current protocol, and 2. the information that the actions carry, thus formally capturing the two functions of protocols.
– Epistemic reasoning in presence of unplanned protocol changes is facilitated.
– New protocol changing operators do not drive the logics beyond PDL or PAL.

[1] Here we assume that if a protocol is announced then it is followed by its executor.

Related work. Besides the ones we already talked about earlier, some more related work should be mentioned here. Process logic [21,11] extends PDL in adding modalities to specify *progressive* behaviours like "during the execution of program π, ϕ will be true at some point." In this paper, we not only reason about properties in the middle of an execution of a protocol but also handle the protocol changes during the execution. The later feature also distinguishes our work from the work using regular expressions as protocols [2,17,26]. Moreover, the semantics of our logics will be defined on the states in the models, instead of on paths as in [21,11]. Aucher [1] also proposed an extended Dynamic Epistemic Logic (DEL), where the reasoning of the ongoing events is facilitated, however, in a setting without protocols. Unlike the work of switching strategies in the context of games [19], the change of our protocols can be made at any time unplanned and also we incorporate knowledge in the discussions.

The existing work on protocols in DEL enriches the epistemic models with explicit protocols (sets of sequences of DEL events) such that the possible behaviours of agents are not only restricted by the inherent preconditions of epistemic events but also by protocol information [13,22,12]. This is similar to the treatment of protocols in Epistemic Temporal Logic (ETL) [10,18], where the temporal development of a system is generated from an initial situation by a commonly known protocol[2]. In our work, the semantics of our languages with protocol announcements will be defined on *standard* Kripke models. The extra protocol information is only introduced by protocol announcements while evaluating a formula. Such an approach makes it possible to not only model the "installation" of the initial protocol explicitly but also to handle protocol changes during the execution of the current one.

Our treatment for the events that carry meaning is largely inspired by [18], in which the authors give a very general and elegant semantics for messages (events) according to the underlying protocol in the setting of ETL. Here we can explicitly express the protocols and their changes *in* the logical language. Note that in the standard PAL, the interpretations of announcements are fixed and implicitly assumed to be common knowledge, e.g., in PAL an announcement $!\phi$ is assumed to have an inherent meaning: ϕ is true. This is because the semantic objects (event models) are explicitly included in the syntax as in the general DEL framework. However, the same utterance (syntax) may carry different meanings (semantics) as we have seen in the we-will-see example. A closer look at public announcements should separate the utterances and their meanings, as we will demonstrate later in this work.

Structure of the paper. In this paper we develop two logics featuring protocol changing operators. As an appetizer, we start in Section 2 with the first logic $\mathtt{PDL}^!$, a version of test-free PDL equipped with protocol announcements $[!\pi]$. The semantics is given in a non-standard style by using *modes* of satisfaction relations [8,24]. It is shown that $[!\pi]$ and many other similar protocol announcement operators do not increase the expressive power of the logic. Section 3 extends the

[2] However, the framework in [22] can also handle the protocols which are not common known.

language $\mathtt{PDL}^!$ with knowledge operators and Boolean tests to handle the cases like the above we-will-see example where knowing a protocol gives meanings to actions. We show that this new logic, when interpreted on S5 models, is equally expressive as PAL. We conclude in Section 4 and point out future work.

2 Protocol Announcement Logic $\mathtt{PDL}^!$

The formulas of $\mathtt{PDL}^!$ are built from a set of basic proposition letters $\mathbf{P}$ and a finite set of atomic action symbols $\mathbf{\Sigma}$ as follows:

$$\begin{array}{lll}\phi & ::= & \top \mid p \mid \neg\phi \mid \phi\wedge\phi \mid [\pi]\phi \mid [!\pi]\phi \\ \pi & ::= & \mathbf{1} \mid \mathbf{0} \mid a \mid \pi\cdot\pi \mid \pi+\pi \mid \pi^* \end{array}$$

where $p \in \mathbf{P}$ and $a \in \mathbf{\Sigma}$. Note that π are actually *regular expressions* based on $\mathbf{\Sigma}$ (we denote the set of such regular expressions as $Reg_{\mathbf{\Sigma}}$). The intended meaning of the formulas is mostly as in PDL, but "in context" of the protocol constraints: $[\pi]\phi$ now says that "after any run of the program π which is allowed by the *current protocol*, ϕ holds". The new formula $[!\pi]\phi$ expresses "after the announcement of the new protocol π, ϕ holds".

To give the semantics to $\mathtt{PDL}^!$, we first recall some basic facts about regular expressions.

The language of a regular expression π is defined as follows:

$$\begin{array}{l}\mathcal{L}(\mathbf{0})=\emptyset \qquad \mathcal{L}(\mathbf{1})=\{\epsilon\} \qquad \mathcal{L}(a)=\{a\} \\ \mathcal{L}(\pi\cdot\pi')=\{wv \mid w\in\mathcal{L}(\pi), v\in\mathcal{L}(\pi')\} \\ \mathcal{L}(\pi+\pi')=\mathcal{L}(\pi)\cup\mathcal{L}(\pi') \\ \mathcal{L}(\pi^*)=\{\epsilon\}\cup\bigcup_{n>0}(\mathcal{L}(\pi^n))\end{array}$$

where ϵ is the 'skip' protocol (empty sequence) and $\pi^n = \underbrace{\pi\cdots\pi}_{n}$.

The language of the *input derivative* $\pi\backslash a$ of a regular expression $\pi \in Reg_{\mathbf{\Sigma}}$ is defined as $\mathcal{L}(\pi\backslash a) = \{v \mid av \in \mathcal{L}(\pi)\}$. With the output function $o : Reg_{\mathbf{\Sigma}} \to \{\mathbf{0}, \mathbf{1}\}$ we can axiomatize the operation $\backslash a$ (cf. [4,5]):

$$\begin{array}{ll}\pi = o(\pi)+\sum_{a\in\mathbf{\Sigma}}(a\cdot\pi\backslash a) & \\ \mathbf{1}\backslash a=\mathbf{0}\backslash a=b\backslash a=\mathbf{0} \qquad (a\neq b) & a\backslash a=\mathbf{1} \\ (\pi\cdot\pi')\backslash a=(\pi\backslash a)\cdot\pi'+o(\pi)\cdot(\pi'\backslash a) & (\pi+\pi')\backslash a=\pi\backslash a+\pi'\backslash a \\ (\pi)^*\backslash a=\pi\backslash a\cdot(\pi)^* & o(\pi\cdot\pi)=o(\pi)\cdot o(\pi') \\ o(\pi^*)=\mathbf{1} & o(\mathbf{1})=\mathbf{1} \\ o(\mathbf{0})=o(a)=\mathbf{0} & o(\pi+\pi')=o(\pi)+o(\pi')\end{array}$$

Given $w = a_0a_1\cdots a_n \in \mathbf{\Sigma}^*$, let $\pi\backslash w = (\pi\backslash a_0)\backslash a_1\cdots\backslash a_n$. It is clear that $\pi\backslash w = \{v \mid wv \in \mathcal{L}(\pi)\}$[3]. Together with the axioms of Kleene algebra [14] we can syntactically derive $\pi\backslash w$ which is intuitively the *remaining* protocol of π after executing a run w. For example:

$$(a+(b\cdot c))^*\backslash b = (a\backslash b+(b\cdot c)\backslash b)\cdot(a+b\cdot c)^* = (\mathbf{0}+(\mathbf{1}\cdot c))\cdot(a+b\cdot c)^* = c\cdot(a+(b\cdot c))^*$$

[3] $\pi\backslash w$ is also a regular language cf. [5].

More generally, we can define $\mathcal{L}(\pi\backslash\pi') = \{v \mid \exists w \in \mathcal{L}(\pi') \text{ such that } wv \in \mathcal{L}(\pi)\}$ (cf. [5]). We say $w \in \mathbf{\Sigma}^*$ is *compliant with* π (notation: $w \propto \pi$) if $\pi\backslash w \neq \mathbf{0}$, namely, executing w is allowed by the protocol π.

Intuitively, to evaluate $[\pi]\phi$ we need to memorize the current protocol in some way. Here we employ a trick similar to the ones used in the semantics developed in [8,24,3]: we define the satisfaction relation w.r.t. a *mode* π (notation: $\vDash_\pi$), which is used to record the current protocol. Given the current protocol π, the allowed runs in a program π' w.r.t. π are those $w \in \mathbf{\Sigma}^*$ such that $w \in \mathcal{L}(\pi')$ and $w \propto \pi$. Note that if the current protocol is π, then after executing a run w we have to update π by the remaining protocol $\pi\backslash w$.

As in the standard `PDL`, we interpret `PDL`$^!$ formulas (w.r.t. $\mathbf{P}$,$\mathbf{\Sigma}$) on Kripke models $\mathcal{M} = (S, \to, V)$ where S is a non-empty set of states, $\to \subseteq S \times \mathbf{\Sigma} \times S$ is a set of labelled transitions, and the valuation function $V : S \to 2^{\mathbf{P}}$ assigns to each state a set of basic propositions. Now we are ready to give the semantics as follows:

$$
\boxed{
\begin{array}{rcl}
\mathcal{M}, s \vDash \phi & \Leftrightarrow & \mathcal{M}, s \vDash_{\mathbf{\Sigma}^*} \phi \\
\mathcal{M}, s \vDash_\pi p & \Leftrightarrow & p \in V(s) \\
\mathcal{M}, s \vDash_\pi \neg\phi & \Leftrightarrow & \mathcal{M}, s \nvDash_\pi \phi \\
\mathcal{M}, s \vDash_\pi \phi \wedge \psi & \Leftrightarrow & \mathcal{M}, s \vDash_\pi \phi \text{ and } \mathcal{M}, s \vDash_\pi \psi \\
\mathcal{M}, s \vDash_\pi [\pi']\phi & \Leftrightarrow & \forall (w, s') : w \in \mathcal{L}(\pi'), w \propto \pi, \text{ and } s \xrightarrow{w} s' \implies \mathcal{M}, s' \vDash_{\pi\backslash w} \phi \\
\mathcal{M}, s \vDash_\pi [!\pi']\phi & \Leftrightarrow & \mathcal{M}, s \vDash \langle\pi'\rangle\top \implies \mathcal{M}, s \vDash_{\pi'} \phi
\end{array}
}
$$

where the mode $\mathbf{\Sigma}^*$ stands for the *universal protocol* $(a_0 + a_1 + \cdots + a_n)^*$ if the set of atomic actions $\mathbf{\Sigma}$ is $\{a_0, a_1, \ldots, a_n\}$. The first clause says that initially everything is allowed and the last one says that the newly announced protocol overrides the current one. $[\pi']\phi$ is true w.r.t. the current protocol π iff on each s' that is reachable from s by some run w of π' which is allowed by the current protocol π: ϕ holds w.r.t. the remaining protocol $\pi\backslash w$. Note that it is important to remember w which denotes *how you get to* s' as the following example shows:

Example 1. Consider the following model $\mathcal{M}$:

$$
\begin{tikzcd}
s \arrow[r, dashed, "a"] & \bullet \arrow[r, dashed, bend left, "c"] \arrow[r, dashed, bend right, "d"'] & \bullet
\end{tikzcd}
$$

It can be verified that:

$$\mathcal{M}, s \vDash [!(a \cdot c + b \cdot d)][a + b](\neg\langle d\rangle\top \wedge \langle c\rangle\top \wedge [!(c + d)]\langle d\rangle\top)$$

The intuition behind is as follows. After announcing the protocol $a \cdot c + b \cdot d$, the program $a + b$ can be executed according to this protocol, but actually only a can be executed on the model. Thus after executing $a + b$ only c is possible according to the remaining protocol $(a \cdot c + b \cdot d)\backslash a = c$. However, if we then announce a new protocol $(c + d)$ then d also becomes available. ♣

Recall the standard `PDL` semantics, it is not hard to see that the following proposition holds.

Proposition 1. *For any test-free* PDL *formula ϕ and any pointed Kripke model $(\mathcal{M}, s)$:*

$$\mathcal{M}, s \vDash_{\mathtt{PDL}^!} \phi \iff \mathcal{M}, s \vDash_{\mathtt{PDL}} \phi$$

A natural question to ask is that whether $\mathtt{PDL}^!$ is more expressive than test-free PDL. To answer the question, we now have a closer look at the strings w in the semantics of $[\pi']\phi$. Given π, let $C_{\mathcal{L}(\pi)}$ be the set of all the *pre-sequences* of π: $\{w \mid w \propto \pi\}$.

We first show that we can partition $C_{\mathcal{L}(\pi)}$ into finitely many regular expressions satisfying certain properties.

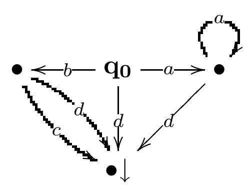
Lemma 1. *For any regular expression π there is a minimal natural number k such that $C_{\mathcal{L}(\pi)}$ can be finitely partitioned into $\pi_0, \dots, \pi_k$ and for any $w, v \in \mathcal{L}(\pi_i) : \pi \backslash w = \pi \backslash v$.*

Proof. By Kleene's theorem we can construct a deterministic finite automaton (DFA) recognizing the language of π. It is well known that DFA can be minimized, thus we obtain a smallest DFA that recognizes $\mathcal{L}(\pi)$:

$$\mathsf{A}_\pi = (\{q_0, \dots, q_k\}, \mathbf{\Sigma}, q_0, \rightarrowtail, F)$$

where $\{q_0, \dots, q_k\}$ is a set of states with q_0 being the start state and a subset F being the set of accept states. For each $i \leq k$ such that q_i can reach a state in F: we let π_i be the regular expression corresponding to the automaton $(\{q_0, \dots, q_k\}, \mathbf{\Sigma}, q_0, \rightarrowtail, \{q_i\})$. Since A_π is deterministic, it is not hard to see that these π_i form a partition that we want. $\square$

In the sequel, we call the above unique partition $\pi_0, \dots, \pi_k$ the *pre-derivatives* of π. For example, the minimal deterministic automaton[4] of $a^* \cdot d + b \cdot (c + d)$ is:



thus the pre-derivatives of $a^* \cdot d + b \cdot (c + d)$ are $\mathbf{1}$, $a \cdot a^*, b, a^* \cdot d + b \cdot (c + d)$.

Now we define the following translation from $\mathtt{PDL}^!$ to the test-free PDL:

$$\begin{array}{rcl}
t(\phi) & = & t_{\mathbf{\Sigma}^*}(\phi) \\
t_\pi(p) & = & p \\
t_\pi(\neg\phi) & = & \neg t_\pi(\phi) \\
t_\pi(\phi_1 \wedge \phi_2) & = & t_\pi(\phi_1) \wedge t_\pi(\phi_2) \\
t_\pi([\pi']\phi) & = & \bigwedge_{i=0}^{k}([\theta_i] t_{\pi\backslash\pi_i}(\phi)) \\
t_\pi([!\pi']\phi) & = & \langle\pi'\rangle\top \to t_{\pi'}(\phi)
\end{array}$$

where $\pi_0, \dots, \pi_k$ are the pre-derivatives of π, θ_i is a regular expression corresponding to $\mathcal{L}(\pi') \cap \mathcal{L}(\pi_i)$, and $\pi\backslash\pi_i = \pi\backslash w$ for any $w \in \mathcal{L}(\pi_i)$ due to Lemma 1.

[4] We omit the transitions to the *"trash"* state which can not reach any accept state.

By this translation we can eliminate the $[!\pi]$ operator in $\mathtt{PDL}^!$ and thus showing that $\mathtt{PDL}^!$ and the test-free $\mathtt{PDL}$ are equally expressive.

Theorem 1. *For any pointed Kripke model* $\mathcal{M}, s$:

$$\mathcal{M}, s \vDash_{\mathtt{PDL}^!} \phi \iff \mathcal{M}, s \vDash_{\mathtt{PDL}} t(\phi).$$

Proof. By induction on ϕ we can show: $\mathcal{M}, s \vDash_\pi \phi \iff \mathcal{M}, s \vDash_{\mathtt{PDL}} t_\pi(\phi)$. The only non-trival case is for $[\pi']\phi$:

$\mathcal{M}, s \vDash_\pi [\pi']\phi$

$\iff \forall(w, s') : w \in \mathcal{L}(\pi'), w \propto \pi,$ and $s \xrightarrow{w} s' \implies \mathcal{M}, s' \vDash_{\pi \backslash w} \phi$

$\iff \forall(w, s')$: if there is a pre-derivative $\pi_i : w \in \mathcal{L}(\pi'), w \in \mathcal{L}(\pi_i),$ and $s \xrightarrow{w} s'$ then $\mathcal{M}, s' \vDash_{\pi \backslash w} \phi$

$\iff$ for all pre-derivatives $\pi_i,$ for all s' : if there is a $w \in \mathcal{L}(\pi') \cap \mathcal{L}(\pi_i)$ and $s \xrightarrow{w} s'$ then $\mathcal{M}, s' \vDash_{\pi \backslash w} \phi$

$\iff \mathcal{M}, s \vDash \bigwedge_{i=0}^{k} [\theta_i] t_{\pi \backslash \pi_i}(\phi)$

□

Discussion. In this section, we take a rather *liberal* view on the "default" protocol, namely we assume that everything is allowed initially. On the other hand, we can well start with a *conservative* initialization where nothing is allowed unless announced later. It is not hard to see that we can also translate this conservative version of $\mathtt{PDL}^!$ to $\mathtt{PDL}$ if we let $t(\phi) = t_{\mathbf{1}}(\phi)$ where $\mathbf{1}$ is the constant for empty sequence i.e., the *skip* protocol. For example, $t_{\mathbf{1}}([a]\bot \wedge [!a]\langle a + b\rangle\top) = [\mathbf{0}]\bot \wedge t_a(\langle a + b\rangle\top) = \langle a\rangle\top$.

Moreover, $[!\pi]$ is rather *radical* in the sense that it changes the protocol completely. We may define a more general operation as follows: Let $\pi(x) \in Reg_{\mathbf{\Sigma} \cup \{x\}}$, namely, $\pi(x)$ is a regular expression with a variable x. Now we define:

$$\boxed{\mathcal{M}, s \vDash_\pi [!\pi'(x)]\phi \Leftrightarrow (\mathcal{M}, s \vDash \langle \pi'(\pi)\rangle\top \implies \mathcal{M}, s \vDash_{\pi'(\pi)} \phi)}$$

We can then concatenate, add, insert and repeat protocols by announcing $x \cdot \pi'$, $x{+}\pi'$, $\pi'{+}x$, and x^* respectively. It is easy to see that the announcement operator $[!\pi]$ introduced previously is a special case of $[!\pi(x)]$. We can still translate the logic with the generalized protocol announcements to $\mathtt{PDL}$ with an easy revision of the translation:

$$t_\pi([!\pi'(x)]\phi) = \langle \pi'(\pi)\rangle\top \rightarrow t_{\pi'(\pi)}(\phi)$$

Similarly, without adding the expressive power, we can introduce a *refinement operator* $[!(a/\pi')]$ for each $a \in \mathbf{\Sigma}$ with the following semantics:

$$\boxed{\mathcal{M}, s \vDash_\pi [!(a/\pi')]\phi \Leftrightarrow \mathcal{M}, s \vDash \langle \pi[a/\pi']\rangle\top \implies \mathcal{M}, s \vDash_{\pi[a/\pi']} \phi}$$

where $\pi[a/\pi']$ is the regular expression obtained by substituting each a in π with π'. Intuitively the operator $[!(a/\pi')]$ refines the current protocol by making the atomic step a more complicated.

3 Public Event Logic $\mathtt{PDL}^{!?_b}$

In this section, we allow tests in the protocol announcements and study how agents update their knowledge according to the protocols and their observations of the public events. We shall see that by announcing a protocol with tests, we can let actions carry propositional information as motivated in the introduction.

Given a finite set $\mathbf{P}$ of basic propositions, a finite set $\mathbf{\Sigma}$ of atomic actions and a set $\mathbf{I}$ of agents, the language of $\mathtt{PDL}^{!?_b}$ is defined as follows:

$$\begin{array}{lcl} \phi & ::= & \top \mid p \mid \neg\phi \mid \phi\wedge\phi \mid [\pi']\phi \mid [!\pi]\phi \mid K_i\phi \\ \pi & ::= & ?\phi_b \mid a \mid \pi\cdot\pi \mid \pi+\pi \mid \pi^* \end{array}$$

where $i \in \mathbf{I}$, ϕ_b are Boolean formulas based on $\mathbf{P}$, and π' are a test-free regular expressions. Note that we do not include $\mathbf{1}$ and $\mathbf{0}$ as atomic actions in π since they can be expressed by the Boolean tests $?\top$ and $?\bot$. We call the programs with possibly Boolean tests *guarded regular expressions.*

In this section, we assume that all the $a \in \mathbf{\Sigma}$ are *public events* which can be observed by all the agents, while the tests, unless announced, are not observable to the agents. Here $[\pi']\phi$ is intended to express that "ϕ holds after the agents observe any sequence of public events which is not only allowed in π' but also complaint with the current protocol." Therefore only test-free programs are considered in the modality $[\pi']$, since the tests can not be observed anyway. Now we can express the Häagen-Dazs slogan mentioned in the introduction by the protocol: $\pi_{H\text{-}D} = ?p_{love}\cdot a_{buy}$. A suitable semantics should let $[!\pi_{H\text{-}D}][a_{buy}]K_ip_{love}$ be valid. However, without the announcement $!\pi_{H\text{-}D}$, buying an ice cream does not mean anything: $[a_{buy}]K_ip_{love}$ should not be valid.

To prepare ourselves for the definition of the semantics, we first interpret guarded regular expressions as the languages of *guarded strings* following the definitions in [15]. A *(uniform) guarded string* over finite sets $\mathbf{P}$ and $\mathbf{\Sigma}$ is a sequence $\rho a_1\rho a_2\rho\ldots\rho a_n\rho$ where $a_i \in \mathbf{\Sigma}$ and $\rho\subseteq\mathbf{P}$ representing the valuations of basic propositions in $\mathbf{P}$ ($p\in\rho$ iff p is true according to ρ as a valuation). For any $\rho\subseteq\mathbf{P}$, let ϕ_ρ be the characteristic formula $\phi_\rho = \bigwedge_{p\in\rho}p\wedge\bigwedge_{p\in\mathbf{P}-\rho}\neg p$. On the other hand, for any Boolean formula ψ, let $X_\psi\subseteq 2^{\mathbf{P}}$ be the corresponding set of valuations, represented by subsets of $\mathbf{P}$, that make ψ true.

Now we can define the language of guarded strings associated with a guarded regular expression over $\mathbf{\Sigma}$ and $\mathbf{P}$:

$$\begin{array}{l} \mathcal{L}_g(a) = \{\rho a\rho \mid \rho\subseteq\mathbf{P}\} \\ \mathcal{L}_g(?\psi) = \{\rho \mid \rho\in X_\psi\} \\ \mathcal{L}_g(\pi_1\cdot\pi_2) = \{w\diamond v \mid w\in\mathcal{L}_g(\pi_1), v\in\mathcal{L}_g(\pi_2)\} \\ \mathcal{L}_g(\pi_1+\pi_2) = \mathcal{L}_g(\pi_1)\cup\mathcal{L}_g(\pi_2) \\ \mathcal{L}_g(\pi^*) = \{\epsilon\}\cup\bigcup_{n>0}(\mathcal{L}_g(\pi^n)) \end{array}$$

where $\pi^n = \underbrace{\pi\cdots\pi}_{n}$, and $\diamond$ is the *fusion product*: $w\diamond v = w'\rho v'$ when $w = w'\rho$ and $v = \rho v'$.

We write $\pi_1 \equiv_g \pi_2$ if $\mathcal{L}_g(\pi_1) = \mathcal{L}_g(\pi_2)$. For example, we have:

$$?p \cdot ?q \cdot a \equiv_g ?(p \wedge q) \cdot a \equiv_g ?(p \wedge p) \cdot ?q \cdot a$$

$$?(p \wedge q) \cdot a + ?(p \wedge \neg q) \cdot a \equiv_g ?p \cdot a \text{ and } ?p \cdot a \cdot a \not\equiv_g ?p \cdot a.$$

We now define the language of input derivative $\pi\backslash w$ for a guarded string w as:

$$\mathcal{L}_g(\pi\backslash w) = \{v \mid w \diamond v \in \mathcal{L}_g(\pi)\}$$

and we say $w \propto_g \pi$ if $\mathcal{L}_g(\pi\backslash w) \neq \emptyset$. As in the previous section, we let $C_{\mathcal{L}_g(\pi)} = \{w \mid w \propto_g \pi\}$ and let $\mathcal{L}_g(\pi_1\backslash\pi_2) = \{v \mid \exists w \in \mathcal{L}_g(\pi_2) \text{ and } w \diamond v \in \mathcal{L}_g(\pi_1)\}$. For example, if p is the only proposition letter then $\mathcal{L}_g((?p{\cdot}a{\cdot}b{\cdot}b + ?\neg p{\cdot}a{\cdot}b{\cdot}c)\backslash(a{\cdot}b)) = \{\{p\}b\{p\}, \emptyset c\emptyset\} = \mathcal{L}_g(?p \cdot b + ?\neg p \cdot c)$.

Before defining the semantics of $\mathtt{PDL}^{!?_b}$ formally, let us recall the semantics of PAL (cf. e.g., [20,9]). Given a set of agents $\mathbf{I}$, the language of PAL extends the propositional logic with the standard knowledge operators K_i for each $i \in \mathbf{I}$ and propositional announcement operators $[!\psi]$ with the following semantics based on S5 Kripke models $(S, \{\sim_i\}_{i \in \mathbf{I}}, V)$[5]:

$$\boxed{\begin{array}{l} \mathcal{M}, s \vDash K_i\phi \Leftrightarrow \text{for all } t, \text{ if } s \sim_i t \text{ then } \mathcal{M}, t \vDash \phi \\ \mathcal{M}, s \vDash [!\psi]\phi \Leftrightarrow \text{ if } \mathcal{M}, s \vDash \psi \text{ then } \mathcal{M}|_\psi, s \vDash \phi \end{array}}$$

where $\mathcal{M}|_\psi = (S', \{\sim'_i\}_{i \in \mathbf{I}}, V')$ with $S' = \{s \in S \mid \mathcal{M}, s \vDash \psi\}$; $\sim'_i = \sim_i \cap (S' \times S')$; and $V' = V|_{S'}$ (i.e. the restriction of $V_\mathcal{M}$ on the domain S'). Intuitively the effect of announcing a formula ψ is to restrict the model to the ψ-worlds. In our setting, observing a sequence of public events is similar to hearing a sequence of public announcements, but the propositional information carried by the public events are given by the previously announced protocols, not by the syntactic forms of the public events. What we need to do in the semantics is to let agents match the protocol knowledge with their observations and find out what tests have been done when executing the protocol so far. According to the information about the tests, the agents can eliminate some possible worlds as in PAL.

Given a sequence v of atomic actions and Boolean tests, let $\mathcal{L}_p(v)$ be the subsequence of v obtained by ignoring all the tests but keeping all the public events $a_0 \ldots a_k$, e.g., $\mathcal{L}_p(?p \cdot a \cdot b) = \mathcal{L}_p(?q \cdot a \cdot ?p \cdot b) = a \cdot b$. Now we interpret $\mathtt{PDL}^{!?_b}$ on the $S5$ models $(S, \{\sim_i\}_{i \in \mathbf{I}}, V)$ as follows:

$$\boxed{\begin{array}{rcl} \mathcal{M}, s \vDash \phi & \Leftrightarrow & \mathcal{M}, s \vDash_{\mathbf{\Sigma}^*} \phi \\ \mathcal{M}, s \vDash_\pi p & \Leftrightarrow & p \in V(s) \\ \mathcal{M}, s \vDash_\pi \neg\phi & \Leftrightarrow & \mathcal{M}, s \nvDash_\pi \phi \\ \mathcal{M}, s \vDash_\pi \phi \wedge \phi' & \Leftrightarrow & \mathcal{M}, s \vDash_\pi \phi \text{ and } \mathcal{M}, s \vDash_\pi \phi' \\ \mathcal{M}, s \vDash K_i\phi & \Leftrightarrow & \text{for all } t\text{: } s \sim_i t \implies \mathcal{M}, t \vDash \phi \\ \mathcal{M}, s \vDash_\pi [\pi']\phi & \Leftrightarrow & \text{for all } w \in \mathcal{L}(\pi') : \mathcal{M}, s \vDash \phi^w_\pi \implies \mathcal{M}|_{\phi^w_\pi}, s \vDash_{\pi\backslash w} \phi \\ \mathcal{M}, s \vDash_\pi [!\pi']\phi & \Leftrightarrow & (\exists w : w = \rho v \in \mathcal{L}_g(\pi') \text{ and } V(s) = \rho) \implies \mathcal{M}, s \vDash_{\pi'} \phi \end{array}}$$

[5] An S5 model is a Kripke model where the relations $\sim_i$ are equivalence relations.

where:

$$\phi^w_\pi = \bigvee\{\phi_\rho \mid v = \rho a_1 \rho a_2 \rho \cdots \rho a_k \rho, \mathcal{L}_p(v) = w, v \propto_g \pi\}$$

Note that we do not include the transitions labelled by $a \in \mathbf{\Sigma}$ in the models since we assume that each public event is executable at each state unless it is not compliant with the current protocol (e.g., you can talk about anything in public unless constrained by some law or conventions). Since the public events are intended to be announcement-like events, we also assume that executing a protocol of such event does not result in changing the facts on the real state. This explains the uniformity of ρ in guarded strings.

Intuitively ϕ^w_π in the above clause for $[\pi']\phi$ is the propositional information agents can derive from observing $w \in \mathbf{\Sigma}^*$ in the context of protocol π. To see this, consider an observable sequence $w = a_1 a_2 \cdots a_k \in \mathbf{\Sigma}^*$, each $v = \rho a_1 \rho a_2 \rho \cdots \rho a_k \rho$ such that $v \propto_g \pi$ is a possible actual execution of the protocol consistent to the observation w. However, agents can not distinguish $v, v' \propto_g \pi$ if $\mathcal{L}_p(v) = \mathcal{L}_p(v') = w$. Therefore the disjunction ϕ^w_π is then the information which can be derived from the observation of w according to the protocol π. The intuition behind the last clause is that the new protocol is updated only if it is executable at the current pointed model.

Consider the Häagen-Dazs example, let $\mathcal{M}$ be a two-world model representing that a girl i does not know whether a boy loves her or not (she is not sure between a p_{love}-world s and a $\neg p_{love}$ world t). Let $\pi_0 =?p_{love} \cdot a_{buy}$. Note that $\mathcal{L}_p(\{p_{love}\}a_{buy}\{p_{love}\}) = \mathcal{L}_p(\emptyset a_{buy} \emptyset) = \{a_{buy}\}$, thus $\phi^{a_{buy}}_{\mathbf{\Sigma}^*} = p \vee \neg p = \top$. On the other hand $\phi^{a_{buy}}_{\pi_0}$ is clearly p_{love}. We now show $\mathcal{M}, s \nvDash [a_{buy}]K_i p_{love}$:

$$\begin{aligned}
& \mathcal{M}, s \vDash [a_{buy}]K_i p_{love} \\
\iff & \mathcal{M}, s \vDash_{\mathbf{\Sigma}^*} [a_{buy}]K_i p_{love} \\
\iff & \text{for all } w \in \mathcal{L}(a_{buy}), \mathcal{M}, s \vDash_{\mathbf{\Sigma}^*} \phi^w_{\mathbf{\Sigma}^*} \implies \mathcal{M}|_{\phi^w_{\mathbf{\Sigma}^*}}, s \vDash_{\mathbf{\Sigma}^* \setminus w} K_i p_{love} \\
\iff & \mathcal{M}, s \vDash \phi^{a_{buy}}_{\mathbf{\Sigma}^*} \implies \mathcal{M}|_{\phi^{a_{buy}}_{\mathbf{\Sigma}^*}}, s \vDash_{\mathbf{\Sigma}^* \setminus a_{buy}} K_i p_{love} \\
\iff & \mathcal{M}, s \vDash_{\mathbf{\Sigma}^*} K_i p_{love}
\end{aligned}$$

Since $s \sim_i t$ and $\mathcal{M}, t \vDash \neg p_{love}$ then $\mathcal{M}, s \nvDash [a_{buy}]K_i p_{love}$. On the other hand:

$$\begin{aligned}
& \mathcal{M}, s \vDash [!\pi_0][a_{buy}]K_i p_{love} \\
\iff & \mathcal{M}, s \vDash_{\pi_0} [a_{buy}]K_i p_{love} \\
\iff & \mathcal{M}, s \vDash_{\pi_0} \phi^{a_{buy}}_{\pi_0} \implies \mathcal{M}|_{\phi^{a_{buy}}_{\pi_0}}, s \vDash_{\pi_0 \setminus a_{buy}} K_i p_{love} \\
\iff & \mathcal{M}|_{p_{love}}, s \vDash_{?p_{love}} K_i p_{love}
\end{aligned}$$

It is clear that $\mathcal{M}, s \vDash [!\pi_0][a_{buy}]K_i p_{love}$.

Similarly, for the we-will-see scenario mentioned in the introduction, if $\mathcal{M}$ is a two-world model representing that a Westerner i does know whether p_{no} (state s) or $\neg p_{no}$ (state t) then we can show that:

$$\mathcal{M}, s \vDash [!(?\top \cdot a_{will-see})]([a_{will-see}]\neg K_i p_{no} \wedge [!(?p_{no} \cdot a_{will-see})][a_{will-see}]K_i p_{no})$$

where $?\top \cdot a_{will-see}$ is the default protocol a Westerner may have as the standard interpretation for the sentence "we will see" which does not carry any useful information. As a more complicated example, the reader can check the model validity of the following formula on a model where agent i is not sure about p:

$$[!(?p \cdot a \cdot b + ?\neg p \cdot a \cdot c)]([a]\neg(K_i p \vee K_i \neg p) \wedge [a \cdot (b + c)](Kp \vee K\neg p)).$$

Note that the semantics of $[\pi']\phi$ is very similar to the one for public announcement, but with protocol updates and a quantification over $w \in \mathcal{L}(\pi')$. It is not hard to see that each PAL formula with Boolean announcements only can be mimicked by a $\mathtt{PDL}^{!?_b}$ formula by setting the meaning of announcements by protocols at the beginning. For example $[!p](K_i p \wedge [!q]q)$ can be reinterpreted in $\mathtt{PDL}^{!?_b}$ as $[!(?p \cdot a + ?q \cdot b)^*][a](K_i p \wedge [b]q)$. In this way we separate an announcement as an action with its meaning.

In the rest of this section we will show that $\mathtt{PDL}^{!?_b}$ can be translated back to PAL. We will follow a similar strategy as in the previous section for the expressivity of $\mathtt{PDL}^{!}$. This time we need to use automata on guarded strings.

Given $\mathbf{P}$ let $\mathcal{B}(\mathbf{P})$ be the set $2^{2^{\mathbf{P}}}$. Intuitively, $X \in \mathcal{B}(\mathbf{P})$ represent Boolean formulas over $\mathbf{P}$. We denote the corresponding formula of $X \in \mathcal{B}(\mathbf{P})$ by ϕ_X. Based on the exposition in [15], we define the automata which recognize uniform guarded strings.

Definition 1. ***(Automata on guarded strings)*** *A finite automaton on (uniform) guarded strings (or simply guarded automaton) over a finite set of actions* $\mathbf{\Sigma}$ *and a finite set of atomic tests* $\mathbf{P}$ *is a tuple* $\mathsf{A} = (Q, \mathbf{\Sigma}, \mathbf{P}, q_0, \rightarrowtail, F)$ *where the transitions are labelled by atomic actions in* $\mathbf{\Sigma}$ *(action transitions) and sets* $X \in \mathcal{B}(\mathbf{P})$ *(test transitions).* A *accepts a finite string* w *over* $\mathbf{\Sigma} \cup \mathcal{B}(\mathbf{P})$ *(notation:* $w \in \mathcal{L}_{\mathbf{\Sigma} \cup \mathcal{B}(\mathbf{P})}(\mathsf{A})$*), if it accepts* w *as a standard finite automaton over label set* $\mathbf{\Sigma} \cup \mathcal{B}(\mathbf{P})$. *The acceptance for guarded strings is defined based on the acceptance of normal strings and the following transformation function* G *which takes a string over* $\mathbf{\Sigma} \cup \mathcal{B}(\mathbf{P})$ *and outputs a set of uniform guarded strings.*

$$\begin{aligned} G(a) &= \{\rho a \rho \mid \rho, \rho \subseteq \mathbf{P}\} \\ G(X) &= \{\rho \mid \rho \in X\} \\ G(ww') &= \{v\rho v' \mid v\rho \in G(w) \text{ and } \rho v' \in G(w')\} \end{aligned}$$

We say A *accepts a finite guarded string* $v : \rho a_0 \rho \ldots a_{k-1} \rho$ *over* $\mathbf{\Sigma}$ *and* $\mathbf{P}$, *if* $v \in G(w)$ *for some string* $w \in \mathcal{L}_{\mathbf{\Sigma} \cup \mathcal{B}(\mathbf{P})}(\mathsf{A})$. *Let* $\mathcal{L}_g(\mathsf{A})$ *be the language of guarded strings accepted by* A. ♠

We say a guarded automaton is *deterministic* if the following hold (cf. [15]):

- Each state is either a state that only has outgoing action transitions (*action state*) or a state that only has outgoing test transitions (*test state*).
- The outgoing action transitions are deterministic: for each action state q and each $a \in \mathbf{\Sigma}$, q has one and only one a-successor.
- The outgoing test transitions are deterministic: they are labelled by $\{\{\rho\} \mid \rho \subseteq \mathbf{P}\}$ and for each test state q and each ρ, q has one and only one $\{\rho\}$-successor. Clearly these tests ρ at a test state are logically pairwise exclusive and altogether exhaustive (viewing ρ as the Boolean formula ϕ_ρ).

- The start state is a test state and all accept states are action states.
- Each cycle contains at least one action transition.

The Kleene theorem between guarded automata and guarded regular expressions is proved in [15][6].

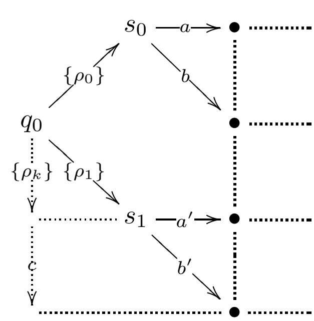
Theorem 2. *[15, Theorem 3.1, 3.4] For each guarded regular expression π over $\mathbf{P}$ and $\mathbf{\Sigma}$ there is a (deterministic) guarded automaton A over $\mathbf{P}$ and $\mathbf{\Sigma}$ such that $\mathcal{L}_g(\pi) = \mathcal{L}_g(\mathsf{A})$, and vice versa.*

Given a guarded regular expression π, we let $C_p(\pi) = \{\mathcal{L}_p(w) \mid w \in C_{\mathcal{L}_g(\pi)}\}$. Namely, $C_p(\pi)$ is the collection of all the possible observations of public events (no tests) according to π. Following the idea in the previous section, we need to finitely partition it.

Lemma 2. *Given a guarded regular expression π over $\mathbf{\Sigma}$ and $\mathbf{P}$, we can finitely partition $C_p(\pi)$ into test-free regular expressions $\pi_0, \ldots, \pi_n$ such that for any $i \leq n : w, v \in \mathcal{L}(\pi_i) \implies \phi_\pi^w = \phi_\pi^v$ and $\pi \backslash w = \pi \backslash v$.*

Proof. (Sketch) The strategy for the proof is as follows: we first partition $C_p(\pi)$ into $\pi_0, \ldots, \pi_k$ such that for any $i \leq k$, for any $w, v \in \mathcal{L}(\pi_i) \implies \phi_\pi^w = \phi_\pi^v$, then we further partition each π_i according to the shared derivatives like in the proof of Lemma 1.

From Theorem 2, we can build a deterministic guarded automaton A_π such that $\mathcal{L}_g(\mathsf{A}_\pi) = \mathcal{L}_g(\pi)$. Based on A_π it is easy to build a deterministic automaton A such that $\mathcal{L}_g(\mathsf{A}) = C_{\mathcal{L}_g(\pi)}$ by setting new accept states. From the definition of deterministic guarded automata, the start state in A has only test transitions labelled by $\{\rho\}$ for each $\rho \subseteq \mathbf{P}$. Since we only consider uniform guarded strings in the language, then an accepting path starting with a transition $\{\rho\}$ can never go through any other test transition labelled by $\{\rho'\}$ for any $\rho' \neq \rho$. Then we can prune and massage A into the following shape while keeping the accepting language intact:



[6] In [15], the author considered general guarded strings whose guards need not to be uniform throughout the string. Our definitions of the languages of the guarded regular expressions and the languages of guarded automata are essentially restrictions of the corresponding definitions of languages in [15] to uniform guarded strings. Therefore the Kleene theorem proved in [15] also applies to our restricted setting.

where $k = |2^{\mathbf{P}}|$, the start state q_0 is the only test state, and there is no incoming transition at q_0.

Let B_{s_i} be the standard finite automaton over the action set $\mathbf{\Sigma}$: $(Q_{act}, \mathbf{\Sigma}, s_i, \rightarrowtail, F)$ where Q_{act} is the set of action states in Q, F is the set of accept states in A, and $q \overset{a}{\rightarrowtail} q'$ in $\mathsf{B}_{s_i} \iff q \overset{a}{\rightarrowtail} q'$ in A. Given $Z \subseteq \{\rho_0, \ldots, \rho_k\}$ (intuitively a Boolean formula), let D_Z be the product automaton $\mathit{\Pi}_{\rho_i \in Z} \mathsf{B}_{s_i} \times \mathit{\Pi}_{\rho_i \notin Z} \overline{\mathsf{B}_{s_i}}$ where $\overline{\mathsf{B}_{s_i}}$ is the complement automaton of B_{s_i}. We can show that D_Z recognizes all the sequences w based on $\mathbf{\Sigma}$ such that $\{\rho \mid w = \mathcal{L}_p(v), v = \rho a_1 \rho \cdots \rho a_k \rho \in \mathcal{L}_g(\mathsf{A})\} = Z$. By Kleene Theorem, we can turn each D_Z into a regular expression.

Thus, we can finitely partitioned $C_p(\pi)$ into $\pi_0, \ldots, \pi_n$ such that for any $i \leq n$, for any $w, v \in \mathcal{L}(\pi_i) \implies \phi^w_\pi = \phi^v_\pi$. By the similar techniques as in the proof of Lemma 1, we can further partition each of these regular expressions π_i into finitely many regular expressions $\pi_{i0} \ldots \pi_{im}$ such that for any $w, v \in \mathcal{L}(\pi_{ij})$: $\pi \backslash w = \pi \backslash v$. This gives us the desired final partition. $\square$

Now we define the following translation from $\mathtt{PDL}^{!?_b}$ to $\mathtt{PAL}$:

$$
\begin{array}{lcl}
t(\phi) & = & t_{\mathbf{\Sigma}^*}(\phi) \\
t_\pi(p) & = & p \\
t_\pi(\neg\phi) & = & \neg t_\pi(\phi) \\
t_\pi(\phi_1 \wedge \phi_2) & = & t_\pi(\phi_1) \wedge t_\pi(\phi_2) \\
t_\pi(K_i \phi) & = & K_i t_\pi(\phi) \\
t_\pi([\pi']\phi) & = & \bigwedge\{[!\psi_j] t_{\theta_j}(\phi) \mid \mathcal{L}(\pi_j) \cap \mathcal{L}(\pi') \neq \emptyset\} \\
t_\pi([!\pi']\phi) & = & \chi_{\pi'} \rightarrow t_{\pi'}(\phi)
\end{array}
$$

where:

- all the π_j form a partition of $C_p(\pi)$ satisfying the desired properties stated in the above lemma,
- $\psi_j = \phi^w_\pi$ for some $w \in \mathcal{L}(\pi_j)$,
- $\theta_j = \pi \backslash w$ for some $w \in \mathcal{L}(\pi_j)$,
- $\chi_{\pi'} = \bigvee\{\phi_\rho \mid \rho v \in \mathcal{L}_g(\pi')$ for some sequnece $v\}$.

Note that by the properties of the partition, ψ_j and θ_j are well-defined. Intuitively, $\chi_{\pi'}$ is the precondition of π' to be executed. Based on the above translation t, it is not hard to prove:

Theorem 3. *For any pointed $S5$ Kripke model $\mathcal{M} = (S, \{\sim_i\}_{i \in \mathbf{I}}, V, s)$:*

$$\mathcal{M}, s \vDash_{\mathtt{PDL}^{!?_b}} \phi \iff \mathcal{M}, s \vDash_{\mathtt{PAL}} t(\phi).$$

Since $\mathtt{PAL}$ is equally expressive as the standard epistemic logic ($\mathtt{EL}$) without any announcement operators (cf. e.g., [9]) and $\mathtt{PDL}^{!?_b}$ is clearly at least as expressive as $\mathtt{EL}$ then $\mathtt{PAL}$ and $\mathtt{PDL}^{!?_b}$ are equally expressive.

4 Conclusions and Future Work

In this paper we proposed two $\mathtt{PDL}$-style logics with simple and natural operators for reasoning about protocol changes and knowledge updates: Logic $\mathtt{PDL}^{!}$ handles protocol changes in a context without knowledge; $\mathtt{PDL}^{!?_b}$ includes tests in the protocols and knowledge operators to deal with the situations where events carry information for agents according to the knowledge of the protocols. We showed that $\mathtt{PDL}^{!}$ is equally expressive as the test-free $\mathtt{PDL}$ and $\mathtt{PDL}^{!?_b}$ is equally expressive as $\mathtt{PAL}$. By using the new languages, what we gain is the explicitness and convenience in modelling scenarios with protocol changes and knowledge updates, as we demonstrated by various examples. In [25] we also investigated another closely related logic $\mathtt{PDL}^{\boxtimes}$, which extends the $\mathtt{DEL}$ framework with more general product update operations taking general guarded automata as update models. For interested readers who want to see more applications of the protocol changing operations, we refer to [27] where we integrated a similar protocol changing operator in a specific setting of communications over channels.

It is shown in [16] that the public announcement logic, though equally expressive as epistemic logic, is exponentially more succinct than the pure epistemic logic in expressing certain properties on K models. Here we conjecture that similar results apply to our new logics as well. However, we leave out the succinctness and complexity analysis for future work.

Also note that the logic $\mathtt{PDL}^{!?_b}$ is interpreted on S5 epistemic models which do not have action transitions, since we implicitly assume all the public events are like public communications by words which are always executable unless constrained by protocols. We can well consider more general models with action transitions or consider actions that can change the facts in models as discussed in [23]. Another restriction in $\mathtt{PDL}^{!?_b}$ is that we only consider Boolean tests for simplicity. To make the logic more interesting we would like to include epistemic tests in the future. Such extensions may essentially increase the expressive power of the logic.

Acknowledgements. The author would like to thank Ram Ramanujam and Hans van Ditmarsch for their insightful remarks on various versions of this work, and the anonymous referees for their useful comments.

References

1. Aucher, G.: BMS revisited. In: Proceedings of TARK 2009, pp. 24–33 (2009)
2. Benevides, M., Schechter, L.: A propositional dynamic logic for CCS programs. In: Hodges, W., de Queiroz, R. (eds.) WoLLIC 2008. LNCS (LNAI), vol. 5110, pp. 83–97. Springer, Heidelberg (2008)
3. Bonnay, D., Égré, P.: Inexact knowledge with introspection. Journal of Philosophical Logic 38(2), 179–227 (2009)
4. Brzozowski, J.A.: Derivatives of regular expressions. Journal of the ACM 11(4), 481–494 (1964)
5. Conway, J.H.: Regular Algebra and Finite Machines. Chapman and Hall, Boca Raton (1971)

6. Fagin, R., Halpern, J.Y., Moses, Y., Vardi, M.Y.: Knowledge-based programs. Distributed Computing 10(4), 199–225 (1997)
7. Fischer, M.J., Ladner, R.E.: Propositional dynamic logic of regular programs. Journal of Computer and System Sciences 18(2), 194–211 (1979)
8. Gabbay, D.M.: A theory of hypermodal logics: Mode shifting in modal logic. Journal of Philosophical Logic 31(3), 211–243 (2002)
9. Gerbrandy, J., Groeneveld, W.: Reasoning about information change. Journal of Logic, Language and Information 6(2), 147–169 (1997)
10. Halpern, J.Y., Fagin, R.: Modelling knowledge and action in distributed systems. Distributed Computing 3(4), 159–177 (1989)
11. Harel, D., Kozen, D., Parikh, R.: Process logic: Expressiveness, decidability, completeness. Journal of Computer and System Sciences 25(2), 144–170 (1982)
12. Hoshi, T.: Epistemic Dynamics and Protocol Information. PhD thesis, Stanford (2009)
13. Hoshi, T., Yap, A.: Dynamic epistemic logic with branching temporal structures. Synthese 169(2), 259–281 (2009)
14. Kozen, D.: A completeness theorem for kleene algebras and the algebra of regular events. In: Proceedings of LiCS 1991, pp. 214–225 (1991)
15. Kozen, D.: Automata on guarded strings and applications. Matématica Contemporânea 24, 117–139 (2003)
16. Lutz, C.: Complexity and succinctness of public announcement logic. In: Proceedings of AAMAS 2006, pp. 137–143. ACM, New York (2006)
17. Meyer, J.J.: A different approach to deontic logic: deontic logic viewed as a variant of dynamic logic. Notre Dame Journal of Formal Logic 29(1), 109–136 (1987)
18. Parikh, R., Ramanujam, R.: A knowledge based semantics of messages. Journal of Logic, Language and Information 12(4) (2003)
19. Paul, S., Ramanujam, R., Simon, S.: Stability under strategy switching. In: Ambos-Spies, K., Löwe, B., Merkle, W. (eds.) CiE 2009. LNCS, vol. 5635, pp. 389–398. Springer, Heidelberg (2009)
20. Plaza, J.A.: Logics of public communications. In: Emrich, M.L., Pfeifer, M.S., Hadzikadic, M., Ras, Z.W. (eds.) Proceedings of the 4th International Symposium on Methodologies for Intelligent Systems, pp. 201–216 (1989)
21. Pratt, V.R.: Process logic. In: Proceedings of POPL 1979, pp. 93–100. ACM, New York (1979)
22. van Benthem, J., Gerbrandy, J., Hoshi, T., Pacuit, E.: Merging frameworks for interaction. Journal of Philosophical Logic 38(5), 491–526 (2009)
23. van Benthem, J., van Eijck, J., Kooi, B.: Logics of communication and change. Information and Computation 204(11), 1620–1662 (2006)
24. Wang, Y.: Indexed semantics and its application in modelling interactive unawareness. Master's thesis, University of Amsterdam (2006)
25. Wang, Y.: Epistemic Modelling and Protocol Dynamics. PhD thesis, University of Amsterdam (2010)
26. Wang, Y., Kuppusamy, L., van Eijck, J.: Verifying epistemic protocols under common knowledge. In: Proceedings of TARK 2009, pp. 257–266. ACM, New York (2009)
27. Wang, Y., Sietsma, F., van Eijck, J.: Logic of information flow on communication channels (extended abstract). In: van der Hoek, Kaminka, Lespérance, Luck, Sen (eds.) Proceedings of AAMAS (2010)

Becoming Aware of Propositional Variables

Hans van Ditmarsch[1,⋆] and Tim French[2]

[1] University of Sevilla, Sevilla, Spain
hvd@us.es
[2] University of Western Australia, Perth, Australia
tim@csse.uwa.edu.au

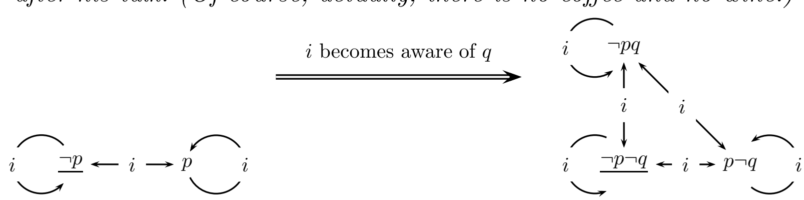
Abstract. We examine a logic that combines knowledge, awareness, and change of awareness. Change of awareness involves that an agent becomes aware of propositional variables. We show that the logic is decidable, and we present a complete axiomatization.

1 Introduction

Awareness and knowledge. Modal logic has long been used to reason about knowledge and belief in multi-agent systems. In modal logics we model *uncertainty* by allowing the value of propositions to vary between the so-called possible worlds. An agent *knows* a proposition in a given world if the proposition is true in all worlds accessible from that world. The logics require that the agents are *aware* of all propositional variables in the model. Thus reasoning in these models is undertaken under a closed world assumption: the relevant propositional variables are known to all agents. For every propositional variable in every world, every agent assigns a value to that variable.

While agents may be *uncertain* about the value of propositions, they may also be *unaware* of these propositions, and they may *become aware* of propositions. Uncertainty and incompleteness (i.e., unawareness) are different issues in modelling multi-agent systems. Without taking awareness into account, it seems difficult to explain the following transition, wherein the epistemic complexity of the model increases:

> *Initially, Hans (i) does not know whether coffee is served (p) after his talk. (Actually, no coffee will be served—$\neg p$, underlined.) Hans is unaware of it that wine is not served ($\neg q$) after his talk. Now, someone mentions that wine and coffee will not both be served. This makes Hans aware that wine is an issue. After this, Hans does not know whether coffee is served after his talk and also does not know whether wine is served after his talk. (Of course, actually, there is no coffee and no wine.)*



⋆ Hans van Ditmarsch is also affiliated to the Institute of Mathematical Sciences Chennai as a research associate.

M. Banerjee and A. Seth (Eds.): ICLA 2011, LNAI 6521, pp. 204–218, 2011.

We find that there are many subtleties and intricacies involved in defining the semantics for such dynamics of awareness. In this paper we will discuss these intricacies and in doing so make the following contributions:

1. We will introduce a new form of model equivalence modulo the agents' awareness and uncertainty, called *awareness bisimulation*.
2. We will define a new type of knowledge, referred to as *intrinsic knowledge*. Intrinsic knowledge is essential to express the dynamic interactions between awareness and knowledge. It relates to implicit and explicit knowledge.
3. We will introduce an logical operator for *becoming aware* of propositional variables and give semantics for this operator that is consistent with our intuitions of awareness and knowledge.

Prior research. Our work is rooted in: the tradition of epistemic logic [10] and in particular multi-agent epistemic logic [13,2]; in various research since the 1980s on the interaction of awareness and knowledge [1,14,15,8] — including a relation to recent works like [9,5,7]; and in modal logical research in propositional quantification, starting in the 1970s with [3] and followed up by work on bisimulation quantifiers [18,11,4].

Works treating awareness either follow a more *semantically* flavoured approach, where awareness concerns propositional variables in the valuation [1,15,8], or a more *syntactically* flavoured approach. In the latter, awareness concerns all formulas of the language in a given set, in order to model 'limited rationality' of agents. It is (also) pursued in [1] and in recent work like [5]. We are straight into the semantic corner: within the limits of their awareness, agents are fully rational.

For the static part of the logical language we follow [1]. For the dynamic part, it is remarkable that levels of 'interactive unawareness' in [8] can be described in terms of the awareness bisimulation introduced in our work (at the end of our paper). The insights made clear in their paper were very motivating for us. Our work builds on [17], which focusses on a special case (public global awareness) of the current paper, but unlike the present paper also treats awareness of other agents and forgetting (i.e., becoming unaware).

2 Structures

Given are a countably infinite set of propositional variables (facts) P and a (disjoint) finite set of agents N. Propositional variables are named p, q, r, and agent variables are named i, j, k, possibly indexed or quoted.

Definition 1 (Epistemic awareness model). *An* epistemic awareness model *for N and P is a tuple $M = (S, R, \mathcal{A}, V)$ that consists of a* domain S *of (factual)* states *(or 'worlds'), an* accessibility function $R : N \to \mathcal{P}(S \times S)$*, an* awareness function $\mathcal{A} : N \to S \to \mathcal{P}(P \cup N)$ *and a* valuation function $V : P \to \mathcal{P}(S)$*. For $R(i)$ we write R_i and for $\mathcal{A}(i)$ we write $\mathcal{A}_i$; accessibility function R can be seen as a set of* accessibility relations R_i*, and V as a set of* valuations $V(p)$*. A pointed epistemic awareness model (M, s) is an* epistemic awareness state.

Given an arbitrary model M we will refer to the elements of the tuple as $(S^M, R^M, \mathcal{A}^M, V^M)$. The awareness function $\mathcal{A}$ may be varied to reflect different logics. *Public global awareness* results if the value of $\mathcal{A}$ is the same for all agents and for all states. *Individual global awareness* results if the awareness function is the same in all states, but may vary among agents. These logics are discussed in [17]. In this work we focus on the logic of *individual local awareness* where there are no constraints placed on the awareness function $\mathcal{A}$. For the sake of generality we will assume no restrictions on the accessibility function R_i, either. However, we will sometimes require that the relation satisfies some simple properties (such as reflexivity, transitivity, etc.). The property of *awareness introspection* [8] holds if all agents know when they are aware of a fact or of another agent: "If $(s,t),(s,u) \in R_i$, then $\mathcal{A}_i(t) = \mathcal{A}_i(u)$."

Awareness bisimulation. Consider the following scenario: in state s agent i is aware of proposition p, state u is accessible for agent i from state s, and in state u agent j is aware of proposition p and also of proposition q. That agent j is also aware of q in u should leave agent i indifferent, as she is not aware of q in s! This sort of similarity is captured in the following notion, named *awareness bisimulation*. Informally, given a model and a set of propositional variables $P' \subseteq P$, another model is a P' awareness bisimulation if it cannot be distinguished from the first by formulas consisting only of the propositional variables in P', in the scope of agents who are aware of those propositions.

Definition 2 (Awareness bisimulation). *Let epistemic awareness models $M = (S, R, \mathcal{A}, V)$ and $M' = (S', R', \mathcal{A}', V')$ be given. For all $P' \subseteq P$ we define the relation $\mathfrak{R}[P']$ by $(s, s') \in \mathfrak{R}[P']$ iff:*

atoms *for all $p \in P'$, $s \in V(p)$ iff $s' \in V'(p)$;*
aware *for all $i \in N$, $\mathcal{A}_i(s) \cap P' = \mathcal{A}'_i(s') \cap P'$;*
forth *for all $i \in N$, if $t \in S$ and $R_i(s,t)$ then there is a $t' \in S'$ such that $R'_i(s',t')$ and $(t,t') \in \mathfrak{R}[P' \cap \mathcal{A}_i(s)]$;*
back *for all $i \in N$, if $t' \in S'$ and $R'_i(s',t')$ then there is a $t \in S$ such that $R_i(s,t)$ and $(t,t') \in \mathfrak{R}[P' \cap \mathcal{A}'_i(s')]$.*

Epistemic awareness state (M', s') is a P'-awareness bisimulation *of epistemic awareness state (M, s) (written $(M', s') \leftrightarrows_{P'} (M, s)$) iff $(s, s') \in \mathfrak{R}[P']$.*

The 'aware' clause can be considered as an additional basic structural requirement besides 'atoms', only due to the nature of our models where states have more structure than merely factual truth. If we were to replace $\mathfrak{R}[P' \cap \mathcal{A}_i(s)]$ in the **back** and **forth** clauses with $\mathfrak{R}[P']$, we would have the definition of a standard (restricted) bisimulation over labelled transition structures [16]. (Restricted to $P' \subseteq P$.) Thus every bisimulation is an awareness bisimulation. Vice versa, if all agents are aware of all propositional variables, the awareness bisimulation is a standard bisimulation (for the relations R_i). This is what we desire: we then revert to the standard multi-agent epistemic situation, where awareness plays no role.

Proposition 1. *The relation $\underline{\leftrightarrow}_{P'}$ is an equivalence relation.*

Proof. This can be easily seen by examining the Definition 2.

Definition 2 is more complex than the definition of standard bisimulation, however its motivation is very simple. Two worlds are P'-awareness bisimilar if, for any observer aware only of the propositions in P', the worlds appear identical. It gives us the "P'-perspective" of a world. We also call it *observational equivalence*. Let that observer be agent i in state s, then the required P' is $\mathcal{A}_i(s)$ and her perspective is that of $\mathcal{A}_i(s)$-awareness-bisimilarity. We might also say that her view of the model is that of its $\mathfrak{R}[\mathcal{A}_i(s)]$ equivalence class.

The crucial part of the definition is that in 'forth', in the requirement "$(t, t') \in \mathfrak{R}[P' \cap \mathcal{A}_i(s)]$", the bisimulation for **state t** is (further) restricted to the propositional variables that agent i is aware of in **state s**, the i-predecessor of t. (And similarly for 'back'.) An honoured principle (also in economics, and in artificial intelligence) is that incompleteness precedes uncertainty. The awareness function of an agent in a given state (incompleteness) determines what the agent can 'see' in all accessible states (uncertainty), *and so on*. This chaining of awareness is expressed with awareness bisimulation. This chaining requirement was present in epistemic awareness structures since its inception in [1]. We have merely employed it to the full and in the *one and only* way, for structural similarity.

Example. In Figure 1 agent i is aware of p but unaware of q in state s. In the figure, names of states are followed, separated by a dot, by values of propositional variables. Unaware variables are between parentheses. For example, $s.p(\neg q)$ means that in state s p is true and q is false, and the agent is aware of p and not of q. The three depicted epistemic states, wherein she (from left to right) implicitly knows q, knows $\neg q$, or does not know whether q, are observationally indistinguishable for the agent: they are p-awareness bisimilar. A p-awareness bisimulation between (e.g.) the left and the right picture is $\mathfrak{R} = \{(s, s''), (t, t''), (t, t''')\}$.

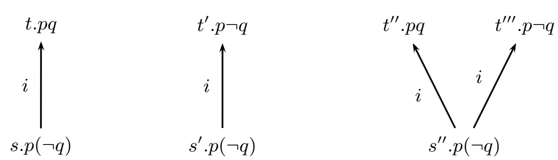


Fig. 1. Agent i is aware of p but unaware of q in state s

3 Language and Semantics

We augment multi-agent epistemic logic with three new operators: $A_i\varphi$, to mean that agent i *is* aware of all the propositional variables in φ; and $A_i^+ p\varphi$ for agent i *becoming aware* of propositional variable p, after which φ is true. The construct $K_i\varphi$, "agent i knows φ" stands in our case for "agent i *intrinsically* knows φ"—the meaning of intrinsic will be explained later.

Definition 3 (Language). *Given are a countably infinite set of propositional variables (facts/atoms) P, and a (disjoint) countably infinite set of agents N. The language $\mathcal{L}$ of individual local awareness is defined as*

$$\varphi ::= \top \mid p \mid \varphi \wedge \varphi \mid \neg\varphi \mid K_i\varphi \mid A_i\varphi \mid A_i^+ p\varphi$$

where $i \in N$ and $p \in P$. Implication $\rightarrow$, disjunction $\vee$, and equivalence $\leftrightarrow$ are defined by abbreviation. For $\neg K_i \neg\varphi$ we write $L_i\varphi$.

The semantics of the awareness operator A_i is purely syntax-based, namely using the *free variables* of a formula. These are defined as: $v(\top) = \emptyset$, $v(p) = \{p\}$, $v(\varphi \wedge \psi) = v(\varphi) \cup v(\psi)$, $v(\neg\varphi) = v(\varphi)$, $v(K_i\varphi) = v(\varphi)$, $v(A_i\varphi) = v(\varphi)$, and $v(A_i^+ p\varphi) = v(\varphi) \setminus \{p\}$. Note that $v(\varphi) \subseteq P$. We explicitly include $\top$ in the language, as the usual abbreviation $p \vee \neg p$ complicates cases where not all agents are aware of p (an agent unaware of p would then not explicitly know truth).

Definition 4 (Semantics). *Let $M = (S, R, \mathcal{A}, V)$ be given.*

$$\begin{array}{ll}
(M,s) \models \top & \\
(M,s) \models p & \textit{iff } s \in V(p) \\
(M,s) \models \varphi \wedge \psi & \textit{iff } (M,s) \models \varphi \textit{ and } (M,s) \models \psi \\
(M,s) \models \neg\varphi & \textit{iff } (M,s) \not\models \varphi \\
(M,s) \models K_i\varphi & \textit{iff } \forall t \in sR_i,\ \forall (M',t') \leftrightarrows_{\mathcal{A}_i(s)} (M,t), (M',t') \models \varphi \\
(M,s) \models A_i\varphi & \textit{iff } v(\varphi) \subseteq \mathcal{A}_i(s) \\
(M,s) \models A_i^+ p\varphi & \textit{iff } (M^{i \mapsto p}, s) \models \varphi
\end{array}$$

where $M^{i \mapsto p} = (S, R, \mathcal{A} \cup \{(i,(t,p)) \mid t \in S\}, V)$. The set of validities (and the logic) is called DLILA (Dynamic Logic of Individual Local Awareness).

Intrinsic knowledge. The treatment of knowledge in this semantics is novel. An agent knows φ only if in all accessible states φ remains true for *every* possible interpretation of all propositional variables that she is unaware of. We achieve this by composing the accessibility relation for an agent with the bisimulation relation modulo the propositional variables of which the agent is unaware. Because the constraints in this composition are interdependent, we have *one* K_i operator in the logical language and not, instead, *two* independent operators, one for standard modal accessibility and another one for bisimulation quantification. If the agent is aware of every propositional variable in the formula φ, the interpretation of knowledge is as for epistemic logic.

Awareness dynamics. Compared to knowledge, the semantics of becoming aware is simple. The complexity of becoming aware can only be seen in the context of intrinsic knowledge. Suppose that the agent is unaware of p and that p is true in all accessible states. We then have that $A_i^+ pK_ip$ is true: after the agent becomes aware of p, p is true. But although the agent considers that as a possibility, she does not know that, and she also considers it possible that after becoming aware of p, she knows that p is false, or that she is uncertain about p: all true are $L_iA_i^+ pK_ip$, $L_iA_i^+ pK_i\neg p$, and $A_i^+ p\neg(K_ip \vee K_i\neg p)$.

In this paper, we made one of three possible choices for awareness dynamics. All three consist of making an unaware variable into an aware variable, i.e., changing the set $\mathcal{A}$ in a model but leaving all other parameters the same. Given state s, one can make agent i aware of the propositional variable p:

– in the actual state (only): $\mathcal{A} \cup \{(i,(s,p))\}$.
– in the actual state and all states accessible for agent i: $\mathcal{A} \cup \{(i,(s,p))\} \cup \{(i,(t,p)) \mid t \in S \text{ and } R_i(s,t)\}$.
– in all states of the model: $\mathcal{A} \cup \{(i,(t,p)) \mid t \in S\}$.

All three are bisimulation invariant (with for the 'actual state only' version the restriction that the operation is performed on a bisimulation contraction, this requires a further adjustment of the definition). You might see the public version of becoming aware as the 'public announcement' version of awareness dynamics: just as in information dynamics, more complex dynamics have more complex axiomatizations, and this is on our future agenda.

KD45 and S5 Apart from the logic $DLILA$ we also consider the logics $DLILA_{\mathrm{L}}$, where every modal operator K_i satisfies the axioms of the logic L. Typical choices of L are S5 and KD45. One should be careful to note that this is *not* a simple case of restriction. Restricting the underlying logic to L (for example KD45) means that in interpreting the formula $K_i\varphi$, we may only consider pointed models (M',t') that satisfy the constraints of L (so transitive, serial and euclidean for KD45). The validities of $DLILA_{\mathrm{L}}$ therefore do not necessarily extend those of $DLILA$. And indeed, each axiomatization also poses new problems.

Specific logics require us to vary the semantics of the operator A_i^+p. For example, given awareness introspection and $S5$, the minimal way of becoming aware makes an agent aware of a propositional variable in the current world *and* in every indistinguishable world (the second option, before). In this paper we show completeness for the logic $DLILA_K$ namely for awareness models M where (S^M, R^M) is a tree, and where becoming aware means becoming aware in every world.

Where to put the complexity? An alternative interaction between knowledge and becoming aware is embodied in the following semantics (presented in [17]):

$$(M,s) \models K_i\varphi \quad \text{iff } \forall t \in sR_i,\ (M,t) \models \varphi$$
$$(M,s) \models A_i^+p\varphi \text{ iff } \exists (N,t) \leftrightarrows_{\mathcal{A}_i(s)} (M,s), (N^{i\mapsto p},t) \models \varphi$$

Here, the epistemic operator K_i remains the 'classical' one, whereas the becoming aware operator A_i^+p is the complex one. The advantage is obvious: the novel operator is the only addition to a well-known logic (namely that of [1]). The disadvantage is that a propositional variable may change its value in the process of the agent becoming aware of it; p may be true, but in the transition to a $(P \setminus \{p\})$-bisimilar state it may become false. So, e.g., the agent may become aware that she knows p to be false, even if prior to that she 'implicitly knew' p to be true. In that semantics, K_i does not mean implicit knowledge at all.

It seemed better to stock all the factual change into the mind of the agent only, as in the complex K_i operator, such that the becoming aware operation is merely revealing the veil of incompleteness. For $KD45$ and $S5$ structures that also satisfy 'awareness introspection' the distinction is immaterial, as the two semantics then are identical with respect to explicit knowledge. So, from an agent's point of view, there is no difference.

Proposition 2. *The semantics of $DLILA_K$ are invariant to bisimulation.*

Proof. This is straightforward because relation $\leftrightarrows_A$ is closed under bisimulation.

Examples

1. The introductory example about coffee and wine can be explained by seeing the model on the left as the *equivalence class modulo unawareness of* q of the model on the right. The agent can speculate over all models in that class (cf. the semantics of K_i, with bisimulation except for q). Becoming aware means that a model identical to the right model but with q unaware in all states, is transformed into the right model. On the left, in the actual state where p is false, it is e.g. true that: $\neg A_i q \wedge A_i^+ q \neg K_i q$.
2. Consider again Figure 1, and the roots of the models. In all three cases agent i knows that p. But she does not know in state s that q, because accessible state t is p awareness bisimilar to (e.g.) t' wherein q is false. After becoming aware of q in state s, she knows q: then, any state that is $\{p, q\}$ awareness bisimilar to t *must* satisfy q. So $A_i^+ q K_i q$ is true. Consider a $KD45$ extension of these models, i.e., add access (t, t) on the left, (t', t') in the middle, and $(t'', t''), (t'', t'''), (t''', t''), (t''', t''')$ on the right. Now we have that the agent considers it possible that: after becoming aware of q, she knows that q, or she knows that $\neg q$, or she does not know whether q.
3. Consider the case of $DLILA_{KD45}$, where every agent's accessibility relation is transitive, serial and euclidean. Crucially, in $KD45$, strong beliefs may be mistaken, but you do not consider that possible: to yourself, your beliefs appear knowledge. So $L_i(\neg p \wedge K_i p)$ is inconsistent. However, in $DLILA_{KD45}$ it is valid that an agent i considers it possible that she becomes aware of a propositional variable p that is false and that she believes to be true. That is nothing but speculating about becoming aware of false information that you had reason to accept! A validity of the language is $\neg A_i p \rightarrow L_i A_i^+ p(\neg p \wedge K_i p)$. The interpretation of this formula is shown in Figure 2. The crucial aspect is that the pair $(s, t) \in \mathfrak{R}[\emptyset]$ (the dashed line): agent i cannot a priori distinguish the reality of p being true in the believed world from the speculative option that p is false there but believed true. However, after becoming aware of p (in both s and t) this option is out of reach, as $(s, t) \notin \mathfrak{R}[p]$.

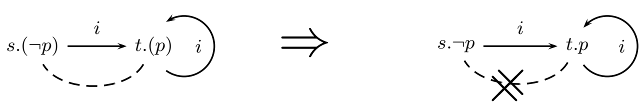


Fig. 2. You can become aware of a false belief

4 Intrinsic, Explicit and Implicit Knowledge

Past literature on knowledge and awareness has focused on the difference between implicit knowledge ("knowing" something without being fully aware of that thing) and explicit knowledge ("knowing" something as well as being fully aware of that thing). Intrinsic knowledge is strictly weaker than explicit knowledge and strictly stronger than implicit knowledge. It allows us to reason about the process of becoming aware, and that is our reason to complicate the existing picture. Implicit knowledge and explicit knowledge are definable in our framework, and we can compare those definitions with the traditional definitions.

Definition 5 (Explicit knowledge K_i^E and implicit knowledge K_i^I)
- $K_i^E \varphi$ *iff* $A_i \varphi \wedge K_i \varphi$ *(explicit knowledge)*
- $K_i^I \varphi$ *iff* $A_i^+ v(\varphi) K_i \varphi$ *(implicit knowledge)*

Expression $A_i^+ v(\varphi)$ means 'becoming aware of a finite set of propositional variables' and is defined in the obvious way. We also have that $K_i^I \varphi$ is equivalent to $A_i^+ v(\varphi) K_i^E \varphi$. The [1] definitions (in bold) are that $(M, s) \models \mathbf{K}_i^I \varphi$ iff $\forall t \in sR_i,\ (M, t) \models \varphi$, and that $(M, s) \models \mathbf{K}_i^E \varphi$ iff $(M, s) \models A_i \varphi$ and $\forall t \in sR_i$, $(M, t) \models \varphi$. We now observe that $K_i^E \varphi$ iff $\mathbf{K}_i^E \varphi$, and that $\mathbf{K}_i^I \varphi$ implies $K_i^I \varphi$ but not vice versa (e.g., if i is unaware of p, but j is aware of p, then i implicitly knows j to know that i is aware of p: $K_i^I K_j^E A_i p$ — this may come closer to implicit knowledge as in [12]). Intrinsic knowledge is clearly not definable in terms of implicit and explicit knowledge, given its semantics employing bisimulation quantification! Interaction between the three kinds of knowledge includes:

Proposition 3. $\models K_i^E \varphi \rightarrow K_i \varphi$ *and* $\models K_i \varphi \rightarrow K_i^I \varphi$*.*

On the other hand, $\not\models K_i^I \varphi \rightarrow K_i \varphi$. For example, you can implicitly know that p but, as you are unaware of p, you do not intrinsically know that p.

Proposition 4. *Awareness bisimilar states satisfy the same explicit knowledge: If* $(M, s) \models K_i^E \varphi$ *and* $(M, s) \leftrightarrows_{\mathcal{A}_i(s)} (M', s')$*, then* $(M', s') \models K_i^E \varphi$*.*

Proof. Note that $A_i \varphi$ means $v(\varphi) \subseteq \mathcal{A}_i(s)$. In the language restricted to $\mathcal{A}_i(s)$ the epistemic awareness states (M, s) and (M', s') are therefore bisimilar in the standard sense, from which follows logical equivalence, thus equivalence of $A_i \varphi \wedge K_i \varphi$ in both states.

5 Decidability

In this section we show decidability via an embedding into bisimulation-quantified modal logics [4]. Bisimulation-quantified modal logic is an extension of multi-modal (such as multi-agent) modal logic with the bisimulation quantifier, $\exists p \varphi$, which is interpreted as: "there is some model bisimilar to the current model except for the atom p, and in which φ is true". We recall the notion of restricted bisimulation already apparent in Definition 2. These logics are interpreted on models without the awareness function but that are otherwise similar.

Definition 6 (Bisimulation Quantified Modal Logic). *Let $L_{\mathcal{C}}$ be the set of validities for a model class $\mathcal{C}$. We define the bisimulation-quantified extension of $L_{\mathcal{C}}$ to be $QL_{\mathcal{C}}$ with the syntax:*

$$\varphi ::= p \mid \neg\varphi \mid \varphi \wedge \varphi \mid [i]\varphi \mid \exists p\varphi$$

where $p \in P$ and $i \in N$, and with the crucial semantic clause:

$$M, s \models_{\mathcal{C}} \exists p\varphi \iff \textit{for some } M', t \in \mathcal{C},\ M, s \underline{\leftrightarrow}_{P-p} M', t \textit{ and } M', t \models_{\mathcal{C}} \varphi.$$

It is shown in [4] that bisimulation quantified logics are decidable where $L_{\mathcal{C}}$ is an *idempotent transduction logic*; multi-agent **S5** and multi-agent **K** describe such idempotent transduction logics, and consequently have decidable bisimulation quantified extensions. We will give the embedding for $DLILA_K$ in QL_K (where K is the class of models satisfying all **K** validities).

Definition 7. *Let $\varphi \in \mathcal{L}$, and for every agent $i \in N$ and for every propositional variable (atom) $p \in v(\varphi)$, let a^i_p be an atom not appearing in φ, (referred to as an* awareness atom*, where $\mathcal{A}^\varphi$ is the set of awareness atoms for φ). The* embedding of $DLILA_K$ into QL_K *is given by the recursive function $\psi|_\varphi$ such that:*

$$\begin{array}{rl@{\qquad}rl}
\top|_\varphi = \top & & (A_i\psi)|_\varphi = \bigwedge_{p\in v(\psi)} a^i_p & \\
p|_\varphi = p & & (K_i\psi)|_\varphi = \bigwedge_{C\subseteq v(\psi)}(A_iC \to [i]\forall\overline{C^\varphi}\psi|_\varphi) & \\
(\neg\psi)|_\varphi = \neg(\psi|_\varphi) & & (A^+_i p\psi)|_\varphi = \psi|_\varphi[\top\backslash a^i_p] & \\
(\psi \wedge \chi)|_\varphi = (\psi|_\varphi) \wedge (\chi|_\varphi) & & &
\end{array}$$

where $\forall C$ is an abbreviation for $\forall p_0 ... \forall p_n$ for the set of atoms $C = \{p_0, ..., p_n\}$; C^φ is the set of all atoms in C along with the awareness atoms a^i_p where p appears in C; $\overline{C}$ is the complement of C with respect to the set of atoms in φ and the set $\mathcal{A}^\varphi$; and for $C \subseteq v(\varphi)$, A_iC is an abbreviation for $\bigwedge_{p\in C}[i]a^i_p \wedge \bigwedge_{p\in v(\varphi)\backslash C}[i]\neg a^i_p$. Also, $\psi[\chi\backslash p]$ is an abbreviation for the replacement of every free occurrence of the atom p in ψ with the formula χ.

This embedding is a direct encoding of the semantics for $DLILA_K$ into the logic QL_K. The only non-trivial recursions are for $K_i\psi$, which uses bisimulation quantifiers to encode an awareness bisimulation, and $A^+_i p\psi$, that simply sets agent i's awareness of p to true at every state.

Proposition 5. *Let $M = (S, R, \mathcal{A}, V)$ be a model such that for all atoms $p \in v(\varphi)$, and for all agents $i \in N$, for all $s \in S$, we have $s \in V(a^i_p)$ if and only if $p \in \mathcal{A}_i(s)$. Then for all $s \in S$: $M, s \models \varphi \iff M, s \models \varphi|_\varphi$. (On the right hand side, ignore the awareness parameter of the model in order to interpret $\varphi|_\varphi$.)*

Proof (Sketch). We give this proof by induction over the complexity of formulas. The induction hypothesis holds for $\psi \subset \varphi$ if and only if for all models N where the awareness atoms match the agents' awareness in M we have for every $s \in S^N$, $N, s \models \psi$ if and only if $N, s \models \psi|_\psi$. The base of the induction is the propositional atoms of φ and the truth symbol, $\top$, and these may be seen to support the induction hypothesis. The inductive cases for $\neg\alpha$ and $\alpha \wedge \beta$ follow directly from their semantic definitions, and the case for $A_i\alpha$ follows immediately from the fact that the awareness atoms agree with the agent's awareness.

For $K_i\alpha$, suppose that (N, s) is a pointed model such that $N, s \models K_i\alpha$. Let C be the set of propositional variables $\mathcal{A}_i^N(s) \cap v(\varphi)$. We may proceed as follows:

$$\begin{array}{ll} & N, s \models K_i\alpha \\ \Leftrightarrow [Def] & \forall t \in sR_i^N, \forall (N', t') \leftrightarroweq_C (N, t),\ N', t' \models \alpha \\ \Leftrightarrow [I.H.] & \forall t \in sR_i^N, \forall (N', t') \leftrightarroweq_C (N, t),\ N', t' \models \alpha|_\varphi \\ \Leftrightarrow [*] & \forall t \in sR_i^N, \forall (N', t') \cong_{C^\varphi} (N, t),\ N', t' \models \alpha|_\varphi \\ \Leftrightarrow [Def] & \forall t \in sR_i^N, N, t \models \forall \overline{C^\varphi}\alpha|_\varphi \\ \Leftrightarrow [Def] & N, s \models [i]\forall \overline{C^\varphi}\alpha|_\varphi \end{array}$$

The equivalences here are all straightforward, except for the one labelled $*$. In the forward direction this is trivial: every C^φ-bisimulation *is* a C-awareness bisimulation. In the reverse direction, we must note the construction of $\alpha|_\varphi$. Here modalities $[i]$ only appear in the form $\bigwedge_{C \subseteq v(\alpha)}(A_i c \rightarrow [i]\forall\overline{C^\varphi}\alpha|_\varphi)$. With respect to this form we can see that C^φ-bisimulations and C-awareness-bisimulations indeed are equivalent since every application of a modality extends the bisimulation according to awareness function at that point. Since C is defined to be the set of atoms of which agent i is aware, the result follows. Finally, we note that for the interpretation of $A_i^+ p\alpha$ it is enough to manually fix the interpretation of every free occurrence of the atom a_i^p in $\alpha|_\varphi$ to true, as $A_i p$ will be true in every state.

Thus we have a translation from $DLILA_K$ to the bisimulation quantified logic QL_K that preserves the meaning of formulas (given a set of awareness atoms in the model). Decidable satisfiability and model-checking follow. We have not yet investigated the lower bound for complexity of the translation. The translation given is quite general, as it also suffices for QL_{KD45} or QL_{S5}.

6 Axiomatization

We provide an axiomatization **DLILA** for $DLILA_K$, and we show it to be sound and complete. The propositional rules and axioms, and those for knowledge (only involving K_i), are standard. The axioms for awareness (for A_i) simply capture the syntactic definition. The interaction between knowledge and awareness is governed by the axioms **AK1**–**AK4**.

C0	All tautologies of prop. logic	**K**	$K_i(\varphi \rightarrow \psi) \rightarrow K_i\varphi \rightarrow K_i\psi$
MP	From φ and $\varphi \rightarrow \psi$ infer ψ	**Nec**	From φ infer $K_i\varphi$
A1	$A_i(\varphi \wedge \psi) \leftrightarrow A_i\varphi \wedge A_i\psi$	**A2**	$A_i\neg\varphi \leftrightarrow A_i\varphi$
A3	$A_iK_j\varphi \leftrightarrow A_i\varphi$	**A4**	$A_iA_j\varphi \leftrightarrow A_i\varphi$
A5	$A_iA_j^+p\varphi \leftrightarrow A_ip \wedge A_i\varphi$	**A6**	$A_i\top$

AK1 $K_i\varphi \wedge \neg A_ip \rightarrow K_i\varphi[\psi\backslash p]$
AK2 From $A_i\varphi \wedge K_i\varphi \rightarrow K_i\psi$ infer $A_i\psi \wedge K_i\varphi \rightarrow K_i\psi$
AK3 $(K_i(p \rightarrow \varphi) \vee K_i(\neg p \rightarrow \varphi)) \wedge \neg A_ip \rightarrow K_i\varphi$
AK4 $(K_i(A_jp \rightarrow \varphi) \vee K_i(\neg A_jp \rightarrow \varphi)) \wedge \neg A_ip \rightarrow K_i\varphi$.

In **AK1** we require that ψ is free for p in φ, and the axioms **AK3** and **AK4** may only be applied in the case where the atom p does not appear outside the scope of a modal (knowledge) operator. Axioms **AK1** and **AK2** are not required in the completeness proof, but we have left them in as they represent important principles that hold in *all* semantic variations of $DLILA$: **AK1** shows that if an agent is not aware of an atom, then the agent may not distinguish the interpretation of that atom from the interpretation of an arbitrary proposition; **AK2** states that if intrinsic knowledge of ψ can be derived from explicit knowledge of φ, then intrinsic knowledge of ψ may also be derived from intrinsic knowledge of φ and awareness of ψ. Axioms **AK3** and **AK4** are specific to the **K** semantics: they capture the intrinsic nature of the knowledge operator: if an agent is unaware of an atom, he does not refute any interpretation of that atom, nor does he refute the interpretation of any agent's awareness of that atom.

Finally we present axioms for becoming aware. We note from the semantics that if an agent i becomes aware of an atom this will only affect the interpretation for formulas $A_i\varphi$ or $K_i\varphi$. Consequently A_i^+p commutes with all other operators.

B0	$A_i^+p\top$	**B1**	$A_i^+pq \leftrightarrow q$
B2	$A_i^+p(\varphi \wedge \psi) \leftrightarrow A_i^+p\varphi \wedge A_i^+p\psi$	**B3**	$A_i^+p\neg\varphi \leftrightarrow \neg A_i^+p\varphi$
B4a	$A_jp \rightarrow (A_i^+pK_j\varphi \leftrightarrow K_jA_i^+p\varphi)$	**B4b**	$K_iA_i^+p\varphi \rightarrow A_i^+pK_i\varphi$
B5a	$A_i^+pA_j\varphi \leftrightarrow A_j\varphi$ where $i \neq j$	**B5b**	$A_i^+pA_i\varphi \leftrightarrow A_i\varphi[\top\backslash p]$
B6	$A_i^+pA_j^+q\varphi \leftrightarrow A_j^+qA_i^+p\varphi$		

Soundness and completeness. The soundness is straightforward. We show completeness for **DLILA** by constructing a canonical model for any formula using maximal consistent sets of formulas in $\mathcal{L}$—proofs are in the appendix.

Definition 8. *The* canonical model *is built from the set $\mathcal{S}$ of all maximal consistent sets of formulas with respect to the system* **DLILA**. *Further we define $\mathcal{M} = (\mathcal{S}, \mathcal{R}, \mathcal{A}, \mathcal{V})$ where:*

- *for all $i \in N$, for all maximal consistent sets $\sigma, \tau \in \mathcal{S}$, $(\sigma, \tau) \in \mathcal{R}_i$ if and only if for all formulas $A_i^+v(\varphi)K_i\varphi \in \sigma$, we have $\varphi \in \tau$;*
- *for all $\sigma \in \mathcal{S}$ for all $i \in N$, for all $p \in P$, we have $p \in \mathcal{A}_i(\sigma)$ if and only if $A_ip \in \sigma$;*
- *for all $\sigma \in \mathcal{S}$, for all $p \in P$, we have $\sigma \in \mathcal{V}(p)$ if and only if $p \in \sigma$.*

Proposition 6. *Every* canonical model *is an epistemic awareness model.*

Proof. In the presence of complete awareness, intrinsic knowledge is equivalent to explicit knowledge, and the logic of explicit knowledge is canonical. We note that the awareness function $\mathcal{A}$ is constant for each agent's local state because of the axiom **AK3**.

Lemma 1 (Truth Lemma). *For every $\sigma \in \mathcal{S}$, for every formula φ, we have $\varphi \in \sigma$ if and only if $\mathcal{M}, \sigma \models \varphi$.*

(Proof in Appendix.) It follows that for every consistent formula φ we may construct a model so the axiomatization **DLILA** is complete for the logic $DLILA$.

7 Comparison

Our approach is in some respects simpler and more constrained than [8]. From the epistemic awareness structure we are able to implicitly derive a complete lattice of spaces via awareness bisimulation, whereas in [8] this structure is given explicitly. In other words, we have a succinct, technical tool to derive that result.

The principles $A1, \ldots, A6$ in **DLILA** straightforwardly correspond to (a multi-agent version of) L^{KXA} in [6] and Proposition 3 in [8]—epistemic operators K_i in the scope of awareness operators can be replaced by the *explicit* knowledge operators K_i^E assumed by those authors; $A5$ is a 'mix' axiom relating to dynamics. Principles $AK1$ and $AK2$ were conceived using results for bisimulation-quantified logics and are strictly about *intrinsic* knowledge only.

Although we do not explicitly have propositional quantifiers, they are indirectly present in intrinsic knowledge operators. Propositional quantification is integrated with awareness and knowledge in [7] (and in various precursors). This concerns quantification over the set of formulas of which an agent is aware. They interestingly mention that "Using semantic valuations [for quantification] does not work in the presence of awareness" [7, p.506]; although of course correct, we are wondering if our work may make the authors reconsider the suggested scope of that remark.

Dynamics of (factual) awareness is presented in [9,5,19]. In [9] becoming aware means (initially) becoming ignorant about that proposition. It uses an algebraic approach. Becoming ignorant is also the approach in the recent [19], that contains various other novelties. In [5], the approach in Section 3 is similarly dynamic modal as ours, and it provides an integrated combination of syntactic and semantic awareness.

Acknowledgment

We thank the ICLA reviewers for their helpful and detailed comments. We acknowledge helpful interactions with Thomas Ågotnes, Chitta Baral, Oliver Board, Jan Broersen, Kim-Sau Chung, Valentin Goranko, Joe Halpern, Sophie Pinchinat, Burkhard Schipper, and Fernando Velazquez.

References

1. Fagin, R., Halpern, J.Y.: Belief, awareness, and limited reasoning. Artificial Intelligence 34(1), 39–76 (1988)
2. Fagin, R., Halpern, J.Y., Moses, Y., Vardi, M.Y.: Reasoning about Knowledge. MIT Press, Cambridge (1995)
3. Fine, K.: Propositional quantifiers in modal logic. Theoria 36(3), 336–346 (1970)
4. French, T.: Bisimulation quantifiers for modal logic. PhD thesis, University of Western Australia (2006)
5. Grossi, D., Velázquez-Quesada, F.R.: Twelve Angry Men: A study on the fine-grain of announcements. In: He, X., Horty, J., Pacuit, E. (eds.) LORI 2009. LNCS, vol. 5834, pp. 147–160. Springer, Heidelberg (2009)

6. Halpern, J.Y.: Alternative semantics for unawareness. Games and Economic Behavior 37(2), 321–339 (2001)
7. Halpern, J.Y., Rego, L.C.: Reasoning about knowledge of unawareness. Games and Economic Behavior 67(2), 503–525 (2009)
8. Heifetz, A., Meier, M., Schipper, B.C.: Interactive unawareness. Journal of Economic Theory 130, 78–94 (2006)
9. Hill, B.: Awareness dynamics. Journal of Philosophical Logic (to appear, 2009)
10. Hintikka, J.: Knowledge and Belief. Cornell University Press, Ithaca (1962)
11. Hollenberg, M.: Logic and bisimulation. PhD thesis, University of Utrecht (1998)
12. Levesque, H.J.: A logic of implicit and explicit belief. In: Brachman, R.J. (ed.) Proceedings of AAAI, pp. 198–202. AAAI Press, Menlo Park (1984)
13. Meyer, J.-J.C., van der Hoek, W.: Epistemic Logic for AI and Computer Science. Cambridge Tracts in Theoretical Computer Science, vol. 41. Cambridge University Press, Cambridge (1995)
14. Modica, S., Rustichini, A.: Awareness and partitional information structures. Theory and Decision 37, 107–124 (1994)
15. Modica, S., Rustichini, A.: Unawareness and partitional information structures. Games and Economic Behavior 27, 265–298 (1999)
16. Stirling, C.: The joys of bisimulation. In: Brim, L., Gruska, J., Zlatuška, J. (eds.) MFCS 1998. LNCS, vol. 1450, p. 142. Springer, Heidelberg (1998)
17. van Ditmarsch, H., French, T.: Awareness and forgetting of facts and agents. In: Proceedings of WI-IAT Workshops 2009, pp. 478–483. IEEE Press, Los Alamitos (2009)
18. Visser, A.: Bisimulations, model descriptions and propositional quantifiers, Logic Group Preprint Series 161, Department of Philosophy, Utrecht University (1996)
19. Wang, Y.: Epistemic Modelling and Protocol Dynamics. PhD thesis, Universiteit van Amsterdam (2010)

Appendix: Proof of Truth Lemma 1

For a convenient proof, we give a syntactic version of awareness bisimulation.

Definition 9. *We say a formula of DLILA is* explicit *if it is built from the following syntax:* $\varphi ::= \top \mid p \mid \varphi \wedge \varphi \mid \neg\varphi \mid K_i\varphi \wedge A_i\varphi \mid A_i\varphi$*. For every* $C \subseteq P$*, let* $\mathcal{B}(C)$ *be a binary relation on* $\mathcal{S}$ *satisfying for all* $\sigma, \tau \in \mathcal{S}$*,* $(\sigma, \tau) \in \mathcal{B}(C)$ *if and only if for every explicit formula* φ *containing only the atoms in* C*, we have we have* $\varphi \in \sigma$ *implies* $\varphi \in \tau$*. We refer to such formulas* φ *as* C*-explicit.*

The following lemma is a strengthening of Proposition 4 and shows the correspondence between Definition 9 and Definition 2.

Lemma 2. *For every* $\sigma, \tau \in \mathcal{S}$*, for every* $C \subseteq P$ *we have* $(\mathcal{M}, \sigma) \leftrightarroweq_C (\mathcal{M}, \tau)$ *if and only if* $(\sigma, \tau) \in \mathcal{B}(C)$*.*

Proof (Sketch). ($\Longrightarrow$) We show by induction over the complexity of formulas that for any σ, τ where $(\mathcal{M}, \sigma) \leftrightarroweq_C (\mathcal{M}, \tau)$, we have, for any C-explicit formula φ, $\varphi \in \sigma$ if and only if $\varphi \in \tau$. In the case that φ is a propositional atom or $\top$ it is clear that $\varphi \in \sigma$ iff $\varphi \in V(\sigma)$ iff $\varphi \in V(\tau)$ iff $\varphi \in \tau$. The inductions for the propositional operators $\wedge$ and $\neg$ are similarly straightforward. For $A_i\varphi$, by application of the axioms **A1-A6** we have $A_i\varphi \in \sigma$ iff for all atoms p in φ we have $A_ip \in \tau$. By the **aware** clause of Definition 2, this is equivalent to $A_ip \in \tau$ so we must have $A_i\varphi \in \tau$. Finally, if $K_i\varphi \wedge A_i\varphi \in \sigma$, then for all $\sigma' \in \sigma\mathcal{R}_i$ we have $\varphi \in \sigma'$. By Definition 2 for all $\tau' \in \tau\mathcal{R}_i$, there exists $\sigma' \in \sigma\mathcal{R}_i$ such that $\tau' \leftrightarrows_{C\cap\mathcal{A}_i(\sigma)}\sigma'$. By the induction hypothesis it follows that $\varphi \in \tau'$, so $K_i\varphi \wedge A_i\varphi \in \sigma$ iff $K_i\varphi \wedge A_i\varphi \in \tau$.

($\Longleftarrow$) Here we show that the relations $\mathcal{B}(C)$ satisfy the properties specified in Definition 2. Clearly the clauses **atom** and **aware** hold since if $(\sigma,\tau) \in \mathcal{B}(C)$ then σ and τ agree on all C-explicit formulas which includes the atoms in C, and the awareness of those atoms. To see **forth** holds, suppose that $(\sigma,\tau) \in \mathcal{B}(C)$. Then for all agents i, for all $\sigma' \in \sigma\mathcal{R}_i$, for all $C \cap \mathcal{A}_i(\sigma)$-explicit formulas $\varphi \in \sigma'$, we have $L_i\varphi \wedge A_i\varphi \in \sigma$. By the definition of $\mathcal{B}(C)$ we have, for all $C \cap \mathcal{A}_i(\sigma)$ formulas $\varphi \in \sigma'$, $L_i\varphi \wedge A_i\varphi \in \tau$. By the axiom **B4a** every finite subset of the $C \cap \mathcal{A}_i(\sigma)$-explicit formulas in σ' is consistent with the set of implicit knowledge formulas, $\{\psi \mid A^+v(\psi)K_i\psi \in \tau\}$. As there is no finite proof of inconsistency we may conclude that the set of $C \cap \mathcal{A}_i(\sigma)$-explicit formulas in σ' is consistent with the set of implicit knowledge formulas in τ. By Definition 8 there is some $\tau' \in \tau\mathcal{R}_i$ such that $(\sigma,\tau') \in \mathcal{B}(C \cap \mathcal{A}_i(\sigma)$, as required. The case for **back** is handled symmetrically.

Lemma 2 provides a compelling justification for the notion of awareness bisimulation. Two states are C-awareness bisimilar exactly when they agree on all C-explicit formulas. We continue with the proof of the Truth Lemma 1 proper.

Proof (Sketch). This lemma is given by induction over the complexity of formulas. The base case, where $\varphi \in P$ or $\varphi = \top$ is a direct application of the definition of $\mathcal{V}$, so we may assume for all $\psi \subset \varphi$, for all $\sigma \in \mathcal{S}$ we have $\psi \in \sigma$ if and only if $\mathcal{M},\sigma \models \psi$. The induction proceeds as follows:

- $\neg$ Suppose $\varphi = \neg\psi$. Then since $\psi \in \sigma$ if and only if $\mathcal{M},\sigma \models \psi$, from the consistency of σ we have $\varphi \in \sigma$ if and only if $\mathcal{M},\sigma \models \varphi$.
- $\wedge$ Suppose $\varphi = \psi_1 \wedge \psi_2$. Then since $\psi_i \in \sigma$ if and only if $\mathcal{M},\sigma \models \psi_i$, from the consistency of σ we have $\varphi \in \sigma$ if and only if $\mathcal{M},\sigma \models \varphi$.
- A_i Suppose $\varphi = A_i\psi$. Then clearly by the axioms **A1-A7**, $A_i\psi \in \sigma$ if and only if, for all atomic propositions $p \in \psi$ we have $A_ip \in \sigma$. This is equivalent to $(\mathcal{M},\sigma) \models A_ip$ for all atoms, p in ψ, which is equivalent to $(\mathcal{M},\sigma) \models A_i\psi$.
- K_i Suppose $L_i\psi \in \sigma$. Let $\Psi = \{\alpha \mid K_i\alpha \wedge A_i\alpha \in \sigma\}$. Now $\Psi \cup \{\psi\}$ is consistent (since the conjunction γ of every finite subset appears in σ as $L_i\gamma$), and furthermore, it must be consistent with the $\mathcal{A}_i(\sigma)$-explicit formulas that appear in some $\tau \in \sigma\mathcal{R}_i$. Therefore, we may find a maximal consistent set τ' such that $\psi \in \tau' \leftrightarrows_{\mathcal{A}_i(\sigma)}\tau \in \sigma\mathcal{R}_i$, so $(\mathcal{M},\sigma) \models L_i\psi$ as required.
  Conversely, suppose that $(\mathcal{M},\sigma) \models L_i\psi$. We proceed by induction over the knowledge-depth of ψ, where the induction hypothesis is, that for all ψ of knowledge depth n:
  1. for all $\tau \in \mathcal{S}$, $\vdash L_i(\tau_\psi^\Gamma) \wedge \overline{A_i\Gamma}^\psi \to L_i\tau_\psi$, and
  2. for all $\tau \in \mathcal{S}$, $(\mathcal{M},\tau) \models \psi$ if and only if $\psi \in \tau$.

Here, Γ is a set of propositional variables; τ_ψ^Γ is the set of subformulas of ψ containing only atoms from Γ that appear in the set τ; $\overline{A_i\Gamma}^\psi$ is an abbreviation for $\bigwedge\{\neg A_i p \mid p \in \psi \backslash \Gamma\}$; and τ_ψ is the set of subformulas of ψ that appear in τ. This is sufficient to show $\mathcal{M}, \sigma) \models L_i\psi$ implies $L_i\psi \in \sigma$ since if $(\mathcal{M}, \sigma) \models L_i\psi$, we have $\psi \in \tau$ for some $\tau \leftrightarroweq_{A_i(\sigma)} \tau' \in \sigma\mathcal{R}_i$. By Lemma 2, $\tau_\psi^{A_i(\sigma)} \in \tau'$, and hence $L_i(\tau_\psi^{A_i(\sigma)}) \in \sigma$. Applying the inductive hypothesis we have $L_i\psi \in \sigma$ as required.

For the base case, iff ψ has knowledge depth 0, we can see from axioms **B0-B7** and **A1-A6** that ψ is effectively a propositional formula where the atoms are either propositional atoms, or agents' awareness of propositional atoms. Now there are two cases: if $A_i\psi \in \sigma$, then by the axiom **B4b** we have for every $\tau \in \sigma\mathcal{R}_i$, for every $\tau' \leftrightarroweq_{\mathcal{A}_i(\sigma)} \tau$, $\psi \in \tau'$, so $(\mathcal{M}, \sigma) \models K_i\psi$ and we are done. Alternatively, if $A_i\psi \notin \sigma$, then there are some atoms in ψ that agent i is not aware of at σ. Let $\mathcal{T} = \{p, A_j p \mid p, j \in v(\psi)\}$. For any τ, τ' where $\tau \leftrightarroweq_{\mathcal{A}_i(\sigma)} \tau' \in \sigma\mathcal{R}_i$ there is a subset of $\mathcal{T}$ true at τ. We may apply the axioms **AK3** and **AK4** to derive

$$L_i\chi(\tau) = L_i\left(\bigwedge_{\alpha\in\mathcal{T}\cap\tau} \alpha \wedge \bigwedge_{\alpha\in\mathcal{T}\backslash\tau} \neg\alpha \wedge \psi\right)$$

Therefore there is some maximal consistent set ρ containing $\chi(\tau)$, and as τ' agrees with ρ on the interpretation of all atoms up to the depth of ψ, we must have $(\mathcal{M}, \tau') \models \psi$ as required. As this is the case for every $\tau \in \mathcal{S}$ it must be that $\vdash L_i(\tau_\psi^\Gamma) \wedge \overline{A_i\Gamma}^\psi \to L_i\tau_\psi$.

For the inductive step we proceed in a similar fashion. Suppose $\tau \leftrightarroweq_{\mathcal{A}_i}(\sigma)\tau' \in \sigma\mathcal{R}_i$. Given that we may apply the axiom **AK4** to replicate the awareness state of agents at τ'. We may then apply the inductive hypothesis to infer $K_j\psi_k$ at $(\mathcal{M}, \tau)$, where ψ_k has knowledge depth less than n. Finally we may again apply **AK3** and **AK4** to replicate the interpretation of atoms, and other agents' awareness of the atoms at τ. As ψ may be written as a Boolean combination of atoms, agents' awareness of atoms and formulas $K_j\psi_k$ (where the knowledge depth of ψ_k is less than n), the result follows.

A_i^+p Suppose that $\varphi = A_i^+p\psi$, and $\varphi \in \sigma$. From axiom **B3** we can see the set $\tau = \{\alpha \mid A_i^+p\alpha \in \sigma\}$ is maximal and consistent. Furthermore from Definition 8 we can see that for all j, for every $\rho \in \mathcal{S}$ we have $\rho \in \sigma\mathcal{R}_j$ if and only if $\rho \in \tau\mathcal{R}_j$. Thus the successors of τ are exactly the successors of σ. Also, from the axioms **B0-B6** we have for all atoms p, $\tau \in \mathcal{V}(p)$ iff $\sigma \in \mathcal{V}(p)$, for all agents $j \neq i$ we have $\mathcal{A}_j(\sigma) = \mathcal{A}_j(\tau)$ and finally $\mathcal{A}_i(\sigma) \cup \{p\} = \mathcal{A}_i(\tau)$. From this we can see $(\mathcal{M}, \tau)$ is bisimilar to $(\widehat{\mathcal{M}}^{i\mapsto p}, \sigma)$, where $\widehat{\mathcal{M}}$ is the tree unwinding of $\mathcal{M}$. As $\psi \in \tau$, the result follows.

Conversely, suppose that $(\mathcal{M}, \sigma) \models A_i^+p\psi$ and for contradiction, suppose that $\neg A_i^+p\psi \in \sigma$. Applying **B3** we have $A_i^+p\neg\psi \in \sigma$. From the argument above it follows that a model bisimilar to $(\widehat{\mathcal{M}}^{i\mapsto p}, \sigma)$ satisfies $\neg\psi$, so the contradiction follows from the bisimulation invariance of $DLILA_K$ (Proposition 2).

Author Index