Logic and Interactive RAtionality Yearbook 2012

Volume I

Editors: Zoé Christoff Paolo Galeazzi Nina Gierasimczuk Alexandru Marcoci Sonja Smets

Collecting Editors: Alexandru Baltag (Europe) Wesley Holliday (United States) Fenrong Liu (China)

The printing of this book was supported by the UP fund of Johan van Benthem. Cover design by Nina Gierasimczuk.

Foreword

It is our great pleasure to present you with the two volumes of the "Logic and Interactive Rationality" (LIRa) Yearbook 2012, this time officially a joint-venture of Amsterdam-BayArea-Beijing. Following tradition, the Yearbook reflects the activities of the LIRa seminar held regularly at the ILLC in Amsterdam. We are very happy that Tsinghua University in Beijing and Stanford University and UC Berkeley in the Bay Area have joined our efforts. Submissions from China and the US show the global nature of research into interactive rationality and logical dynamics. We expect that the Yearbook can serve as a platform for future collaborations among congenial colleagues working in distant parts of the world.

We want to thank all the contributors for allowing us to include the fruits of their work. We would like to give recognition to Nina Gierasimczuk for her cover design, but also for coordinating the work on the volume. We also would like to direct our special thanks to Alexandru Marcoci for typesetting yet another edition of the LIRa Yearbook. Finally, we want to thank Johan van Benthem for sponsoring the seminar's activities, for his efforts to keep us all connected, and for being a constant source of scientific inspiration.

The Editors Amsterdam-Berkeley-Beijing February 21st, 2014

Contents of Volume I

| 1 | Preface by Johan van Benthem | 1 |
|----|--|-----|
| 2 | Dynamic Epistemic Logic as a Substructural Logic by Guillaume Aucher | 3 |
| 3 | Strategic Voting and the Logic of Knowledge by Hans van Ditmarsch, Jérôme Lang, and Abdallah Saffidine | 27 |
| 4 | What Does it Mean to Know an Action? by Jiahong Guo | 53 |
| 5 | What You Can Do Depends on What You Have Done by Fengkui Ju and Li Liang | 81 |
| 6 | Public Announcements under Sheaves by Kohei Kishida | 105 |
| 7 | Transition Semantics. The Dynamics of Dependence Logic by Pietro Galliani | 121 |
| 8 | Dynamic Measure Logic <i>by Tamar Lando</i> | 145 |
| 9 | Surprise in Probabilistic Dynamic Epistemic Logic by Lorenz Demey | 177 |
| 10 | A Geo-logical Solution to the Lottery Paradox, with Applications to Con- ditional Logic | - |
| | by Hanti Lin and Kevin T. Kelly | 201 |

| 11 | Changing Types: Information Dynamics for Qualitative Type Spaces by Dominik Klein and Eric Pacuit | 253 |
|----|---|-----|
| 12 | Dependent Type Semantics: An Introduction by Daisuke Bekki | 277 |
| 13 | A Medieval Epistemic Puzzle by Sara L. Uckelman | 301 |

Preface

Johan van Benthem

University of Amsterdam, Stanford University johan.vanbenthem@uva.nl

Rocking All Over the World

Though the song playing on the radio right now is from a band whimsically named the Status Quo, 'Rocking All Over The World' seems the perfect background when reading these exciting volumes. So many dynamic themes combining the tested rigor of logic with attunement to the latest that is happening in the international world today.

In this book, you can drift along on many currents, that meet and separate all the time. There are logical analyses of social networks, pluralistic ignorance, strategic voting, and bubbles, trying to come to grips with the burning issues of society right around us. But there are also investigations into the foundations of games as played by rational but bounded agents, who must do their planning based on fallible information and beliefs that sometimes run into surprises and the unexpected, and whose interactions depend on the pervasive phenomenon of dependence. Computation and complexity are further basic themes that are never far away in many contributions, but so is human cognitive behavior, and the intricacies of our common uses of natural language, both logical and probabilistic. But there is also a lot of methodological innovation to be found here. Mathematicians may appreciate several more technical papers that provide powerful new links with proof theory and substructural logics, as well as new forms of outreach toward category-theoretic sheave models, probability theory, and measurement theory. And philosophers may be drawn to papers studying the entanglement of knowledge, evidence, action, learning, and ability, some of them bridging between epistemology and philosophy of science. And if, after all this, you need a trip to quite different realms, just read the papers on dynamic logical foundations of the quantum world below us, or on medieval history: a past every bit as exciting as the future.

This Yearbook is now a production of editors from three sites some eight time zones apart: Amsterdam, Beijing, and the Bay Area. Just multiply these two figures, and you will see that the sun never sets in the realm of dynamics.

Dynamic Epistemic Logic as a Substructural Logic

Guillaume Aucher

University of Rennes 1, INRIA guillaume.aucher@irisa.fr

Abstract

Dynamic Epistemic Logic (DEL) is an influential logical framework for reasoning about the dynamics of beliefs and knowledge. It has been related to older and more established logical frameworks but, despite these connections, DEL remains, arguably, a rather isolated logic in the vast realm of non-classical logics and modal logics. This is problematic if logic is to be viewed ultimately as a unified and unifying field and if we want to avoid that DEL goes on "riding off madly in all directions" (a metaphor used by van Benthem about logic in general). In this article, we show that DEL can be redefined naturally and meaningfully as a two-sorted substructural logic. In fact, it is even one of the most primitive substructural logics since it does not preserve any of the structural rules. Moreover, the ternary semantics of DEL and its dynamic interpretation provides a conceptual foundation for the Routley & Meyer's semantics of substructural logics.

1 Introduction

Dynamic Epistemic Logic (DEL) is an influential logical framework for reasoning about the dynamics of beliefs and knowledge, which has drawn the attention of a number of researchers ever since the seminal publication of Batlag et al. (1998). A number of contributions have linked DEL to older and more established logical frameworks: it has been embedded into (automata) PDL (van Eijck 2004, van Benthem and Kooi 2004), it has been given an algebraic semantics (Baltag et al. 2005; 2007), and it has been related to epistemic temporal logic (van Benthem et al. 2009, Aucher and Herzig

2011) and the situation calculus (van Benthem 2011a, van Ditmarsch et al. 2009). Despite these connections, DEL remains, arguably, a rather isolated logic in the vast realm of non-classical logics and modal logics. This is problematic if logic is to be viewed ultimately as a unified and unifying field and if we want to avoid that DEL goes on "riding off madly in all directions" (a metaphor used by van Benthem (2008; 2011b) about logic in general). In this article, we will show that DEL can be redefined naturally and meaningfully as a two-sorted substructural logic. In fact, it is even one of the most primitive substructural logics since it does not preserve any of the structural rules.

Substructural logics will also benefit from this interaction with DEL. The wellknown semantics for substructural logics is based on a ternary relation introduced by Routley & Meyer for relevance logic in the 1970's (Routley and Meyer 1972a;b; 1973, Routley et al. 1982). However, the introduction of this ternary relation was originally motivated by technical reasons, and it turns out that providing a non-circular and conceptually grounded interpretation of this relation remains problematic (Beall et al. 2012). As we shall see, the ternary semantics of DEL provides a conceptual foundation for Routley & Meyer's semantics. In fact, the dynamic interpretation induced by the DEL framework turns out to be not only meaningful, but also consistent with the interpretations of this ternary relation proposed in the substructural literature.

The article is structured as follows. In Section 2 we recall the core of DEL viewed from a semantic perspective. In Section 3 we briefly recall elementary notions of relevance and substructural logics and we observe that the ternary relation of relevance logic can be interpreted as a sort of update. In Section 4 we proceed further to define a substructural language based on this idea. This substructural language extends the DEL language with operators stemming from the Lambek calculus (a substructural logic), but we show that these different substructural operators actually correspond to the DEL operators of (Aucher 2011; 2012). This allows us to show that DEL *is* a (two-sorted) substructural logic. In Section 5, we conclude.

2 Dynamic Epistemic Logic

Dynamic epistemic logic (DEL) is a relatively recent non-classical logic (Batlag et al. 1998) which extends ordinary modal epistemic logic (Hintikka 1962) by the inclusion of *event/action models* (called \mathcal{L}_{α} -models in this article) to describe actions, and a *product update* operator that defines how *epistemic models* (called \mathcal{L} -models in this article) are updated as the consequence of executing actions described through event models (see Baltag and Moss 2004, van Ditmarsch et al. 2007, van Benthem 2011b, for more details). So, the methodology of DEL is such that it splits the task of representing the agents' beliefs and knowledge into three parts: first, one repre-

sents their beliefs/knowledge about an initial situation; second, one represents their beliefs/knowledge about an event taking place in this situation; third, one represents the way the agents update their beliefs/knowledge about the situation after (or during) the occurrence of the event. Following this methodology, we also split the exposition of the DEL framework into three sections.

2.1 Representation of the initial situation: *L*-model

In the rest of this article, ATM is a countable set of propositional letters called *atomic facts* which describe static situations, and $AGT := \{1, ..., m\}$ is a finite set of agents.

Definition 2.1 (Language \mathcal{L} and \mathcal{L} -structure). We define the language \mathcal{L} inductively as follows:

$$\mathcal{L}:\varphi ::= p \mid \neg\varphi \mid \varphi \land \varphi \mid \varphi \lor \varphi \mid \Box_i \varphi$$

where *p* ranges over *ATM* and *j* over *AGT*. We define $\bot := p \land \neg p$ for a chosen $p \in ATM$ and we also define $\top := \neg \bot$. The formula $\diamondsuit_j \varphi$ is an abbreviation for $\neg \Box_j \neg \varphi$, the formula $\varphi \rightarrow \psi$ is an abbreviation for $\neg \varphi \lor \psi$, and the formula $\varphi \leftrightarrow \psi$ is an abbreviation for $(\varphi \rightarrow \psi) \land (\psi \rightarrow \varphi)$.

A *L*-structure is defined inductively as follows, with φ ranging over *L*:

$$X ::= \varphi \mid (X, X).$$

We abusively write $\varphi \in X$ when the formula $\varphi \in \mathcal{L}$ is a substructure of *X*.

A (pointed) \mathcal{L} -model (\mathcal{M} , w) represents how the actual world represented by w is perceived by the agents. Atomic facts are used to state properties of this actual world.

Definition 2.2 (\mathcal{L} -model). A \mathcal{L} -model is a tuple $\mathcal{M} = (W, R_1, \dots, R_m, I)$ where:

- W is a non-empty set of possible worlds,
- $R_j \subseteq W \times W$ is an accessibility relation on W, for each $j \in AGT$,
- $I: W \to 2^{ATM}$ is a function assigning to each possible world a subset of ATM. The function I is called an *interpretation*.

We write $w \in \mathcal{M}$ for $w \in W$, and (\mathcal{M}, w) is called a pointed \mathcal{L} -model (*w* often represents the actual world). We denote by *C* the set of pointed \mathcal{L} -models. If $w, v \in W$, we write $wR_i v$ or $(\mathcal{M}, w)R_i(\mathcal{M}, v)$ for $(w, v) \in R_i$, and $R_i(w)$ denotes the set $\{v \in W \mid wR_i v\}$.

Intuitively, wR_jv means that in world *w* agent *j* considers that world *v* might correspond to the actual world. Then, we define the following epistemic language that can be used to describe and state properties of \mathcal{L} -models:

Definition 2.3 (Truth conditions of \mathcal{L}). Let \mathcal{M} be a \mathcal{L} -model, $w \in \mathcal{M}$ and $\varphi \in \mathcal{L}$. $\mathcal{M}, w \models \varphi$ is defined inductively as follows:

 $\begin{array}{ll} \mathcal{M}, w \models p & \text{iff} \quad p \in I(w) \\ \mathcal{M}, w \models \neg \psi & \text{iff} \quad \text{not } \mathcal{M}, w \models \varphi \\ \mathcal{M}, w \models \varphi \land \psi & \text{iff} \quad \mathcal{M}, w \models \varphi \text{ and } \mathcal{M}, w \models \psi \\ \mathcal{M}, w \models \varphi \lor \psi & \text{iff} \quad \mathcal{M}, w \models \varphi \text{ or } \mathcal{M}, w \models \psi \\ \mathcal{M}, w \models \Box_i \varphi & \text{iff} \quad \text{for all } v \in R_i(w), \mathcal{M}, v \models \varphi \end{array}$

We write $\mathcal{M} \models \varphi$ when $\mathcal{M}, w \models \varphi$ for all $w \in \mathcal{M}$, and $\models \varphi$ when for all \mathcal{L} -model \mathcal{M} , $\mathcal{M} \models \varphi$. A \mathcal{L} -formula φ is said to be *valid* if $\models \varphi$. We extend the scope of the relation \models to also relate pointed \mathcal{L} -models to structures:

$$\mathcal{M}, w \models X, Y$$
 iff $\mathcal{M}, w \models X$ and $\mathcal{M}, w \models Y$

Let *C* be a class of pointed \mathcal{L} -models, let *X*, *Y* be \mathcal{L} -structures. We say that *X* entails *Y* in the class *C*, written $X \models Y$, when the following holds:

 $X \models_{\overline{C}} Y$ iff for all pointed \mathcal{L} -model $(\mathcal{M}, w) \in C$, if for all $\varphi \in X \mathcal{M}, w \models \varphi$, then there is $\psi \in Y$ such that $\mathcal{M}, w \models \psi$.

We also write $X \models Y$ for $X \models C$, where *C* is the class of all pointed \mathcal{L} -models.

The formula $\Box_j \varphi$ reads as "agent *j* believes φ ". Its truth conditions are defined in such a way that agent *j* believes φ is true in a possible world when φ holds in all the worlds agent *j* considers possible.

Example 1. Assume that agents A, B and C play a card game with three cards: a white one, a red one and a blue one. Each of them has a single card but they do not know the cards of the other players. At each step of the game, some of the players show their/her/his card to another player or to both other players, either privately or publicly. We want to study and represent the dynamics of the agents' beliefs/knowledge in this game. The initial situation is represented by the pointed \mathcal{L} -model (\mathcal{M} , w) of Figure 1.

In this example, $AGT := \{A, B, C\}$ and $ATM := \{r_j, b_j, w_j \mid j \in AGT\}$ where r_j stands for 'agent *j* has the *r*ed card', b_j stands for 'agent j has the *b*lue card' and w_j stands for 'agent *j* has the *w*hite card'. The boxed possible world corresponds to the actual world. The propositional letters not mentioned in the possible worlds do not hold



Figure 1: Cards Example

in these possible worlds. The accessibility relations are represented by arrows indexed by agents between possible worlds. Reflexive arrows are omitted in the figure, which means that for all worlds $v \in M$ and all agents $j \in AGT$, $v \in R_j(v)$. In this model, we have for example the following statement: $M, w \models (w_B \land \neg \Box_A w_B) \land \Box_C \neg \Box_A w_B$. It states that player A does not 'know' that player B has the white card and player C 'knows' it.

2.2 Representation of the event: \mathcal{L}_{α} -model

The language \mathcal{L}_{α} was introduced by Baltag et al. (1999). The propositional letters p_{ψ} describing events are called *atomic events* and range over $ATM_{\alpha} = \{p_{\psi} \mid \psi \text{ ranges over } \mathcal{L}\}$. The reading of p_{ψ} is "an event of precondition ψ is occurring".

Definition 2.4 (Language \mathcal{L}_{α} and \mathcal{L}_{α} -structure). We define the language \mathcal{L}_{α} inductively as follows:

$$\mathcal{L}_{\alpha}: \alpha ::= p_{\psi} \mid \neg \alpha \mid \alpha \land \alpha \mid \alpha \lor \alpha \mid \Box_{j} \alpha$$

where ψ ranges over \mathcal{L} and j over AGT. We define $\bot := p_{\psi} \land \neg p_{\psi}$ for a chosen $\psi \in \mathcal{L}$ and we define $\top := \neg \bot$. The formula $\diamondsuit_j \alpha$ is an abbreviation for $\neg \Box_j \neg \alpha$, the formula $\alpha \to \beta$ is an abbreviation for $\neg \alpha \lor \beta$, and the formula $\alpha \leftrightarrow \beta$ is an abbreviation for $(\alpha \to \beta) \land (\beta \to \alpha)$.

A \mathcal{L}_{α} -structure is defined inductively as follows, with β ranging over \mathcal{L}_{α} :

$$S_{\alpha}$$
: X_{α} ::= $\beta \mid (X_{\alpha}, X_{\alpha})$

We abusively write $\alpha \in X_{\alpha}$ when the formula $\alpha \in \mathcal{L}_{\alpha}$ is a substructure of X_{α} .

A pointed \mathcal{L}_{α} -model (\mathcal{E}, e) represents how the actual event represented by e is perceived by the agents. Intuitively, $f \in R_j(e)$ means that while the possible event represented by e is occurring, agent j considers possible that the possible event represented by f is actually occurring.

Definition 2.5 (\mathcal{L}_{α} -model, Batlag et al. 1998). A \mathcal{L}_{α} -model is a tuple $\mathcal{E} = (W_{\alpha}, R_1, \dots, R_m, I)$ where:

- W_{α} is a non-empty set of possible events,
- $R_j \subseteq W_{\alpha} \times W_{\alpha}$ is an accessibility relation on W_{α} , for each $j \in AGT$,
- *I*: *W_α* → *L* is a function assigning to each possible event a formula of *L*. The function *I* is called the *precondition function*.

Let *P* be a subset of \mathcal{L} . A *P*-complete \mathcal{L}_{α} -model is a \mathcal{L}_{α} -model which satisfies moreover the following condition:

• $I(e) \in P$, for each $e \in W_{\alpha}$. (*P*-complete)

We write $e \in \mathcal{E}$ for $e \in W_{\alpha}$, and (\mathcal{E}, e) is called a *pointed* \mathcal{L}_{α} -model (e often represents the actual event). We denote by C_{α} the set of pointed \mathcal{L}_{α} -models, by C_{α}^{P} the set of pointed *P*-complete event models. If $e, f \in W_{\alpha}$, we write $eR_{j}f$ or $(\mathcal{E}, e)R_{j}(\mathcal{E}, f)$ for $(e, f) \in R_{j}$, and $R_{j}(e)$ denotes the set $\{f \in W_{\alpha} \mid eR_{j}f\}$.

The truth conditions of the language \mathcal{L}_{α} are identical to the truth conditions of the language \mathcal{L} :

Definition 2.6 (Truth conditions of \mathcal{L}_{α}). Let \mathcal{E} be a \mathcal{L}_{α} -model, $e \in \mathcal{E}$ and $\alpha \in \mathcal{L}_{\alpha}$. $\mathcal{E}, e \models \alpha$ is defined inductively as follows:

$$\begin{array}{lll} \mathcal{E}, e \models p_{\psi} & \text{iff} & I(e) = \psi \\ \mathcal{E}, e \models \neg \alpha & \text{iff} & \text{not } \mathcal{E}, e \models \alpha \\ \mathcal{E}, e \models \alpha \land \beta & \text{iff} & \mathcal{E}, e \models \alpha \text{ and } \mathcal{E}, e \models \beta \\ \mathcal{E}, e \models \alpha \lor \beta & \text{iff} & \mathcal{E}, e \models \alpha \text{ or } \mathcal{E}, e \models \beta \\ \mathcal{E}, e \models \Box_{j} \alpha & \text{iff} & \text{for all } f \in R_{j}(e), \mathcal{E}, f \models \alpha \end{array}$$

Let *C* be a class of pointed \mathcal{L}_{α} -models, let X_{α}, Y_{α} be \mathcal{L}_{α} -structures. We say that *X* entails *Y* in the class *C*, written $X_{\alpha} \models Y_{\alpha}$, when the following holds:

$$X_{\alpha} \models Y_{\alpha}$$
 iff for all pointed \mathcal{L}_{α} -model $(\mathcal{E}, e) \in C$,
if for all $\alpha \in X_{\alpha} \mathcal{E}, e \models \alpha$, then there is $\beta \in Y_{\alpha}$ such that $\mathcal{E}, e \models \beta$.

We also write $X_{\alpha} \models Y_{\alpha}$ for $X_{\alpha} \models_{C_{\alpha}} Y_{\alpha}$, where C_{α} is the class of all pointed \mathcal{L}_{α} -models.



Figure 2: Players A and B show their cards to each other in front of player C

Example 2. Let us resume Example 1 and assume that players A and B show their card to each other. As it turns out, C noticed that A showed her card to B but did not notice that B did so to A. Players A and B know this. This event is represented in the \mathcal{L}_{α} -model (\mathcal{E}, e) of Figure 2. The boxed possible event *e* corresponds to the actual event 'player A shows her *r*ed card' (with precondition r_A), *f* stands for the event 'player A shows her *w*hite card' (with precondition w_A) and *g* stands for the atomic event 'players A and B show their *r*ed and *w*hite cards respectively to each other' (with precondition $r_A \wedge w_A$). The following statement holds in the example of Figure 2:

$$\mathcal{E}, e \models p_{r_A} \land (\Diamond_A p_{r_A} \land \Box_A p_{r_A}) \land (\Diamond_B p_{r_A} \land \Box_B p_{r_A}) \land (\Diamond_C p_{w_A} \land \Diamond_C p_{r_A \land w_A} \land \Box_C (p_{w_A} \lor p_{r_A \land w_A})).$$
(1)

It states that players A and B show their cards to each other, players A and B 'know' this and consider it possible, while player C considers possible that player A shows her white card and also considers possible that player A shows her red card, since he does not know her card. In fact, that is all that player C considers possible since he believes that either player A shows her red card or her white card.

2.3 Update of the initial situation by the event: Product Update

The DEL product update of Batlag et al. (1998) is defined as follows. This update yields a new \mathcal{L} -model $(\mathcal{M}, w) \otimes (\mathcal{E}, e)$ representing how the new situation which was previously represented by (\mathcal{M}, w) is perceived by the agents after the occurrence of the event represented by (\mathcal{E}, e) .

Definition 2.7 (Product Update). Let $(\mathcal{M}, w) = (W, R_1, \dots, R_m, I, w)$ be a pointed \mathcal{L} -model and let $(\mathcal{E}, e) = (W_\alpha, R_1, \dots, R_m, I, e)$ be a pointed \mathcal{L}_α -model such that



Figure 3: Situation after the update of the situation represented in Figure 1 by the event represented in Figure 2

 $\mathcal{M}, w \models I(e)$. The product update of (\mathcal{M}, w) and (\mathcal{E}, e) is the pointed \mathcal{L} -model $(\mathcal{M} \otimes \mathcal{E}, (w, e)) = (W^{\otimes}, R_1^{\otimes}, \dots, R_m^{\otimes}, I^{\otimes}, (w, e))$ defined as follows: for all $v \in W$ and all $f \in W_{\alpha}$,

- $W^{\otimes} = \{(v, f) \in W \times W_{\alpha} \mid \mathcal{M}, v \models I(f)\},\$
- $R_j^{\otimes}(v, f) = \{(u, g) \in W^{\otimes} \mid u \in R_j(v) \text{ and } g \in R_j(f)\},\$
- $I^{\otimes}(v, f) = I(v)$.

Example 3. As a result of the event described in Example 2, the agents update their beliefs. We get the situation represented in the \mathcal{L} -model $(\mathcal{M}, w) \otimes (\mathcal{E}, e)$ of Figure 3. In this \mathcal{L} -model, we have for example the following statement:

$$(\mathcal{M}, w) \otimes (\mathcal{E}, e) \models (w_B \wedge B_A w_B) \wedge B_C \neg B_A w_B.$$

It states that player A 'knows' that player B has the white card but player C believes that it is not the case.

3 Substructural logics

Substructural logics are a family of logics lacking some of the structural rules of classical logic. A structural rule is a rule of inference which is closed under substitution of formulas. We shall see in this article that DEL invalidates all of them.

3.1 A substructural language

Our exposition of substructural logics is based on (Restall 2000; 2006, Dunn and Restall 2002). The logical framework presented in (Restall 2000) is much more general and studies a wide range of substructural logics: relevance logic, linear logic, lambek calculus, display logic, etc... For what concerns us in this article, we will only introduce a fragment of this general framework. The semantics of this fragment is based on the ternary relation of the frame semantics for relevant logic originally introduced by Routley & Meyer (Routley and Meyer 1972a;b; 1973, Routley et al. 1982).

Definition 3.1 (Language \mathcal{L}_{Sub} and \mathcal{L}_{Sub} -structure). The *language* \mathcal{L}_{Sub} is defined inductively as follows:

where p ranges over ATM.

A \mathcal{L}_{Sub} -structure is defined inductively as follows, with φ ranging over \mathcal{L}_{Sub} :

$$X ::= \varphi \mid (X, X) \mid (X; X)$$

Definition 3.2 (Point set, plump accessibility relation). A *point set* $\mathcal{P} = (P, \sqsubseteq)$ is a set *P* together with a partial order \sqsubseteq on *P*. The set $Prop(\mathcal{P})$ of *propositions* on \mathcal{P} is the set of all subsets *X* of *P* which are *closed upwards*: that is, if $x \in X$ and $x \sqsubseteq x'$ then $x' \in X$. We abusively write $x \in \mathcal{P}$ for $x \in P$.

- A binary relation S is a *positive two-place accessibility relation* on the point set \mathcal{P} iff for any $x, y \in \mathcal{P}$ where xSy, if $x' \sqsubseteq x$ then there is a $y' \sqsupseteq y$, where x'Sy'. Similarly, if xSy and $y \sqsubseteq y'$ then there is some $x' \sqsubseteq x$, where x'Sy'.
- A ternary relation \mathcal{R} is a *three-place accessibility relation* iff whenever $\mathcal{R}xyz$ and $z \sqsubseteq z'$ then there are $y' \sqsupseteq y$ and $x' \sqsubseteq x$, where $\mathcal{R}x'y'z'$. Similarly, if $x' \sqsubseteq x$ then there are $y' \sqsubseteq y$ and $z' \sqsupseteq z$, where $\mathcal{R}x'y'z'$, and if $y' \sqsubseteq y$ then there are $x' \sqsubseteq x$ and $z' \sqsupseteq z$, where $\mathcal{R}x'y'z'$.
- A ternary relation \mathcal{R} is a *plump accessibility relation* on the point set \mathcal{P} if and only if for any $x, y, z, x', y', z' \in \mathcal{P}$ such that $\mathcal{R}xyz$, if $x' \sqsubseteq x, y' \sqsubseteq y$ and $z \sqsubseteq z'$, then $\mathcal{R}x'y'z'$.

Our definition of \mathcal{L}_{Sub} -model corresponds to the definition of a *model* in (Restall 2000, Chap. 11) stripped out from all its truth sets. These other features are not needed for what concerns us here.

Definition 3.3 (\mathcal{L}_{Sub} -model). A \mathcal{L}_{Sub} -model is a tuple $\mathcal{M}_{\mathcal{R}} = (\mathcal{P}, \mathcal{S}, \mathcal{R}, \mathcal{I})$ where:

- $\mathcal{P} = (P, \sqsubseteq)$ is a point set;
- $S \subseteq \mathcal{P} \times \mathcal{P}$ is a positive two-place accessibility relation on \mathcal{P} ;
- $\mathcal{R} \subseteq \mathcal{P} \times \mathcal{P} \times \mathcal{P}$ is a three–place accessibility relation on \mathcal{P} ;
- $I: P \to 2^{ATM}$ is an interpretation function.

We abusively write $x \in \mathcal{M}_{\mathcal{R}}$ for $x \in \mathcal{P}$, and $(\mathcal{M}_{\mathcal{R}}, x)$ is called a *pointed* \mathcal{L}_{Sub} -model.

Note that in the above definition, there could be multiple positive two–place accessibility relations S_1, \ldots, S_n corresponding to multiple modalities $\Box_1, \ldots \Box_n$. We refrain from defining \mathcal{L}_{Sub} –models in their full generality in order to ease the readability of the article.

Definition 3.4 (Truth conditions of \mathcal{L}_{Sub}). Let $\mathcal{M}_{\mathcal{R}}$ be a \mathcal{L}_{Sub} -model, $x \in \mathcal{M}_{\mathcal{R}}$ and $\varphi \in \mathcal{L}_{Sub}$. The relation $\mathcal{M}_{\mathcal{R}}, x \parallel - \varphi$ is defined inductively as follows:

| $\mathcal{M}_{\mathcal{R}}, x \models \top$ | | always |
|---|-----|---|
| $\mathcal{M}_{\mathcal{R}}, x \parallel \perp$ | | never |
| $\mathcal{M}_{\mathcal{R}}, x \Vdash p$ | iff | $p \in \mathcal{I}(x)$ |
| $\mathcal{M}_{\mathcal{R}}, x \Vdash \neg \varphi$ | iff | not $\mathcal{M}_{\mathcal{R}}, x \Vdash \varphi$ |
| $\mathcal{M}_{\mathcal{R}}, x \models \varphi \land \psi$ | iff | $\mathcal{M}_{\mathcal{R}}, x \models \varphi$ and $\mathcal{M}_{\mathcal{R}}, x \models \psi$ |
| $\mathcal{M}_{\mathcal{R}}, x \models \varphi \lor \psi$ | iff | $\mathcal{M}_{\mathcal{R}}, x \models \varphi \text{ or } \mathcal{M}_{\mathcal{R}}, x \models \psi$ |
| $\mathcal{M}_{\mathcal{R}}, x \models \Box \varphi$ | iff | for all $y \in \mathcal{M}_{\mathcal{R}}$, where $xSy, \mathcal{M}_{\mathcal{R}}, y \models \varphi$ |
| $\mathcal{M}_{\mathcal{R}}, x \models \varphi \supset \psi$ | iff | for all $y, z \in \mathcal{P}$ where $\mathcal{R}xyz$, |
| | | if $\mathcal{M}_{\mathcal{R}}, y \models \varphi$ then $\mathcal{M}_{\mathcal{R}}, z \models \psi$ |
| $\mathcal{M}_{\mathcal{R}}, x \models \psi \subset \varphi$ | iff | for all $y, z \in \mathcal{P}$ where $\mathcal{R}yxz$ |
| | | if $\mathcal{M}_{\mathcal{R}}, y \models \varphi$ then $\mathcal{M}_{\mathcal{R}}, z \models \psi$ |
| $\mathcal{M}_{\mathcal{R}}, x \models \varphi \circ \psi$ | iff | there are $y, z \in \mathcal{P}$ such that $\mathcal{R}yzx$, |
| | | $\mathcal{M}_{\mathcal{R}}, y \models \varphi$ and $\mathcal{M}_{\mathcal{R}}, z \models \psi$ |
| | | |

We extend the scope of the relation \parallel to also relate points to \mathcal{L}_{Sub} -structures:

 $\begin{array}{lll} \mathcal{M}_{\mathcal{R}}, x \models X, Y & \text{iff} & \mathcal{M}_{\mathcal{R}}, x \models X \text{ and } \mathcal{M}_{\mathcal{R}}, x \models Y \\ \mathcal{M}_{\mathcal{R}}, x \models X, Y & \text{iff} & \text{there are } y, z \in \mathcal{M}_{\mathcal{R}} \text{ such that } \mathcal{R}yzx, \\ \mathcal{M}_{\mathcal{R}}, y \models X \text{ and } \mathcal{M}_{\mathcal{R}}, z \models Y \end{array}$

We say that $\mathcal{M}_{\mathcal{R}}$ validates a \mathcal{L}_{Sub} -structure X when for all $x \in \mathcal{M}_{\mathcal{R}}, \mathcal{M}_{\mathcal{R}}, x \models X$. Let X be a structure and let $\varphi \in \mathcal{L}_{Sub}$. We say that X entails φ , written X $\models \varphi$, when the following holds:

 $X \parallel - \varphi$ iff for all pointed \mathcal{L}_{Sub} -model $(\mathcal{M}_{\mathcal{R}}, x)$, if $\mathcal{M}_{\mathcal{R}}, x \parallel - X$, then $\mathcal{M}_{\mathcal{R}}, x \parallel - \varphi$.

3.2 Updates as ternary relations

The ternary relation \mathcal{R} of the Routley & Meyer semantics was introduced originally for technical reasons: any 2-ary (*n*-ary) connective of a logical language can be given a semantics by resorting to a 3-ary (resp. n + 1-ary) relation on worlds. Subsequently, a number of philosophical interpretations of this ternary relation have been proposed and we will briefly recall some of them at the end of this section (see Beall et al. 2012, Restall 2006, Mares and Meyer 2001, for more details). However, one has to admit that providing a non-circular and conceptually grounded interpretation of this relation remains problematic. In this article, we propose a new *dynamic* interpretation of this relation, inspired by the ternary semantics of DEL.

First, one should observe that the DEL product update \otimes of Definition 2.7 can be seen as a partial function \uparrow from a pair of pointed \mathcal{L} -model and pointed \mathcal{L}_{α} -model to another pointed \mathcal{L} -model:

$$\uparrow: C \times C_{\alpha} \to C. \tag{2}$$

There is a formal similarity between this abstract definition of the DEL product update and the function \sqcup introduced by Urquhart in the early 1970s for providing a semantics to the implication of relevance logic. This similarity is not only formal but also intuitively meaningful. Indeed, the intuitive interpretation of the DEL product update operator is very similar to the intuitive interpretation of the function \sqcup of Urquhart. Points are sometimes also called "worlds", "states", "situation", "set-ups", and as explained by Restall:

"We have a class of points (over which x and y vary), and a function \sqcup which gives us new points from old. The point $x \sqcup y$ is supposed, on Urquhart's interpretation, to be the body of information given by combining x with y." (Restall 2006, p. 363)

and also,

"To be committed to $A \supset B$ is to be committed to *B* whenever we gain the information that *A*. To put it another way, a body of information warrants $A \supset B$ if and only if whenever you *update* that information with new information which warrants *A*, the resulting (perhaps new) body of information warrants *B*." (my emphasis) (Restall 2006, p. 362)

From these two quotes, it is natural to interpret the DEL product update \otimes of Definition 2.7 as a specific kind of Urquhart's function \sqcup . Moreover, as explained by Restall, this substructural "update" can be nonmonotonic and may correspond to some sort of *revision*:

"[C]ombination is sometimes nonmonotonic in a natural sense. Sometimes when a body of information is combined with another body of information, some of the original body of information might be lost. This is simplest to see in the case motivating the failure of $A \models B \supset A$. A body of information might tell us that A. However, when we combine it with something which tells us B, the resulting body of information might no longer warrant A (as A might with B). Combination might not simply result in the addition of information. It may well warrant its *revision*." (my emphasis) (Restall 2006, p. 363)

Our dynamic interpretation of the ternary relation is consistent with the above considerations: sometimes, updating beliefs amounts to *revise* beliefs. As it turns out, belief revision has also been extensively studied within the DEL framework and DEL has been extended to deal with this phenomenon (Aucher 2004, van Ditmarsch 2005, van Benthem 2007a, Baltag and Smets 2008a;b, Liu 2008, Aucher 2008).

More generally, an update can be seen as a *partial* function \uparrow from a pair of pointed \mathcal{L} -model and pointed \mathcal{L}_{α} -model to a *set* of pointed \mathcal{L} -model:

$$\uparrow: C \times C_{\alpha} \to \mathcal{P}(C) \tag{3}$$

Equivalently, an update can be seen as a ternary relation \mathcal{R} defined on $\mathcal{C} \cup \mathcal{C}_{\alpha}$ between three pointed models $((\mathcal{M}, w), (\mathcal{E}, e), (\mathcal{M}_f, w_f))$ where (\mathcal{M}, w) is a pointed \mathcal{L}_{-} model, (\mathcal{E}, e) is a pointed \mathcal{L}_{α} -model and (\mathcal{M}_f, w_f) is another pointed \mathcal{L} -model:

$$\mathcal{R} \subseteq \mathcal{C} \times \mathcal{C}_{\alpha} \times \mathcal{C} \tag{4}$$

The ternary relation of Equation (4) then resembles the ternary relation of the Routley & Meyer semantics. This is not surprising since the Routley & Meyer semantics generalizes the Urquhart semantics (they are essentially the same, since as we explained it in the previous section, an operational frame is a Routley & Meyer frame where $\Re xyz$ holds if and only if $x \sqcup y = z$). Viewed from the perspective of DEL, the ternary relation then represents a particular sort of *update*. With this interpretation in mind, $\Re xyz$ reads as 'the occurrence of event y in world x results in the world z' and the corresponding conditional $\alpha \supset \varphi$ reads as 'the occurrence in the current world of an event satisfying property α results in a world satisfying φ' .

The dynamic reading of the ternary relation and its corresponding conditional is very much in line with the so-called "Ramsey Test" of conditional logic. The Ramsey test can be viewed as the very first modern contribution to the logical study of conditionals and much of the contemporary work on conditional logic can be traced back to the famous footnote of Ramsey (1929). Roughly, it consists in defining a counterfactual conditional in terms of belief revision: an agent currently believes that φ would

be true if ψ were true (i.e. $\psi \supset \varphi$) if and only if he should believe φ after *learning* ψ . A first attempt to provide truth conditions for conditionals, based on Ramsey's ideas, was proposed by Stalnaker. He defined his semantics by means of selection functions over possible worlds $f : W \times 2^W \to W$. As one can easily notice, Stalnaker's selection functions could also be considered from a formal point of view as a special kind of ternary relation, since a relation $\mathcal{R}_f \subseteq W \times 2^W \times W$ can be canonically associated to each selection function f. Moreover, like the ternary relation corresponding to a product update (Equation (4)), this ternary relation is 'two-sorted': the antecedent of a conditional takes value in a set of worlds (instead of a single world).¹ So, the dynamic reading of the ternary semantics is consistent with the dynamic reading of conditionals proposed by Ramsey.

This dynamic reading was not really considered and investigated by substructural logicians when they connected the substructural ternary semantics with conditional logic (Beall et al. 2012). On the other hand, the dynamic reading of inferences has been stressed to a large extent by van Benthem (2007b; 2011b) (we will come back to this point in Section 4.2), and also by Baltag and Smets who distinguished *dynamic* belief revision from *static* (standard) belief revision (Baltag and Smets 2006; 2008a;b). What distinguishes *dynamic* belief revision from *static* belief revision is that the latter is a revision of the agent's beliefs about the state of the world as it was before an event, and the former is a revision of the state of the world as it is after the event. Note, however, that this important distinction between static belief revision and dynamic belief revision collapses in the case of relevant logic, because in that case we only deal with propositional formulas. This shows again that a *dynamic* interpretations proposed by substructural logicians.

To summarize our discussion, the DEL product update provides substructural logics with an intuitive and consistent interpretation of its ternary relation. This interpretation is consistent in the sense that the intuitions underlying the definitions of the DEL framework are coherent with those underlying the ternary semantics of substructural logic, as witnessed by our quotes and citations from the substructural literature.

Other interpretations of the ternary relation One interpretation, due to Barwise (1993) and developed by Restall (1996), takes worlds to be 'sites' or 'channels', a site being possibly a channel and a channel being possibly a site. If x, y and z are sites, $\Re xyz$ reads as 'x is a channel between y and z'. Hence, if $\varphi \supset \psi$ is true at channel

¹Note that Burgess (1981) already proposed a ternary semantics for conditionals, but his truth conditions and his interpretation of the ternary relation were quite different from ours.

x, it means that all sites y and z connected by channel x are such that if φ is information available in y, then ψ is information available in z. Another similar interpretation due to Mares (1996) adapts Perry and Israel (1990)'s theory of information to the relational semantics. In this interpretation, worlds are situations in the sense of Barwise and Perry (1983)'s situation semantics and pieces of information – called *infons* – can carry information about other infons: an infon might carry the information that a red light on a mobile phone carries the information that the battery of the mobile phone is low. In this interpretation, the ternary relation \mathcal{R} represents the informational links in situations: if there is an informational link in situation x that says that an infon σ carries the information that the infon π also holds, then if $\mathcal{R}xyz$ holds and y contains the infon σ , then z contains the infon π . Other interpretations of the ternary relation have been proposed by Beall et al. (2012), with a particular focus on their relation to conditionality.

4 DEL is a substructural logic

In this section, we will extend the languages \mathcal{L} and \mathcal{L}_{α} of Section 2 with the substructural operators \circ, \supset and \subset . We will also provide a substructural semantics for this language based on the idea to view an update as a ternary relation of a substructural frame (\mathcal{L}_{Sub} -model). This idea is motivated and intuitively grounded in the analysis of the previous section.

4.1 An extended DEL language

Our language extends both the language \mathcal{L} and the language \mathcal{L}_{α} of Section 2. Like our semantics, it is two-sorted: it contains both formulas of \mathcal{L} and formulas of \mathcal{L}_{α} .

Definition 4.1 (Language $\mathcal{L}_{\mathcal{R}}$). The *language* $\mathcal{L}_{\mathcal{R}}$ is two-sorted and is defined by a double induction as follows:

where *p* ranges over ATM, ψ ranges over $\mathcal{L}^{1}_{\mathcal{R}}$ and *j* over AGT. The abbreviations $\varphi \to \psi, \varphi \leftrightarrow \psi$ and $\alpha \to \beta, \alpha \leftrightarrow \beta$ are defined as in Definitions 2.1 and 2.4.

Definition 4.2 ($\mathcal{L}_{\mathcal{R}}$ -structure and $\mathcal{L}_{\mathcal{R}}$ -sequent). The $\mathcal{L}_{\mathcal{R}}$ -structures are defined inductively as follows:

where φ ranges over $\mathcal{L}_{\mathcal{R}}$ and X_{α} ranges over \mathcal{L}_{α} -structures. A $\mathcal{L}_{\mathcal{R}}$ -sequent is a \mathcal{L}_{α} -sequent or an expression of the form $X \vdash Y$, where $X \in S_1, Y \in S_2$.

Definition 4.3 (DEL product update model). The *DEL product update model* is the tuple $\mathcal{M}_{\otimes} = (\mathcal{P}, \mathcal{R}_1, \dots, \mathcal{R}_m, \mathcal{R}_{\otimes}, \mathcal{I})$ where:

- $\mathcal{P} := (\mathcal{C} \cup \mathcal{C}_{\alpha}, \rightleftharpoons)$ where \Leftrightarrow is the bisimilarity relation;
- $\mathcal{R}_j \subseteq \mathcal{P} \times \mathcal{P}$ is a positive two-place accessibility relation on \mathcal{P} for each $j \in AGT$ such that for all $x, y \in \mathcal{P}$, where $x = (\mathcal{M}_x, w_x)$ and $y = (\mathcal{M}_y, w_y)$:

 $x \in \mathcal{R}_i(y)$ iff $\mathcal{M}_x = \mathcal{M}_y$ and $w_x \in \mathcal{R}_i(w_y)$.

- $\mathcal{R}_{\otimes} := \{(x, y, z) \in C \times C_{\alpha} \times C \mid x \otimes y = z\}$ is a plump ternary relation on \mathcal{P} ;
- I(x) := I(x), for all $x \in C \cup C_{\alpha}$.

The DEL product update model is a \mathcal{L}_{Sub} -model where points are pointed \mathcal{L}_{-} models and pointed \mathcal{L}_{α} -models. The ternary relation \mathcal{R}_{\otimes} is defined and motivated by the explanations of the previous section. Note that the accessibility relations R_j of \mathcal{L} -models and \mathcal{L}_{α} -models are seen in this definition as positive two-place accessibility relations \mathcal{R}_j . The truth conditions are the same as the ones for $\mathcal{L}_{\mathcal{R}}$ -models:

Definition 4.4 (Truth conditions of $\mathcal{L}_{\mathcal{R}}$). Let \mathcal{M}_{\otimes} be the DEL product update model, $x \in \mathcal{M}_{\otimes}$ and $\varphi \in \mathcal{L}_{\mathcal{R}}$. The relation $\mathcal{M}_{\otimes}, x \models \varphi$ is defined inductively as follows:

| $\mathcal{M}_{\otimes}, x \Vdash p$ | iff | $p \in \mathcal{I}(x)$ |
|---|-----|--|
| $\mathcal{M}_{\otimes}, x \Vdash \neg \varphi$ | iff | not $\mathcal{M}_{\otimes}, x \Vdash \varphi$ |
| $\mathcal{M}_{\otimes}, x \models \varphi \land \psi$ | iff | $\mathcal{M}_{\otimes}, x \Vdash \varphi$ and $\mathcal{M}_{\otimes}, x \Vdash \psi$ |
| $\mathcal{M}_{\otimes}, x \models \varphi \lor \psi$ | iff | $\mathcal{M}_{\otimes}, x \models \varphi \text{ or } \mathcal{M}_{\otimes}, x \models \psi$ |
| $\mathcal{M}_{\otimes}, x \models \Box_j \varphi$ | iff | for all $y \in \mathcal{P}$ such that $x\mathcal{R}_j y, \mathcal{M}_{\otimes}, y \models \varphi$ |
| $\mathcal{M}_{\otimes}, x \models \alpha \supset \psi$ | iff | for all $y, z \in \mathcal{P}$ such that $\mathcal{R}_{\otimes} xyz$, |
| | | if $\mathcal{M}_{\otimes}, y \models \alpha$ then $\mathcal{M}_{\otimes}, z \models \psi$ |
| $\mathcal{M}_{\otimes}, x \models \psi \subset \varphi$ | iff | for all $y, z \in \mathcal{P}$ such that $\mathcal{R}_{\otimes} yxz$, |
| | | if $\mathcal{M}_{\otimes}, y \models \varphi$ then $\mathcal{M}_{\otimes}, z \models \psi$ |
| $\mathcal{M}_{\otimes}, x \models \varphi \circ \alpha$ | iff | there are $y, z \in \mathcal{P}$ such that $\mathcal{R}_{\otimes}yzx$, |
| | | $\mathcal{M}_{\otimes}, y \models \varphi$ and $\mathcal{M}_{\otimes}, z \models \alpha$ |

We extend the scope of the relation \parallel to also relate points to $\mathcal{L}_{\mathcal{R}}$ -structures:

$$\begin{array}{lll} \mathcal{M}_{\otimes}, x \Vdash X, Y & \text{iff} & \mathcal{M}_{\otimes}, x \Vdash X \text{ and } \mathcal{M}_{\otimes}, x \Vdash Y \\ \mathcal{M}_{\otimes}, x \Vdash X; Y & \text{iff} & \text{there are } y, z \in \mathcal{M}_{\mathcal{R}} \text{ such that } \mathcal{R}yzx, \\ \mathcal{M}_{\otimes}, y \Vdash X \text{ and } \mathcal{M}_{\otimes}, z \Vdash Y \end{array}$$

Let $C \subseteq C \cup C_{\alpha}$ be a class of pointed \mathcal{L} -models or \mathcal{L}_{α} -models, and let $X \models \varphi$ be a $\mathcal{L}_{\mathcal{R}}$ -sequent. We say that X entails φ in the class C, written $X \parallel_{\overline{C}} \varphi$, when the following holds:

 $X \parallel_{\overline{C}} \varphi$ iff for all $x \in C$, if $\mathcal{M}_{\otimes}, x \Vdash X$ then $\mathcal{M}_{\otimes}, x \Vdash \varphi$.

We also write $X \parallel -\varphi$ for $X \parallel_{\overline{C \cup C_{\alpha}}} \varphi$.

4.2 DEL operators are substructural operators

In this section, we will show that the DEL operators introduced in (Aucher 2011; 2012) correspond to the substructural operators \circ, \supset and \subset .

Recently again, van Benthem (2010) expressed some worries about interpreting the Lambek Calculus (the paradigmatic substructural logic) as a base logic of information flow while trying to connect the operators \circ , \supset and \subset of substructural logic to some sort of DEL operators. Indeed, the DEL operators usually rely on the regular algebra of sequential composition, choice and iteration which are of a quite different nature. Recently, I introduced some DEL operators called progression, regression and epistemic planning (Aucher 2011; 2012), the operator of regression being a natural generalization of the standard and original action modality [\mathcal{E} , e] φ of DEL (Batlag et al. 1998). It turns out that these operators can all be identified with connectives of the substructural language $\mathcal{L}_{\mathcal{R}}$. We first briefly recall their definitions below and then we give our correspondence results between the two kinds of operators.

Progression The operator of *progression* is denoted \otimes in (Aucher 2011). In (Aucher 2012, Def. 41), a constructive definition of this operator is provided using characteristic formulas (called "Kit Fine" formulas). Here, we provide an alternative and non–constructive definition of the *progression* of φ by α , denoted $\varphi \otimes \alpha$:

Theorem 1. Let (\mathcal{M}_f, w_f) be a pointed \mathcal{L} -model and let $\varphi \in \mathcal{L}$ and $\alpha \in \mathcal{L}_{\alpha}$. Then,

$$\begin{aligned} \mathcal{M}_{f}, w_{f} \models \varphi \otimes \alpha & \text{iff} \quad \text{there is a pointed } \mathcal{L}\text{-model} \left(\mathcal{M}, w\right) \text{ and a pointed} \\ \mathcal{L}_{\alpha}\text{-model} \left(\mathcal{E}, e\right) \text{ such that } \left(\mathcal{M}, w\right) \otimes \left(\mathcal{E}, e\right) & \cong \left(\mathcal{M}_{f}, w_{f}\right), \\ \mathcal{M}, w \models \varphi \text{ and } \mathcal{E}, e \models \alpha. \end{aligned}$$

Proof. It follows from Lemmata 43 and 44 of (Aucher 2011).

Epistemic planning The operator of *epistemic planning* is denoted \bigotimes_P in (Aucher 2012). It is defined relatively to a finite set *P* of formulas/preconditions/atomic events. In (Aucher 2012, Def. 14–15), a constructive definition of this operator is provided

using characteristic formulas (called "Kit Fine" formulas). As it turns out, an alternative and non–constructive definition of the *epistemic planning from* φ *to* φ_f , denoted $\varphi \otimes_P \varphi_f$, exists as well:

Theorem 2 (Aucher 2012). Let $\varphi, \varphi_f \in \mathcal{L}$ and let P be a finite subset of \mathcal{L} . Then, for all P-complete \mathcal{L}_{α} -model (\mathcal{E}, e), it holds that

$$\mathcal{E}, e \models \varphi \otimes_P \varphi_f \quad iff \quad \begin{array}{l} \text{there is } (\mathcal{M}, w) \text{ such that } \mathcal{M}, w \models \varphi, \\ \mathcal{M}, w \models I(e) \text{ and } (\mathcal{M}, w) \otimes (\mathcal{E}, e) \models \varphi_f. \end{array}$$

The dual of the operator $\varphi \otimes_P \varphi_f$ is defined by:

$$\varphi[\otimes]_P \varphi_f := \neg(\varphi \otimes_P \neg \varphi_f). \tag{5}$$

Theorem 2 entails that $\varphi[\odot]_P \varphi_f$ can be alternatively defined as follows: for all *P*-complete \mathcal{L}_{α} -model (\mathcal{E}, e), it holds that

$$\mathcal{E}, e \models \varphi[\bigcirc]_P \varphi_f \quad \text{iff} \quad \text{for all} \quad (\mathcal{M}, w) \text{ such that } \quad \mathcal{M}, w \models \varphi, \text{ if} \qquad (6)$$
$$\mathcal{M}, w \models I(e) \text{ then } (\mathcal{M}, w) \otimes (\mathcal{E}, e) \models \varphi_f.$$

Example 4. In the situation depicted in the \mathcal{L} -model of Figure 1, agent B does not know that agent A has the red card and does not know that agent C has the blue card: $\mathcal{M}, w \models (\diamond_B r_A \land \diamond_B \neg r_A) \land (\diamond_B b_C \land \diamond_B \neg b_C)$. Our problem is therefore the following:

What sufficient and necessary property (i.e. 'minimal' property) an event should fulfill so that its occurence in the initial situation (\mathcal{M} , w) results in a situation where agent B knows the true state of the world, i.e. agent B knows that agent A has the red card and that agent C has the blue card?

The answer to this question obviously depends on the kind of atomic events we consider. In this example, the events $P = \{p_{b_C}, p_{r_A}, p_{w_B}\}$ under consideration are the following. First, agent C shows her blue card (p_{b_C}) , second, agent A shows her red card (p_{r_A}) , and third, agent B herself shows her white card (p_{w_B}) . Answering this question amounts to compute the formula $(M, w) \otimes_P \Box_B (r_A \wedge b_C \wedge w_B)$. Applying the algorithm of (Aucher 2012, Definition 15), we obtain that

$$(\mathcal{M}, w) \otimes_P \Box_B (r_A \wedge b_C \wedge w_B) \leftrightarrow \Box_B (p_{b_C} \vee p_{r_A})$$
 is valid.

In other words, this result states that agent B should believe either that agent A shows her red card or that agent C shows her blue card in order to know the true state of the world. Indeed, since there are only three different cards which are known by the agents and agent B already knows her card, if she learns the card of (at least) one of the other agents, she will also be able to infer the card of the third agent.

Regression The operator of *regression* is denoted \oslash in (Aucher 2011). In (Aucher 2012, Def. 41), a constructive definition of this operator is provided using characteristic formulas (called "Kit Fine" formulas) by adapting and translating the reduction axioms of Batlag et al. (1998). As it turns out, an alternative and non–constructive definition of the *regression* of φ_f by α , denoted $\alpha \oslash \varphi_f$, exists as well:

Theorem 3. Let $\alpha \in \mathcal{L}_{\alpha}$ and $\varphi_f \in \mathcal{L}$. Then, for all \mathcal{L} -model (\mathcal{M}, w) , it holds that

$$\mathcal{M}, w \models \alpha \oslash \varphi_f \quad iff \quad \begin{array}{l} \text{there is } (\mathcal{E}, e) \text{ such that } \mathcal{E}, e \models \alpha, \\ \mathcal{M}, w \models I(e) \text{ and } (\mathcal{M}, w) \otimes (\mathcal{E}, e) \models \varphi_f. \end{array}$$

Note that we could define a dual operator of $\alpha \oslash \varphi_f$ as follows:

$$\alpha[\oslash]\varphi_f = \neg \left(\alpha \oslash \neg \varphi_f\right). \tag{7}$$

Then, the counterpart of Theorem 3 for this dual operator is as follows:

$$\mathcal{M}, w \models \alpha[\emptyset]\varphi_f \quad \text{iff} \quad \begin{cases} \text{for all } (\mathcal{E}, e) \text{ such that } \mathcal{E}, e \models \alpha, \\ \text{if } \mathcal{M}, w \models I(e) \text{ then } (\mathcal{M}, w) \otimes (\mathcal{E}, e) \models \varphi_f. \end{cases}$$
(8)

As shown in (Aucher 2012, Sec. 6), the operator $\alpha[\emptyset]\varphi_f$ is a generalization of the original and more standard DEL operator $[\mathcal{E}, e]\varphi$ almost exclusively used in the DEL literature (Batlag et al. 1998).

Correspondence between DEL and substructural operators As one can easily notice, there is a strong similarity between the operations of progression, epistemic planning and regression and the operations of substructural logic, more precisely of the Lambek Calculus. In fact, there exists a rigorous mapping between them, as the following theorem shows:

Theorem 4. Let P be a finite subset of \mathcal{L} , let $x = (\mathcal{M}, w) \in C$ and let $y = (\mathcal{E}, e) \in C_{\alpha}^{P}$ be a P-complete pointed event model. Let $\varphi, \psi \in \mathcal{L}$ and let $\alpha \in \mathcal{L}_{\alpha}$. Then,

$$\begin{array}{ll} \mathcal{M}_{\otimes}, x \models \varphi \circ \alpha & \textit{iff} & \mathcal{M}, w \models \varphi \otimes \alpha \\ \mathcal{M}_{\otimes}, x \models \alpha \supset \varphi & \textit{iff} & \mathcal{M}, w \models \alpha[\oslash] \varphi \\ \mathcal{M}_{\otimes}, y \models \psi \subset \varphi & \textit{iff} & \mathcal{E}, e \models \varphi[\odot]_{P} \psi \end{array}$$

Moreover, for all $\alpha, \alpha_1, \ldots, \alpha_n \in \mathcal{L}_{\alpha}$ *, for all* $\varphi, \psi, \varphi_0, \varphi_1, \ldots, \varphi_n \in \mathcal{L}$ *, we have:*

 $\varphi; \alpha \Vdash \psi \qquad iff \quad \varphi, \alpha \models \psi \\ (((\varphi_0; \alpha_1), \varphi_1); \dots; \alpha_n), \varphi_n \Vdash \psi \qquad iff \quad \varphi_0, \alpha_1, \varphi_1, \dots, \alpha_n, \varphi_n \models \psi$

| Substructural operators | DEL operators |
|-------------------------|---------------|
| 0 | \otimes |
| \supset | [Ø] |
| C | $[\bigcirc]$ |

Figure 4: Correspondence between DEL and substructural operators

The key Theorem 42 of (Aucher 2011) relates DEL–sequents and the operator of progression: for all $\varphi, \varphi_f \in \mathcal{L}$ and $\alpha \in \mathcal{L}_{\alpha}$, it holds that

$$\varphi, \alpha \models \varphi_f \text{ iff } \varphi \otimes \alpha \models \varphi_f. \tag{9}$$

As it turns out, this theorem is also valid in any substructural logic: it corresponds to a theorem of the Lambek calculus. More generally, all the theorems of the nonassociative Lambek calculus hold in our DEL setting if we use the translation given in Figure 4. In particular, if *P* be a finite subset of \mathcal{L} , then for all $\varphi, \varphi_f \in \mathcal{L}$ and $\alpha \in \mathcal{L}_{\alpha}$, it holds that

$$\varphi; \alpha \Vdash \varphi_f \quad \text{iff} \quad \varphi \Vdash \alpha[\emptyset] \varphi_f \tag{10}$$

$$\varphi \Vdash \alpha[\oslash] \varphi_f \quad \text{iff} \quad \varphi \otimes \alpha \Vdash \varphi_f \tag{11}$$

$$\varphi \otimes \alpha \models \varphi_f \quad \text{iff} \quad \alpha \models_{C^P} \varphi[\heartsuit]_P \varphi_f$$

$$\tag{12}$$

$$\varphi \Vdash \alpha[\oslash] \varphi_f \quad \text{iff} \quad \alpha \parallel_{\overline{C^P}} \varphi[\oslash]_P \varphi_f \tag{13}$$

5 Conclusion

We proved in this article that DEL is a two-sorted substructural logic. Also, we argued in Section 3.2 that our embedding of DEL within the framework of substructural logic is intuitively consistent, in the sense that in this embedding the intuitions underlying the DEL framework are coherent with the intuitive interpretations proposed for the ternary semantics of substructural logics. This may explain to a certain extent *why* some substructural phenomena arise in the dynamic inferences of van Benthem (2008): "it seemed that structural rules address mere *symptoms* of some underlying phenomenon" (van Benthem 2011b, p. 297). I claim that these "symptoms" are caused at a deeper semantic level by the fact that an update, and in that case the DEL product update, can be represented by the ternary relation of substructural logics.

In a certain sense, this article is in line with (van Benthem 2008; 2011b) and contributes to relate even more closely the programs of Logical Pluralism (Beall and Restall 2006) and Logical Dynamics (van Benthem 2011b). Roughly, the informal idea underpinning the connection between these two logical paradigms is to consider different reasoning styles and their corresponding consequence relations as the result of different sorts of updates induced by various informational tasks (such as observation, memory, questions and answers, dialogue, or general communication). We showed that this approach is not only meaningful from an intuitive point of view, but it can also be realized at a formal level if the ternary relation of substructural logic is interpreted intuitively as a sort of update. So, we hope that our embedding will strengthen the connections between the two areas of research represented by Logical Pluralism (and substructural logics) on the one hand and Logical Dynamics on the other hand. In fact, our point of view is also very much in line with the claim of (Gärdenfors 1991, Makinson and Gärdenfors 1989) that non-monotonic reasoning and belief revision are "two sides of the same coin": as a matter of fact, non-monotonic reasoning is a reasoning style and belief revision is a sort of update. Likewise, the formal connection in this case also relies on a similar idea based on the Ramsey test.

In this article, we focused on the DEL product update. It is, however, a particular kind of update operator and the ternary relation of substructural logics could actually be a representation of any sort of update, including the various revision and update operators which have been studied in the logics of "common sense reasoning" of artificial intelligence and philosophical logic, such as conditional logic (Nute and Cross 2001), default and non-monotonic logics (Makinson 2005, Gabbay et al. 1998), belief revision theory (Gärdenfors 1988), etc. Different kinds of updates, induced by different informational tasks, define different kinds of reasoning styles. If one adheres to our interpretation of the ternary relation, the dynamic notion of update then becomes the foundational concept of substructural logics.²

References

G. Aucher. A combined system for update logic and belief revision. In M. Barley and N. K. Kasabov, editors, *PRIMA*, volume 3371 of *Lecture Notes in Computer Science*, pages 1–17. Springer, 2004.

G. Aucher. *Perspectives on belief and change*. PhD thesis, University of Otago – University of Toulouse, 2008.

²This article is a short version of (Aucher 2013).

G. Aucher. DEL-sequents for progression. *Journal of Applied Non-Classical Logics*, 21(3-4):289–321, 2011.

G. Aucher. DEL-sequents for regression and epistemic planning. *Journal of Applied Non-Classical Logics*, 22(4):337–367, 2012.

G. Aucher. *Outstanding Contributions: Johan F. A. K. van Benthem on Logical and Informational Dynamics*, chapter DEL as a substructural logic. Trends in Logic. Springer, forthcoming, 2013.

G. Aucher and A. Herzig. Exploring the power of converse events. *Dynamic Formal Epistemology*, pages 51–74, 2011.

A. Baltag and L. Moss. Logic for epistemic programs. *Synthese*, 139(2):165–224, 2004.

A. Baltag and S. Smets. Conditional doxastic models: A qualitative approach to dynamic belief revision. *Electronic Notes in Theoretical Computer Science*, 165:5–21, 2006.

A. Baltag and S. Smets. *Texts in Logic and Games*, volume 4, chapter The Logic of Conditional Doxastic Actions, pages 9–31. Amsterdam University Press, 2008a.

A. Baltag and S. Smets. *Texts in Logic and Games*, volume 3, chapter A Qualitative Theory of Dynamic Interactive Belief Revision, pages 9–58. Amsterdam University Press, 2008b.

A. Baltag, L. Moss, and S. Solecki. The logic of public announcements, common knowledge and private suspicions. Technical report, Indiana University, 1999.

A. Baltag, B. Coecke, and M. Sadrzadeh. Algebra and sequent calculus for epistemic actions. *Electronic Notes in Theoretical Computer Science*, 126:27–52, 2005.

A. Baltag, B. Coecke, and M. Sadrzadeh. Epistemic actions as resources. *Journal of Logic and Computation*, 17(3):555–585, 2007.

J. Barwise. Constraints, channels, and the flow of information. *Situation theory and its applications*, 3:3–27, 1993.

J. Barwise and J. Perry. *Situations and Attitudes*. Cambridge, Massachusetts. MIT Press, 1983.

A. Batlag, L. S. Moss, and S. Solecki. The logic of public announcements and common knowledge and private suspicions. In I. Gilboa, editor, *TARK*, pages 43–56. Morgan Kaufmann, 1998.

J. Beall, R. Brady, J. M. Dunn, A. Hazen, E. Mares, R. K. Meyer, G. Priest, G. Restall, D. Ripley, J. Slaney, et al. On the ternary relation and conditionality. *Journal of philosophical logic*, 41(3):595–612, 2012.

J. C. Beall and G. Restall. Logical pluralism. Oxford University Press, 2006.

J. van Benthem. Exploring logical dynamics. CSLI publications Stanford, 1996.

J. van Benthem. Dynamic logic for belief revision. *Journal of Applied Non-Classical Logics*, 17(2):129–155, 2007a.

J. van Benthem. Inference in action. *Publications de l'Institut Mathématique-Nouvelle Série*, 82(96):3–16, 2007b.

J. van Benthem. Logical dynamics meets logical pluralism? *The Australasian Journal of Logic*, 6:182–209, 2008.

J. van Benthem. Modal logic for open minds. CSLI publications, 2010.

J. van Benthem. McCarthy variations in a modal key. *Artificial intelligence*, 175(1): 428–439, 2011a.

J. van Benthem. *Logical Dynamics of Information and Interaction*. Cambridge University Press, 2011b.

J. van Benthem and B. Kooi. Reduction axioms for epistemic actions. In R. Schmidt, I. Pratt-Hartmann, M. Reynolds, and H. Wansing, editors, *AiML-2004: Advances in Modal Logic*, number UMCS-04-9-1 in Technical Report Series, pages 197–211, University of Manchester, 2004.

J. van Benthem, J. Gerbrandy, T. Hoshi, and E. Pacuit. Merging frameworks for interaction. *Journal of Philosophical Logic*, 38(5):491–526, 2009.

J. P. Burgess. Quick completeness proofs for some logics of conditionals. *Notre Dame Journal of Formal Logic*, 22(1):76–84, 1981.

H. van Ditmarsch. Prolegomena to dynamic logic for belief revision. *Synthese*, 147: 229–275, 2005.

H. van Ditmarsch, W. van der Hoek, and B. Kooi. *Dynamic Epistemic Logic*, volume 337 of *Synthese library*. Springer, 2007.

H. P. van Ditmarsch, A. Herzig, and T. D. Lima. From situation calculus to dynamic epistemic logic. *Journal of Logic and Computation*, 21(2):179–204, 2009.

J. M. Dunn and G. Restall. Relevance logic. *Handbook of philosophical logic*, 6: 1–128, 2002.

J. van Eijck. Reducing dynamic epistemic logic to PDL by program transformation. Technical Report SEN-E0423, CWI, 2004.

D. M. Gabbay, C. J. Hogger, J. A. Robinson, J. Siekmann, and D. Nute, editors. *Handbook of logic in artificial intelligence and logic programming*, volume Nonmonotonic reasoning and uncertain reasoning (Volume 3). Clarendon Press, 1998.

P. Gärdenfors. *Knowledge in Flux (Modeling the Dynamics of Epistemic States)*. Bradford/MIT Press, Cambridge, Massachusetts, 1988.

P. Gärdenfors. Belief revision and nonmonotonic logic: Two sides of the same coin? In *Logics in AI*, pages 52–54. Springer, 1991.

J. Hintikka. *Knowledge and Belief, An Introduction to the Logic of the Two Notions*. Cornell University Press, Ithaca and London, 1962.

F. Liu. *Changing for the Better: Preference Dynamics and Agent Diversity*. PhD thesis, ILLC, University of Amsterdam, 2008.

D. Makinson. Bridges from classical to nonmonotonic logic. King's College, 2005.

D. Makinson and P. Gärdenfors. Relations between the logic of theory change and nonmonotonic logic. In A. Fuhrmann and M. Morreau, editors, *The Logic of Theory Change*, volume 465 of *Lecture Notes in Computer Science*, pages 185–205. Springer, 1989.

E. D. Mares. Relevant logic and the theory of information. *Synthese*, 109(3):345–360, 1996.

E. D. Mares and R. K. Meyer. *The Blackwell guide to philosophical logic*, chapter Relevant Logics. Wiley-Blackwell, 2001.

D. Nute and C. B. Cross. *Handbook of philosophical logic*, volume 4, chapter Conditional logic, pages 1–98. Kluwer Academic Pub, 2001.

J. Perry and D. Israel. What is information? *Information, Language, and Cognition*, 1, 1990.

F. Ramsey. *Philosophical Papers*, chapter General Propositions and Causality. Cambridge University Press, Cambridge, 1929.

G. Restall. Information flow and relevant logics. In *Logic, Language and Computation: The 1994 Moraga Proceedings. CSLI*, pages 463–477. csli Publications, 1996.

G. Restall. An Introduction to Substructural Logics. Routledge, 2000.

G. Restall. Relevant and substructural logics. *Handbook of the History of Logic*, 7: 289–398, 2006.

R. Routley and R. Meyer. The semantics of entailment. *Studies in Logic and the Foundations of Mathematics*, 68:199–243, 1973.

R. Routley and R. K. Meyer. The semantics of entailment—II. *Journal of Philosophical Logic*, 1(1):53–73, 1972a.

R. Routley and R. K. Meyer. The semantics of entailment—III. *Journal of philosophical logic*, 1(2):192–208, 1972b.

R. Routley, V. Plumwood, and R. K. Meyer. *Relevant logics and their rivals*. Ridgeview Publishing Company, 1982.

Strategic Voting and the Logic of Knowledge

Hans van Ditmarsch, Jérôme Lang, and Abdallah Saffidine

LORIA - CRNS / Université de Lorraine LAMSADE - CNRS / Université Paris Dauphine University of New South Wales hans.van-ditmarsch@loria.fr, lang@lamsade.dauphine.fr, abdallahs@cse.unsw.edu.au

Abstract

We propose a general framework for strategic voting when a voter may lack knowledge about other votes or about other voters' knowledge about her own vote. In this setting we define notions of manipulation and equilibrium. We also model action changing knowledge about votes, such as a voter revealing its preference or as a central authority performing a voting poll. Some forms of manipulation are preserved under such updates and others not. Another form of knowledge dynamics is the effect of a voter declaring its vote. We envisage Stackelberg games for uncertain profiles. The purpose of this investigation is to provide the epistemic background for the analysis and design of voting rules that incorporate uncertainty.

1 Introduction

A well-known fact in social choice theory is that strategic voting, also known as manipulation, becomes harder when voters know less about the preferences of other voters. Standard approaches to manipulation in social choice theory (Gibbard 1973, Satterthwaite 1975) as well as in computational social choice (Bartholdi et al. 1989) assume that the manipulating voter knows perfectly how the other voters will vote. Some approaches (Duggan and Schwartz 2000, Barbera et al. 1998) assume that voters have a probabilistic prior belief on the outcome of the vote, which encompasses the case where each voter has a probability distribution over the set of profiles. A recent paper (Conitzer et al. 2011) extends coalitional manipulation to incomplete knowledge, by distinguishing manipulating from non-manipulating voters and by considering that the manipulating coalition has, for each voter outside the coalition, a set of possible votes encoded in the form of a partial order over candidates. Still, we think that the study of strategic voting under complex belief states has received little attention so far, especially when voters are uncertain about the uncertainties of other voters, i.e., when we model higher-order beliefs of voters.

An extreme case of uncertainty is when a voter is completely ignorant about other votes. In that case, if a manipulation under incomplete knowledge is defined in a pessimistic way, i.e., if it is said to be successful if it succeeds for all possible votes of other voters, voting rules may well be non-manipulable. For the special case where all other voters are non-strategic this is shown for most common voting rules in (Conitzer et al. 2011).

In the first place we model how uncertainty about the preferences of other voters may determine a strategic vote, and how a reduction in this uncertainty may change a strategic vote. We restrict ourselves to the case where uncertainty is over a number of well-described alternatives, including the true state of affairs, between which the voter is unable to distinguish.

We also investigate the dynamics of uncertainty. The uncertainty reduction may be due to receiving information on voting intentions in polls or to voters directly telling you their preference. For simplicity we assume that received information is correct, or rather, we only model the consequences of incorporating new information after the decision to consider the information reliable. Such informative actions can then be modelled as truthful public announcements (Plaza 1989).

Another form of dynamics is the dynamics of declaring votes. Declaring votes can be modeled as assignments (ontic / factual change). Just as there may be uncertainty about truthful votes, there may also be uncertainty about declared votes. Consider the following. Half of the votes are declared. It is not known whether candidate x or yhas taken the lead, but z has clearly lost. You still have to vote. Does this influence your strategy? Another example is that of *safe manipulation* (Slinko and White 2008), where the manipulating voter announces her vote to a (presumably large) set of voters sharing her preferences but is unsure of how many will follow her. Finally, consider Stackelberg voting games, wherein voters declare their votes in sequence, following a fixed, exogeneously defined order. Our framework applies to Stackelberg voting games with uncertainty about profiles.

There are several ways of expressing incomplete knowledge about the linear order of a voter. The literature on possible and necessary winners assumes that it is expressed by a collection of partial strict orders (one for each voter), while Hazon *et al.* (2008) consider it to consist of a collection of probability distributions, or a collection of sets
of linear orders (one for each voter). Whereas the latter is more expressive (some sets of linear orders do not correspond to the set of extensions of a partial order), the former is more succinct. Ours is a more expressive modelling than both modes of representation, because an uncertain profile can be any set of profiles. A set of profiles such as $\{(a >_1 b >_1 c, a >_2 b >_2 c), (b >_1 a >_1 c, b >_2 a >_2 c)\}$ expresses uncertainty (ignorance) which candidate voters 1 and 2 rank first, but knowledge (certainty) that voters 1 and 2 have identical preferences — which is not possible in (Hazon et al. 2008), and *a fortiori* also not in (Konczak and Lang 2005) and subsequent works on the possible winner problem. Of course, this mode of representation is also the less succinct of all. However, succinctness and complexity issues will play no role yet in this paper, where we focus on modelling and expressivity.

Somewhat surprisingly, there are yet more complex scenarios that cannot be seen as uncertainty between a number of given profiles: it may be that a voter cannot distinguish between two situations with identical profiles, because in the first case yet another voter has some uncertainty about the profile, but in the other case not.

Our investigation is restricted in various ways: (i) we model uncertainty and manipulability of individuals but not of coalitions, (ii) we model knowledge but not belief, and, in the dynamics, truthful announcements but not lying, (iii) we model incomplete knowledge (uncertainty) but not other forms of incompleteness, and (iv) as already said, we have not investigated complexity and succinctness. The reason for these restrictions is our desire to, first, present this complete logical framework for voters uncertain about profiles. Later we wish to broaden our scope. Let us briefly comment on these issues here.

Epistemic and voting notions for coalitions are treated in Section 8 in some detail.

There are many scenarios wherein voters may have incorrect beliefs about preferences, or where information changing actions are intended to deceive. I may incorrectly believe that you prefer a over b, whereas you really prefer b over a. I may tell you that I prefer a over b, but I may be lying. Such scenarios can also be modelled in epistemic logic, with the same tools and techniques as presented in this paper, but we have restricted ourselves to knowledge: reliable beliefs. This is already a far and high enough jump from the typical social choice theory perspective of reliable common knowledge of preferences, and we think that the variety of phenomena described within the restriction of knowledge and reliable information already sufficiently demonstrate the expressive power of the extension of voting with uncertainty.

The study of uncertain votes is different from the study of other forms of incompleteness, e.g., when the number of voters or candidates may be unknown — the only form of incompleteness that we model is incomplete knowledge in the form of inability to determine which of a number of well-defined alternatives is the case. Here, we also restrict ourselves. Complexity issues will be occassionally referred to in running text and in the concluding Section 4.

A link between epistemic logic and voting has first been given, as far as we know, in (Chopra et al. 2004)—they use knowledge graphs to indicate that a voter is uncertain about the preference of another voter. A more recent approach, within the area known as social software, is (Parikh et al. 2011). The recent (Conitzer et al. 2011) walks a middle way namely where equivalence classes are called information sets, as in treatments of knowledge and uncertainty in economics, but where the uncertain voter does not take the uncertainty of other voters into account.

2 Voting

This section recalls standard voting terminology.

Assume a finite set $N = \{1, ..., n\}$ of *n* voters (or agents), and a finite set $C = \{a, b, c, ...\}$ of *m* candidates (or alternatives). Voter variables are *i* and *j*, and candidate variables are *x* and *y* (and $x_1, x_2, ...$).

Definition 2.1 (Vote). For each voter *i* a *vote* $>_i \subseteq C \times C$ is a linear order on *C*.

If voter *i* prefers candidate *a* to candidate *b* in vote \succ_i , we write $a \succ_i b$. Vote variables are \succ_i, \succ'_i , etc. Instead of $x_1 \succ_i \cdots \succ_i x_n$ we also write $i : x_1 \dots x_n$, or depict it vertically in a table.

Definition 2.2 (Profile). A *profile P* is a collection $\{\succ_1, \ldots, \succ_n\}$ of *n* votes.

Let O(C) be the set of linear orders of C. Then $O(C)^n$ is the set of all profiles for \mathcal{N} . Profile variables are P, P', \dots If $P \in O(C)^n, \succ_i \in P$, and $\succ'_i \in O(C)$, then $P[\succ_i/\succ'_i]$ is the profile wherein \succ_i is substituted by \succ'_i in P.

Definition 2.3 (Voting rule). A *voting rule* is a function $F : O(C)^n \to C$ from the set of profiles to the set of candidates.

The voting rule determines which candidate wins the election — F(P) is the *winner*. A *voting correspondence* $C : O(C)^n \to 2^C \setminus \{\emptyset\}$ maps a profile to a nonempty set of *tied cowinners*. To obtain a voting rule from a voting correspondence (to obtain a unique winner from a non-empty set of cowinners) we assume an exogeneously specified *tie-breaking mechanism*, that is a total order > over candidates.

Voters cannot be assumed to vote according to their preferences. Relative to a given profile *P*, a vote $\succ_i \in P$ can be called the *truthful vote* or *preference*. A voter may change her truthful vote if this improves the outcome of the voting. This is called a *manipulation* or *strategic vote*.

Definition 2.4 (Manipulation). Let $i \in N$, $P \in O(C)^n$ and $\succ_i \in P$, and let $\succ'_i \in O(C)$. If $F(P[\succ_i/\succ'_i]) \succ_i F(P)$, then \succ'_i is a *successful manipulation* by voter *i*.

Of course some votes that are not truthful still do not improve the outcome — relative to the truthful vote $>_i \in P$, any $>'_i \in O(C)$ can be called a *possible vote*. Finally, there is the case of the *declared vote*, after which a voter can no longer change her vote. Information on declared votes may be available to other voters (such as in Stackelberg games), and that may change their subsequent strategic votes. This is an overview of different votes:

- truthful vote / preference;
- strategic vote / successful manipulation;
- possible vote;
- · declared vote.

We now define stable outcomes of the voting rule. The combination of a profile P and a voting rule F defines a strategic game: a player is a voter, an individual strategy for a player is a vote (an individual strategy for a player in the game theoretical sense may not be a strategic vote in the social choice theoretical sense), a strategy profile (of players) is therefore a profile in our defined sense (of voters), and the preference of a player among the outcomes is according to his preferred vote: given voter i with truthful vote $\geq_i \in P$, and profiles P', P'', i prefers outcome F(P') over outcome F(P'') in the game theoretical sense iff $F(P') \geq_i F(P'')$. The relevant equilibrium notion is:

Definition 2.5 (Equilibrium profile). Given a profile P, a profile P' is an *equilibrium profile* iff no agent has a successful manipulation.

In the view of a voting process as a game, an equilibrium profile corresponds to a Nash equilibrium.

Manipulation and equilibrium for coalitions will be addressed in Section 8, later.

3 Knowledge profiles

We model uncertainty about voting in the sense of incomplete knowledge about votes. The terminology to describe such uncertainty that we introduce in this section is fairly standard in modal logic (Fagin et al. 1995), but its application to social choice theory is novel. The novelty consists in taking models with *profiles* instead of *valuations of propositional variables*. An expression like $b >_i a$ is a proposition 'voter *i* prefers

candidate b over candidate a', which is true or false for any given profile; and from that perspective, a profile is nothing but a collection where for all voters all such variables are given a value true or false: a valuation.

Definition 3.1 (Knowledge profile). Given is the set $O(C)^n$ of all profiles for a set $\mathcal{N} = \{1, \ldots, n\}$ of *n* voters. A *profile model* is a structure $\mathcal{P} = (S, \{\sim_1, \ldots, \sim_n\}, \pi)$, where *S* is a *domain* of abstract objects called states; where for $i = 1, \ldots, n, \sim_i$ is an *indistinguishability relation* that is an equivalence relation; and where *valuation* $\pi : S \to O(C)^n$ assigns a profile to each state. A *knowledge profile* is pointed structure \mathcal{P}_s where \mathcal{P} is a profile model and *s* is a state in the domain of \mathcal{P} .

If $s \sim_i s'$, $\pi(s) = P$, and $\pi(s') = P'$, then voter *i* is uncertain if the profile is *P* or *P'*; e.g. if j : bca in *P* and j : cba in *P'*, then voter *i* is uncertain if voter *j* prefers *b* over *c* or *c* over *b*. Instead of 'voter *i* is uncertain if' we also say 'voter *i* does not know that'. We can do this formally in a logical language interpreted on knowledge profiles.

Definition 3.2 (Logical language). The language \mathcal{L} over the set of voters $\mathcal{N} = \{1, ..., n\}$ and the set of preferences is defined as follows, where *i* is an agent and $a, b \in C$:

$$\varphi ::= a \succ_i b \mid \neg \varphi \mid \varphi \land \varphi \mid K_i \varphi$$

A profile *P* is defined in \mathcal{L} by abbreviation as the description of the valuation (the conjunction of all its terms $a >_i b$ and all its excluded terms $\neg(a >_i b)$). Similarly, a vote $>_i$ is defined in \mathcal{L} by abbreviation as the *i*-part of that.

An element of the language is called a formula, φ is a formula variable. Formula $K_i\varphi$ stands for 'voter *i* knows that φ '. We have allowed ourselves to overload the meaning of $a >_i b$, as it is really the name for the atomic proposition uniquely interpreted (below) as the truth of $a >_i b$.

Definition 3.3 (Semantics). The interpretation of formulas in a knowledge profile is defined as follows:

$$\mathcal{P}_{s} \models a \succ_{i} b \quad \text{iff} \quad a \succ_{i} b, \text{ where } \succ_{i} \in \pi(s) \\ \mathcal{P}_{s} \models \neg \varphi \quad \text{iff} \quad \text{it is not the case that } \mathcal{P}_{s} \models \varphi \\ \mathcal{P}_{s} \models \varphi \land \psi \quad \text{iff} \quad \mathcal{P}_{s} \models \varphi \text{ and } \mathcal{P}_{s} \models \psi \\ \mathcal{P}_{s} \models K_{i}\varphi \quad \text{iff} \quad \text{for every } t \text{ such that } s \sim_{i} t, \mathcal{P}_{t} \models \varphi$$

Given a knowledge profile \mathcal{P}_s and a proposition φ , *agent i knows that* φ if and only if φ holds for all states in \mathcal{P} indistinguishable for *i* from *s* (i.e., for all $s' \in \mathcal{P}$ such that $s \sim_i s'$). If $\mathcal{P}_s \models \varphi$ for all $s \in S$, we write $\mathcal{P} \models \varphi$ (φ is valid on \mathcal{P}) and if this is the case for all \mathcal{P} , we say that φ is valid, and we write $\models \varphi$. Propositions like 'voter *i* knows the profile' now have a precise description.

Example 1. Consider the following \mathcal{P} consisting of three states s, t, u and for two voters 1 and 2. State *s* is assigned to profile *P*, wherein $a >_1 c >_1 b >_1 d$ and $d >_2 c >_2 b >_2 a$, etc. States that are indistinguishable for a voter *i* are linked with an *i*-labelled edge. The partition for 1 on the domain is therefore {{*s*, *t*}, {*u*}}, and the partition for 2 on the domain is {{*s*}, {*t*, *u*}}.



States *s* and *t* have been assigned the same profile *P* but have different epistemic properties. In *s*, 2 knows that 1 prefers *a* over *d*, whereas in *t* 2 does not know that. We list some such relevant formulas:

- $\mathcal{P}_s \models K_2 a \succ_1 d$
- $\mathcal{P}_t \not\models K_2 a \succ_1 d$
- $\mathcal{P} \models (\succ_1 \to K_1 \succ_1) \land (\succ_2 \to K_2 \succ_2)$ (Both voters know their preference.)

The example demonstrates than we cannot do away with states. Sometime, different states are being assigned the same profile. But in many typical scenarios different states are assigned different profiles, and then we can truly say that the uncertainty of a voter is about a collection of profiles.

We now define the notion of 'voter *i* changes her vote' in \mathcal{L} .

Definition 3.4 (Changing a vote). We define $P \leftrightarrow_i P'$ as

$$P \to \succ_i \land P' \to \succ'_i \land \bigvee_{j \neq i, a, b \in C} (a \succ_j b \leftrightarrow a \succ'_j b).$$

Given the abbreviations defined, $P \rightarrow >_i$ stands for $>_i \in P$. Formula $P \leftrightarrow_i P'$ says that there is a vote $>'_i$ such that $P' = P[>_i/>'_i]$.

Surprisingly, our logic of knowledge and voter preferences, that we extend with dynamics in the next sections, is not in fact a dynamic logic of preference (Liu 2011). Given that, the following perspective may be of interest. In our models, the preferences are modelled as propositional variables. These induce preferences between states by

enriching the model with total orders expressing that: one state is more preferred than another one, if the outcome of the truthful vote for the profile of the first state is more preferred than the outcome of the vote for the profile of the second state.

Definition 3.5 (Models for knowledge and preference). Given a knowledge profile \mathcal{P}_s with $\mathcal{P} = (S, \{\sim_1, \ldots, \sim_n\}, \pi)$ the induced *preference knowledge profile* \mathcal{P}_s^{\succ} is defined as $\mathcal{P}^{\succ} = (S, \{\sim_1, \ldots, \sim_n\}, \{\succ_1, \ldots, \succ_n\}, \pi)$ where \succ_i is defined as: for all $s, t \in S$, $s \succ_i t$ iff $F(\pi(s)) \succ_i F(\pi(t))$.

Thus we reclaim the epistemic plausibility models of (Baltag and Smets 2008) and therefore, indirectly, approaches as (Liu 2011), although not in the meaning of 'agent i considers state s more plausible than state t', but in the sense of 'voter i prefer the outcome of voting of the profile in s to the outcome of voting of the profile in t'. As there, one has a choice between global preferences or 'local' preferences (intersection of global preferences with equivalence classes). This embedding seems important enough to mention as a result:

Proposition 1. The epistemic logic of votes can be embedded into epistemic plausibility logic.

Proof. We refer to the embedding of Definition 3.5.

4 Manipulation and knowledge

In a knowledge profile it may be that a voter can manipulate the vote but does not know that, because she considers it possible that another profile is the case in which she cannot manipulate the vote. Such situations call for more refined notions of manipulation that also involve knowledge. They can be borrowed from the knowledge and action literature (van Benthem 2001, Jamroga and van der Hoek 2004).

Given is a knowledge profile \mathcal{P}_s where $\pi(s) = P$. If voter *i* can manipulate *P*, then voter *i* also can manipulate \mathcal{P}_s . The uncertainty is about what the profile is. But this does not affect that *P* is the actual profile.

In our modelling, if the voter can manipulate P, she always considers it possible that she can manipulate P. This is a consequence of modelling uncertain knowledge instead of uncertain belief. However, there are situations wherein she considers it possible that she can manipulate, but where in fact she cannot manipulate, namely if she considers a state possible with a profile that is not the profile in the actual state.

A curious situation is the one wherein in all states that the voter considers possible there is a successful manipulation, but where, unfortunately, this is not the same strategic vote in all such states! So she knows that she has a successful manipulation, but she does not know what the manipulation is. This is called *de dicto knowledge* of manipulation.

A stronger form of knowing is when there is a vote that is strategic in the profile for any state that the voter considers possible. This is called *de re knowledge* of manipulation.

A further situation of interest for voting theory is when (a) in any profile that the voter considers possible she can vote such that the outcome is either the same or better than when she had voted sincerely, and when (b) for at least one possible profile the outcome is better. This can be called *weakly successful manipulation*. (It is somewhat unclear if the qualification weak should apply to the manipulation or to the knowledge, as it is a property of a set of profiles.)

Definition 4.1 (Knowledge of manipulation). Given a knowledge profile \mathcal{P}_s ,

- voter *i* can successfully manipulate \mathcal{P}_s if she can successfully manipulate the profile $\pi(s)$;
- voter *i* considers possible that she can successfully manipulate *P_s* if there is a *t* such that *s* ∼_i *t* and she can successfully manipulate π(*t*);
- voter *i* knows 'de dicto' that she can successfully manipulate \mathcal{P}_s , if for all *t* such that $s \sim_i t$ she can successfully manipulate $\pi(t)$;
- voter *i* knows 'de re' that she can successfully manipulate \mathcal{P}_s if there is a vote \succ'_i such that for all *t* such that $s \sim_i t, \succ'_i$ is a successful manipulation for profile $\pi(t)$;
- voter *i* knows 'de re' that she can weakly successfully manipulate \mathcal{P}_s if: (a) there is a vote \succ'_i such that for all *t* such that $s \sim_i t$, either \succ'_i is a successful manipulation for profile $\pi(t)$ or the outcome of that vote in $\pi(t)$ does not change, and (b) there is a *t* such that $s \sim_i t$ and \succ'_i is a successful manipulation for profile $\pi(t)$.

There is also a weakly successful version of 'de dicto' knowledge of manipulation.

These notions of knowledge of manipulation do not assume that voters know their own vote, although to apply them under these circumstances could lead to counterintuitive results.

If voter *i* knows 'de re' that she can manipulate the election, she has the ability to manipulate, namely by strategically voting \succ'_i . On the other hand, 'de dicto' manipulations do not have any practical interest, since the voter does not seem to have the ability to manipulate the election. It is akin to 'game of chicken' type equilibria in game theory (Osborne and Rubinstein 1994). Therein, for each strategy of a player there is a complementary strategy of the other player such that the pair is an equilibrium. This

cannot be guaranteed without coordination. Example 2, below, illustrates 'de dicto' manipulability.

Example 2. We consider manipulation with voting according to the Borda voting rule. Consider three agents, four candidates, and two profiles P and P' that are indistinguishable for agent 1, but that agents 2 and 3 can tell apart; as follows.



There is also a tie-breaking preference b > c > d > a. The difference between the profiles *P* and *P'* is that 3 prefers *d* over *a* in *P* but *a* over *d* in *P'*. We prove that 1 can manipulate the election if the profile is *P*, and that 1 can manipulate the election if the manipulation for *P* gives a worse outcome for *P'*, and that the manipulation for *P'* gives a worse outcome for *P'*. Therefore she is not effectively able to manipulate the outcome of the election.

In Borda, the ranks for each candidate in each vote are added, and the candidate with the highest sum wins, modulo the tie-breaking preference. The preferred candidate gets 3 points, the 2nd choice 2 points, etc. First, the outcome when all three agents give their truthful vote. We write xyzw when there are x points for a, y for b, z for c, w for d.

| profile | count | observation | outcome |
|---------|-------|------------------|---------|
| Р | 3555 | b, c, d are tied | b |
| P' | 5553 | a, b, c are tied | b |

Voter 1 can manipulate P or P' by downgrading b. But this is tricky, because it comes at the price of making a or d, or both, more preferred. This price is indeed too high:

In *P*, 1 can achieve a better outcome by \succ'_1 defined as 1 : *cabd*. Let $Q = P[\succ_1/\succ'_1]$, and $Q' = P[\succ_1/\succ'_1]$. Although 1 prefers the winner in *Q* over the winner in *P*, the winner in *Q'* is less preferred by her than the winner in *P'*:

| profile | count | observation | outcome |
|---------|-------|---------------|---------|
| Q | 4455 | c, d are tied | С |
| Q' | 6453 | | а |

In *P'*, 1 can achieve a better outcome by \succ_1'' defined as 1 : *cdba*. Let $R = P[\succ_1/\succ_1'']$, and $R' = P[\succ_1/\succ_1'']$. Now, 1 prefers the winner in *R'* over the winner in *P'*, but the

winner in *R* is less preferred by her than the winner in *P*:

| profile | count | observation | outcome |
|-----------|-------|-----------------|---------|
| R | 2457 | 1's worst dream | d |
| <i>R'</i> | 4455 | c, d are tied | С |

For the record, the winners for all different votes for voter 1 where c is most preferred.

| 1: cbad | 1 : <i>cabd</i> | 1 : <i>cdba</i> | 1: cadb | 1 : cdab | 1 : <i>cbda</i> |
|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|
| <i>b</i> (3555) | <i>c</i> (4455) | <i>d</i> (2457) | <i>d</i> (4356) | <i>d</i> (3357) | <i>d</i> (2556) |
| <i>b</i> (5553) | a(6453) | <i>c</i> (4455) | <i>a</i> (6354) | <i>c</i> (5355) | <i>b</i> (4554) |

In the language \mathcal{L} we cannot say that the outcome of the election in *P* is preferred by a voter to the outcome of the election in *P'*. For that, we need to add primitives $P >_i P'$ to the language. These act as background knowledge. They encode the voting function so that its results are available in all states and in all profile models.

Definition 4.2 (Language \mathcal{L}^+). We expand the set of propositional variables with $P \succ_i P'$ for any $P, P' \in O(C)^n$, and we add the following clause to the semantics:

$$\mathcal{P}_s \models P \succ_i P' \quad \text{iff} \quad F(P) \succ_i F(P').$$

The variables $P >_i P'$ mean that voter *i* prefers the candidate chosen by the votes in *P* over the candidate chosen by the votes in *P'*. This is a(n) (inefficient) way to encode the voting function. We observe that the semantics is indeed independent from state *s* and profile model *P*. These are model validities $\models P >_i P'$.

All notions of manipulation in Definition 4.1 are definable in the language \mathcal{L}^+ .

Definition 4.3. Let \mathcal{P}_s be a knowledge profile with profile *P*.

• Voter *i* has a successful manipulation:

$$P \land (P \to \succ_i) \land \bigvee_{P'} (P' \succ_i P \land (P' \leftrightarrow_i P)).$$

• Voter *i* has a successful manipulation \geq_i :

$$P \land (P \to \succ_i) \land (P' \to \succ'_i) \land (P' \leftrightarrow_i P) \land P' \succ_i P.$$

• Voter *i* knows de dicto that she has a successful manipulation:

$$P \land (P \to \succ_i) \land K_i \bigvee_{P'} ((P' \leftrightarrow_i P)) \land P' \succ_i P).$$

• Voter *i* knows de re that she has a successful manipulation:

$$P \land (P \to \succ_i) \land \bigvee_{\succ'_i} [((P' \leftrightarrow_i P) \land P' \succ_i P \land P' \to \succ'_i)) \land K_i(P'' \to ((P' \leftrightarrow_i P'') \land P' \succ_i P'')).$$

De re knowledge of weak manipulation is similarly defined.

Proposition 2. *Knowledge of manipulation is definable in* \mathcal{L}^+ *.*

Proof. As evidenced in Definition 4.3.

5 Equilibrium and knowledge

Determining equilibria under incomplete knowledge comes down to decision taking under incomplete knowledge. Therefore we have to choose a decision criterion. Expected utility makes no sense here, because we didn't start with probabilities over profiles in the first place, nor with utilities. In the absence of prior probabilities, the following three criteria make sense. (i) The insufficient reason (or Laplace) criterion considers all possible states in a given situation as equiprobable. This criterion was used in (Ågotnes and van Ditmarsch 2011) to determine equilibria of certain (Bayesian) games of imperfect information. (ii) The maximum regret criterion selects the decision minimizing the maximum utility loss, taken over all possible states, compared to the best decision, had the voter known the true state. (iii) The pessimistic (or Wald, or maximin) criterion compares decisions according to their worst possible consequences. The latter criterion, that we also call risk averse, is one that fits well our probability-free and utility-free model; this was also the criterion chosen in (Conitzer et al. 2011). The only assumption here is that the probability distribution is positive in all states. We now fix this criterion for the rest of the paper. (Pessimistic, optimistic, and yet other criteria only assuming positive probability are applied to social choice settings in the recent (Parikh et al. 2011). We think their interesting results can be modelled as games using our setting.)

In the presence of knowledge, the definition of an equilibrium extends naturally. The trick is that for each agent, the combination of an agent *i* and an equivalence class $[s]_{\sim i}$ for that agent (for some state *s* in the knowledge profile) defines a so-called virtual agent (we model these imperfect information games as Bayesian games (Harsanyi 1967–1968)). Thus, agent *i* is multiplied in as many virtual agents as there are equivalences classes for \sim_i in the model.

In our setting we can almost think of these equivalence classes as sets of indistinguishable profiles. Almost but not quite: we recall that states with different properties in a given equivalence class, or states in different equivalence classes, may be assigned the same profile.

An equilibrium is then a combination of votes such that none of the virtual agents has an interest to deviate. A intuitively more appealing solution than virtual agents, also applied in (Ågotnes and van Ditmarsch 2011), is to stick to the agents we already have, but change the set of votes into a larger set of *conditional votes* — where the conditions are the equivalence classes for the agents. This we will now follow in the definition below. For risk-averse voters we can effectively determine if a conditional profile is an equilibrium without taking probability distributions into account, unlike in the more general setting of Bayesian games that it originates with.

Definition 5.1 (Conditional equilibrium). Given is a knowledge profile model \mathcal{P} such that every voter knows her preference (truthful vote). For each agent *i*, a *conditional vote* is a function $[>]_i : S/\sim_i \rightarrow O(C)$, i.e., a function that assigns a vote to each equivalence class for that agent. A *conditional profile* is a collection of *n* conditional votes, one for each agent. A *conditional voting game* is then a (standard) strategic game where voters declare conditional votes. A conditional profile is an *equilibrium* iff no agent has a successful manipulation in any of its equivalence classes.

The outcome of a conditional profile consisting of conditional votes is a *n*-tuple of vectors (x_1, \ldots, x_m) where voter *i* has *m* equivalence classes. The definition of equilibrium for the conditional voting game is derived from the Bayesian game form. It is not the standard form of strategic games! Consider a case for two equivalence classes for a voter 1 where two outcome vectors for 1 are (a, d) and (d, a), and $a >_i d$. We cannot say which of these two are preferred: therefore, the outcomes for 1 are not ordered, and therefore, it does not define a standard strategic game. However, if we only vary 1's vote in the first argument (equivalence class) or in the second argument, the outcomes are ordered. This is the Bayesian game computation of equilibrium, where we determine manipulability for each virtual agent. Therefore, in the definition we did not write 'A conditional profile is an equilibrium iff no agent has a successful manipulation *in any of its equivalence classes*.'

The requirement in Def. 5.1 that voters need to know their preference (truthful vote), is because the value they associate with that class is the worst outcome. This might otherwise be undefined.

Example 3. We recall Example 1. There are two voters 1, 2, and four candidates a, b, c, d. Consider a plurality vote with a tie-breaking rule b > a > c > d.

First consider the profile P defined as

| 1 | 2 |
|---|---|
| а | d |
| с | С |
| b | b |
| d | a |

If 1 votes for her preference a and 2 votes for his preference d, then the tie prefers a, 2's least preferred candidate. If instead 2 votes c, a will still win. But if 2 votes b, b wins. We observe that (a, b) and (b, b) are equilibria pairs of votes, and that for 1 voting a is dominant.

This is also apparent from the voting matrix (wherein equilibria are boxed), and even more so when we express the payoffs for both voters by their ranking for the winner, as on the right.

| 1\2 | а | b | С | d | 1\2 | а | b | С | d |
|-----|---|---|---|---|-----|----|----|----|----|
| а | а | b | а | a | а | 30 | 11 | 30 | 30 |
| b | b | b | b | b | b | 11 | 11 | 11 | 11 |
| С | а | b | С | d | С | 30 | 11 | 22 | 03 |
| d | а | b | d | d | d | 30 | 11 | 03 | 03 |

Example 4. We now add uncertainty to the setting of Example 3. Consider another profile P', that is as P, but where 1's vote is 1 : dcba. Now consider a knowledge profile as follows. It remains the case that the actual profile is P; voter 2 is uncertain which of P and P' is the case; whereas voter 1 knows that. (It is tempting to add: voter 1 *of course* knows that, as he knows his own vote; but our framework equally applies to situations where he does not, e.g., because he has not yet made up his mind.) And, as one should always add: 1 and 2 know that this is the uncertainty about the profile. This knowledge profile \mathcal{P}_P consists of states t and u.



What are the conditional equilibria of \mathcal{P} ? Votes (a, b) and (b, b) still lead to elect b and are the equilibria in state t with profile P. The only equilibrium vote for for state

u with profile P' is (d, d)—the preferences are identical for 1 and 2, and *d* is their top candidate.

We argue our way towards the equilibria of this conditional voting game. There are two. Of course, alternatively to this argument one can directly determine these are equilibria by applying Definition 5.1 in a 16×4 matrix (below). Recall that we assumed that voters are risk-averse.

| 1\2 | a | b | С | d |
|-----|-----|-----|-----|-----|
| aa | aaa | bbb | aaa | aaa |
| ab | aba | bbb | aaa | aaa |
| ac | aaa | bbb | aca | aca |
| ad | aaa | bbb | aaa | ada |
| ba | baa | bbb | baa | baa |
| bb | bbb | bbb | bbb | bbb |
| bc | baa | bbb | bcb | bcb |
| bd | baa | bbb | bbb | bdb |
| ca | aaa | bbb | caa | caa |
| cb | aaa | bbb | cbb | cbb |
| сс | aaa | bbb | ссс | ccc |
| cd | aaa | bbb | ссс | cdc |
| da | aaa | bbb | caa | daa |
| db | aaa | bbb | cbb | dbb |
| dc | aaa | bbb | ссс | dcc |
| dd | aaa | bbb | ссс | ddd |

First, consider voter 1. For each equivalence class of 1, we have to determine her optimal vote. If the profile is P, 1's vote for a is dominant, so no matter what strategic considerations 2 may have due to the additional uncertainty about the profile, does not make a difference. Voter 1 votes a. If the profile is P', d is dominant for 1.

Next, consider voter 2. Because 2 is risk-averse he will vote *b*. Because if 2 votes *d* and the profile is *P*, *a* wins because 1 votes *a*, as this is dominant for 1 (or *b* wins because 1 votes *b*); whereas if the profile is P' and 2 votes *d*, then *d* wins because 1 votes *d*, which is dominant there. The worst outcome of these two is *a* (or *b*). Whereas if 2 votes *b*, the worst outcome is *b*. (The votes *c* and *a* can be eliminated from consideration as well.)

The two equilibria that we can associate with this knowledge profile are below. The conditional vote for 1 in the first equilibrium actually is actually defined as: $[>]_1(\{t\}) = >_1$ and $[>]_1(\{u\}) = >'_1$; and the vote for 2 is conditional to one equiva-

lence class — in other words, it is unconditional. The equivalent verbose formulation is more intelligible.

- (if 1 prefers a then a and if 1 prefers d then d, b),
- (if 1 prefers a then b and if 1 prefers d then d, b).

In particular, 2 does not know that d is his equilibrium vote in P', because he considers it possible that the profile is P, where, if 2 votes d, 1 votes a (or 1 can improve her outcome by voting a), in which case 2 is worse off than d.

In the 16×4 matrix, a conditional vote *ab* for 1 means: in *t* she votes *a* and in *u* she votes *b*. The outcome triples *xyz* represent: (worst and only) outcome for 1 in equivalence class of *t*, (worst and only) outcome for 1 in equivalence class of *u*; (worst) outcome for 2 in equivalence class of {*t*, *u*}. The table contains much symmetry. We omitted the table in terms of ranked outcomes. A triple like *aaa* corresponds to ranked outcome 144: the equal winners *a* for voter 1 are ranked according to different profiles, *a* is preferred in state *t* / in profile *P*, hence 1, but *a* is least preferred in state *u* / in profile *P'*, hence 4. In the table, the third of a triple *xyz* is necessarily equal to the least preferred of *x* and *y*, but this is an artifact of the example (namely, that the two equivalence classes for 1 together comprise the equivalence class for 2).

| Example 5. We | can add further | uncertainty to | Example 5 |
|---------------|-----------------|----------------|-----------|
|---------------|-----------------|----------------|-----------|

| 1 | 2 | | 1 | 2 | | 1 | 2 |
|---|------|---|---|-----|---|------|------------|
| a | d | | a | d | | d | d |
| c | С | 1 | c | С | 2 | c | c |
| b | b | | b | b | | b | b |
| d | a | | d | а | | a | a |
| | s, 1 | D | t | , P | | u, I | o ′ |

Consider a third state that has the same profile P as the actual state, but that has different epistemic properties: 2 is not uncertain about the profile there, but 1 cannot distinguish this from the other state for P wherein 2 is uncertain about the profile. This is the profile model from Example 1.

Will 1 vote differently in s and t? In fact, she will not, nor will 2, and the conditional equilibria votes remain the same; strictly, 2's vote should depend on his equivalence class, but as 2's choice is the same either way, namely b, his vote is more succinctly described as an unconditional: b.

We did not yet attempt to characterize conditional equilibria in the logic of the previous sections, as we did for manipulation and knowledge of manipulation (Def. 3.4 and 4.3). This might be interesting for epistemic game theory (Aumann and Brandenburger 1995, Perea 2012), but even so we only deal with the special case of voting games.

6 Dynamics: revealing preference

We can extend the modal logical setting for voting and knowledge of the previous sections with logical operations that are dynamic in character. In the context of voting, two obvious choices here are *public announcement of a proposition* (such as an agent revealing her true preference), and *declaring a vote*. Such actions can be modelled as semantic operations $\mathcal{P}_s \mapsto \mathcal{P}_s | \varphi$ (for propositions φ , e.g., respectively, $\varphi = >_i$ for revealing her preference) and $\mathcal{P}_s \mapsto \mathcal{P}_s^{\gg_i := \top}$ (for voter *i* declaring vote \gg_i). In this section we deal with public announcement, in the next section, with public assignment.

A well-known dynamic feature of epistemic logics is *truthful public announcement* (Plaza 1989). Given a knowledge profile \mathcal{P}_s , the requirement for execution of public announcement of φ is that φ is true in \mathcal{P}_s , and the way to execute it is to restrict the model \mathcal{P} to all the states where φ is true. We can then investigate the truth of propositions in that model restriction: we can evaluate formulas of form $[\varphi]\psi$, for 'After announcement of φ , ψ (is true)', such as: 'After 1 reveals her preference (truthful vote) to 2, 2 knows that he has a successful manipulation'. We need to add a clause to the logical language for these announcements and define their semantics. The model restriction to the φ -states is denoted as $\mathcal{P}_s | \varphi$.

Definition 6.1 (Public announcement). We add an inductive clause $[\varphi]\varphi$ to the logical language \mathcal{L} (i.e., a dynamic modal operator with an argument of type formula followed by a postcondition also of type formula). Its semantics is:

 $\mathcal{P}_s \models [\varphi] \psi$ iff $\mathcal{P}_s \models \varphi$ implies $\mathcal{P}_s | \varphi \models \psi$,

where $\mathcal{P}_s | \varphi = (S', \sim'_1, \dots, \sim'_n, \pi')$ such that $S' = \{t \in S : \mathcal{P}_t \models \varphi\}, \sim'_i = \sim_i \cap (S' \times S'),$ and $\pi'(a >_i b) = \pi(a >_i b) \cap S'.$

Example 6. Consider again Examples 1 and 4, with plurality voting. In state t (for profile P), after voter 1 informs voter 2 of her true preference (a public announcement), the uncertainty in the model disappears and 1 and 2 commonly know that the profile is P. The equilibrium vote remains (b, b). So this seems not a big deal.

On the other hand, in state u voter 1 has an incentive to make her preference known to 2: after that, 2's equilibrium vote changes from b to d, and the equilibrium profile is now (d, d). And that is a big deal.

The transitions can be depicted as follows:

| 4 | d | a | | d | a | | a | a | | a | a |
|---|--------|---|---|--|--|---|--------|--------|---|--------|--------|
| | c b | c | ⇐ | $\begin{vmatrix} c \\ b \end{vmatrix}$ | $\begin{vmatrix} c \\ b \end{vmatrix}$ | 2 | c b | с b | ⇒ | с b | с b |
| 6 | a | d | | a | d | | d | d | | d | d |
| | 1 | 2 | | 1 | 2 | | 1 | 2 | | 1 | 2 |

We can now formalize statements as

$$\mathcal{P}_t \models \neg K_2 \ a \succ_1 c \land [a \succ_1 c] K_2 \ a \succ_1 c.$$

There are two obvious ways to interpret such public announcements in voting theory: (i) when voters make announcements about their own preferences (and such that these announcements are trusted by other voters), and, more properly from the viewpoint of public announcement logic, (ii) when external observers, such as a central authority, reveal preferences to voters. The last can be interpreted as holding a voting poll. Successive voting polls reduce the uncertainty for the individual voter of the preferences (truthful vote) of *other* voters. And this may determine the strategic vote.

Two obvious results are that:

Proposition 3. *Knowledge of weakly successful manipulation is not preserved after update.*

Proof. We recall Definition 4.1. For the weak form of manipulation there were two requirements: (a) the profile of at least one state in a given equivalence class for voter i needs to have a manipulation, and (b) the profiles of all states in that equivalence class must have either equal or better outcome. The state with a manipulation need not be the actual state, therefore, after model restriction the existential requirement (a) may no longer hold. This holds for 'de re' as well as 'de dicto' knowledge.

Proposition 4. *Knowledge of successful manipulation is preserved after update.*

Proof. The profiles of all states have a manipulation, a universal property that is preserved after update.

7 Dynamics of declaring votes

A voter *i* declaring a vote \succ_i can be modelled in dynamic epistemic terms as an *assignment* (a.k.a. ontic change, in contrast to an informative change like an announcement

and coalition deliberation). A succinct way to model this is to expand the knowledge profiles with a *duplicate set of propositional variables* expressing voter preference, initially all set to false. To distinguish the preference (truthful vote) from the declared vote we keep writing $>_i$ for the former whereas we write \gg_i for the latter. So, the set of variables $a >_i b$ encode the preferences of the voters, whereas variables $a \gg_i b$ encode their declared votes.

The action of declaring a vote \gg_i , defined by preferences $a \gg_i b$, sets the value of the propositions encoding \gg_i in the model to true: these are the assignments $a \gg_i b := \top$ executed for all $a \gg_i b$ in \gg_i . If we assume that the declared vote is public, then this assignment can be executed in all states of the knowledge profile. The dynamic epistemic logic equivalent to achieve that is a public assignment (van Ditmarsch et al. 2005, van Benthem et al. 2006).

Definition 7.1 (Public assignment). We add an inductive clause $[a \gg_i b := \top]\varphi$ to the logical language. For the semantics, given a knowledge profile \mathcal{P}_s , $\mathcal{P}_s \models [a \gg_i b := \top]\varphi$ iff $(\mathcal{P}^{a\gg_i b})_s \models \varphi$, where $\mathcal{P}^{a\gg_i b}$ is as \mathcal{P} except that $\pi(a \gg_i b) = \mathcal{D}(\mathcal{P})$. By abbreviation we define $\gg_i := \top$ as the sequential execution of all assignments $a \gg_i b := T$ for all terms $a \gg_i b$ in \gg_i .

Assignments need not be to 'true' (\top) but can be to any formula. Such an assignment $a \gg_i b := \psi$ has semantics $\pi(a \gg_i b) = \{t \in \mathcal{D}(\mathcal{P}) \mid \mathcal{P}_t \models \psi$. Declaring one's preference, the truthful vote, can then be seen as the assignment $\gg_i := \succ_i$.

Example 7. Consider $a \gg_1 b \gg_1 c$. The assignment declaring this vote is the sequence of three assignments $a \gg_1 b := \top, b \gg_1 c := \top, a \gg_1 c := \top$, abbreviated as $\gg_i := \top$.

Example 8. Another continuation of Example 4 is with declaring votes. If in state *t* voter 2 declares his vote, i.e., fixes *d* as the candidate of his choice, 1 votes *a*, because with the given tie b > a > d > c, her preference *a* now gets elected. We can simulate this assignment as the sequence of $d \gg_2 c := \top$, $d \gg_2 b := \top$, $d \gg_2 a := \top$ (or as the assignment of preference to the declared vote: $\gg_2 := >_2$). For simplicity this is depicted as making *d* bold.



We have no results yet for the interaction of declaring votes and revealing voter preference, but Stackelberg games are the obvious games of interest here.

Axiomatization and completeness All four logics proposed in this work have sound and complete axiomatizations with respect to the class of profile models. However, this is not remarkable. We have therefore omitted these axiomatizations, for that see the cited references.

8 Chair and coalitions

We have some modelling results concerning matters relevant for social choice theory that we have chosen not to incorporate in the main story, as not to lose focus there: how to model the central authority, and group notions of preference and knowledge.

8.1 Central authority

Apart from the *n* voters, it seems convenient to distinguish yet another agent: a designated agent named 0, the *central authority*, or *chair*. We recall that the tie-breaking preference $>_{tie}$ is a linear order on candidates. Apart from applying the tie, the central authority may perform other kinds of actions such as fixing the agenda. This also opens the door to the logical modelling of well-studied problems in computational social choice, such as control by the chair, or determining possible winners. The main reason *not* to model the chair it that her role is uniform throughout the model (throughout any knowledge profile model). We assume that there is no uncertainty on what the voting rule (and the tie-breaking preference) is. So in that sense it is exogenous.

The universal relation $S \times S$ on a knowledge profile model can be seen as the indistinguishability relation of the agent 0, the central authority. On a connected model (i.e., when there is always a path between any two states in the model) this is the same as common knowledge of the voters. The computational tasks of the central authority, be it determining the possible winners or finding strategic actions such as agenda fixing or any other form of control, can only be harder on knowledge profiles as it has to take uncertainty into account. By identifying the central authority with an agent with universal ignorance we can be precise about how much harder.

A partial profile in the social choice literature corresponds in a profile model to the set of profiles completing it, with identity access for all voters, and indistinguishable for the central authority, as in the following example. (The set of partial profiles then seems to consist of such disconnected parts.)

Example 9. The following depicts the partial profile $(b >_1 a >_1 c, a >_2 \{b, c\})$. Voters 1 and 2 have identity access on the profile model. The central authority is agent 0.



8.2 Coalitional manipulation

Group notions play an important role in social choice theory. We consider coalitions $G \subseteq N$. As straightforward generalizations of (individual) preference $>_i$, (individual) manipulation, (weak) equilibrium, and (weak) equilibrium of a conditional voting game, we can also define: *coalitional preference* $>_G$, and *successful manipulation by a coalition G*. A profile P' is a *strong equilibrium profile* iff no coalition has a successful manipulation.

Group notions also play an important role in epistemic logic. Two notions useful in our setting are common knowledge and distributed knowledge. Given a knowledge profile, a proposition is commonly known if it is true in all states reachable (from the actual state of the knowledge profile) by arbitrarily long finite paths in the model (reflexive transitive closure of access for all voters in the coalition). With the interpretation of common knowledge of coalition *G* we can thus associate an equivalence relation \sim_G (defined as $(\bigcup_{i \in G} \sim_i)^*$). A proposition is distributedly known in a knowledge profile, if it is true in the intersection of accessibility relations in the actual state (the relation $\bigcap_{i \in G} \sim_i$).

If there is no uncertainty about the profile, the voters have common knowledge about the profile. This assumption is almost always made in social choice theory. It is important to observe that in the presence of uncertainty this strong form of common knowledge disappears, but that still some form of common knowledge remains: all agents have common knowledge of the structure of the profile model. This means that they have common knowledge of the set of states, the accessibility relations of the knowledge model, and what profiles these states stand for. The only thing they do (or rather, may) not know is the designated point of the profile model: what the preferences (truthful votes) are.

Coalitions play a big role in voting, partly because in realistic settings the power of individual voters is very limited. Now by analogy, just as the vote of an individual agent depends on her knowledge, the vote of a coalition would seem to depend on the common knowledge of that coalition. But that seems wrong. In voting theory, the power of a coalition means the power of a set of agents that can decide on a joint action as a result of communication between them. Communication makes the uncertainty about each others' profiles disappear. In terms of knowledge profiles, this means that we are talking about another model, namely the model where for all agents $i \in G$, \sim_i is refined to $\bigcap_{i \in G} \sim_i$. What determines the voting power of a coalition seems rather its distributed knowledge.

We are still exploring the implications of these observation, and should note that also other choices can be made to model the power of a coalition in voting.

Knowledge of manipulation and equilibria of conditional voting games can also be defined for coalitions but have been left out of this presentation.

9 Conclusion, further research

We presented a formal logical semantics for the interaction of voting and knowledge. The semantic primitive is the knowledge profile: a profile including uncertainty of voters about what the actual profile is. This reveals different notions for knowledge of manipulation, such as de re knowledge of manipulation and de dicto knowledge of manipulation, and novel notions for equilibria, such as conditional equilibrium for riskaverse voters. Dynamic operations on such knowledge profiles can also be modelled, and their effects on manipulation, where we distinguished public announcements, such as revealing true preferences, from public assignments, i.e., declaring votes.

As far as the formalization is concerned, our setting is very similar to that of the recent literature on robust mechanism design (Bergemann and Morris 2005), which generalizes classical mechanism design by weakening the common knowledge assumptions of the environment among the players and the planner. In (Bergemann and Morris 2005) uncertainty is modelled with information partitions. The main technical difference is that in our setting, as in classical social choice theory, preferences are ordinal, whereas in (robust) mechanism design preferences are numerical payoffs, which allows for payments (which we don't). This connection with mechanism design, however, is certainly worth exploring further. (We are very grateful to an anonymous reviewer for pointing this connection to us.)

The logical setting defined in the paper allows us to represent various classes of situations already studied specifically in (computational) social choice, thus offering a general representation framework in which, of course, new classes of problems will be representable as well, thus providing an homogeneous, unified representation framework. In some of the classes of problems we need one more agent, the chair. The chair may have preferences, but does not vote. In some classes of problems the dynamics plays a crucial role in defining these problems, both as announcements (revealing preference) and assignments (declaring votes). Here are a few such problems:

- 1. *Possible and necessary winners* (Konczak and Lang 2005): there is one more agent (the chair), who has an incomplete knowledge of each of the votes; the voters' knowledge is does not matter. *x* is a possible winner if the chair does not know that *x* is not a (co)winner, and a necessary winner if the chair knows that *x* is a (co)winner.
- 2. *Stackelberg voting games* (Xia and Conitzer 2010): voters express their votes in sequence, in a commonly known order. Their preferences are common knowledge. The votes are announced publicly and each voter thus know the vote of the voters which speak before him.
- 3. *Sequential voting games with abstention* (Desmedt and Elkind 2010): voters express their votes in sequence, preferences are common knowledge; the voting rule is plurality; voters have the choice to vote or to abstain; voting is costly.
- 4. *Control by adding or removing voters or candidates* (Bartholdi III et al. 1992): the chair has a perfect knowledge of the voters' preferences; voters have no knowledge (and thus are supposed to vote truthfully); the chair may add or remove some candidates as well as register or unregister voters.
- 5. Sequential voting on multi-issue domains (Lang and Xia 2009): the set of alternatives is a combinatorial domains, therefore the valuations are preference relations over tuples of values; voters vote in sequence, issue by issue, and the value for the (binary) issue is chosen by majority, and then communicated to the voters.

10 Acknowledgements

This contribution to the LIRa yearbook is identical to (van Ditmarsch et al. 2013). The work was previously presented at AAMAS 2012 Valencia and at the ESSLLI 2012 Opole workshop 'Strategies for Learning, Belief Revision and Preference Change'. Hans van Ditmarsch is also affiliated to IMSc (Institute of Mathematical Sciences), Chennai, as a research associate. He acknowledges support from European Research Council grant EPS 313360.

References

R. Aumann and A. Brandenburger. Epistemic conditions for nash equilibrium. *Econometrica*, 63:1161–1180, 1995.

A. Baltag and S. Smets. A qualitative theory of dynamic interactive belief revision. In *Proc. of 7th LOFT*, Texts in Logic and Games 3, pages 13–60. Amsterdam University Press, 2008.

S. Barbera, A. Bogomolnaia, and H. van der Stel. Strategy-proof probabilistic rules for expected utility maximizers. *Mathematical Social Sciences*, 35(2):89–103, 1998.

J. Bartholdi, C. Tovey, and M. Trick. The computational difficulty of manipulating an election. *Social Choice and Welfare*, 6(3):227–241, 1989.

J. Bartholdi III, C. Tovey, and M. Trick. How hard is it to control an election? *Mathematical and Computer Modelling*, 16(8/9):27–40, 1992.

J. van Benthem. Games in dynamic epistemic logic. *Bulletin of Economic Research*, 53(4):219–248, 2001.

J. van Benthem, J. van Eijck, and B. Kooi. Logics of communication and change. *Information and Computation*, 204(11):1620–1662, 2006.

D. Bergemann and S. Morris. Robust mechanism design. *Econometrica*, 73(6):1771–1813, 2005.

S. Chopra, E. Pacuit, and R. Parikh. Knowledge-theoretic properties of strategic voting. In *Proc. of 9th JELIA*, LNCS 3229, pages 18–30, 2004.

V. Conitzer, T. Walsh, and L. Xia. Dominating manipulations in voting with partial information. In *Proc. of AAAI*, 2011.

Y. Desmedt and E. Elkind. Equilibria of plurality voting with abstentions. In *ACM Conference on Electronic Commerce*, pages 347–356, 2010.

H. van Ditmarsch, W. van der Hoek, and B. Kooi. Dynamic epistemic logic with assignment. In *Proc. of 4th AAMAS*, pages 141–148. ACM, 2005.

H. van Ditmarsch, J. Lang, and A. Saffidine. Strategic voting and the logic of knowledge. In *Proc. of 14th TARK – Chennai*, 2013.

J. Duggan and T. Schwartz. Strategic manipulability without resoluteness or shared beliefs: Gibbard-Satterthwaite generalized. *Social Choice and Welfare*, 17(1):85–93, 2000.

R. Fagin, J. Halpern, Y. Moses, and M. Vardi. *Reasoning about Knowledge*. MIT Press, Cambridge MA, 1995.

A. Gibbard. Manipulation of voting schemes: A general result. *Econometrica*, 41: 587–601, 1973.

J. Harsanyi. Games with Incomplete Information Played by 'Bayesian' Players, Parts I, II, and III. *Management Science*, 14:159–182, 320–334, 486–502, 1967–1968.

N. Hazon, Y. Aumann, S. Kraus, and M. Wooldridge. Evaluation of election outcomes under uncertainty. In *Proc. of AAMAS '08*, pages 959–966, 2008.

W. Jamroga and W. van der Hoek. Agents that know how to play. *Fundamenta Informaticae*, 63:185–219, 2004.

K. Konczak and J. Lang. Voting procedures with incomplete preferences. In *Proc. IJCAI Multidisciplinary Workshop on Advances in Preference Handling*, 2005.

J. Lang and L. Xia. Sequential composition of voting rules in multi-issue domains. *Mathematical Social Sciences*, 57(3):304–324, 2009.

F. Liu. *Reasoning about Preference Dynamics*. Springer, 2011. Synthese Library, Vol. 354.

M. Osborne and A. Rubinstein. A Course in Game Theory. MIT Press, 1994.

R. Parikh, C. Tasdemir, and A. Witzel. The power of knowledge in games. In *Proc. of the Workshop on Reasoning About Other Minds*, 2011. CEUR Workshop Proceedings. Volume: 751.

A. Perea. Epistemic game theory. Cambridge University Press, 2012.

J. Plaza. Logics of public communications. In *Proc. of the 4th ISMIS*, pages 201–216. Oak Ridge National Laboratory, 1989.

M. A. Satterthwaite. Strategy-proofness and Arrow's conditions: Existence and correspondence theorems for voting procedures and social welfare functions. *Journal of Economic Theory*, 10(2):187–217, April 1975.

A. Slinko and S. White. Is it ever safe to vote strategically? Technical report, Auckland University, 2008. Dep. of Math. Research Report 563.

L. Xia and V. Conitzer. Stackelberg voting games: Computational aspects and paradoxes. In *Proc. of AAAI*, 2010.

T. Ågotnes and H. van Ditmarsch. What will they say? - Public announcement games. *Synthese*, 179(S.1):57–85, 2011.

What Does it Mean to Know an Action?

Jiahong Guo

School of Philosophy and Sociology, Beijing Normal University jiahong.guo@bnu.edu.cn

Abstract

This paper focuses on knowing actions as knowing their denotations in terms of successful transitions. We view actions (programs) as black boxes exhibiting inputoutput behavior. This view ignores all intermediate details, and we investigate how much it can do for us. Knowledge of an action is then defined with a new notation in a combined language of dynamic and epistemic logic. The main purpose of this article is to propose a first approximation of knowing actions by their external input-output behavior, implemented in a first order epistemic setting. The full version of this idea takes place in the framework of first order epistemic models, with the help of a standard translation on program expressions. Some basic logical principles of reasoning are explored with respect to validity in this way, and we apply them in particular to general epistemic properties of knowing actions, and to knowledge of different ways of generating actions, through tests, serial combination, choice, and iteration. Beyond that, because of the shape of our definition, knowing an action basically has the same introspection properties as those assumed in the basic propositional epistemic logic. The general base logic of the system is also demonstrated. Inside this full system, the logic of knowing actions here involves just a smaller fragment of the full language, but we have not determined its special properties yet.

1 Introduction

It is customary in Cognitive Psychology and Epistemology to make at least a rough division between *declarative knowledge* and *procedural knowledge*. Declarative knowledge is about knowing propositions, about *knowing that*. Procedural knowledge is about interaction with the world, about *knowing how*. It is also common in Philosophy community that a finer distinction of knowledge is made with the above categories *knowing that*, *knowing how* and *knowing to do* (Tang 2011). Some philosophers claim that *knowing how* is just a species of *knowing that* (Stanley 2011, Stanley and Williamson 2001). We are not going to explore and make justifications for the actual division of different kinds of knowledge and relationships between them. Perhaps only a kind of procedural knowledge - *knowing an action* will be considered in a formal representation here.

There is a rich literature about declarative knowledge. The history of the study of what it means to have factual knowledge starts in Antiquity, and makes a restart with the advent of Kripke semantics and the proposal of Hintikka to analyze knowledge and belief in terms of access to possible worlds (Hintikka 1962). This analysis was taken up in cognitive science (Gärdenfors 1988), computer science (Halpern 1987, Halpern et al. 2009, Fagin et al. 1995) and game theory (Aumann 1976, Battigalli and Bonanno 1999, Perea 2012).

The analysis of knowledge in computer science, however, also brings in a dynamic component. It is just a small step from knowledge based programming to knowledge of programs, knowledge of procedures, or knowledge of actions. Porgrams and actions have long been the area of dynamic logics that capture the effects of explicitly defined procedures (Pratt 1982, Harel et al. 2000). A natural combination of this dynamics with knowledge emerges in dynamic epistemic logic or DEL (van Benthem 1996, Baltag et al. 1998, van Benthem 2011, Ditmarsch et al. 2006). But this does still not directly deal with the analysis of what it means to know a procedure.

In fact, several perspectives on the analysis of "knowing how" are possible:

- 1. Knowing how is about how to achieve a proposition φ via some actions or procedures: e.g, to win a game.
- 2. Knowing how is about maintaining φ as an invariant through some actions or procedures that can be controlled by the agent.
- 3. Knowing how is about all (or some of) the denotations of procedures or actions.
- 4. Knowing how is for an agent to be able to perform certain actions or procedures as a whole: being able to swim, being able to play chess.
- 5. More variants are possible and useful, e.g., for the analysis of what it means for a person to know a foreign language, to know a book (the Bible, say), to know the special theory of relativity, or even to know one's spouse.

Guo

In this paper, we will take only one of these lines, focussing on knowing actions as knowing their denotations in terms of successful transitions. Thus, we view actions or procedures (complex actions) as a black box exhibiting input-output behavior. This means viewing a procedure or action as a relation, whose denotation is a set of pairs of input/output states. This view ignores all intermediate details, and we shall investigate how much it can do for us, while eventually also arriving at a better understanding of its limitations, and the need for richer views of actions.

There is some previous literature on logics for knowing actions (for some approaches different from ours, see Singh 1999). In a separate section at the end of this paper, we also briefly discuss some approaches closer to ours, and state our reasons for going beyond these earlier attempts.

Technically speaking, our proposal will be as follows. In dynamic logic, a program α denotes a set of ordered pairs in some given state space, where all intermediate events are ignored. We will say that an agent knows an action α if he can distinguish the set of input-output pairs that constitutes α from all other possible sets of input-output pairs (possibly within some given class of relations).

Here the propositional notion of knowledge is standard, given in terms of the accessibility relations *R* of epistemic logic. Here $K\varphi$ is true in a world *w* when φ is true in all *R*-successors of *w*. Now we will define knowledge of an action as follows, using a new notation in a combined language of dynamic and epistemic logic, and following the above idea of knowing the input-output behavior:

 $K\alpha$ is true at a world w iff for any two states x, y, if $xR_{\alpha}y$ then the agent knows it, otherwise the agent knows $\neg xR_{\alpha}y$.

Practically speaking, then, the agent can make a judgment whether a given pair of states is a possible transition in the action relation or not.

To bring out the logical principles governing this notion, it seems natural to explain this a bit further in a first order modal language, where we can quantify over a suitable domains of states and transitions. This is the technical framework that we shall develop below.

In summary, the main purpose of this article is to propose a first approximation of knowing actions by their external input-output behavior, implemented in a first order epistemic-dynamic setting. We will discuss in particular, what reasoning about knowing-that and knowing-how becomes available in this way, with a major focus on knowledge of complex actions. At the end, we evaluate where we stand, and discuss what further structure of actions and logics would be needed to give a fuller account of action and procedure.

2 Syntax and semantics of Propositional Dynamic Logic

We start with the language of propositional dynamic logic (PDL) for acquiring a suitable language of knowing actions, the basis of our later definition of the essential semantic item $\dot{K}\alpha$ in a first order epistemic framework. All the formulas in the original language of PDL have a corresponding "standard translation" into first-order logic, enriched with fixed-point operators where needed. We will use this tool, too, toward setting up our eventual proposed system.

It is clear that the language for PDL is defined in two aspects: formulas and actions (van Benthem et al. 2012). The language of propositional part (denoted by L_{PDL}) is over some set of basic propositions Φ is given by (for the simplicity, only single agent is considered here):

$$\varphi ::= \top | p | \neg \varphi | \varphi \lor \varphi | [\alpha] \varphi$$

where *p* ranges over Φ . We do not yet introduce knowledge operators *K* and the crucial new item $\dot{K}\alpha$ of knowing action α here: they will be settled later in the first order epistemic framework. The dual of $[\alpha]\varphi$ is denoted by $\langle \alpha \rangle \varphi$. Other Boolean connectives are defined as usual.

If we assume that a set of basic action symbols *A* is given, then the language of actions for regular programs can be formally defined as:

$$\alpha ::= a |?\varphi| \alpha^* | \alpha; \alpha | \alpha \cup \alpha | \alpha^*$$
 where a ranges over A.

For example, $p \wedge [a; b]q \rightarrow q$ is a well-formed formula. There are two kinds of atomic actions need to be mentioned later, they are *abort* ι , and *skip* \downarrow , representing empty relation and identity relation on states respectively.

Models *M* for PDL are triples

 $(S, R_a \{ \text{for each atomic action symbol a } \}, V),$

where *S* is a set of states or worlds (a state is commonly denoted by *s*). And *V* is a valuation map assigning each propositional letter to a set of possible states, that is, $V(p) \subseteq S$, meaning that *p* is true in every state of V(p). Binary relation R_a on *S* interprets the respective atomic action *a* in the model *M*. The interpretation of an arbitrary action α in a model *M* can be explained inductively by the structure of actions as usual in dynamic logic. Now the truth of an arbitrary formula of PDL can be defined in a world of a model as follows (van Benthem et al. 2012):

 $\begin{array}{lll} M, s \models \top & \text{always} \\ M, s \models p & \Longleftrightarrow & s \in V(p) \\ M, s \models \neg \varphi & \Longleftrightarrow & M, s \not\models \varphi \\ M, s \models \varphi \lor \psi & \longleftrightarrow & M, s \models \varphi \text{ or } M, s \models \psi \\ M, s \models \varphi \land \psi & \longleftrightarrow & M, s \models \varphi \text{ and } M, s \models \psi \\ M, s \models \langle \alpha \rangle \varphi & \longleftrightarrow & \text{for some } t, (s, t) \in [\![\alpha]\!]^M \text{ and } M, t \models \varphi \\ M, s \models [\alpha] \varphi & \longleftrightarrow & \text{for all } t \text{ with } (s, t) \in [\![\alpha]\!]^M \text{ it holds that } M, t \models \varphi \end{array}$

where the binary relation $[\![\alpha]\!]^M$ (also denoted by R_α if the context *M* is clear) interpreting the action α in the model is defined as in (van Benthem et al. 2012):

$$\llbracket a \rrbracket^{M} = R_{a}$$

$$\llbracket ?\varphi \rrbracket^{M} = \{(s,s) \mid M, s \models \varphi \}$$

$$\llbracket \alpha^{*} \rrbracket^{M} = (\llbracket \alpha \rrbracket^{M})^{*}$$

$$\llbracket \alpha \cup \beta \rrbracket^{M} = \llbracket \alpha \rrbracket^{M} \cup \llbracket \beta \rrbracket^{M}$$

$$\llbracket \alpha^{*} \rrbracket^{M} = (\llbracket \alpha \rrbracket^{M} \cup \llbracket \beta \rrbracket^{M}$$

It is clear that the natural denotation of *abort* action ι is \emptyset , and $skip \Downarrow is \Delta$, the identity relation on *S*. If $(s,t) \in R_{\alpha}$ in a model *M*, we may write this as $M, (s,t) \models \alpha$ for convenience, but α itself is not a formula in the language.

3 Standard translation of PDL into FOL

As can be seen in any good textbook on modal logic (Blackburn et al. 2001), all PDL formulas (without Kleene star) can be translated into a fragment of first order language (FOL hereafter) with at most two free variables, and actually even with Kleene star, the PDL formulas can be translated into a fragment of FOL where infinitely disjunctions are allowed. Now have a quick look at standard translation of PDL with regular programs.

Definition 3.1. Let L_{PDL} be given as in the above. The target language FOL which is used for translating PDL formulas has unary predicate symbols P, Q, \ldots corresponding to propositional letters p, q, \ldots in Φ and a binary relation symbol R_a for each atomic action a. We write F(x) to denote a first order formula F with free variable x.

Let x be a first-order variable. Following (Blackburn et al. 2001), the standard translation ST_x taking propositional modal formulas in the system PDL to first-order formulas in FOL can be defined inductively as follows:

- $ST_x(\top) = (x = x)$
- $ST_x(p) = P_x$
- $ST_x(\neg \varphi) = \neg ST_x(\varphi)$
- $ST_x(\varphi \lor \psi) = ST_x(\varphi) \lor ST_x(\psi)$
- $ST_x(\varphi \land \psi) = ST_x(\varphi) \land ST_x(\psi)$
- $ST_x([\alpha]\varphi) = \forall y(ST_{xy}(\alpha) \to ST_x(\varphi))$
- $ST_x(\langle \alpha \rangle \varphi) = \exists y(ST_{xy}(\alpha) \land ST_x(\varphi))$

The formula translation calls on ST_{xy} to start recursively decomposing the action α . This requires two free variables to define binary relations. First for the *-free fragment of PDL:

- $ST_{xy}(a) = xR_ay$
- $ST_{xy}(?\varphi) = xR_{?\varphi}y$ where φ is a PDL formula and $R_{?\varphi} = \{(s, s) \mid M, s \models \varphi\}$
- $ST_{xy}(\alpha;\beta) = \exists z(ST_{xz}(\alpha) \land ST_{zy}(\beta))$
- $ST_{xy}(\alpha \cup \beta) = ST_{xy}(\alpha) \vee ST_{xy}(\beta)$

Still, the intended interpretation of α^* is the reflexive, transitive closure of R_{α} and this kind of closure of a binary relation is not expressible in FOL. Since the meaning of α^* is defined as

$$R^*_{\alpha} = \bigcup_{n \in \mathbb{N}} R^n_{\alpha},$$

one way to go here is to use infinitely long disjunctions to capture the meaning of an iterated action α :

$$x(R_{\alpha})^*y \iff (x=y) \lor xR_{\alpha}y \lor \bigvee_{n\geq 1} \exists z_1 \dots z_n (xR_{\alpha}z_1 \land \dots \land z_nR_{\alpha}y).$$

So, in an infinitary modal logic allowing countably infinite disjunctions and conjunctions, we can standardly translate all the formulas of PDL. The clauses for the *-free fragment has been given as in the above, and the following clause settles the Kleene star:

$$ST_{xy}(\alpha^*) = (x = y) \lor xR_{\alpha}y \lor \bigvee_{n \ge 1} \exists z_1 \dots z_n (xR_{\alpha}z_1 \land \dots \land z_nR_{\alpha}y).$$

Alternatively, we can translate PDL into the logic LFP(FO) which extends first order logic with fixed-point operators for recursive definitions using monotone operations.

Either way, like with the standard translation of classical modal logic, we can define natural corresponding first order models.

Definition 3.2. A corresponding first order model M^F can be constructed as $\langle S, I \rangle$ where *S* is the domain, it is the same as in *M*. *I* is an interpretation which interprets unary predicate symbols P, Q, \ldots as subsets of *S* with $P^I = V(p), Q^I = V(q), \ldots$ respectively. And *I* interprets the binary relation symbol R_a just as R_a in PDL for each atomic action *a*.

We may understand the states in the PDL model M intuitively as individuals (objects) here in the model M^F . Likewise, transition relations between worlds become relations between individuals. And, getting ahead of ourselves, while a single PDL model allows of no variation here, it is easy to see how we could give this an epistemic twist, by varying universes of states and denotations of programs across worlds, allowing for significant knowledge or ignorance of agents about what a given action really does.

Now we are ready to state some correspondence results between PDL and FOL (allowing countably infinite disjunctions and conjunctions) with the help of standard translation:

Fact 3.1. For every $\varphi \in L_{PDL}$,

$$M, s \models \varphi \iff M^F \models S T_x(\varphi)[s];$$

$$M \models \varphi \iff M^F \models \forall x S T_x(\varphi).$$

Proof. Proofs are similar as in classical textbook of modal logic (Blackburn et al. 2001). We only check the first one with formulas in case of $\langle ?\varphi \rangle \psi$. First suppose $M, s \models \langle ?\varphi \rangle \psi$. This means for some $t \in S$ with $sR_{?\varphi}t$ and $M, t \models \psi$. $sR_{?\varphi}t$ means that s = t and $M, s \models \varphi$, showing that $M, s \models \psi$ as well. Then by induction hypothesis, we have $M^F \models ST_x(\psi)[s]$. And $ST_x(\langle ?\varphi \rangle \psi) = \exists y(ST_{xy}(?\varphi) \land ST_y(\psi)) = \exists y(xR_{?\varphi}y \land ST_y(\psi))$. It's clear to have that $M^F \models (\exists y(xR_{?\varphi}y) \land ST_y(\psi))[s]$ since $M, s \models \varphi$ (then $sR_{?\varphi}s$) and $M^F \models ST_x(\psi)[s]$ (then $M^F \models \exists yST_y(\psi)[s]$). It shows that $M^F \models ST_x(\langle ?\varphi \rangle \psi)[s]$, as required. Next we prove the direction from right to left. Suppose that $M^F \models ST_x(\langle ?\varphi \rangle \psi)[s]$. And $ST_x(\langle ?\varphi \rangle \psi) = \exists y(xR_{?\varphi}y \land ST_y(\psi))$ as above. This

means $M^F \models \exists y(sR_{2\varphi}y \land ST_y(\psi))$ holds. It intuitively shows that there exists some $t \in S$ satisfying $sR_{2\varphi}t$ and $ST_x(\psi)$ is true under the assignment assigning t to x. The former shows that $M, s \models \varphi$ and s = t. The latter means that $M^F \models ST_x(\psi)[t]$. By induction hypothesis, we have $M, t \models \psi$. Those results show precisely that $M, s \models \langle ?\varphi \rangle \psi$, as required.

4 Extension to a first order epistemic framework

Next the formalization of knowledge of actions can be considered. We add knowledge operators K, \dot{K} and the formula $\dot{K}\alpha$ into the language of PDL directly, but it seems not easy to settle the semantics of this item purely inside the usual propositional setting.

By contrast, our intuitive idea is to consider this as knowing a binary relation of the interpreted α : for any two states *x*, *y*, if $xR_{\alpha}y$ then the agent knows it, otherwise the agent knows $\neg xR_{\alpha}y$. Practically speaking, the agent can make a judgment whether the given pair of states is in the action relation or not. It seems natural to explain this intuition in an extended first order modal setting.

4.1 Language

First the alphabet of our first order epistemic language (FOEL, hereafter) is defined as follows (it is based on countably infinite FOL, we add constant symbols for convenience in proving some results later):

- constant symbols: *a*, *b*, *c*, . . .
- variables: x, y, z, \ldots
- unary predicate symbols: P, Q, ... with respect to the propositional letters p, q, ... in PDL
- a special binary relation symbol: =
- binary relation symbols: R_a, R_b, \ldots with respect to the atomic actions a, b, \ldots in PDL
- quantifier symbols: \forall , \exists
- knowledge operators: K, < K >

Only unary and binary predicates are concerned here. Like in FOL with countably infinite disjunctions and conjunctions, we can have the standard translation from PDL

to FOEL. Then formulas and actions in PDL are translated into first order formulas (maybe infinitely long) with several free variables. It seems useful to generate new binary relation symbols for compound actions, according to the rules for compound actions shown in the following, while the subscripted atomic actions can be standardly translated into countably infinite FOL (and FOEL) as done in the last section:

$$R_{?\varphi}, R_{\alpha;\beta}, R_{\alpha\cup\beta}, R_{\alpha^*}.$$

This means that, essentially, every action has a corresponding binary relation symbol in FOEL; although strictly speaking, only binary relation symbols for atomic actions are in the language. We introduce this trick in order to have a simple expression of our later proposal. And the language does not have functional symbols either, then terms only contain constants and variables here, where the set of terms may be represented as $t_1, t_2, ...$

Now the FOEL formulas (denoted by L_{FOEL}) can be generated as

$$\varphi \quad ::= \quad P_t \mid t_1 = t_2 \mid t_1 R_\alpha t_2 \mid \neg \varphi \mid \varphi \lor \varphi \mid K\varphi \mid \forall x\varphi$$

Other Boolean cases can be defined as usual.

Now we have enough syntax to formally introduce the formula $\dot{K}\alpha$ (for every action α) as an abbreviation of a first order epistemic formula in the next definition:

Definition 4.1. Formula $\dot{K}\alpha$ representing knowing the arbitrary action α is defined as the following first order epistemic sentence (no free variables contained)

$$\forall x \forall y ((xR_{\alpha}y \to KxR_{\alpha}y) \land (\neg xR_{\alpha}y \to K \neg xR_{\alpha}y))$$

in our first order epistemic language L_{FOEL} .

In what follows, we will pay close attention to the first order formulas that are standard translations from PDL and that abbreviate statements of knowing actions since exploring the logical properties of knowing actions is a central theme here.

Moreover, when talking with PDL formula φ in FOEL, we use a first order sentence $\forall xST_x(\varphi)$ to refer to φ , but perhaps directly apply φ as an abbreviation of $\forall xST_x(\varphi)$ in more situations for several reasons: in some cases, such as analyzing properties of tests, we don't want the problem to be much complicated, then only original PDL formulas are allowed to be arguments of a test; φ is simpler than $\forall xST_x(\varphi)$, just as $\dot{K}\alpha$, it can be more intuitively understood when used to form a valid result with $\dot{K}\alpha$; every φ has a correspondence with $\forall xST_x(\varphi)$ in the respective models M and M^F , it's easy to deal with semantically.

4.2 Models

Having given the language, we now look at the semantic structures where it is supposed to be interpreted.

Definition 4.2. A first order epistemic model can be constructed from the respective original dynamic model M and first order model M^F , we augment M^F to a quintuple form $M^C = \langle W, R, D, I, w^* \rangle$ where W is a set of non-empty possible worlds and R is a usual epistemic relation (commonly equivalence relation) on W. D is just S in M^F (same as the set of possible states in M). I is an interpretation, technically it interprets all the unary predicate symbols as sets of individuals with respect to a world ($P^{I,w} \subseteq D$ for every $w \in W$), and the binary relation symbols as binary relations over domain D in a world w: $R_a^{I,w} \subseteq D \times D$. $w^* \in W$ is the actual world which reflects (inherits) all the information from original PDL model M with $P^{I,w^*} = V(p), Q^{I,w^*} = V(q), \ldots$ for the respective proposition letters p, q, \ldots in Φ and $R_a^{I,w^*} = R_a, R_b^{I,w^*} = R_b, \ldots$ for the respective atomic actions a, b, \ldots in A.

Here we may consider D intuitively as states with respect to every world $w \in W$ and the set D is fixed (that is also called constant domains in literature) over different worlds. All the interpretations of binary relation symbols (we added them into the language for simplicity) for an arbitrary (complex) action α can be inductively defined in each world, just as we did in dynamic logic:

$$\begin{array}{lll} R_{a}^{I,w} &=& \text{a binary relation } R \subseteq D \times D \\ R_{\gamma\varphi}^{I,w} &=& \{(d,d) \in D \times D \mid M,d \models \varphi\} \\ R_{\alpha^{\vee}}^{I,w} &=& (R_{\alpha}^{I,w})^{\vee} \\ R_{\alpha^{\vee}\beta}^{I,w} &=& R_{\alpha}^{I,w} \circ R_{\beta}^{I,w} \\ R_{\alpha\cup\beta}^{I,w} &=& R_{\alpha}^{I,w} \cup R_{\beta}^{I,w} \\ R_{\alpha^{\vee}\beta}^{I,w} &=& (R_{\alpha}^{I,w})^{*} \end{array}$$

It is clear that we have $R_{\alpha}^{I,w} \subseteq D \times D$ for an arbitrary action α with each world w.

Here are some further explanations. Every constant symbol *c* is interpreted as the same individual $c \in D$ in different worlds. The superscript *I*, *w* will be omitted for convenience if the context is clear. Just as in first order semantics, the valuation function σ will assign a respective individual in *D* to each variable and constant symbol at each world. Also as in first order logic, we can define the notion of terms *t*: all constants and variables (recall there are no function symbols in our language).

We have given an intuitive motivation for these semantic structures earlier. Basically, they are a first-order way of viewing what the state space of a process can be like. Here the possible variation across worlds may arise naturally when we think of agents who do not know exactly how the process runs, for instance because of observational limitations. In multi-agent settings (not defined here, but analogous to our setting), epistemic variation may arise because agents who know their own actions may still not know those of other agents.

Now we come to the truth definition for our language.

Definition 4.3. Truth of a formula φ is defined in a world *w* under some assignment σ of a given model M^C inductively as usual (just as in first order semantics, it's easy to have that for every term $t, \sigma(t) \in D$):

$$\begin{array}{lll}
 & M^{C}, w, \sigma \models t = t & \iff & \text{always} \\
 & M^{C}, w, \sigma \models P_{t} & \iff & \sigma(t) \in P^{I,w} \\
 & M^{C}, w, \sigma \models t_{1}R_{a}t_{2} & \iff & (\sigma(t_{1}), \sigma(t_{2})) \in R_{a}^{I,w} \\
 & M^{C}, w, \sigma \models \sigma & \iff & M^{C}, w, \sigma \models \varphi \\
 & M^{C}, w, \sigma \models \varphi \lor \psi & \iff & M^{C}, w, \sigma \models \varphi \text{ or } M^{C}, w \models \psi \\
 & M^{C}, w, \sigma \models \varphi \land \psi & \iff & M^{C}, w \models \varphi \text{ and } M^{C}, w \models \psi \\
 & M^{C}, w, \sigma \models K\varphi & \iff & \text{for every } w' \in W \text{ with } wRw', M^{C}, w', \sigma \models \varphi \\
 & M^{C}, w, \sigma \models \forall x\varphi & \iff & \text{for all } d \in D, M^{C}, w, \sigma[x := d] \models \varphi.
 \end{array}$$

If a first order formula *F* is true under every assignment σ at world *w*, we say that it is true in the world *w*, denoted by M^C , $w \models F$ (also written as $w \models F$ if the context M^C is clear). If the formula is true at all worlds of M^C , we claim that it is valid in the model M^C , denoted by $M^C \models F$. If *F* is valid in all models of the above first order epistemic setting, it is called valid, denoted by $\models F$.

The duals of *K* and \forall can be defined as usual in epistemic logic and first order logic. It is clear that if a first order formula does not contain free variables, it has nothing to do with assignment σ . Then we may consider the truth of $\dot{K}\alpha$ just in a world *w* of M^C simpliciter, that is,

$$M^{C}, w \models \forall x \forall y ((xR_{\alpha}y \rightarrow KxR_{\alpha}y) \land (\neg xR_{\alpha}y \rightarrow K\neg xR_{\alpha}y))$$

This will be just true or false at the world, since the first order epistemic formula abbreviating $\dot{K}\alpha$ does not contain free variables.

4.3 Logic

The next natural theme is studying the logic of these models and language in FOEL. For a start, the classical epistemic logic for knowledge of propositions with single agent is widely accepted to be **S5**. Here are its three well-known axioms with prominent epistemic interpretations over and above the axiom system **K** based on basic modal logic:

Veridicality $K\varphi \rightarrow \varphi$.

Positive Introspection $K\varphi \rightarrow KK\varphi$.

Negative Introspection $\neg K\varphi \rightarrow K\neg K\varphi$.

According to basic modal logic, **S5** is sound and complete with respect to the class of all equivalence relational models (frames as well).

As for an intuitive interpretation, the first of these axioms seems uncontroversial: knowledge has to be true, or it does not deserve to be called knowledge according to the tradition of literature in western philosophy. Although introspective properties have been widely and deeply discussed by philosophers, many objections as well as supports can be found in literature, it is a kind of reasonable characterization of propositional knowledge. We will not focus on this philosophical issue here. Our treatment of knowing an action would also work, we believe, if other philosophical positions are taken.

Next, for standard first order modal logic, it is well known that all the theorems of minimal first order normal modal logic with constant domains (the system is also called first order \mathbf{K}) as done in (Priest 2008) are also valid in all of the above models. In our setting, the assumption of a constant domain means that agent knows the state space of the relevant process, though perhaps not its transition structure. In particular, one can check that both the Barcan formula (BF) and the Converse Barcan Formula (CBF)

$$BF: \forall xKF(x) \to K \forall xF(x)$$
$$CBF: K \forall xF(x) \to \forall xKF(x)$$

are valid in standard relational first order modal models with constant domains (Arló-Costa and Pacuit 2006).

This fixes our basic logic. The minimal logic derived by all of the above first order epistemic structures is first order S5 plus BF and CBF with a countably infinite language, since R is an equivalence relation and the domain remains constant across different worlds.

5 Reasoning about knowledge of actions

Now we can study the logical properties of reasoning about knowledge of actions. We will do so in several steps. First, we consider the simplest atomic actions, namely, tests.
After that, our second topic is general epistemic properties of knowing an action, such as it being veridical or introspective. Finally, we take up the topic of knowing complex actions, and how this knowledge relates to knowing their components.

5.1 Knowing tests

Consider the test action $?\varphi$ (without loss of generality, we assume proposition φ here to be pure Boolean in order to avoid entangled iterations). In regular programs, action $?\varphi$ can only be executed in a state where φ is true, and it does not change the original state. Our intuition would be this. When someone says that a person knows an test $?\varphi$, this means that the person knows whether φ . Formally, then, we expect the following assertion to hold:

Fact 5.1.

$$\dot{K}(?\varphi) \leftrightarrow K\varphi \lor K\neg\varphi$$
 is valid

Interestingly we can indeed prove this, in a somewhat roundabout way with the help of standard translation. The rest of this subsection is a somewhat detailed investigation of this fact, and several basic observations around it.

The test $?\varphi$ denotes a corresponding binary relation of $?\varphi$ as $\{(a, a)|a \in D\}$ where φ is true in the state *a* of model *M*. We only consider the test of φ in PDL formulas, the standard translation of $?\varphi$ is $ST_{xy}(?\varphi) = ST_x(?\varphi) = xR_{?\varphi}x$. Recall that we can think of that for every world *w* in the model M^C , φ is true at a state *s* of *M*, if and only if the standard translation of φ is true in the corresponding state (individual) assigned to the free variable with respect to the world *w*. That is, given a world *w*, for every first order assignment σ and state (individual) *a*, $M, a \models \varphi$ if and only if $M^C, w, \sigma[x := a] \models ST_x(\varphi)$. In this above setting, we may first expect to have the following property of the knowledge of test: for a PDL formula φ , if a rational agent knows that (denoted by $K\varphi$, actually it's $K \forall xST_x(\varphi)$ in FOEL, recall we abbreviate $\forall xST_x(\varphi)$ as φ for simplicity), then he knows its test.

This means that the following claim should hold:

Fact 5.2.

The formula
$$K\varphi \to K(?\varphi)$$
 is valid.

Proof. Now we check its validity. Suppose $K\varphi$ is true at an arbitrary world w of an arbitrary first order epistemic model M^C . We can conclude that for every $w' \in W$ satisfying wRw', $\forall xST_x(\varphi)$ is true in w'. And we know R is an equivalence relation, so $\forall xST_x(\varphi)$ is true at w. Now consider arbitrary $a \in D$ (recall that a is just an element of S in PDL model M) that satisfy $aR_{2\varphi}a$ in w. Since $\forall xST_x(\varphi)$ is true in w, it

then follows that for every assignment σ , M^C , w, $\sigma \models \forall xST_x(\varphi)$ (by the definition of truth at a world of the model M^C) and actually every world w' that satisfies wRw' has M^C , $w', \sigma \models \forall xST_x(\varphi)$ holds as well since $K\varphi$ is true in w. By the properties of standard translation, we have, $M, a \models \varphi$ for every $a \in S$. This means $aR_{2\varphi}a$ holds in every world w'. Therefore, $KaR_{2\varphi}a$ is true in w. We can then conclude, $\forall x(xR_{2\varphi}x \to KxR_{2\varphi}x)$ is true at w of M^C , as required. Next consider the case of arbitrary $a \in D$ that do not satisfy $aR_{2\varphi}a$ (that is, $\neg aR_{2\varphi}a$) is true in w. It must be that $M, a \not\models \varphi$, and it follows by properties of standard translation that $M^C, w, \sigma[x := a] \not\models ST_x(\varphi)$ for each such kind of assignments σ . But we have shown $M^C, w \models \forall xST_x(\varphi)$, a contradiction is deduced now, and then we conclude that this case is impossible to hold. It is vacuously to have that $\neg aR_{2\varphi}a \to K \neg aR_{2\varphi}a$ is true at w. Since a is arbitrary, It's save to have $\forall x(\neg xR_{2\varphi}x \to K \neg xR_{2\varphi}x)$ is true at w. The two cases of the above show that $\dot{K}(?\varphi)$ is true at w. Since w is arbitrary and $R_{2\varphi}$ is interpreted the same in different models with a constant domain S, we conclude that the formula $K\varphi \to \dot{K}(?\varphi)$ is valid.

Interestingly, the converse of our observation does not hold.

Fact 5.3.

The formula $\dot{K}(?\varphi) \rightarrow K\varphi$ is not valid in general in our first order epistemic models.

Proof. To see this, consider a counter model M^C with $W = \{w, w'\}$, $D = \{a, b\}$, $R = \{(w, w), (w, w'), (w', w), (w', w')\}$ and a valuation V that makes p false at a and b in the original model M, other things remain the same). The interpretation I at w of M^C makes unary predicate P as $V(p) = \emptyset$. It's clear then $R_{2p}^{I,w} = \emptyset$ and $R_{2p}^{I,w'} = \emptyset$. This means $M^C, w \models \forall xy(xR_{2p}y \rightarrow KxR_{2p}y)$ and $M^C, w \models \forall xy(\neg xR_{2p}y \rightarrow K \neg xR_{2p}y)$, that is, $M^C, w \models K(?p)$. But it is obvious $M^C, w \nvDash Kp$ (i.e., $K \forall xP_x$) since we can find a world w' with wRw' such that $M^C, w' \nvDash \forall xP_x$.

This formal result agrees with our intuition that the knowledge of an action of testing some proposition φ does not necessarily imply the knowledge that φ holds. However, we can obtain a following weaker version:

Fact 5.4.

The formula $\dot{K}(?\varphi) \rightarrow K\varphi \lor K \neg \varphi$ is valid.

Proof. Suppose that M^C , $w \models \dot{K}(?\varphi)$ for arbitrary M^C and w. First consider the case $R_{?\varphi}^{I,w} = \emptyset$. This means for every $x \in D$, $\neg x R_{?\varphi} x$ is true at w. Since $\dot{K}(?\varphi)$ is true at w, it follows that for arbitrary x, $K \neg x R_{?\varphi} x$ is true at w as well, that means for all $x \in D$, $\neg x R_{?\varphi} x$ is true at each w' with w Rw'. It clear to get now, for an arbitrary σ and all $a \in D$, M^C , w', $\sigma[x := a] \models S T_x(\neg \varphi)$, and then M^C , w', $\sigma[= \forall x S T_x(\neg \varphi)$, that is,

 $M^C, w' \models \neg \varphi$ by abbreviation and σ is arbitrary. Now we have φ must be false (hence $\neg \varphi$ is true) at each w', and so $K \neg \varphi$ is true at w. Next consider the case $R_{2\omega}^{I,w} = \{(d,d) \mid d\}$ $d \in D$. This means for all $x \in D$, $xR_{2\varphi}x$ is true at w. Since $\dot{K}(2\varphi)$ is true there, it follows that for every x, $KxR_{2\omega}x$ is also true at w. Then we conclude that, for every x, $xR_{2\omega}x$ is true at each world w' that is a *R*-successor of w. It is clear to get now, for an arbitrary σ and all $a \in D, M^C, w', \sigma[x := a] \models ST_x(\varphi)$, and then $M^C, w', \sigma \models \forall xST_x(\varphi)$, that is, M^C , $w' \models \varphi$ by abbreviation and σ is arbitrary. Now we have φ is true at each w', and so $K\varphi$ is true at w. The last case here is $\emptyset \subset R^{I,w}_{2\varphi} \subset \{(d,d) \mid d \in D\}$. This means there exists an object a and c in D such that $aR_{2\varphi}a$ is true but $cR_{2\varphi}c$ is false $(cR_{2\neg\varphi}c)$ is true) at w. Since $\dot{K}(?\varphi)$ is true at w, we can also get $KaR_{?\varphi}a$ and $KcR_{?\neg\varphi}c$ are both true there, and further to have $aR_{?\varphi}a$ is true but $cR_{?\varphi}c$ is false at each w' with wRw'. Then it shows for an arbitrary σ , M^{C} , w', $\sigma[x := a] \models ST_{x}(\varphi)$ but M^{C} , w', $\sigma[x := c] \models ST_{x}(\neg \varphi)$ for each w', that is, M^C , w', $\sigma[x := c] \models \neg ST_x(\varphi)$. It follows that M^C , w', $\sigma \models \neg \forall xST_x(\varphi)$. Since σ is arbitrary, we have M^C , $w' \models \neg \forall xST_x(\varphi)$ for each w' with wRw', and further to have M^C , $w \models K \neg \forall x S T_x(\varphi)$. That is just M^C , $w \models K \neg \varphi$ by abbreviation, as required.

The above proof shows intuitively that knowledge of a test φ leads to knowledge of φ or knowledge of $\neg \varphi$. By the help of first rule, we can deduce a more general result, perhaps surprising at first sight, which describes the connection between knowledge of $\varphi \varphi$ and knowledge of φ which is a PDL formula (here it is also the abbreviation of $\forall xST_x(\varphi)$ in FOEL).

Fact 5.5.

The formula $\dot{K}(?\varphi) \land \varphi \leftrightarrow K\varphi$ is valid.

Proof. Proving the direction from right to left is trivial since we have shown $K\varphi \rightarrow \dot{K}(?\varphi)$ is valid and $K\varphi \rightarrow \varphi$ is valid in classical epistemic logic. Now have a look from left to right: suppose $\dot{K}(?\varphi) \land \varphi$ is true in an arbitrary w of a model M^C . Then we get for arbitrary assignment σ , M^C , w, $\sigma \models \forall xST_x(\varphi)$ and then in every $a \in D$, M^C , w, $\sigma[x := a] \models ST_x(\varphi)$. It follows that for every a, $aR_{?\varphi}a$ is true at w. Then by $\dot{K}(?\varphi)$ true at w, we have for every a, $KaR_{?\varphi}a$ is also true at w. This means $aR_{?\varphi}a$ is true at each w' with wRw' for every a. That is, in an arbitrary σ and every $a \in D$, M^C , $w', \sigma[x := a] \models ST_x(\varphi)$ and then we have $M^C, w', \sigma \models \forall xST_x(\varphi)$. It follows $M^C, w', \sigma[x := a] \models ST_x(\varphi)$ and then we have $M^C, w', \sigma \models \forall xST_x(\varphi)$. It follows $M^C, w' \models \forall xST_x(\varphi)$ since σ is arbitrary, that is just $M^C, w' \models \varphi$ by abbreviation. Now we have $K\varphi$ true in w.

It may be also interesting to check the following.

Fact 5.6.

The formula
$$K(?\varphi) \leftrightarrow K(?\neg\varphi)$$
 is valid.

Proof. We only prove the left to right and the other direction can be done similarly. Suppose M^C , $w \models \dot{K}(?\varphi)$ for arbitrary M^C and w. We check $\dot{K}(?\neg\varphi)$ also holds at w. For arbitrary $a \in D$, first consider the case $aR_{?\neg\varphi}a$: it means M^C , $w, \sigma[x := a] \models ST_x(\neg\varphi)$ for every σ , that is, M^C , $w, \sigma[x := a] \nvDash ST_x(\varphi)$. It follows that $\neg aR_{\varphi}a$ is true in w. By the supposition of truth of $\dot{K}(?\varphi)$, we have $K \neg aR_{?\varphi}a$ is true at w. But $\neg aR_{?\varphi}a$ is equivalent to $aR_{?\neg\varphi}a$ (it can be shown at original model M), hence we get M^C , $w \models KaR_{?\neg\varphi}a$. Now we have M^C , $w \models \forall x(xR_{?\neg\varphi}x \to KxR_{?\neg\varphi}x)$, as required. Then consider the case $\neg aR_{?\varphi}a$ for arbitrary $a \in D$: it means M^C , $w, \sigma[x := a] \models ST_x(\varphi)$ for every σ , that is, $aR_{?\varphi}a$ is true at w. By the supposition of truth of $\dot{K}(?\varphi)$, we have $KaR_{?\varphi}a$ is true at w. By the supposition of truth of $\dot{K}(?\varphi)$, we have $KaR_{?\varphi}a$ is true at w. By the supposition of truth of $\dot{K}(?\varphi)$, we have $KaR_{?\varphi}a$ is true at w. But $aR_{?\varphi}a$ is equivalent to $\neg aR_{?\neg\varphi}a$, hence we get $w \models K \neg aR_{?\varphi}a$. Now we have M^C , $w \models \forall x(\neg xR_{?\neg\varphi}x \to K \neg xR_{?\neg\varphi}a)$, as required. The result of two cases shows that $\dot{K}(?\neg\varphi)$ is also true in w, meaning that $\dot{K}(?\varphi) \to \dot{K}(?\neg\varphi)$ is valid.

With the help of all the above results, we can derive the statement made at the beginning of this subsection, which said that 'knowing whether a proposition φ is just knowing it or knowing its negation':

$$\dot{K}(?\varphi) \leftrightarrow K\varphi \lor K \neg \varphi.$$

Proof. It is not difficult to show this: the left to right direction has already been shown. Let's do the other direction. Suppose $K\varphi \lor K\neg\varphi$ is true at an arbitrary world w of a model M^C . It's clear to have $K\varphi$ is true at w or $K\neg\varphi$ is true at w. In any case, we can have $\dot{K}(?\varphi)$ is true at w or $\dot{K}(?\neg\varphi)$ is true at w or $\dot{K}(?\varphi)$ has been shown valid. Application with the result $\dot{K}(?\varphi) \leftrightarrow \dot{K}(?\neg\varphi)$ we have just proved, it's easy to have $\dot{K}(?\varphi)$ is true at w, as required.

5.2 General properties of knowing actions

Just as in classical epistemic logic for propositions, we are interested in observing general properties in reasoning about knowing actions. Interestingly, many sound general results, such as Positive and Negative Introspections, as similar correspondences to classical epistemic logic can be proved valid. First, we have the following similar corresponding Positive Introspection property for knowing actions:

Fact 5.7.

The formula
$$K\alpha \rightarrow KK\alpha$$
 is valid.

Proof. Suppose for an arbitrary world w of a given first order epistemic model M^C , $M^C, w \models \dot{K}\alpha$. We need to show in every $w' \in W$ with $wRw', w' \models \dot{K}\alpha$, that is, $w' \models \forall x \forall y ((xR_\alpha y \rightarrow KxR_\alpha y) \land (\neg xR_\alpha y \rightarrow K \neg xR_\alpha y)).$

Now consider two arbitrary objects $c, d \in D$ and first suppose the case $w' \models cR_{\alpha}d$. There are two subcases need to be observed. One is $w \models cR_{\alpha}d$: according to the supposition $w \models K\alpha$, we can have $w \models KcR_{\alpha}d$, meaning that for every $u \in W$ with wRu, $u \models cR_{\alpha}d$. But we know wRw' as well and R is an equivalence relation, it follows that w'Ru for every $u \in W$. This means $w' \models KcR_{\alpha}d$, as required. The other subcase is $w \not\models cR_{\alpha}d$: it is equivalent to $w \models \neg cR_{\alpha}d$ since $cR_{\alpha}d$ is nothing to do with first order assignments. Then by the supposition of $w \models K\alpha$ we get $w \models K \neg cR_{\alpha}d$. It follows that $w' \models \neg cR_{\alpha}d$ since wRw', contradicting to $w' \models cR_{\alpha}d$. So this subcase is impossible.

Next suppose the case $w' \models \neg cR_{\alpha}d$. Similarly we need to consider two subcases. One is $w \models cR_{\alpha}d$: this case can be similarly excluded by deriving a contradiction. We only observe the other subcase $w \not\models cR_{\alpha}d$: it is equivalent to $w \models \neg cR_{\alpha}d$ for the same reason in above. Then by the supposition of $w \models K\alpha$ we get $w \models K \neg cR_{\alpha}d$. It means that for every $u \in W$ with wRu, we have $u \models \neg cR_{\alpha}d$. But we know wRw' and R is an equivalence relation, it follows that w'Ru holds for every $u \in W$ as well. This means $w' \models K \neg cR_{\alpha}d$, as required. Those two cases just show the result what we wanted. \Box

Interestingly, we have also a similar corresponding property of Negative Introspection:

Fact 5.8.

The formula
$$\neg K\alpha \rightarrow K \neg K\alpha$$
 is valid.

Proof. Suppose for an arbitrary world w of a given first order epistemic model M^C , M^C , $w \models \neg \dot{K}\alpha$. We need to show in every $w' \in W$ with wRw', $w' \models \neg \dot{K}\alpha$, that is, $w' \models \neg \forall x \forall y((xR_\alpha y \rightarrow KxR_\alpha y) \land (\neg xR_\alpha y \rightarrow K\neg xR_\alpha y))$, equivalent to $w' \models \exists x \exists y((xR_\alpha y \land \neg KxR_\alpha y) \lor (\neg xR_\alpha y \land \langle K \rangle xR_\alpha y)).$

Since $w \models \neg \dot{K}\alpha$ by supposition, we can find some objects *c*, *d* which satisfy the following condition:

- 1. $w \models cR_{\alpha}d \land \neg KcR_{\alpha}d$
- 2. $w \models \neg cR_{\alpha}d \wedge \langle K \rangle cR_{\alpha}d$

Now consider the first case:

- 1. $w' \models cR_{\alpha}d$. By $w \models \neg KcR_{\alpha}d$ in case 1, we have that there exists some $u \in W$ with wRu such that $u \models \neg cR_{\alpha}d$. But wRw' and R is an equivalence relation, then w'Ru. Now we have $w' \models \neg KcR_{\alpha}d$, and then $w' \models cR_{\alpha}d \land \neg KcR_{\alpha}d$, as required.
- 2. $w' \models \neg cR_{\alpha}d$. By $w \models cR_{\alpha}d$ and w'Rw (since wRw' and R is symmetric), we have $w' \models \langle K \rangle cR_{\alpha}d$, and then $w' \models \neg cR_{\alpha}d \land \langle K \rangle cR_{\alpha}d$.

In those two subcases of case 1, we can have $w' \models (cR_ad \land \neg KcR_ad) \lor (\neg cR_ad \land \langle K \rangle cR_ad)$ holds.

Next consider the second case, we also need to observe two subcases:

- 1. $w' \models cR_{\alpha}d$. By $w \models \neg cR_{\alpha}d$ in case 2 and w'Rw (since wRw' and R is symmetric), we have $w' \models cR_{\alpha}d \land \neg KcR_{\alpha}d$, as required.
- 2. $w' \models \neg cR_{\alpha}d$. By $w \models \langle K \rangle cR_{\alpha}d$ in case 2, we have that there exists some $u \in W$ with *wRu* such that $u \models cR_{\alpha}d$. But *wRw'* and *R* is an equivalence relation, then *w'Ru*. Now we have $w' \models \langle K \rangle cR_{\alpha}d$, and then $w' \models \neg cR_{\alpha}d \land \langle K \rangle cR_{\alpha}d$.

In those two subcases of case 2, we can have $w' \models (cR_ad \land \neg KcR_ad) \lor (\neg cR_ad \land \langle K \rangle cR_ad)$ holds as well.

Hence in any cases, there exist some objects $c, d \in D$ satisfying $w' \models (cR_{\alpha}d \land \neg KcR_{\alpha}d) \lor (\neg cR_{\alpha}d \land \langle K \rangle cR_{\alpha}d)$ with every world $w' \in W$ with wRw'. This just says $w \models K \neg K \alpha$ holds for the arbitrary given world w and model M^C with supposition $w \models \neg K \alpha$, as required.

Finally, we define the dual of $\dot{K}\alpha$ as $\langle \dot{K} \rangle \alpha$ by adding two negations into the universal first order sentence to get an existential version, that is,

$$\langle K \rangle \alpha = \exists x \exists y ((R_{\alpha} xy \rightarrow KR_{\alpha} xy) \land (\neg R_{\alpha} xy \rightarrow K \neg R_{\alpha} xy))$$

Then it is possible to show the following general properties (for a non-empty domain *D*):

$$\dot{K} \alpha \rightarrow \langle \dot{K} \rangle \alpha$$

 $\langle \dot{K} \rangle \alpha \rightarrow K \langle \dot{K} \rangle \alpha$

These may be seen as correspondents of the Axioms ${\bf D}$ and ${\bf 5}$ in classical doxastic logic.

5.3 Knowing complex actions

Now we provide an extensive technical discussion of how knowing complex actions relates to knowing their components. We follow the inductive steps in defining complex *PDL* programs, in a number of subsections.

Abort and Skip First consider the simplest case of the *abort* action ι . The relation symbol of *abort* action R_{ι} should be interpreted as \emptyset in every model. It is intuitive that the *abort* action should be known in every world by rational agents, and $\dot{K}\iota$ can indeed be checked to be valid in our first order epistemic setting. The argument is this. For

the first clause of definition of $K\iota$, given an arbitrary world w, since no pair of states (individuals) is in the empty relation, then $\forall x \forall y (R_\iota xy \rightarrow KR_\iota xy)$ holds vacuously. Then for the second clause of definition, in every world, every pair of two states (a, b) does not belong to \emptyset , that is, $\neg aR_\iota b$ is true in every world. It's clear that $w \models K \neg aR_\iota b$.

Next, consider the atomic *skip* action \Downarrow . The corresponding relation symbol R_{\Downarrow} is interpreted as $\Delta_D = \{(a, b) \in D^2 | a = b\} = \{(a, a) | a \in D\}$ for every world in each of the above first order model. In this way, we can again safely assume that $\dot{K} \Downarrow$ is valid. Here is the proof in more formal detail. Suppose for arbitrary states $c, d \in D$, if $(c, d) \in R_{\Downarrow}$ in a world w, then c = d holds as well and actually in every world since D = S is fixed and R_{\Downarrow} is interpreted same (identity relation) in every world. It follows that $Kc\Delta_Dd$) in w. If $(c, d) \notin R_{\Downarrow}$ is true in world w, then $c \neq d$ holds in world w as well and then at every world by the same reason that D is fixed. Similarly we have $K \neg c\Delta_D d$) in w. Those just show $\dot{K} \Downarrow$ holds in w. But w is arbitrary, it means that $\dot{K} \Downarrow$ is true in every world of an arbitrary model.

Converse actions A natural next step is the relationship between knowledge of an action α and its converse action. Suppose an agent knows α in a world. It means that for each pair of states, the agent can decide whether it satisfies R_{α} or not. For the agent is rational and has sufficient inference ability, he can decide whether the pair satisfies the converse of R_{α} as well. And vice versa. The following fact bears this out:

Fact 5.9.

The formula $\dot{K}\alpha \leftrightarrow \dot{K}\alpha^{*}$ is valid.

Proof. It is not difficult to prove it formally. In every world w of each first order model M^C , R_{α^*} is interpreted as the converse relation of $R_{\alpha}^{I,w}$. And $\dot{K}\alpha$ is just $\forall x \forall y ((R_{\alpha}xy \rightarrow KR_{\alpha}xy) \land (\neg R_{\alpha}xy \rightarrow K\neg R_{\alpha}xy))$. It is equivalent to $\forall y \forall x ((R_{\alpha^*}yx \rightarrow KR_{\alpha^*}yx) \land (\neg R_{\alpha^*}yx \rightarrow K\neg R_{\alpha^*}yx))$ by the definition of converse relation and first order logic. The latter is just $\dot{K}\alpha^*$ by the definition of knowing actions. And vice versa.

Sequential combination Next we take a look at knowledge of the crucial sequential combination of two actions α and β given knowledge of α and β . Intuitively speaking, for a rational agent, if he can grasp all the denotations of R_{α} and R_{β} , it should be possible for him to grasp the denotation of new relation which is combined from the relations R_{α} and R_{β} . That is, we expect

Fact 5.10.

The formula
$$K\alpha \wedge K\beta \rightarrow K(\alpha;\beta)$$
 is valid.

Proof. Interestingly, this can be proved formally in our first order epistemic setting: suppose we have M^C , $w \models \dot{K} \alpha \land \dot{K} \beta$ in an arbitrary world w of a given model M^C , and first consider the case of arbitrary a and b in D of M^{C} (we will directly use variables such as x, y to represent individuals a, b in later proofs for convenience, and M^C , w may also be omitted if the context is clear), $aR_{\alpha} \circ R_{\beta}b$ (the added relation symbol $R_{\alpha;\beta}$ is interpreted as a binary relation $R_{\alpha} \circ R_{\beta}$ in our first order epistemic models just as the meaning in classical logic of actions) holds. It means that there exists z such that $aR_{\alpha}z\wedge$ $zR_{\beta}b$ holds at w. And then we have, there exists z such that $KaR_{\alpha}z \wedge KzR_{\beta}a$ is true at w as well by the definition of knowing an action and supposition of $\dot{K}\alpha \wedge \dot{K}\beta$. It follows that there exists z, such that $K(aR_{\alpha z} \wedge zR_{\beta}b)$ is true in w by classical epistemic logic. According to first order modal logic K with constant domains expressed as in (Priest 2008) and (Arló-Costa and Pacuit 2006), we can safely conclude that $K \exists z(aR_{\alpha}z \wedge zR_{\beta}b)$, that is $KaR_{\alpha} \circ R_{\beta}b$ ($KaR_{\alpha;\beta}b$) is true in w. Since a, b are arbitrary, it follows that $\forall xy(KxR_{\alpha} \circ R_{\beta}y)$ is true at w, as required. Now consider the case of $\neg xR_{\alpha} \circ R_{\beta}y$ for arbitrary x, y that is true in w. It means that $\neg \exists z(xR_{\alpha}z \wedge zR_{\beta}y)$. By first order logic, it is equivalent to $\forall z(\neg xR_{\alpha}z \lor \neg zR_{\beta}y)$. Since we have $\dot{K}\alpha \land \dot{K}\beta$ true at w, it is clear to get $\forall z(K \neg xR_{\alpha z} \lor K \neg zR_{\beta y})$ true at w, implying $\forall zK(\neg xR_{\alpha z} \lor \neg zR_{\beta y})$ in basic modal logic, and that implies $\forall z K \neg (x R_{\alpha} z \land z R_{\beta} y)$ by propositional logic. As we know that BF and CBF are valid in the standard first order modal logic (relational semantics) with constant domains, it's safe to conclude that w satisfies $K \forall z \neg (xR_{\alpha}z \land zR_{\beta}y)$ which is equivalent to $K \neg \exists z (xR_{\alpha}z \land zR_{\beta}y)$, that is, $K \neg xR_{\alpha} \circ R_{\beta}y$ is true at w, as required.

We may expect the converse holds as well. But that does not work:

Fact 5.11.

The formula $\dot{K}(\alpha;\beta) \rightarrow \dot{K}\alpha \wedge \dot{K}\beta$ is not valid.

Proof. From our experience of properties about binary relations, we can conclude that it should not be valid in general. Intuitively speaking, if an agent knows all pairs in the extension of two combined binary relations, it does not mean that he knows all pairs in extensions of separate relations, because the combination can be generated from different possibilities. Even with more information such as knowing one of the combined actions, the agent cannot necessarily know the other. This means $\dot{K}(\alpha;\beta) \wedge$ $\dot{K}\alpha \rightarrow \dot{K}\beta$ does not valid either. Suppose we have $R_a^{I,w} = \{(1,2), (1,3)\}$ and $R_b^{I,w} =$ $\{(2,2), (3,2)\}$ to represent denotations of atomic actions *a* and *b* respectively. It follows that $R_{a;b}^{I,w} = \{(1,2)\}$. Then the agent cannot make a certain conclusion about R_b from the information of R_a and R_a . It is possible for him to get results which are different from the original $R_b^{I,w}$.

A fact pointing in the same direction is that the formula is not valid in general at the tableau framework as done in (Priest 2008).

Choice Turning to combinations of actions by choice, we can make similar observations.

Fact 5.12.

The formula
$$\dot{K}\alpha \wedge \dot{K}\beta \rightarrow \dot{K}(\alpha \cup \beta)$$
 is valid.

Proof. Suppose $\dot{K}\alpha \wedge \dot{K}\beta$ is true at an arbitrary w of M^C , and for arbitrary x, y, consider the case that $xR_{\alpha\cup\beta}y$ holds at w. In dynamic logic for actions, we know that $R_{\alpha\cup\beta} = R_{\alpha} \cup R_{\beta}$, so we have $xR_{\alpha} \cup R_{\beta}y$. It follows that $xR_{\alpha}y$ or $xR_{\beta}y$ and then $KxR_{\alpha}y$ or $KxR_{\beta}y$ is true at w since $\dot{K}\alpha \wedge \dot{K}\beta$ is true there. According to normal modal logic, we conclude $K(xR_{\alpha}y \vee xR_{\beta}y)$, that is, $K(xR_{\alpha\cup\beta}y)$ is true at w, as required. Now consider the case $\neg xR_{\alpha\cup\beta}y$ for arbitrary x, y. It means that $\neg xR_{\alpha}y$ and $\neg xR_{\beta}y$ are true at w. With the help of supposition that $\dot{K}\alpha \wedge \dot{K}\beta$ is true there, we can conclude $K(\neg xR_{\alpha}y \wedge \neg xR_{\beta}y)$ in propositional epistemic logic, and then equivalent to $K(\neg xR_{\alpha}y \wedge \neg xR_{\beta}y)$ in propositional logic. Since $xR_{\alpha}y \vee xR_{\beta}y$ is just $xR_{\alpha\cup\beta}y$ by the definition of \cup , the latter is just $K\neg (xR_{\alpha\cup\beta}y)$. Hence, it is also true at w, as required.

Like with serial combination, the converse does not hold.

Fact 5.13.

The formula
$$\dot{K}(\alpha \cup \beta) \rightarrow \dot{K}\alpha \wedge \dot{K}\beta$$
 is not valid.

Proof. It is easy to understand this intuitively. As we know, knowing the denotation of $R_{\alpha} \cup R_{\beta}$ is not helpful enough to know its components separately. And actually we can check an even weaker version,

$$\dot{K}(\alpha \cup \beta) \rightarrow \dot{K}\alpha \vee \dot{K}\beta$$

is not valid in the above first order epistemic framework by tableaux. It may help us understand this through the following counter example. Suppose in a model M^C , we have $R_{\alpha}^{I,w} = \{(1,2), (1,3)\}$ and $R_{\beta}^{I,w} = \{(2,1), (1,3)\}$. It clear that $R_{\alpha}^{I,w} \cup R_{\beta}^{I,w} = \{(1,2), (1,3),$ $(2,1)\}$. We cannot decide where are the elements in the union set from (a particular component set), such as (1,3), it could be from R_{α} or R_{β} or both. So in general the agent can not make a clear boundary between the extensions of R_{α} and R_{β} within the denotation of $R_{\alpha} \cup R_{\beta}$, that means he may not know any one of α and β .

Interestingly the converse of the above weaker version,

$$(\dot{K}\alpha \lor \dot{K}\beta) \to \dot{K}(\alpha \cup \beta)$$

is not valid either just as we expected since knowing one of component programs is not enough to get knowing the whole big program. Furthermore,

$$\dot{K}\alpha \to \dot{K}(\alpha \cup \beta)$$

and

$$\dot{K}(\alpha \cup \beta) \land \neg \dot{K}\alpha \to \dot{K}\beta$$

are both invalid in our first order epistemic framework. It would appear that they are in keeping with our intuitions about knowledge of random combination of binary relations.

Kleene star Now we continue with observing properties of knowledge over iterations α^* , arguably the most complex form of structured action. Recall that in classical dynamic logic for regular programs (actions), we interpreted R_{α^*} as $(R_{\alpha})^*$ which is the reflexive transitive closure of R_{α} (We can also see this in (Doets and van Eijck 2004, van Benthem et al. 2012). It means that

$$(R_{\alpha})^* = \Delta \cup R_{\alpha} \cup R_{\alpha}^2 \cup R_{\alpha}^3 \cup \dots$$

Here R_{α}^{n} is defined (interpreted) as follows in our first order epistemic model with countably infinite language:

$$R^{0}_{\alpha} = \Delta_{D}$$

$$R^{n}_{\alpha} = R_{\alpha} \circ R^{n-1}_{\alpha} \text{ for } n > 0.$$

This can be expressed without the ..., as mentioned in the standard translation:

$$R^*_{\alpha} = \bigcup_{n \in \mathbb{N}} R^n_{\alpha}.$$

Perhaps the first natural question on this issue is: if a rational agent knows an action α , is it reasonable for him to know the compound action α^* that is generated from α ? The intuitive answer seems "yes", at least at the level of denotations where we are operating. And indeed we can show

Fact 5.14.

The formula
$$\dot{K}\alpha \rightarrow \dot{K}\alpha^*$$
 is valid.

Proof. In order to prove this, we first define a notion of α^n (for $n \in \mathbb{N}$).

$$\alpha^0 = \Downarrow$$

 $\alpha^n = \alpha; \alpha^{n-1} \text{ for } n > 0$

Then we are going to prove a following lemma:

For every $n \in \mathbb{N}, \dot{K}\alpha \to \dot{K}\alpha^n$.

For the case of n = 0, α^0 means there is no execution of α , then α^0 is just \Downarrow . It has been shown that $\dot{K} \Downarrow$ is valid. For the case of n = 1, $\dot{K}\alpha \rightarrow \dot{K}\alpha$ is trivially valid as well. Now suppose $\dot{K}\alpha \rightarrow \dot{K}\alpha^k$ is valid for $k \in \mathbb{N}$. It is clear to get $\dot{K}\alpha \rightarrow \dot{K}\alpha \wedge \dot{K}\alpha^k$ is also valid by the induction hypothesis. But we have already shown that $\dot{K}\alpha \wedge \dot{K}\alpha^k \rightarrow K(\alpha; \alpha^k)$ is valid, it follows that $\dot{K}\alpha \rightarrow K(\alpha; \alpha^k)$, that is, $\dot{K}\alpha \rightarrow \dot{K}\alpha^{k+1}$ is valid as well.

Now we can prove $\dot{K}\alpha \rightarrow \dot{K}\alpha^*$ is valid: suppose $\dot{K}\alpha$ is true in an arbitrary world w of an arbitrary model M^C . It follows from the lemma that M^C , $w \models \dot{K}\alpha^n$ for every $n \in \mathbb{N}$. For any x, y, first consider the case $x(R_\alpha)^* y$ true in w. It means that at least one of the following situations holds in w: $x\Delta_D y$, $xR_\alpha y$, ..., $xR_{\alpha^n} y$, ... (for some n > 1). Without loss of generality, suppose it is $xR_{\alpha^i} y$ (for i > 1). Then by the lemma we have just proved, we get $KxR_{\alpha^i} y$ is true in w. This means for all w' satisfying wRw', $xR_{\alpha^i} y$ is true at w'. It's clear to conclude that $xR_{\alpha}^* y$ is true at every w' since $R_{\alpha^i} \subseteq R_{\alpha}^*$. Hence we have $KxR_{\alpha}^* y$ is true at w, as required. Next consider the case $\neg x(R_\alpha)^* y$ true in w. It means none of the following situations holds in w: $x\Delta_D y$, $xR_\alpha y$, ..., $xR_{\alpha^n} y$, ... (for some n > 1), that is, $\neg x\Delta_D y$, $\neg xR_\alpha y$, ..., $\neg xR_{\alpha^n} y$, ... are all true at w. By the above lemma again, we can get $K \neg x\Delta_D y$, $K \neg xR_\alpha y$, ..., $K \neg xR_\alpha^n y$, ... are all true at w. It follows that $\neg x\Delta_D y$, $\neg xR_\alpha y$, ..., $\neg xR_\alpha^n y$, ... are all true at w. It means $xR_\alpha^* y$ is false (that is, $\neg xR_\alpha^* y$ is true) in every w'. This shows $M, w \models K \neg xR_\alpha^* y$, as required.

The converse principle, that is, the implication

$$\dot{K}\alpha^* \to \dot{K}\alpha$$

seems not valid in general. It is not difficult to check this by providing a counter model, but we omit this here.

From all the above results on knowing complex actions, we can draw two general conclusions. It can be proved in all cases that, if an agent knows the components of an action, then he also knows the complex action. But the converses do not hold in general, and several counter-examples have been found. What this means to us is that the level of mere denotations does not provide enough information for the agent to recapture how the components worked. This seems a natural limit to what can be achieved by letting knowledge just operate on the transition relation associated with complex actions or procedures. If we want to go further, more structure will have to be given for the objects of knowledge.

6 Some related work

An earlier logic for knowing actions can be found in (Li 2005, Liu and Li 2005). Their work is mainly based on a propositional modal language and mainly epistemic actions are considered. The definition of knowing an action $K\alpha$ in Li (2005) is

$$M, s \models K\alpha$$

if and only if for any $t, t' \in S$, if sRt and tRt', then $(t, t') \in R_{\alpha}$.

Here $M = (S, R, R_{\alpha} \{ \text{for each action symbol } \alpha \}, V \}$ is a dynamic epistemic model, and *s* is a state similar to those defined in section 2 here (an epistemic relation *R* has been added). In this way, knowing an action α at *s* can be understood as 'in every *R*-successor *t* knows all the output states of α '.

While this has some technical charm and virtues in its formal development, we doubt the motivation of this definition on at least two points. First, the input states of actions are ignored, even though these seem crucial to understanding what an action does. And next, agents know something in *R*-successors, but not directly at *s*, which seems strange. We can also derive several undesired logical validities in this framework, which we do not pursue here.

For earlier preliminary work on logic of knowing complex actions, see (Wang and Cui 2010). The authors define a kind of denotation for actions α as a non-empty subset of W, making actions degenerate into propositions. And some principles that come out of this are dubious. For example, $K(\alpha;\beta) \rightarrow K\alpha$ is valid, making knowledge of a sequential combination of two actions lead to the knowledge of first action. This seems implausible.

Even though we see the value of these two pioneering attempts, we decided to set out in a different direction in this paper, for the reasons stated.

7 Conclusions and further directions

In this paper, we have proposed and investigated an analysis of knowing an action for a rational agent. Using some plausible intuitions from dynamic logic, we defined knowledge of an action as knowing the transition relation of the action. Our full version of this idea took place in the framework of first order epistemic models, with the help of a standard translation on program expressions. We investigated some basic logical principles of reasoning validated in this way, and applied them in particular to general epistemic properties of knowing actions, and to knowledge of different ways of generating actions, through tests, serial combination, choice, and iteration. We found some interesting principles, such as the equivalence of knowing a test with knowing whether the tested proposition holds. Beyond that, because of the shape of our definition, knowing an action basically had the same introspection properties as those assumed in the base logic adopted for propositional knowledge.

The general base logic of our system is first order epistemic predicate logic **S5** plus the axioms BF and CBF for constant domains. Inside this full system, the logic of knowing actions in our sense will involve just a smaller fragment of the full language, but we have not determined its special properties.

Another obvious desideratum concerns a sort of imbalance in our system. We have propositional knowledge on knowing programs, but these programs themselves were knowledge-free. But of course, it makes eminent sense to also look at programs that themselves involve epistemic structure. For instance, 'knowledge programs' with epistemic test conditions are studied in (Fagin et al. 1995, Baltag and Moss 2004). For instance, knowing the program $\pi = \text{IF } Kp$ THEN α ELSE *skip* seems intuitively equivalent to $\dot{K}((?Kp; \alpha) \cup \Downarrow))$ in our language. Clearly more needs to be done here, to bring this within the scope of our analysis. Concrete examples of settings where our proposal applies may be imperfect information games (van Benthem 2001, van Benthem and Liu 2004), where one often speaks of knowing a strategy, leading to knowledge programs for "uniform strategies".

And even knowledge programs are just one instance of what can be seen as a more general interest in "epistemizing" notions from traditional modal and dynamic logic. To mention one other example, just think of an issue like the following, connecting our various notions in a yet different way. Consider the basic invariance of *bisimulation* underlying modal logic. What does it mean to know a bisimulation between two models (a notion of knowledge referring to a relation once more), and what follows from that for our propositional knowledge of what is true in those models?

Next, our proposal is a simplest start, with a very sparse notion of action. One can easily increase the level of detail here. One simple obvious line left unexplored is the kind of denotation that we have taken. There are also semantics for dynamic logic (and its extensions such as the modal μ -calculus) that do not give just input-output relations, but also intermediate "traces". We think that our proposal will also work on such more structured trace semantics for actions, but the precise logical effects of this remain to be explored.

But in the end, we think that even this is not enough, and we need much more fine-grained views of the inner structure of actions or procedures, to give the knowledge operator more structure to work on. These richer denotations might come from other process theories in computer science, or also from games as models of structured interactive computation. Intuitions about validity and non-validity of the principles discussed in this paper may then reverse. We have nothing to offer in this direction here. But we do hope that we have provided a starting point that lends itself to further investigation, while mapping out issues that will return even as we turn on the magnification in our account of actions and processes.

Acknowledgements This work was sponsored by the youth project of National Social Science Foundation of China (No. 09CZX032) and the major program of National Social Science Foundation of China (No. 12&ZD118). Most of the research for this paper was done while I was visiting ILLC, University of Amsterdam (The Netherlands) from August 2011 to August 2012 funded by China Scholarship Council. I would like to extend my gratitude to Johan van Benthem and Jan van Eijck for their creative and helpful guidance, suggestions, comments on the research, and careful helps in revising, polishing the article as well. Special thanks to Alexandru Baltag, Sonja Smets and other colleagues involved in Amsterdam for their organizing and participating LIRa and LogICCC seminars from those activities I have acquired several views, ideas and techniques related to the research.

References

H. Arló-Costa and E. Pacuit. First-order classical modal logic. *Studia Logica*, 84(2): 171–210, 2006.

R. J. Aumann. Agreeing to disagree. The Annals of Statistics, 4(6):1236–1239, 1976.

A. Baltag and L. Moss. Logics for epistemic programs. *Synthese: Knowledge, Rationality, and Action*, 139(2):165–224, 2004.

A. Baltag, L. Moss, and S. Solecki. The logic of public announcements, common knowledge, and private suspicions. In I. Bilboa, editor, *Proceedings of TARK'98*, pages 43–56, 1998.

P. Battigalli and G. Bonanno. Recent results on belief, knowledge and the epistemic foundations of game theory. *Research in Economics*, 53(2):149–225, 1999.

J. van Benthem. *Exploring Logical Dynamics*. Center for the Study of Language and Information - Studies in Logic, Language, and Information, 1996.

J. van Benthem. Games in dynamic epistemic logic. *Bulletin of Economic Research*, 53:216–248, 2001.

J. van Benthem. *Logical Dynamics of Information and Interaction*. Cambridge University Press, 2011.

J. van Benthem and F. Liu. Diversity of logical agents in games. *Philosophia Scientiae*, 8(2):163–178, 2004.

J. van Benthem, H. van Ditmarsch, J. van Eijck, and J. Jaspars. *Logic in Action*. Open Course Project, Institute for Logic, Language and Computation, University of Amsterdam, 2012.

P. Blackburn, M. de Rijke, and Y. Venema. *Modal Logic*. Cambridge University Press, 2001.

H. van Ditmarsch, W. van der Hoek, and B. Kooi. *Dynamic Epistemic Logic*, volume 337 of *Synthese Library*. Springer, 2006.

K. Doets and J. van Eijck. *The Haskell Road to Logic, Maths and Programming*. College Publications, 2004.

R. Fagin, J. Halpern, Y. Moses, and M. Vardi. *Reasoning about Knowledge*. Cambridge : The MIT Press, 1995.

P. Gärdenfors. *Knowledge in Flux: Modeling the Dynamics of Epistemic States*. The MIT Press, Cambridge, MA, 1988.

J. Halpern. Using reasoning about knowledge to analyse distributed systems. *Annual Review of Computer Science*, 2:37–68, 1987.

J. Halpern, D. Samet, and E. Segev. Defining knowledge in terms of belief: the modal logic perspective. *Review of Symbolic Logic*, 2(3):469–487, 2009.

D. Harel, D. Kozen, and J. Tiuryn. Dynamic Logic. The MIT Press, 1st edition, 2000.

J. Hintikka. Knowledge and Belief. Cornell University Press, 1962.

X. Li. Three kinds of logics of knowing an action. *Logic and Cognition online journal*, 3(3):35–59, 2005.

Z. Liu and X. Li. Cognition of actions. *Journal of Hunan University of Science & Technology (Social Sciecne Edition)*, 8(6):33–38, 2005.

A. Perea. Epistemic Game Theory. Cambridge University Press, 2012.

V. R. Pratt. On the composition of processes. POPL, pages 213-223, 1982.

G. Priest. An Introduction to Non-Classical Logic. Cambridge University Press, second edition, 2008.

M. Singh. Know-how. In A. Rao and M. Wooldridge, editors, *Foundations of Rational Agency, Applied Logic Series*, pages 105–132. Kluwer, 1999.

J. Stanley. Know How. Oxford University Press, 2011.

J. Stanley and T. Williamson. Knowing how. *Journal of Philosophy*, 98(8):411–444, 2001.

R. Tang. Knowing that, knowing how, and knowing to do. *Frontiers of Philosophy in China*, 6(3):426–442, 2011.

J. Wang and J. Cui. Knowing regular action logic. *Journal of Southwest University*(*Social Sciecne Edition*), 36(4):59–65, 7 2010.

What You Can Do Depends on What You Have Done

Fengkui Ju and Li Liang

Department of Philosophy, Beijing Normal University, Beijing, China fengkui.ju@bnu.edu.cn, liangli_logic@163.com

Abstract

We aim to present a deontic logic with updates as an extension of Boolean Modal Logic. The features of this logic include the following: (a) deontic relations are defined on sets of finite sequences of states, called histories, and consequently, formulas are evaluated at histories, not states; and (b) it has two dynamic operators, which tend to update the obligation states of agents in different ways. This logic reflects the distinction between the descriptive and prescriptive use of norm sentences.

1 Introduction

One fundamental issue of deontic logic is *Jorgensen's dilemma* (1937). This dilemma was originally about imperatives. There are inferences involving imperatives in our lives. However, imperatives express orders and do not have truth values, so it is hard to say that there is a logic of imperatives. A dilemma arises. Traditionally, deontic logic does not consider imperatives. However, norm sentences such as "*you should stay*" or "*you may leave*" are similar to imperatives in many cases: they can also be used to change agents' behaviors, and therefore do not have truth values. Hence, this dilemma is also a serious problem in deontic logic. There are two puzzles attached to this dilemma: *Ross's Paradox* and the *Free Choice Permission Paradox*, both of which were identified by Ross (1944). The first puzzle can be illustrated by the inference "*you should mail this letter; therefore, you should mail it or burn it*". This inference is intuitively strange, but valid according to classical logic. The second puzzle is opposite

to the first one; it notes that the inference "you may drink coffee or tea; therefore, you may drink coffee" is not valid in the classical logic but is intuitively plausible.

In order to solve Jorgensen's dilemma, as mentioned in (Hilpinen 2001), many philosophers have proposed a distinction between two different uses of norm sentences: descriptive and prescriptive uses. Norm sentences are descriptively used to state what the agent ought to do or what he is allowed to do, among other actions, etc. These sentences can be true or false in these cases. In the prescriptive way, norm sentences are used to generate norms and do not have truth values. Jorgensen's dilemma would disappear if the prescriptive use of norm sentences were not relevant to deontic logic. Deontic logic is "legalized" this way. We consider this distinction reasonable. However, we do not think that prescriptive norm sentences are irrelevant to deontic logic. We believe that for any moral agent, there is an obligation state regarding his obligations and freedoms. Descriptive norm sentences describe these states, while prescriptive norm sentences change them. In this paper, we present a dynamic deontic logic to realize this concept.

There is a "dynamic" direction in deontic logic, in which works are based on dynamic logics. A fundamental work is that of Meyer (1988), which provided a deontic logic as an extension of Propositional Dynamic Logic. Influenced by Anderson-Kanger Deontic Logic, this work introduced a propositional constant, that intuitively means that the requirements of morality are violated. Deontic operators are defined by this constant, but they are applied to actions, not propositions. There is also a "dynamic" direction in semantics for imperatives and permissions starting from (Veltman 2009). This work is based on update semantics. It proposed a notion *plans*, i.e., a set of todo lists, which can be viewed as sets of actions. Imperatives and permissions update plans in different ways: the former tend to "strengthen" them, while the latter tend to "weaken" them.

This paper attempts to combine the spirits of these two research lines. As a propositional dynamic logic, Boolean Modal Logic contains these three action constructors: *complement, intersection* and *choice*. Our work is an extension of this logic in both language and semantics. The extended language contains a deontic operator, applied to actions, and two dynamic operators, corresponding to the descriptive utterance of obligations and the prescriptive utterance of permissions. The prescriptive utterance of obligations is derived from other utterances. A model is a labeled transition system plus a deontic relation, which is defined on the set of finite sequences of states, called *histories*, not on the set of states. The truth of a formula is defined against histories, not states. Histories represent what the agent has done. In this way, the idea of *what you have done affects what you can do* is reflected semantically. Descriptive norm sentences describe models, while prescriptive norm sentences update models by changing deontic relations. In this logic, Ross's Paradox is not valid, but the Free Choice Permission Paradox is. At the end, we axiomatize the logic.

2 Language and semantics

2.1 Language

Let Π_0 be a countable set of atomic actions and Φ_0 a countable set of atomic propositions. Let *a* range over Π_0 and *p* over Φ_0 . The sets Π of actions and Φ of propositions are defined as follows:

 $\begin{aligned} \alpha &::= a \,|\, 1 \,|\, \overline{\alpha} \,|\, (\alpha \cap \alpha) \,|\, (\alpha \cup \alpha) \\ \varphi &::= p \,|\, \top \,|\, O\alpha \,|\, \neg \varphi \,|\, (\varphi \wedge \varphi) \,|\, \langle \alpha \rangle \varphi \,|\, [\downarrow \alpha] \varphi \,|\, [\uparrow \alpha] \varphi \end{aligned}$

The empty action 0 is defined as $\overline{1}$. Other routine propositional connectives, the falsity \bot , and the dual $[\alpha]\varphi$ of $\langle \alpha \rangle \varphi$ are defined in the usual way. To perform $\overline{\alpha}$ is to do something that is not α . To perform $\alpha \cap \beta$ is to perform α and β at the same time. To perform $\alpha \cup \beta$ is to perform α or β . This language does not have compositions of actions, and all actions are just *one unit deep*. The formula $O\alpha$ means that the agent ought to do α . As the dual of $O\alpha$, $P\alpha$ is defined as $\neg O\overline{\alpha}$, which means that the agent may do α . For any α , $O\alpha$ is called a *pure deontic formula*. The expression $\downarrow \alpha$ denotes the action of descriptive utterance of "you should do α ", and $[\downarrow \alpha]\varphi$ means after this utterance, φ is true. $\uparrow \alpha$ denotes the prescriptive utterance of "you may do α ", and $[\uparrow \alpha]\varphi$ indicates that φ is true after the utterance.

2.2 Models

Let *W* be a set of states. Let Δ_W denote the set of finite non-empty sequences of states in *W*. Each element of Δ_W is called a *history* of *W*. Capitals like *H*, *J* and *K* denote histories. For any $H \in \Delta_W$, let \mathring{H} denote the last state of *H*. A *model* is a tuple $\mathfrak{M} = (W, \{R_\alpha \mid \alpha \in \Pi\}, \mathcal{D}, V)$ where

- 1. *W* is a non-empty set of states;
- 2. $R_{\alpha} \subseteq W \times W$;
- 3. $\mathcal{D} \subseteq \Delta_W \times \Delta_W$ and for any $(H, J) \in \mathcal{D}, J = (H, w)$ for some $w \in W$;
- 4. *V* is a function from Φ_0 to 2^W .

 \mathcal{D} is called the *deontic relation*. There is no loop for \mathcal{D} , i.e., any history H can not reach itself in finite steps. This intuitively means that the agent's history is always going to be his history. A model \mathfrak{M} is standard if it meets such additional constraints:

- 1. $R_1 = W \times W$;
- 2. $R_{\overline{\alpha}} = W \times W R_{\alpha};$
- 3. $R_{\alpha \cap \beta} = R_{\alpha} \cap R_{\beta};$
- 4. $R_{\alpha\cup\beta} = R_{\alpha} \cup R_{\beta};$
- 5. \mathcal{D} is serial.



Figure 1: A Standard Model

Figure 1 depicts what a standard model looks like. A labeled transition system is on the left, and the deontic relation is on the right. Histories are sequences of states. Since actions are transitions of states, histories represent what the agent has done. Suppose he is standing in w_2 with a *blank* history (w_2) behind him, which means that he has done nothing. According to the deontic relation, he now must perform *a*. After *a* is done, he is still in w_2 , however, his history is now (w_2, w_2), and he must perform *b*, which will take him to w_1 . What he is allowed to do is dependent on what he has done. There are three actions possible for the agent to perform in w_1 : *c*, *d* and *e*. However, given the history (w_2, w_2, w_1), as a moral agent, he is not allowed to do *e*, and he must perform *c* or *d*, although he can freely choose which one.

We require the deontic relation to be serial. We make this requirement for the following reasons: we believe that for any action, no matter what the world is and what the agent has done, he is allowed to perform it or the opposite of it, and we do not think a coherent legal system could tolerate the existence of situations in which the

agent is forbidden to do *anything*. In some cases, performing an action in a state might not change this state. For example, consider an agent pushing a revolving door. It is not reasonable to think that what the agent has to do never changes before and after performing this sort of actions, because otherwise if the agent has to push this door, he might have to push it forever. This is one reason we introduce histories as parameters in defining deontic relations. A second reason will be explained later.

2.3 Updates of models

Let $\mathfrak{M} = (W, \{R_{\alpha} \mid \alpha \in \Pi\}, \mathcal{D}, V)$ be a model, *H* a history, and α an action.

Definition 2.1 (Two Updates of Deontic Relations).

- 1. $\mathcal{D}_{\alpha}^{H} = \mathcal{D} \{(H, (H, w)) \mid \neg R_{\alpha}(\mathring{H}, w)\};$
- 2. $\mathcal{D}_{H}^{\alpha} = \mathcal{D} \cup \{(H, (H, w)) | R_{\alpha}(\mathring{H}, w)\}.$

The only difference among \mathcal{D} , \mathcal{D}^{H}_{α} and \mathcal{D}^{α}_{H} lies in that H might "see" less in \mathcal{D}^{H}_{α} and "see" more in \mathcal{D}^{α}_{H} than in \mathcal{D} . For any \mathcal{D} , let $g_{H}(\mathcal{D}) = \{w \in W \mid \mathcal{D}(H, (H, w))\}$, which is called the *goodness set* of H in \mathcal{D} . Let $R^{A}_{\alpha} = \{w \mid R_{\alpha}(\mathring{H}, w)\}$. It can be verified $g_{H}(\mathcal{D}^{H}_{\alpha}) = g_{H}(\mathcal{D}) \cap R^{\mathring{H}}_{\alpha}$ and $g_{H}(\mathcal{D}^{\alpha}_{H}) = g_{H}(\mathcal{D}) \cup R^{\mathring{H}}_{\alpha}$. Essentially, the two updates are two different ways of changing the goodness sets of H in \mathcal{D} . If \mathcal{D} is serial, then \mathcal{D}^{α}_{H} is serial, but \mathcal{D}^{H}_{α} might not be. However, given that \mathcal{D} is serial, if there is a w such that $\mathcal{D}(H, (H, w))$ and $R_{\alpha}(\mathring{H}, w)$, then \mathcal{D}^{H}_{α} is serial. The following proposition includes some results about manipulating updates, which will be used later:

Proposition 1.

- 1. $(\mathcal{D}^H_{\alpha})^J_{\beta} = (\mathcal{D}^J_{\beta})^H_{\alpha};$
- 2. $(\mathcal{D}_{H}^{\alpha})_{I}^{\beta} = (\mathcal{D}_{I}^{\beta})_{H}^{\alpha};$
- 3. $(\mathcal{D}^H_{\alpha})^{\beta}_{I} = (\mathcal{D}^{\beta}_{I})^{H}_{\alpha}$, where $J \neq H$.

Based on the updates of deontic relations, we define updates of models:

Definition 2.2 (Two Updates of Models).

- 1. $\mathfrak{M}^{H}_{\alpha} = (W, \{R_{\alpha} \mid \alpha \in \Pi\}, \mathcal{D}^{H}_{\alpha}, V);$
- 2. $\mathfrak{M}_{H}^{\alpha} = (W, \{R_{\alpha} \mid \alpha \in \Pi\}, \mathcal{D}_{H}^{\alpha}, V).$

The two updates only change the deontic relations of models. We see if \mathfrak{M} is standard, then $\mathfrak{M}_{H}^{\alpha}$ is standard, but $\mathfrak{M}_{\alpha}^{H}$ might not be. $\mathfrak{M}_{\alpha}^{H}$ can be viewed as the result of updating \mathfrak{M} with the descriptive utterance of "you should do α " at the history H, and \mathcal{D}_{H}^{α} as the result of updating \mathcal{D} with the prescriptive utterance of "you may do α " at H. The first update tends to "stop" some transitions, while the second tends to "free" some links. We take the model illustrated in Figure 1 as an example. Uttering "you should do c" in the descriptive way at (w_2, w_2, w_1) would cut the deontic link between (w_2, w_2, w_1) and (w_2, w_2, w_1, w_3) . This means the agent is not allowed to transition to w_3 and must perform c. Prescriptively uttering "you may do e" at (w_2, w_2, w_1) would generate a link between (w_2, w_2, w_1) and (w_2, w_2, w_1, w_5) , which means he can do e now.

2.4 Semantics

Let $\mathfrak{M} = (W, \{R_{\alpha} \mid \alpha \in \Pi\}, \mathcal{D}, V)$ be a model, and *H* a history. Here we do not require \mathfrak{M} to be standard. Truth of formulas at *H* is defined as follows:

- 1. $\mathfrak{M}, H \Vdash p \Leftrightarrow \mathring{H} \in V(p);$
- 2. $\mathfrak{M}, H \Vdash \top$ always holds;
- 3. $\mathfrak{M}, H \Vdash O\alpha \Leftrightarrow$ for any $w \in W$, if $\mathcal{D}(H, (H, w))$, then $R_{\alpha}(\mathring{H}, w)$;
- 4. $\mathfrak{M}, H \Vdash \neg \varphi \Leftrightarrow \operatorname{not} \mathfrak{M}, H \Vdash \varphi;$
- 5. $\mathfrak{M}, H \Vdash (\varphi \land \psi) \Leftrightarrow \mathfrak{M}, H \Vdash \varphi \text{ and } \mathfrak{M}, H \Vdash \psi;$
- 6. $\mathfrak{M}, H \Vdash \langle \alpha \rangle \varphi \Leftrightarrow$ there is a $w \in W$ such that $R_{\alpha}(\mathring{H}, w)$ and $\mathfrak{M}, (H, w) \Vdash \varphi$;
- 7. $\mathfrak{M}, H \Vdash [\downarrow \alpha] \varphi \Leftrightarrow \mathfrak{M}, H \Vdash P \alpha$ implies $\mathfrak{M}^{H}_{\alpha}, H \Vdash \varphi$;
- 8. $\mathfrak{M}, H \Vdash [\uparrow \alpha] \varphi \Leftrightarrow \mathfrak{M}_{H}^{\alpha}, H \Vdash \varphi$.

It can be verified that

- 9. $\mathfrak{M}, H \Vdash P\alpha \Leftrightarrow$ there is a $w \in W$ such that $\mathcal{D}(H, (H, w))$ and $R_{\alpha}(\mathring{H}, w)$;
- 10. $\mathfrak{M}, H \Vdash [\alpha] \varphi \Leftrightarrow$ for any $w \in W$, if $R_{\alpha}(\mathring{H}, w)$, then $\mathfrak{M}, (H, w) \Vdash \varphi$.

The formula $\langle \alpha \rangle \varphi$ being true at *H* means that there is a way to perform α such that after α is done, φ is true at the new history. It can be verified that $\mathfrak{M}, H \Vdash O\alpha$ if $g_H(\mathcal{D}) \subseteq R_{\alpha}^{\mathring{H}}$. This intuitively means that α is obligatory for the agent if whatever he does without violating morality, α would be performed. We can also verify that

 $\mathfrak{M}, H \Vdash P\alpha$ if $g_H(\mathcal{D}) \cap R_{\alpha}^{\mathring{H}} \neq \emptyset$. This means that he is allowed to perform α if there is a way to perform α without violating morality. Similar ideas can be found in (Hilpinen 2001).

We consider only standard models reasonable. As discussed, given that a model \mathfrak{M} is standard, \mathfrak{M}^H_{α} might not be standard, unless there is a $w \in W$ such that $\mathcal{D}(H, (H, w))$ and $R_{\alpha}(\mathring{H}, w)$, that is, $\mathfrak{M}, H \Vdash P\alpha$. This is why we define the truth condition of $[\downarrow \alpha]\varphi$ as conditional. Those updates resulting in non-standard models are unsuccessful ones. The truth of formulas is defined at histories in general models, not just in standard models, so the definition is well-defined. This semantics would collapse to classical relational semantics if the deontic part were ignored; hence, it is a genuine extension of Boolean Modal Logic.

A formula φ is *valid* if for any standard model \mathfrak{M} and history $H, \mathfrak{M}, H \Vdash \varphi$.

3 Valid formulas

Proposition 2. The following formulas are valid:

- 1. $O\alpha \rightarrow P\alpha$;
- 2. $P\alpha \rightarrow \langle \alpha \rangle \top$;
- *3.* $[\downarrow \alpha] O \alpha$;
- 4. $\langle \alpha \rangle \top \rightarrow [\uparrow \alpha] P \alpha$.

From the first two items, we obtain that *Kant's Law*, expressed as $O\alpha \rightarrow \langle \alpha \rangle \top$, is valid. The third indicates that the agent ought to do α after the descriptive utterance of "*you should do* α ". The last item expresses that he is allowed to do α after the prescriptive utterance of "*you may do* α ", given that α is possible to perform.

Hilpinen (2001) proposed a principle to explain why Ross's Paradox seems invalid: in our intuitions, if a norm sentence N_1 entails N_2 , then the normative effects of N_1 entail the normative effects of N_2 . The prescriptive utterance of "you should mail the letter or burn it" gives the agent the permission to burn the mail, but the utterance of "you should mail the letter" does not; therefore, the normative effects of the former do not entail the normative effects of the latter. Then Ross's Paradox is not valid. We consider this principle plausible. Even further, we believe that its converse is also reasonable. In fact, the bi-implication version of this principle underlies update semantics in defining validity. According to the stronger version, the Free Choice Permission Paradox is valid, as the prescriptive utterance of "you may drink coffee" just gives the agent the freedom to drink coffee, whereas the utterance of "you may drink coffee or tea" gives him the freedom to drink tea, in addition to the freedom to drink coffee.

Our language contains dynamic operators and models contain normative factors; thus, the normative effects of utterances can be expressed in this setting. We believe that prescriptive norm sentences generate not only obligations but also permissions. In (Ju and Liu 2011), we have argued that in the aspect of normative effects, prescriptively uttering "you should do α " is equivalent to prescriptively uttering "you may do α " and then descriptively uttering "you should do α ". Define $[\uparrow \downarrow \alpha]$ as $[\uparrow \alpha][\downarrow \alpha]$, which represents the action of prescriptively uttering "you should do α ". Ross's Paradox fails here. Let c denote the action *mailing the letter*, and e the action *burning the letter*. We look at the model illustrated in Figure 1. It can be verified that $[\uparrow\downarrow c][\uparrow\downarrow (c \cup e)]Pe$ is true at the history (w_2, w_2, w_1) , but $[\uparrow\downarrow c]Pe$ is false at it. Therefore, the normative effects of "you should mail the letter" do not entail the normative effects of "you should *mail it or burn it*". One may wonder why $[\downarrow \alpha] \uparrow \alpha$ is not used to denote prescriptive utterances of "you should do α ". The update sequence $[\uparrow \alpha][\downarrow \alpha]$ might be different from $[\downarrow \alpha] \uparrow \alpha$, and the only difference is this: given that α is possible to perform, $[\uparrow \alpha][\downarrow \alpha]$ would always be successful, but $[\downarrow \alpha][\uparrow \alpha]$ might not, as $[\downarrow \alpha]$ might make a standard model not serial. We believe that in real life, given α is possible to perform, prescriptive utterance of "you should do α " is always meaningful. This is the reason. The Free Choice Permission Paradox, $[\uparrow (\alpha \cup \beta)][\uparrow \alpha]\varphi \leftrightarrow [\uparrow (\alpha \cup \beta)]\varphi$, is valid in this semantics, which is easy to check.

The following lemma says that the two updates do not change a model much:

Lemma 1. J is a proper super-sequence of H.

- 1. $\mathfrak{M}^{H}_{\alpha}, J \Vdash \varphi$ if and only if $\mathfrak{M}, J \Vdash \varphi$;
- 2. $\mathfrak{M}_{H}^{\alpha}$, $J \Vdash \varphi$ if and only if \mathfrak{M} , $J \Vdash \varphi$.

By this lemma, we can show these two propositions:

Proposition 3. The following formulas are valid:

- 1. $[\downarrow \alpha]p \leftrightarrow (P\alpha \rightarrow p);$
- 2. $[\downarrow \alpha] \top \leftrightarrow (P\alpha \rightarrow \top);$
- 3. $[\downarrow \alpha] O \beta \leftrightarrow (P \alpha \rightarrow O(\overline{\alpha} \cup \beta));$
- 4. $[\downarrow \alpha] \neg \varphi \leftrightarrow (P\alpha \rightarrow \neg [\downarrow \alpha]\varphi);$
- 5. $[\downarrow \alpha](\varphi \land \psi) \leftrightarrow ([\downarrow \alpha]\varphi \land [\downarrow \alpha]\psi);$

- 6. $[\downarrow \alpha]\langle\beta\rangle\varphi \leftrightarrow (P\alpha \rightarrow \langle\beta\rangle\varphi);$
- 7. $[\downarrow \alpha] [\downarrow \beta] \varphi \leftrightarrow [\downarrow (\alpha \cap \beta)] \varphi$.

Proposition 4. The following formulas are valid:

- *1.* $[\uparrow \alpha]p \leftrightarrow p;$
- 2. $[\uparrow \alpha] \top \leftrightarrow \top;$
- 3. $[\uparrow \alpha] O \beta \leftrightarrow (O \beta \land [\alpha \cap \overline{\beta}] \bot);$
- 4. $[\uparrow \alpha] \neg \varphi \leftrightarrow \neg [\uparrow \alpha] \varphi;$
- 5. $[\uparrow \alpha](\varphi \land \psi) \leftrightarrow ([\uparrow \alpha]\varphi \land [\uparrow \alpha]\psi);$
- 6. $[\uparrow \alpha]\langle\beta\rangle\varphi \leftrightarrow \langle\beta\rangle\varphi;$
- 7. $[\uparrow \alpha][\uparrow \beta]\varphi \leftrightarrow [\uparrow (\alpha \cup \beta)]\varphi$.

From these propositions, we obtain that the formulas containing only one dynamic operator can be equivalently reduced to the formulas not containing any. By introducing histories, we can obtain valid formulas $[\downarrow \alpha]\langle\beta\rangle\varphi \leftrightarrow \langle\beta\rangle\varphi$ and $[\uparrow \alpha]\langle\beta\rangle\varphi \leftrightarrow \langle\beta\rangle\varphi$, and consequently obtain the reduction of dynamic operators. This is the above mentioned second motivation for using the notion of histories.

4 Axiomatization

4.1 Axiomatization

Let Φ_{PC} be the language generated from $\Phi_0 \cup \{\top\}$ under \neg , \land and \lor , where Φ_0 is the set of atomic propositions. Let *f* be a *natural* bijective function from the set Π of actions to Φ_{PC} . We say α and β are *equivalent* if $f(\alpha) \leftrightarrow f(\beta)$ is a tautology. For instance, $\overline{a \cap b}$ is equivalent to $\overline{a} \cup \overline{b}$. The axiomatization of the logic consists of eight classes of axioms:

- A. Basic axioms of normal modal logics:
 - (a) all propositional tautologies;
 - (b) $[\alpha](\varphi \to \psi) \to ([\alpha]\varphi \to [\alpha]\psi).$
- *B*. The axiom for choice: $\langle \alpha \cup \beta \rangle \varphi \leftrightarrow \langle \alpha \rangle \varphi \lor \langle \beta \rangle \varphi$.

- C. Axioms for the universal modality:
 - (a) $\varphi \to \langle 1 \rangle \varphi$;
 - (b) $\varphi \rightarrow [1]\langle 1 \rangle \varphi$;
 - (c) $\langle 1 \rangle \langle 1 \rangle \varphi \rightarrow \langle 1 \rangle \varphi$;
 - (d) $\langle \alpha \rangle \varphi \rightarrow \langle 1 \rangle \varphi$.
- D. The axiom for the empty modality: $[0] \perp$.
- *E.* Axioms for equivalence of actions: $\langle \alpha \rangle \varphi \leftrightarrow \langle \alpha' \rangle \varphi$, if α and α' are equivalent.
- F. Axioms for the deontic operator O:
 - (a) $(O\alpha \land O\beta) \leftrightarrow O(\alpha \cap \beta);$
 - (b) $\neg O\alpha \rightarrow P\overline{\alpha}$;
 - (c) $(O\alpha \land P\beta) \rightarrow P(\alpha \cap \beta);$
 - (d) $O\alpha \rightarrow \langle \alpha \rangle \top$;
 - (e) $P\alpha \rightarrow \langle \alpha \rangle \top$.
- *G*. Axioms for the dynamic operator $[\downarrow \alpha]$:
 - (a) $[\downarrow \alpha]p \leftrightarrow (P\alpha \rightarrow p);$
 - (b) $[\downarrow \alpha] \top \leftrightarrow (P\alpha \to \top);$
 - (c) $[\downarrow \alpha] O \beta \leftrightarrow (P \alpha \rightarrow O(\overline{\alpha} \cup \beta));$
 - (d) $[\downarrow \alpha] \neg \varphi \leftrightarrow (P\alpha \rightarrow \neg [\downarrow \alpha]\varphi);$
 - (e) $[\downarrow \alpha](\varphi \land \psi) \leftrightarrow ([\downarrow \alpha]\varphi \land [\downarrow \alpha]\psi);$
 - (f) $[\downarrow \alpha]\langle\beta\rangle\varphi \leftrightarrow (P\alpha \to \langle\beta\rangle\varphi).$
- *H*. Axioms for the dynamic operator $[\uparrow \alpha]$:
 - (a) $[\uparrow \alpha]p \leftrightarrow p;$
 - (b) $[\uparrow \alpha] \top \leftrightarrow \top;$
 - (c) $[\uparrow \alpha] O \beta \leftrightarrow (O \beta \land [\alpha \cap \overline{\beta}] \bot);$
 - (d) $[\uparrow \alpha] \neg \varphi \leftrightarrow \neg [\uparrow \alpha] \varphi;$
 - (e) $[\uparrow \alpha](\varphi \land \psi) \leftrightarrow ([\uparrow \alpha]\varphi \land [\uparrow \alpha]\psi);$
 - (f) $[\uparrow \alpha] \langle \beta \rangle \varphi \leftrightarrow \langle \beta \rangle \varphi$.

and three inference rules:

- 1. *Modus Ponens*: given φ and $\varphi \rightarrow \psi$, prove ψ ;
- 2. *Generalization*: given φ , prove $[\alpha]\varphi$;
- 3. *Replacement of Validity*: given $\varphi \leftrightarrow \varphi'$, prove $[\downarrow \alpha]\varphi \leftrightarrow [\downarrow \alpha]\varphi'$ and $[\uparrow \alpha]\varphi \leftrightarrow [\uparrow \alpha]\varphi'$.

The class **E** is decidable, as the set of tautologies is decidable. The logic is sound and complete respect to the class of standard models. The soundness is easy to verify. We now show the completeness. Firstly, we show that the logic restricted to Φ' , the sub-language of Φ not containing any dynamic operators, is complete with respect to the class of standard models. Then, by use of the classes **G** and **H** of axioms and the inference rule **RV**, we obtain the completeness of the whole logic in a similar way as (van Ditmarsch et al. 2007) for Public Announcement Logic.

To show the restricted completeness, it suffices to show for any consistent formulas φ in Φ' , there is a standard model \mathfrak{M} and a history H such that $\mathfrak{M}, H \Vdash \varphi$. Let G be a consistent formula. Let Σ be the smallest set of formulas such that $G \in \Sigma$ and Σ is closed under sub-formulas.

4.2 **CINF**

Let a_1, \ldots, a_n be all the atomic actions of Σ and $A = \{a_1, \ldots, a_n, \overline{a_1}, \ldots, \overline{a_n}\}$. Each element of A is called a literal action. Define Θ_A as such a set: $\{X \subseteq A \mid for any i \leq n, exactly one of a_i and \overline{a_i} is in X\}$. Θ_A has 2^n members. For any $X \in \Theta_A$, $\gamma = \bigcap X$ is called a path relative to Σ , which is an intersection of some literal actions. There are 2^n paths, if we do not consider orders of literal actions. Enumerate these paths as $\gamma_1, \ldots, \gamma_{2^n}$. In any standard model, $R_{\gamma_1}, \ldots, R_{\gamma_{2^n}}$ are pairwise disjoint blocks and the union of them is $W \times W$. In other words, $\{R_{\gamma_1}, \ldots, R_{\gamma_{2^n}}\}$ is a partition of $W \times W$. By some refections we can get that for any α built from a_1, \ldots, a_n , if α is not equivalent to 0, R_α is the union of some of these blocks. These blocks are like atomic parts of $W \times W$. Here is an example. Suppose a, b and c are all the atomic actions under considerations. There are 8 paths: $a \cap b \cap c, \ldots, \overline{a} \cap \overline{b} \cap \overline{c}$, and $W \times W$ is divided into 8 parts: $R_{a \cap b \cap c}, \ldots, R_{\overline{a} \cap \overline{b} \cap \overline{c}}$. Each non-empty action whose atomic actions occur in a, b and c is the union of some of these parts. For example, $R_{a \cap \overline{b} \cap \overline{c}} = R_{a \cap b \cap \overline{c}} \cup R_{a \cap \overline{b} \cap \overline{c}}$.

The classes **D** and **E** of axioms guarantee this result:

Lemma 2. For any α occurring in Σ , if α is not equivalent to 0, there are paths $\gamma_{n_1}, \ldots, \gamma_{n_m}$ such that α is equivalent to $\gamma_{n_1} \cup \cdots \cup \gamma_{n_m}$.

For any α not equivalent to 0, we call $\gamma_{n_1} \cup \cdots \cup \gamma_{n_m}$ the *choice-intersection normal* form (CINF) of α relative to Σ . Actions equivalent to 0 such as $a \cap \overline{a}$ do not have corresponding CINFs.

Lemma 3.

- 1. Let $\gamma_{h_1} \cup \cdots \cup \gamma_{h_i}$ and $\gamma_{j_1} \cup \cdots \cup \gamma_{j_k}$ be the CINFs of β and $\overline{\beta}$. Then $\{\gamma_{j_1}, \ldots, \gamma_{j_k}\} = \{\gamma_1, \ldots, \gamma_{2^n}\} \{\gamma_{h_1}, \ldots, \gamma_{h_i}\};$
- 2. Let $\gamma_{h_1} \cup \cdots \cup \gamma_{h_i}$, $\gamma_{j_1} \cup \cdots \cup \gamma_{j_k}$ and $\gamma_{l_1} \cup \cdots \cup \gamma_{l_m}$ be the CINFs of β , π and $\beta \cap \pi$. Then $\{\gamma_{l_1}, \ldots, \gamma_{l_m}\} = \{\gamma_{h_1}, \ldots, \gamma_{h_i}\} \cap \{\gamma_{j_1}, \ldots, \gamma_{j_k}\};$
- 3. Let $\gamma_{h_1} \cup \cdots \cup \gamma_{h_i}$, $\gamma_{j_1} \cup \cdots \cup \gamma_{j_k}$ and $\gamma_{l_1} \cup \cdots \cup \gamma_{l_m}$ be the CINFs of β , π and $\beta \cup \pi$. Then $\{\gamma_{l_1}, \ldots, \gamma_{l_m}\} = \{\gamma_{h_1}, \ldots, \gamma_{h_i}\} \cup \{\gamma_{j_1}, \ldots, \gamma_{j_k}\};$
- 4. If α is equivalent to 1, the CINF of α is $\gamma_1 \cup \cdots \cup \gamma_{2^n}$.

The axiom **D** is used in proving the second item.

4.3 An incomplete model and its generated submodel

Let $\mathfrak{M}^C = (W^C, \{R^C_\alpha \mid \alpha \in \Pi\}, V^C)$ be the structure where

- 1. W^C is the set of maximal consistent sets;
- 2. $R_{\alpha}uv$ if and only if for any $\varphi, \varphi \in v$ implies $\langle \alpha \rangle \varphi \in u$;
- 3. For any $p, V(p) = \{u \in W^C | p \in u\}.$

This structure is not a model, as the deontic relation is missing. Actually, if we ignore the *deontic part* of the language, it is the canonical model.

Lemma 4.

- 1. If $\langle \alpha \rangle \varphi \in u$, there is a $v \in W^C$ such that $\varphi \in v$ and $R^C_{\alpha}uv$;
- 2. $R_{\alpha\cup\beta}^C = R_{\alpha}^C \cup R_{\beta}^C$;
- 3. For any α equivalent to 0, $R_{\alpha}^{C} = \emptyset$.

Let *w* be a maximal consistent set containing *G*. Let $\mathfrak{M} = (W, \{R_{\alpha} | \alpha \in \Pi\}, V)$ be the substructure of \mathfrak{M}^{C} generated from *w* under the relation R_{1}^{C} . The class **C** of axioms guarantee that R_{1} is the universal relation on *W*. Here a similar lemma with Lemma 4:

Lemma 5.

- 1. If $\langle \alpha \rangle \varphi \in u$, there is a $v \in W$ such that $\varphi \in v$ and $R_{\alpha}uv$;
- 2. $R_{\alpha\cup\beta} = R_{\alpha} \cup R_{\beta};$
- *3.* For any α equivalent to 1, $R_{\alpha} = W \times W$;
- 4. For any α equivalent to 0, $R_{\alpha} = \emptyset$.

4.4 Filtration

Define a relation \approx_{Σ} on *W* as this: $u \approx_{\Sigma} v$ if and only if $u \cap \Sigma = v \cap \Sigma$. This is an equivalence relation. Let $\mathfrak{M}^f = (W^f, \{R^f_{\alpha} \mid \alpha \in \Pi\}, V^f)$ be such a structure:

- 1. W^f is the partition of W under \approx_{Σ} ;
- 2. $R_{\alpha}^{f}|u||v|$ if and only if there are $x \in |u|$ and $y \in |v|$ such that $R_{\alpha}xy$;
- 3. for any p, $V^{f}(p) = \{|u| | p \in u\}$.

For any $x, y \in |u|$ and $\varphi \in \Sigma$, $\varphi \in x$ if and only if $\varphi \in y$. We use $\varphi \triangleright |u|$ to express that $\varphi \in x$ for any $x \in |u|$. Here is a similar lemma with Lemma 5:

Lemma 6.

1. If $\langle \alpha \rangle \varphi \triangleright |u|$, there is a $v \in W$ such that $\varphi \triangleright |v|$ and $R_{\alpha}^{f}|u||v|$;

2.
$$R^{f}_{\alpha\cup\beta} = R^{f}_{\alpha} \cup R^{f}_{\beta}$$

- 3. For any α equivalent to 1, $R_{\alpha}^{f} = W^{f} \times W^{f}$;
- 4. For any α equivalent to 0, $R_{\alpha}^{f} = \emptyset$.

With the help of Lemma 3 and 6, it is not hard to show the following lemma:

Lemma 7.

- 1. For any α of Σ not equivalent to 0, $R_{\alpha}^{f} = R_{\gamma_{n_{1}} \cup \cdots \cup \gamma_{n_{m}}}^{f}$, where $\gamma_{n_{1}} \cup \cdots \cup \gamma_{n_{m}}$ is the CINF of α ;
- 2. $R_{\gamma_1}^f \cup \cdots \cup R_{\gamma_{2^n}}^f = W^f \times W^f;$
- 3. For any $\langle \alpha \rangle \top$ and $u \in W$, if $\langle \alpha \rangle \top \triangleright |u|$, there is a $v \in W$ such that $R_{\alpha}^{f} |u||v|$.

We present some observations on the situation confronting us. Our purpose is to show the consistent formula G is satisfiable in a standard model. Two things are important: a standard model and satisfiability. The structure $\mathfrak{M}^f = (W^f, \{R^f_\alpha \mid \alpha \in \Pi\}, V^f)$ might not be standard even if we ignore the deontic part, as $R_{\overline{\alpha}}^f = W^f \times W^f - R_{\alpha}^f$ and $R_{\alpha\cap\beta}^f = R_{\alpha}^f \cap R_{\beta}^f$ might not be satisfied. We hope to transform it to a standard model. Those actions not built from a_1, \ldots, a_n are irrelevant, and we can freely manipulate the interpretations of their parts not involving a_1, \ldots, a_n ; therefore, these actions do not present a problem. However, we are using the Henkin method, so at least to some extent, we should "respect" the interpretations of the actions built from a_1, \ldots, a_n , if we want to obtain satisfiability. Via some reflections with the help of Lemma 3 and 6, we can see that if $R_{\gamma_1}^f, \ldots, R_{\gamma_{2^n}}^f$ are real atomic blocks, \mathfrak{M}^f is standard. That $R_{\gamma_1}^f, \ldots, R_{\gamma_{2^n}}^f$ are real atomic blocks means that $R_{\gamma_1}^f, \ldots, R_{\gamma_{2^n}}^f$ are pairwise disjoint and $R_{\nu_1}^{f} \cup \cdots \cup R_{\nu_m}^{f} = W^f \times W^f$. The second condition holds by Lemma7. The situation is now clear: to achieve our goal, we only need to achieve two things: making $R_{\gamma_1}^J, \ldots, R_{\gamma_2}^J$ pairwise disjoint and respecting the interpretations of the actions built from a_1, \ldots, a_n . The copy method given in (Gargov and Passy 1990) can perform both at the same time, although it made a few mistakes, which will be explained later in a footnote.

4.5 A standard model

Let $\mathfrak{M}_1 = (W_1, \{R^1_{\alpha} \mid \alpha \in \Pi\}, V_1), \ldots, \mathfrak{M}_{2^n} = (W_{2^n}, \{R^{2^n}_{\alpha} \mid \alpha \in \Pi\}, V_{2^n})$ be 2^n pairwise disjoint structures which are isomorphic to \mathfrak{M}^f . We now build up a standard model from these structures.

For any $i \leq 2^n$, let f_i be an isomorphism from \mathfrak{M}_i to \mathfrak{M}^f . Let $f = f_1 \cup \cdots \cup f_{2^n}$ and $U = W_1 \cup \cdots \cup W_{2^n}$. Let $g : U \to \{1, \ldots, 2^n\}$ be this function: for any $s \in U$, g(s) is the index of the set from which *s* comes, i.e., for any $s \in U$, $s \in W_{g(s)}$. For any $s \in U$, let $\varphi \ltimes s$ denote $\varphi \triangleright f(s)$. For any $i \leq 2^n$, we define a relation B_{γ_i} on U:

Definition 4.1 (Atomic Relations). Let $s, t \in U$. Let $\gamma_{k_1}, \ldots, \gamma_{k_m}$ be the sequence such that (i) it consists of all paths γ such that $R_{\gamma}^f f(s)f(t)$ and (ii) $k_1 < \cdots < k_m$. $B_{\gamma_i}st$ if and only if there is a $j \leq m$ such that $i = k_j$ and $j = (g(t) \mod m) + 1$.

The sequence $\gamma_{k_1}, \ldots, \gamma_{k_m}$ is never empty, which is guaranteed by Lemma 7. If a path γ_i is not occurring in $\gamma_{k_1}, \ldots, \gamma_{k_m}$, there is no $j \leq m$ such that $i = k_j$, and so not $B_{\gamma_i} st$. Suppose γ_i is occurring in $\gamma_{k_1}, \ldots, \gamma_{k_m}$. Then there is one and only one $j \leq m$ such that $i = k_j$, which means that γ_i is the *j*-th element in $\gamma_{k_1}, \ldots, \gamma_{k_m}$. In this case, if $j = (g(t) \mod m) + 1$, then $B_{\gamma_i} st$, or else not. Our purpose is to produce atomic relations. To do this, we must get this: for any $s, t \in U$, (s, t) belongs to one and only one path. This definition gives a way to assign (s, t) to the "right" path¹. Here is a concrete example. Let a, b, c be all the atomic actions of Σ . Paths are $\gamma_1 = a \cap b \cap c, \ldots, \gamma_8 = \overline{a} \cap \overline{b} \cap \overline{c}$. Let $s, t \in U$ and g(t) = 4. Then $t \in W_4$. Suppose $R_{\gamma_2}^f f(s)f(t), R_{\gamma_5}^f f(s)f(t), R_{\gamma_7}^f f(s)f(t)$, and no other paths can do this. The sequence satisfying the two conditions in Definition 4.1 is $\gamma_2, \gamma_5, \gamma_7$. Then $k_1 = 2, k_2 = 5$ and $k_3 = 7$. As j = 2 satisfies that $5 = k_j$ and $j = (4 \mod 3) + 1$, we get $B_{\gamma_5} st$. For any $i \leq 8$, if $i \neq 5$, there is no j satisfying that $i = k_j$ and $j = (4 \mod 3) + 1$, and so not $B_{\gamma_1} st$. By the following lemma, $B_{\gamma_1}, \ldots, B_{\gamma_2}$ are real atomic relations:

Lemma 8.

- 1. $B_{\gamma_1}, \ldots, B_{\gamma_{2^n}}$ are pairwise disjoint;
- 2. $B_{\gamma_1} \cup \cdots \cup B_{\gamma_{\gamma_n}} = U \times U.$

Proof. (1) Assume there are $i, j \leq 2^n$ such that $i \neq j$ and $B_{\gamma_i} \cap B_{\gamma_j} \neq \emptyset$. Then there are $s, t \in U$ such that $B_{\gamma_i}st$ and $B_{\gamma_j}st$. Let $\gamma_{k_1}, \ldots, \gamma_{k_m}$ be the sequence such that it consists of all paths γ such that $R_{\gamma}^f f(s) f(t)$ and $k_1 < \cdots < k_m$. Let $x, y \leq m$ be such that $i = k_x$ and $j = k_y$. As $k_x \neq k_y, x \neq y$. By the definitions of B_{γ_i} and B_{γ_j} , we get that $x = (g(t) \mod m) + 1$ and $y = (g(t) \mod m) + 1$. This is impossible. Then $B_{\gamma_1}, \ldots, B_{\gamma_{2^n}}$ are pairwise disjoint.

(2) Trivially, we get $B_{\gamma_1} \cup \cdots \cup B_{\gamma_{2^n}} \subseteq U \times U$. Let $s, t \in U$. Let $\gamma_{k_1}, \ldots, \gamma_{k_m}$ be the sequence such that it consists of all paths γ such that $R^f_{\gamma}f(s)f(t)$ and $k_1 < \cdots < k_m$. Let $x \leq m$ be such that $x = (g(t) \mod m) + 1$. By the definition of $B_{\gamma_{k_1}}$, we have $B_{\gamma_{k_1}}st$. \Box

Definition 4.2 (A Model). $\mathfrak{N} = (U, \{S_{\alpha} \mid \alpha \in \Pi\}, \mathcal{E}, Z)$ is the model where

- 1. U is defined as above;
- 2. For any atomic action *a* occurring in Σ , $S_a = B_{\gamma_{n_1}} \cup \cdots \cup B_{\gamma_{n_m}}$, where $\gamma_{n_1} \cup \cdots \cup \gamma_{n_m}$ is the CINF of a^2 ; For any atomic action *b* not occurring in Σ , $S_b = U \times U$; Interpretations of compound actions are defined from interpretations of atomic actions by corresponding operations;

¹Gargov and Passy (1990) made a mistake at this point: the definition of atomic relations given by it can not guarantee that for any $s, t \in U$, (s, t) belongs to exactly one path. There are two other mistakes in this work: (i) By Lemma 2, those actions equivalent to 0 do not correspond to any CINF, but this paper did not notice this; (ii) Lemma 3 is necessary to the proof of completeness, but it is not mentioned in this paper at all.

²Here a might be the universal action 1.

- 3. $\mathcal{E}(H, J)$ if and only if there is a $s \in U$ such that J = (H, s) and for any $O\alpha \in \Sigma$, if $O\alpha \ltimes \mathring{H}$, then $S_{\alpha}(\mathring{H}, s)$;
- 4. For any $p, Z(p) = \bigcup_{i \le 2^n} V_i(p)$.

Clearly, $S_{\overline{\alpha}} = U \times U - S_{\alpha}$, $S_{\alpha \cap \beta} = S_{\alpha} \cap S_{\beta}$ and $S_{\alpha \cup \beta} = S_{\alpha} \cup S_{\beta}$. As $S_1 = B_{\gamma_1} \cup \cdots \cup B_{\gamma_{2^n}}$, $S_1 = U \times U$ by Lemma 8. If \mathcal{E} is serial, this is a standard model.

By the following lemma, in this model, the interpretations of the actions not equivalent to 0 are unions of some atomic relations.

Lemma 9. For any α of Σ not equivalent to 0, $S_{\alpha} = B_{\gamma_{n_1}} \cup \cdots \cup B_{\gamma_{n_m}}$, where $\gamma_{n_1} \cup \cdots \cup \gamma_{n_m}$ is the CINF of α .

In proving this lemma, we have to use Lemma 3. Now we show a crucial result:

Lemma 10. α occurs in Σ .

- 1. For any $s, t \in U$, if $S_{\alpha}st$, $R_{\alpha}^{f}f(s)f(t)$;
- 2. For any $s \in U$ and $y \in W^f$, if $R^f_{\alpha}f(s)y$, there is a $t \in U$ such that f(t) = y and $S_{\alpha}st$.

Proof. (1) Assume $S_{\alpha}st$. By Definition 4.2, S_{α} is the result of operating on interpretations of atomic actions. Therefore, if α is equivalent to 0, S_{α} is empty. Then α is not equivalent to 0. Let $\gamma_{n_1} \cup \cdots \cup \gamma_{n_m}$ be the CINF of α . By Lemma 9, $S_{\alpha} = B_{\gamma_{n_1}} \cup \cdots \cup B_{\gamma_{n_m}}$. There is an $i \leq m$ such that $B_{\gamma_{n_i}}st$. By the definition of $B_{\gamma_{n_i}}$, $R_{\gamma_{n_i}}^f f(s)f(t)$. By Lemma 7 and 6, $R_{\alpha}^f = R_{\gamma_{n_1}}^f \cup \cdots \cup R_{\gamma_{n_m}}^f$. Then $R_{\alpha}^f f(s)(t)$.

(2) Assume $R_{\alpha}^{f}f(s)y$. By Lemma 6, α is not equivalent to 0. Let $\gamma_{n_{1}} \cup \cdots \cup \gamma_{n_{m}}$ be the CINF of α . Since $R_{\alpha}^{f} = R_{\gamma_{n_{1}}}^{f} \cup \cdots \cup R_{\gamma_{n_{m}}}^{f}$, there is an $i \leq m$ such that $R_{\gamma_{n_{i}}}^{f}f(s)y$. Let $\gamma_{k_{1}}, \ldots, \gamma_{k_{h}}$ be the sequence such that it consists of all the paths γ such that $R_{\gamma}^{f}f(s)y$ and $k_{1} < \cdots < k_{h}$. Then $\gamma_{n_{i}}$ occurs in $\gamma_{k_{1}}, \ldots, \gamma_{k_{h}}$. Let $j \leq h$ be such that $k_{j} = n_{i}$. Suppose j = h. Let $t \in U$ be such that g(t) = 1 and f(t) = y. It can be verified that $B_{\gamma_{k_{j}}}st$, that is, $B_{\gamma_{n_{i}}}st$. Since $S_{\alpha} = B_{\gamma_{n_{1}}} \cup \cdots \cup B_{\gamma_{n_{m}}}$, $S_{\alpha}st$. Suppose j < h. Let $t \in U$ be such that g(t) = j + 1 and f(t) = y. It can also be verified that $B_{\gamma_{k_{j}}}st$, that is, $B_{\gamma_{n_{i}}}st$. Since $S_{\alpha} = B_{\gamma_{n_{1}}} \cup \cdots \cup B_{\gamma_{n_{m}}}$, $S_{\alpha}st$.

Lemma 11 (Existence Lemmas for $\langle \alpha \rangle \varphi$ and $\langle \alpha \rangle \top$).

- 1. For any $\langle \alpha \rangle \varphi \in \Sigma$, if $\langle \alpha \rangle \varphi \ltimes s$, there is a t such that $\varphi \ltimes t$ and $S_{\alpha}st$;
- 2. For any $\langle \alpha \rangle \top$, if $\langle \alpha \rangle \top \ltimes s$, there is a t such that $S_{\alpha}st$.

Lemma 12. *E is serial.*

Proof. Let *H* be a history of \mathfrak{N} . Let $O\alpha_1, \ldots, O\alpha_n$ be all the pure deontic formulas in Σ such that $O\alpha_1, \ldots, O\alpha_n \ltimes \mathring{H}$. Suppose n < 1. Trivially, we have $\mathcal{E}(H, (H, t))$ for any $t \in U$. Suppose $1 \le n$. Then $O\alpha_1, \ldots, O\alpha_n \Join f(\mathring{H})$. Then for any $u \in f(\mathring{H})$, $O\alpha_1, \ldots, O\alpha_n \in u$. By Axiom F1, for any $u \in f(\mathring{H})$, $O(\alpha_1 \cap \cdots \cap \alpha_n) \in u$. By Axiom F4, $\langle \alpha_1 \cap \cdots \cap \alpha_n \rangle \top \in u$ for any $u \in f(\mathring{H})$. Then $\langle \alpha_1 \cap \cdots \cap \alpha_n \rangle \top \Join f(\mathring{H})$. Then $\langle \alpha_1 \cap \cdots \cap \alpha_n \rangle \top \ltimes \mathring{H}$. By Lemma 11, there is a $t \in U$ such that $S_{\alpha_1 \cap \cdots \cap \alpha_n}(\mathring{H}, t)$. Then $S_{\alpha_1}(\mathring{H}, t), \ldots, S_{\alpha_n}(\mathring{H}, t)$. By the definition of \mathcal{E} , we have $\mathcal{E}(H, (H, t))$.

 \mathfrak{N} is a standard model.

4.6 Restricted completeness

Lemma 13 (Existence Lemma for $P\alpha$). If $P\alpha \ltimes \mathring{H}$, there is a $t \in U$ such that $\mathcal{E}(H, (H, t))$ and $S_{\alpha}(\mathring{H}, t)$.

Proof. Suppose $P\alpha \ltimes \mathring{H}$. Let $O\alpha_1, \ldots, O\alpha_n$ be all the pure deontic formulas in Σ such that $O\alpha_1, \ldots, O\alpha_n \ltimes \mathring{H}$. Suppose n < 1. Then trivially, we get $\mathcal{E}(H, (H, t))$ for any $t \in U$. As $P\alpha \ltimes \mathring{H}$, $P\alpha \triangleright f(\mathring{H})$. Then for any $u \in f(\mathring{H})$, $P\alpha \in u$. By Axiom F5, for any $u \in f(\mathring{H}), \langle \alpha \rangle \top \in u$. Then $\langle \alpha \rangle \top \triangleright f(\mathring{H})$. Then $\langle \alpha \rangle \top \ltimes \mathring{H}$. By Lemma 11, there is a $t \in U$ such that $S_{\alpha}(\mathring{H}, (H, t))$.

Suppose $1 \leq n$. Then $O\alpha_1, \ldots, O\alpha_n, P\alpha \geq f(\mathring{H})$. Then for any $u \in f(\mathring{H})$, $O\alpha_1, \ldots, O\alpha_n, P\alpha \in u$. By Axiom F1, for any $u \in f(\mathring{H}), O(\alpha_1 \cap \cdots \cap \alpha_n) \in u$. By Axiom F3, $P(\alpha_1 \cap \cdots \cap \alpha_n \cap \alpha) \in u$ for any $u \in f(\mathring{H})$. By Axiom F5, $\langle \alpha_1 \cap \cdots \cap \alpha_n \cap \alpha \rangle \top \in u$ for any $u \in f(\mathring{H})$. Then $\langle \alpha_1 \cap \cdots \cap \alpha_n \cap \alpha \rangle \top \geq f(\mathring{H})$. Then $\langle \alpha_1 \cap \cdots \cap \alpha_n \cap \alpha \rangle \top \ltimes \mathring{H}$. By Lemma 11, there is a $t \in U$ such that $S_{\alpha_1 \cap \cdots \cap \alpha_n \cap \alpha}(\mathring{H}, t)$. Then $S_{\alpha_1}(\mathring{H}, t), \ldots, S_{\alpha_n}(\mathring{H}, t)$ and $S_{\alpha}(\mathring{H}, t)$. By the definition of $\mathcal{E}, \mathcal{E}(H, (H, t))$.

Lemma 14 (Truth Lemma). For any $\varphi \in \Sigma$, $\varphi \ltimes \mathring{H}$ if and only if $\mathfrak{N}, H \Vdash \varphi$.

Proof. We put an induction on the structure of φ . The cases of $p, \top, \neg \psi$ and $(\psi \land \chi)$ are easy to go through, and we simply skip them.

The case $\varphi = O\alpha$. Suppose $O\alpha \ltimes \mathring{H}$. Let $\mathcal{E}(H, (H, t))$. By the definition of \mathcal{E} , we have $S_{\alpha}(\mathring{H}, t)$. Then $\mathfrak{N}, H \Vdash O\alpha$. Now suppose $\mathfrak{N}, H \Vdash O\alpha$. Then for any $t \in U$, if $\mathcal{E}(H, (H, t)), S_{\alpha}(\mathring{H}, t)$. Assume not $O\alpha \ltimes \mathring{H}$. Then not $O\alpha \rhd f(\mathring{H})$. Since $O\alpha \in \Sigma$, $O\alpha \notin u$ for any $u \in f(\mathring{H})$. Then $\neg O\alpha \in u$ for any $u \in f(\mathring{H})$. By Axiom F2, $P\overline{\alpha} \in u$ for any $u \in f(\mathring{H})$. Then $P\overline{\alpha} \succ f(\mathring{H})$. Then $P\overline{\alpha} \ltimes \mathscr{H}$. By Lemma 13, there is a $t \in U$ such that $\mathcal{E}(H, (H, t))$ and $S_{\overline{\alpha}}(\mathring{H}, t)$. However, since $\mathcal{E}(H, (H, t))$, we have $S_{\alpha}(\mathring{H}, t)$. This is impossible, as $S_{\alpha} \cap S_{\overline{\alpha}} = \emptyset$.

The case $\varphi = \langle \alpha \rangle \psi$. Suppose $\langle \alpha \rangle \psi \ltimes \mathring{H}$. By Lemma 11, there is a $t \in U$ such that $\psi \ltimes t$ and $S_{\alpha}st$. Since Σ is closed under sub-formulas, $\psi \in \Sigma$. By the inductive hypothesis, $\mathfrak{N}, (H, t) \Vdash \psi$. Then $\mathfrak{N}, H \Vdash \langle \alpha \rangle \psi$. Now suppose $\mathfrak{N}, H \Vdash \langle \alpha \rangle \psi$. Then there is a $t \in U$ such that $S_{\alpha}(\mathring{H}, t)$ and $\mathfrak{N}, (H, t) \Vdash \psi$. By the inductive hypothesis, $\psi \ltimes t$. Then $\psi \triangleright f(t)$. Then for any $z \in f(t), \psi \in z$. Since $S_{\alpha}(\mathring{H}, t)$, by Lemma 10, we get $R_{\alpha}^{f}(f(\mathring{H}), f(t))$. Then there are $u \in f(\mathring{H})$ and $v \in f(t)$ such that $R_{\alpha}uv$. Then $\psi \in v$. By the definition of $R_{\alpha}, \langle \alpha \rangle \psi \in u$. Since $\langle \alpha \rangle \psi \in \Sigma, \langle \alpha \rangle \psi \triangleright f(\mathring{H})$. Then $\langle \alpha \rangle \psi \ltimes \mathring{H}$.

Proposition 5. The logic restricted to Φ' is complete with respect to the class of standard models.

4.7 Completeness by translation

Definition 4.3 (Translation). The translation function $t : \Phi \to \Phi'$ is defined as follows:

1.
$$t(p) = p$$

- 2. $t(\top) = \top$
- 3. $t(O\alpha) = O\alpha$
- 4. $t(\neg \varphi) = \neg t(\varphi)$
- 5. $t(\varphi \land \psi) = t(\varphi) \land t(\psi)$
- 6. $t(\langle \alpha \rangle \varphi) = \langle \alpha \rangle t(\varphi)$
- 7. $t([\downarrow \alpha]p) = t(P\alpha \rightarrow p)$
- 8. $t([\downarrow \alpha] \top) = t(\top)$
- 9. $t([\downarrow \alpha] O\beta) = t(P\alpha \rightarrow O(\overline{\alpha} \cup \beta))$
- 10. $t([\downarrow \alpha] \neg \varphi) = t(P\alpha \rightarrow \neg [\downarrow \alpha]\varphi)$
- 11. $t([\downarrow \alpha](\varphi \land \psi)) = t([\downarrow \alpha]\varphi \land [\downarrow \alpha]\psi)$
- 12. $t([\downarrow \alpha]\langle\beta\rangle\varphi) = t(P\alpha \rightarrow \langle\beta\rangle\varphi)$
- 13. $t([\downarrow \alpha][\downarrow \beta]\varphi) = t([\downarrow \alpha]t([\downarrow \beta]\varphi))$
- 14. $t([\downarrow \alpha][\uparrow \beta]\varphi) = t([\downarrow \alpha]t([\uparrow \beta]\varphi))$
- 15. $t([\uparrow \alpha]p) = t(p)$

- 16. $t([\uparrow \alpha] \top) = t(\top)$
- 17. $t([\uparrow \alpha]O\beta) = t(O\beta \land [\alpha \cap \overline{\beta}]\bot)$
- 18. $t([\uparrow \alpha] \neg \varphi) = t(\neg [\uparrow \alpha] \varphi)$
- 19. $t([\uparrow \alpha](\varphi \land \psi)) = t([\uparrow \alpha]\varphi \land [\uparrow \alpha]\psi)$
- 20. $t([\uparrow \alpha]\langle\beta\rangle\varphi) = t(\langle\beta\rangle\varphi)$
- 21. $t([\uparrow \alpha][\uparrow \beta]\varphi) = t([\uparrow \alpha]t([\uparrow \beta]\varphi))$
- 22. $t([\uparrow \alpha][\downarrow \beta]\varphi) = t([\uparrow \alpha]t([\downarrow \beta]\varphi))$

Definition 4.4 (Complexity). The complexity function $c : \Phi \to \mathbb{N}$ is defined as follows:

- 1. $p^c = 1$
- 2. $\top^{c} = 1$
- 3. $(O\alpha)^c = 1$
- 4. $(\neg \varphi)^c = 1 + \varphi^c$
- 5. $(\varphi \wedge \psi)^c = 1 + \max(\varphi^c, \psi^c)$
- 6. $(\langle \alpha \rangle \varphi)^c = 1 + \varphi^c$
- 7. $([\downarrow \alpha]\varphi)^c = 5 \times \varphi^c$
- 8. $([\uparrow \alpha]\varphi)^c = 5 \times \varphi^c$

Lemma 15. *For any* $\varphi \in \Phi$ *, t*(φ) $\in \Phi'$ *.*

This means that *t* can really translate all formulas of Φ into Φ' .

Lemma 16.

- 1. $([\downarrow \alpha]\varphi)^c = ([\downarrow \beta]\varphi)^c;$
- 2. $(t([\downarrow \alpha]\varphi))^c = (t([\downarrow \beta]\varphi))^c$.

Lemma 17. The function c meets the following conditions:

1. If φ is a proper sub-formula of ψ , then $\varphi^c < \psi^c$.

- 2. $([\downarrow \alpha]p)^c > (P\alpha \rightarrow p)^c$
- 3. $([\downarrow \alpha] \top)^c > (\top)^c$
- 4. $([\downarrow \alpha] O\beta)^c > (P\alpha \to O(\overline{\alpha} \cup \beta))^c$
- 5. $([\downarrow \alpha] \neg \varphi)^c > (P\alpha \rightarrow \neg [\downarrow \alpha] \varphi)^c$
- 6. $([\downarrow \alpha](\varphi \land \psi))^c > ([\downarrow \alpha]\varphi \land [\downarrow \alpha]\psi)^c$
- 7. $([\downarrow \alpha] \langle \beta \rangle \varphi)^c > (P\alpha \to \langle \beta \rangle \varphi)^c$
- 8. $([\uparrow \alpha]p)^c > (p)^c$
- 9. $([\uparrow \alpha] \top)^c > (\top)^c$
- 10. $([\uparrow \alpha] O\beta)^c > (O\beta \land [\alpha \cap \overline{\beta}] \bot)^c$
- 11. $([\uparrow \alpha] \neg \varphi)^c > (\neg [\uparrow \alpha] \varphi)^c$
- 12. $([\uparrow \alpha](\varphi \land \psi))^c > ([\uparrow \alpha]\varphi \land [\uparrow \alpha]\psi)^c$
- 13. $([\uparrow \alpha]\langle\beta\rangle\varphi)^c > (\langle\beta\rangle\varphi)^c$
- 14. $([\downarrow \alpha]\varphi)^c > (t([\downarrow \alpha]\varphi))^c$
- 15. $([\uparrow \alpha]\varphi)^c > (t([\uparrow \alpha]\varphi))^c$

Proof. We only show the item (14). The proof for item (15) is similar and other items can be easily proved.

We put an induction on φ^c . Suppose $\varphi^c = 1$. Then $\varphi = p$, $\varphi = \top$ or $\varphi = O\alpha$. By items (2), (3) and (4), we can easily see that $([\alpha]\varphi)^c > (t([\alpha]\varphi))^c$ holds in all these cases. Suppose for any ψ , if $\psi^c < \varphi^c$, then $([\alpha]\psi)^c > (t([\alpha]\psi))^c$. Here we only show that $([\alpha]\varphi)^c > (t([\alpha]\varphi))^c$ if $\varphi = [\downarrow \beta]\psi$ for some ψ .

Since $\psi^c < ([\downarrow \beta]\psi)^c$, we get $([\downarrow \alpha]\psi)^c > (t([\downarrow \alpha]\psi))^c$ by the inductive hypothesis. By Lemma 16, $([\downarrow \beta]\psi)^c > (t([\downarrow \beta]\psi))^c$. Since $([\downarrow \alpha][\downarrow \beta]\psi)^c = 5 \times ([\downarrow \beta]\psi)^c$ and $([\downarrow \alpha]t([\downarrow \beta]\psi))^c = 5 \times (t([\downarrow \beta]\psi))^c$, we know that $([\downarrow \alpha][\downarrow \beta]\psi)^c > ([\downarrow \alpha]t([\downarrow \beta]\psi))^c$. Since $(t([\downarrow \beta]\psi))^c < ([\downarrow \beta]\psi)^c$, we get $([\downarrow \alpha]t([\downarrow \beta]\psi))^c > (t([\downarrow \alpha]t([\downarrow \beta]\psi)))^c$ by the inductive hypothesis again. Since $t([\downarrow \alpha][\downarrow \beta]\psi) = t([\downarrow \alpha]t([\downarrow \beta]\psi))$, we have $([\downarrow \alpha][\downarrow \beta]\psi)^c > (t([\downarrow \alpha][\downarrow \beta]\psi))^c$.

Lemma 18. For any $\varphi \in \Phi$, $\vdash \varphi \leftrightarrow t(\varphi)$.
Proof. We put an induction on φ^c . Suppose $\varphi^c = 1$. Then $\varphi = p$, $\varphi = \top$ or $\varphi = O\alpha$. In these cases, $t(\varphi) = \varphi$, and then $\vdash \varphi \leftrightarrow t(\varphi)$. Suppose for any ψ , if $\psi^c < \varphi^c$, then $\vdash \psi \leftrightarrow t(\psi)$. We want to show that $\vdash \varphi \leftrightarrow t(\varphi)$. It suffices to show that $\vdash \varphi \leftrightarrow t(\varphi)$ in all of the cases that φ is one of the following formulas: $\neg \psi, \psi \land \chi, \langle \alpha \rangle \psi, [\downarrow \alpha]p, [\downarrow \alpha]\top$, $[\downarrow \alpha]O\alpha, [\downarrow \alpha]\neg\psi, [\downarrow \alpha](\psi \land \chi), [\downarrow \alpha]\langle\beta\rangle\psi, [\downarrow \alpha][\downarrow \beta]\psi, [\downarrow \alpha][\uparrow \beta]\psi, [\uparrow \alpha]p, [\uparrow \alpha]\top$, $[\uparrow \alpha]O\alpha, [\uparrow \alpha]\neg\psi, [\uparrow \alpha](\psi \land \chi), [\uparrow \alpha]\langle\beta\rangle\psi, [\uparrow \alpha][\downarrow \beta]\psi$, and $[\uparrow \alpha][\uparrow \beta]\psi$. This can be proved by use of Lemma 16 and 17. Here we only go through the case $\varphi = [\downarrow \alpha][\downarrow \beta]\psi$.

Since $[\downarrow \beta]\psi$ is a proper sub-formula of $[\downarrow \alpha][\downarrow \beta]\psi$, by Lemma 17, we get $([\downarrow \beta]\psi)^c < ([\downarrow \alpha][\downarrow \beta]\psi)^c$. By the inductive hypothesis, $\vdash [\downarrow \beta]\psi \leftrightarrow t([\downarrow \beta]\psi)$. By the inference rule Replacement of Validity, $\vdash [\downarrow \alpha][\downarrow \beta]\psi \leftrightarrow [\downarrow \alpha]t([\downarrow \beta]\psi)$. By Lemma 17, $(t([\downarrow \beta]\psi))^c < ([\downarrow \beta]\psi)^c$. Then $([\downarrow \alpha]t([\downarrow \beta]\psi))^c < ([\downarrow \alpha]t([\downarrow \beta]\psi)^c$. By the inductive hypothesis again, $\vdash [\downarrow \alpha]t([\downarrow \beta]\psi) \leftrightarrow t([\downarrow \alpha]t([\downarrow \beta]\psi))$. Since $t([\downarrow \alpha]t([\downarrow \beta]\psi)) = t([\downarrow \alpha][\downarrow \beta]\psi)$.

Proposition 6 (Completeness). *The logic is complete with respect to the class of standard models.*

Proof. Suppose $\models \varphi$. By Lemma 18, $\vdash \varphi \leftrightarrow t(\varphi)$. The logic is sound, then $\models \varphi \leftrightarrow t(\varphi)$. Then $\models t(\varphi)$. By Lemma 15, $t(\varphi) \in \Phi'$. By Proposition 5, $\vdash t(\varphi)$. By Lemma 18, $\vdash \varphi \leftrightarrow t(\varphi)$. Then $\vdash \varphi$.

5 Future work

Our semantics does not work for descriptive permissions. The descriptive utterances of sentences such as "*you may do* α *or* β " might only inform the agent that he is allowed to do something, but not specify it. After such utterances, the agent might still not know how to act. This sort of utterances raises uncertainties, but our semantics does not have any settings to handle them.

Next on our agenda are two questions. The language in this work contains three action operators: complement, intersection and choice. It is natural to add test and composition to obtain a more powerful language in which conditional and sequential obligations and permissions can be expressed. This is an issue we want to pursue in the future. According to our semantics, the last update always overrides the previous ones. For example, given that α is possible to perform, the prescriptive permission "you may do α " always gives the agent the freedom to do α , regardless of what obligations have been put on him. This is the case only if there is only one speaker or moral source. In real life, there are many moral sources whose authorities are ranked, and only orders and permissions from speakers with higher authorities can overwhelm those from

speakers with lower authorities. Introducing prioritized speakers into this framework is another direction of future work for us.

Notes This paper is an extension of (Ju and Liang 2013), where we only axiomatized the logic restricted to Φ' , which contains no dynamic operators, and did not present an axiomatization for the whole logic.

Acknowledgements This research is supported by the National Social Science Foundation of China (No. 12CZX053) and the Major Bidding Project of the National Social Science Foundation of China (No. 10&ZD073). We would like to thank Maria Aloni, Davide Grossi, Fenrong Liu, Hans van Ditmarsch, Frank Veltman, Yanjing Wang, Yi Wang, Tomoyuki Yamada, and the three anonymous reviewers of the 4th International Workshop of Logic, Rationality and Interaction, for their helpful comments and suggestions.

References

H. van Ditmarsch, W. van Der Hoek, and B. Kooi. *Dynamic Epistemic Logic*. Springer, Heidelberg, 2007.

G. Gargov and S. Passy. A note on boolean modal logic. In P. P. Petkov, editor, *Mathematical Logic*, pages 311–321. Plenum Press, 1990.

R. Hilpinen. Deontic logic. In L. Goble, editor, *The Blackwell Guide to Philosophical Logic*, pages 159–182. Blackwell Publishing, 2001.

J. Jorgensen. Imperatives and logic. Erkenntnis, 7(1):288–296, 1937.

F. Ju and L. Liang. A dynamic deontic logic based on histories. In D. Grossi, O. Roy, and H. Huang, editors, *Logic, Rationality, and Interaction*, volume 8196 of *Lecture Notes in Computer Science*, pages 176–189. Springer Berlin Heidelberg, 2013.

F. Ju and F. Liu. Prioritized imperatives and normative conflicts. *European Journal of Analytic Philosophy*, 7(2):33–54, 2011.

J.-J. C. Meyer. 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, 1988.

A. Ross. Imperatives and logic. *Philosophy of Science*, 11(1):30-46, 1944.

F. Veltman. Imperatives at the borderline of semantics and pragmatics. Manuscript, 2009.

Public Announcements under Sheaves

Kohei Kishida

University of Oxford kishidakohei@gmail.com

The goal of this article is to bring together the frameworks of model-update semantics for (propositional) public-announcement logic (van Ditmarsch et al. 2008) and of sheaf semantics for first-order modal logic (Awodey and Kishida 2008, Gabbay et al. 2009, Kishida 2011), and to thereby obtain a sheaf semantics for first-order publicannouncement logic. The first attempt to extend dynamic epistemic logics to the first order was made by Kooi (2007), who introduced terms to refer to epistemic agents, and an extension of public-announcement logic to the first order was briefly given by Ma (2011);¹ both of these extensions used constant domains for interpreting first-order vocabulary. (A first-order extension of dynamic logic was given in (Harel 1979) and (Harel et al. 2000), also with constant domains.) This article pushes ahead with these extensions by employing a sheaf structure, providing a progress toward a more flexible and useful treatment of first-order notions.

We will first review model-update semantics for propositional publicannouncement logic in Section 1, and sheaf semantics for first-order modal logic in Section 2. Then, in Section 3, we will show how to combine the ideas of these two semantics into sheaf semantics for first-order public-announcement logic. We will conclude the article with a brief discussion of future work and projects in Section 4.

¹Theorem 14 of Ma (2011), the chief theorem in the section on first-order public-announcement logic, is unfortunately incorrect, with an invalid reduction axiom for quantifiers. Hence, even in the constant-domain setting, a correct axiomatization of first-order public-announcement logic has not been given.

1 Reviewing the propositional case

We first review the basics of **PAL**, public-announcement logic, in the propositional case (see van Ditmarsch et al. 2008 or van Benthem 2011 for a detailed exposition of motivations, definitions and results). One characteristic of the approach of this article is to use (monotone) neighborhood semantics, rather than Kripke semantics, as a static basis of **PAL**, although all the results that will follow apply equally to Kripke semantics (which is just a particular subclass of monotone neighborhood semantics). This feature is not standard, but not new to this article either: Neighborhood semantics for propositional **PAL** was first introduced by Demey (2010) and Zvesper (2010).²

A language \mathcal{L} of **PAL** is a propositional language with two sorts of unary modal operators: K_i , for each $i \in I$ for a fixed set I of "agents"; and $[!\sigma]$, for each sentence σ of \mathcal{L} . Given a sentence φ , sentences $K_i\varphi$ is supposed to mean, approximately, that the agent i "knows" that φ , and $[!\sigma]\varphi$ is supposed to mean that φ will be the case after it is "publicly (and truthfully) announced" or "publicly observed" that σ .³ Let us say \mathcal{L} is a *static language* if it contains K_i but not $[!\sigma]$, and a **PAL** *language* if it contains both.

We use (monotone) neighborhood semantics to interpret operators K_i .

Definition 1.1. A monotone neighborhood frame, or *MN*-frame for short, is a tuple $(W, \{\mathcal{B}^i\}_{i \in I})$ of any set $W \neq \emptyset$ of "worlds" and any map $\mathcal{B}^i : W \to \mathcal{PP}W$ for every $i \in I$.⁴ The members $B \subseteq W$ of $\mathcal{B}^i(w)$ are called the *basic i-neighborhoods* of w, and \mathcal{B}^i is called an *i-basic-neighborhood map*.

Moreover, each \mathcal{B}^i induces an operation $\operatorname{int}^i : \mathcal{P}W \to \mathcal{P}W$, called an *i-interior* operation, with

$$w \in \operatorname{int}^{l}(A) \iff B \subseteq A \text{ for some } B \in \mathcal{B}^{l}(w).$$
 (int)

$$w \in int(A) \iff A \in \mathcal{N}(w).$$
 (int')

Clearly, MN-frames in this sense are MN-frames in our sense ((int') is equivalent to (int) when $\mathcal{B} = \mathcal{N}$ is upward closed).

²Both Demey (2010) and Zvesper (2010) discuss the fully general neighborhood semantics, whereas in this article we only use the monotone case. Both Demey's and Zvesper's goal behind their use of neighborhood—as opposed to Kripke—semantics is to falsify principles such as $K_i(\varphi \rightarrow \psi) \rightarrow (K_i\varphi \rightarrow K_i\psi)$ and to avoid the so-called "problem of logical omniscience"; our goal is different, as we lay out at the end of this section.

³See Chs. 2 and 3 of (van Benthem 2011) for discussions of precisely what $K_i\varphi$ and $[!\sigma]\varphi$ mean given the semantics of **PAL**.

⁴Our formulation of monotone neighborhood semantics may appear different from, but is in fact equivalent to, the more common formulation (e.g., in Chellas 1980). Commonly, an MN-frame (W, N) is assumed to have $N : W \to \mathcal{PPW}$ "upward closed" (i.e., to have $A \in \mathcal{N}(w)$ whenever $B \in \mathcal{N}(w)$ and $B \subseteq A \subseteq W$), while \mathcal{N} induces **int** with

It is obvious that a Kripke frame (or model) is just an MN-frame (or model) with each $\mathcal{B}^{i}(w)$ a singleton, so that $B \in \mathcal{B}^{i}(w)$ is just the set of worlds that are *i*-accessible from *w*.

We interpret each sentence φ with a subset $\llbracket \varphi \rrbracket$ of W as the so-called "truth set" of φ , so that $w \in \llbracket \varphi \rrbracket$ means that φ is true at the world w; in this sense subsets of W can be called "propositions". So, each $w \in W$ has a family (perhaps empty) of propositions as its basic *i*-neighborhoods, and w is in the *i*-interior of a proposition $A \subseteq W$ iff some of its basic *i*-neighborhoods "entails" A. Then we interpret the classical connectives classically, and the modal operators K_i with the *i*-interior operations.

Definition 1.2. Given a propositional static or **PAL** language \mathcal{L} , a *monotone neighborhood model*, or *MN-model*, for \mathcal{L} consists of an MN-frame $(W, \{\mathcal{B}^i\}_{i \in I})$ and any map $\llbracket \cdot \rrbracket$ that assigns a set $\llbracket p \rrbracket \subseteq W$ to each atomic sentence p of \mathcal{L} . A *monotone neighborhood interpretation*, or *MN-interpretation*, over an MN-model $(W, \{\mathcal{B}^i\}_{i \in I}, \llbracket \cdot \rrbracket)$ for \mathcal{L} is a map that extends $\llbracket \cdot \rrbracket$ to all the sentences φ of \mathcal{L} following the constraints

| | $\llbracket \neg \rrbracket = W \setminus \cdot,$ | (\neg_{Prop}) |
|---------|--|-------------------------------|
| so that | $\llbracket \neg \varphi \rrbracket = \llbracket \neg \rrbracket \llbracket \varphi \rrbracket = W \setminus \llbracket \varphi \rrbracket,$ | |
| | $[\![\wedge]\!]=\cap,$ | (\wedge_{Prop}) |
| so that | $\llbracket \varphi \land \psi \rrbracket = \llbracket \varphi \rrbracket \llbracket \land \rrbracket \llbracket \psi \rrbracket = \llbracket \varphi \rrbracket \cap \llbracket \psi \rrbracket,$ | |
| | $\llbracket \rightarrow \rrbracket = (W \setminus \cdot) \cup \cdot,$ | $(\rightarrow_{\text{Prop}})$ |
| so that | $\llbracket \varphi \to \psi \rrbracket = \llbracket \varphi \rrbracket \llbracket \to \rrbracket \llbracket \psi \rrbracket = (W \setminus \llbracket \varphi \rrbracket) \cup \llbracket \psi \rrbracket,$ | |
| | $\llbracket K_i \rrbracket = \mathbf{int}^i,$ | (\square_{Prop}) |
| so that | $w \in \llbracket K_i \varphi \rrbracket = \llbracket K_i \rrbracket \llbracket \varphi \rrbracket \iff B \subseteq \llbracket \varphi \rrbracket$ for some $B \in \mathcal{B}^i(w)$. | |

The semantics of **PAL** involves the updating of models by restricting them to submodels.

Definition 1.3. Given an MN-model $M = (W, \{\mathcal{B}^i\}_{i \in I}, \llbracket \cdot \rrbracket)$ for \mathcal{L} and nonempty $S \subseteq W$, let $M_S = (S, \{\mathcal{B}^i_S\}_{i \in I}, \llbracket \cdot \rrbracket_S)$ be the MN-model for \mathcal{L} with

$$A \in \mathcal{N}_{\mathcal{B}}(w) \iff B \subseteq A \text{ for some } B \in \mathcal{B}(w),$$

On the other hand, each MN-frame (W, \mathcal{B}) in our sense "generates" an MN-frame $(W, \mathcal{N}_{\mathcal{B}})$ as commonly defined by setting

and then $\mathcal{N}_{\mathcal{B}}(w)$ induces with (**int**') the same **int** : $\mathcal{P}W \to \mathcal{P}W$ as \mathcal{B} does with (**int**), which means that (W, \mathcal{B}) and $(W, \mathcal{N}_{\mathcal{B}})$ have the same [[K]]. In short, the two formulations give the same semantics of *K*. An advantage of ours is that, in modelling, it enables us to distinguish basic neighborhoods $B \in \mathcal{B}(w)$ from other neighborhoods $A \in \mathcal{N}_{\mathcal{B}}(w)$, a distinction that corresponds to the one between direct observability and verifiability, in terms of the observability interpretation mentioned at the end of this section.

- $\mathcal{B}_{S}^{i}(w) = \{B \cap S \mid B \in \mathcal{B}^{i}(w)\}, 5$ and
- $\llbracket p \rrbracket_S = \llbracket p \rrbracket \cap S$ for every atomic sentence p of \mathcal{L} .

This operation of restriction update provides a semantics for $[!\sigma]$:

Definition 1.4. Given a propositional **PAL** language \mathcal{L} , an MN-interpretation over an MN-model $M = (W, \{\mathcal{B}^i\}_{i \in I}, \llbracket \cdot \rrbracket)$ for \mathcal{L} is called a *monotone neighborhood PALinterpretation*, or *MN-PAL-interpretation*, over *M* if it satisfies

$$[[!\sigma]\varphi]] = [[\sigma]][[\rightarrow]][[\varphi]]_{[\sigma]}, \qquad ([!\cdot]_{\text{Prop}})$$
so that $w \in [[!\sigma]\varphi]] \iff \text{if } w \in [[\sigma]] \text{ then } w \in [[\varphi]]_{[\sigma]},$

where $\llbracket \cdot \rrbracket_{\llbracket \sigma \rrbracket}$ is an MN-PAL-interpretation over $(\llbracket \sigma \rrbracket, \{\mathcal{B}^{i}_{\llbracket \sigma \rrbracket}\}_{i \in I}, \llbracket \cdot \rrbracket_{\llbracket \sigma \rrbracket})$ if $\llbracket \sigma \rrbracket \neq \emptyset$; we write $\llbracket \varphi \rrbracket_{\llbracket \sigma \rrbracket} = \emptyset$ if $\llbracket \sigma \rrbracket = \emptyset$.

Under the condition $([!\cdot]_{Prop})$, $[!\sigma]\varphi$ intuitively means that, if σ is susceptible of announcing truthfully (in the sense of simply being true), then the public announcement that σ will make it the case that φ . It is worth noting that, when we write $\langle !\sigma \rangle \varphi$ for $\neg [!\sigma] \neg \varphi$, $([!\cdot]_{Prop})$ implies

$$\llbracket \langle !\sigma \rangle \varphi \rrbracket = \llbracket \varphi \rrbracket_{\llbracket \sigma \rrbracket}, \qquad (\langle ! \cdot \rangle_{\text{Prop}})$$

because

$$\llbracket \langle !\sigma \rangle \varphi \rrbracket = \llbracket \neg \llbracket !\sigma \rrbracket \neg \varphi \rrbracket = W \setminus (\llbracket \sigma \rrbracket \llbracket \neg \varphi \rrbracket \llbracket \neg \varphi \rrbracket \llbracket \sigma \rrbracket) = \llbracket \sigma \rrbracket \setminus \llbracket \neg \varphi \rrbracket_{\llbracket \sigma \rrbracket} = \llbracket \varphi \rrbracket_{\llbracket \sigma \rrbracket}.$$

The class of MN-PAL-interpretations forms monotone neighborhood semantics for **PAL**. The logic of this semantics, **M-PAL**, has all the axiom or rule schemes of classical logic, as well as the rule scheme

$$\frac{\varphi \to \psi}{K_i \varphi \to K_i \psi}.$$
 M

⁵Zvesper (2010) uses the same definition of restriction to a submodel as ours for **PAL**. Demey (2010), on the other hand, uses a seemingly different definition; it agrees with ours in the monotone case, but disagrees in the non-monotone case. (Related but crucially different notions of submodel in the neighborhood setting are discussed in (Hansen 2003).) Our definition of submodel is a straightforward generalization of a subspace in topology: The family $\{\mathcal{B}_{S}^{i}\}_{i\in I}$ makes the inclusion map $m : S \to W :: w \mapsto w$ "*I*-continuous", in the sense that $m^{-1}[\operatorname{int}_{W}^{i}(A)] \subseteq \operatorname{int}_{S}^{i}(m^{-1}[A])$ for every $A \subseteq W$ and for every $i \in I$. Moreover, $\{\mathcal{B}_{S}^{i}\}_{i\in I}$ is the "coarsest" family from which *m* is *I*-continuous; or, equivalently, any map $f : X \to W$ with $f[X] \subseteq S$ that is *I*-continuous to $(W, \{\mathcal{B}_{i}^{i}\}_{i\in I})$ is *I*-continuous to $(S, \{\mathcal{B}_{S}^{i}\}_{i\in I})$ (when regarded as a map $f : X \to S$). In category-theoretic terms, *m* is a "regular monomorphism" in the category of MN-frames and *I*-continuous maps.

To add more axioms and rules, correspondence results are available: For instance, let us say that an MN-frame $(W, \{\mathcal{B}^i\}_{i \in I})$ is an *MCN-frame* if $B_0, B_1 \in \mathcal{B}^i(w)$ implies $B_2 \subseteq B_0 \cap B_1$ for some $B_2 \in \mathcal{B}^i(w)$; then the logic of the MN-PAL-interpretations over MCN-frames, **MC-PAL**, is given by adding to **M-PAL** the axiom scheme

$$(K_i \varphi \wedge K_i \psi) \to K_i (\varphi \wedge \psi).$$
 C

Moreover, monotone neighborhood semantics for **PAL** validates the same set of axioms of reduction as Kripke semantics for **PAL** does:

| $[!\sigma]p \equiv (\sigma \to p)$ f | for atomic <i>p</i> ; | R _{atom} |
|--------------------------------------|-----------------------|--------------------------|
|--------------------------------------|-----------------------|--------------------------|

 $[!\sigma]\neg\varphi \equiv (\sigma \to \neg [!\sigma]\varphi); \qquad \qquad \mathbf{R}_{\neg}$

$$[!\sigma](\varphi \land \psi) \equiv ([!\sigma]\varphi \land [!\sigma]\psi); \qquad \mathbf{R}_{\land}$$

$$[!\sigma]K_i\varphi \equiv (\sigma \to K_i[!\sigma]\varphi); \qquad \mathbf{R}_{\Box}$$

$$[!\sigma][!\tau]\varphi \equiv [!(\sigma \land [!\sigma]\tau)]\varphi. \qquad \mathsf{R}_{\mathsf{it}}$$

It is worth showing how monotone neighborhood semantics for **PAL**, with a different truth condition for K_i than Kripke's, validates R_{\Box} . First observe that $[\![K_i\varphi]\!]_{[\![\sigma]\!]} = [\![\sigma]\!] \cap [\![K_i[!\sigma]]\varphi]\!]$, since

$$w \in \llbracket K_i \varphi \rrbracket_{\llbracket \sigma \rrbracket} = \operatorname{int}_{\llbracket \sigma \rrbracket}^i \llbracket \varphi \rrbracket_{\llbracket \sigma \rrbracket}$$
$$\iff w \in \llbracket \sigma \rrbracket \text{ and } B \subseteq \llbracket \varphi \rrbracket_{\llbracket \sigma \rrbracket} \text{ for some } B \in \mathcal{B}_{\llbracket \sigma \rrbracket}^i(w)$$
$$\iff w \in \llbracket \sigma \rrbracket \text{ and } B \cap \llbracket \sigma \rrbracket \subseteq \llbracket \varphi \rrbracket_{\llbracket \sigma \rrbracket} \text{ for some } B \in \mathcal{B}^i(w)$$
$$\iff w \in \llbracket \sigma \rrbracket \text{ and } B \subseteq \llbracket \sigma \rrbracket \llbracket \varphi \rrbracket_{\llbracket \sigma \rrbracket} = \llbracket [!\sigma] \varphi \rrbracket \text{ for some } B \in \mathcal{B}^i(w)$$
$$\iff w \in \llbracket \sigma \rrbracket \text{ and } B \subseteq \llbracket \sigma \rrbracket \llbracket \varphi \rrbracket_{\llbracket \sigma \rrbracket} = \llbracket [!\sigma] \varphi \rrbracket \text{ for some } B \in \mathcal{B}^i(w)$$
$$\iff w \in \llbracket \sigma \rrbracket \text{ and } w \in \operatorname{int}^i \llbracket [!\sigma] \varphi \rrbracket = \llbracket K_i [!\sigma] \varphi \rrbracket$$

for each $w \in W$. Therefore

$$\llbracket \llbracket \sigma \rrbracket K_i \varphi \rrbracket = \llbracket \sigma \rrbracket \llbracket \sigma \rrbracket \llbracket K_i \varphi \rrbracket_{\llbracket \sigma \rrbracket} = \llbracket \sigma \rrbracket \llbracket \sigma \rrbracket \llbracket \sigma \rrbracket \llbracket K_i \llbracket \sigma] \varphi \rrbracket = \llbracket \sigma \to K_i \llbracket \sigma] \varphi \rrbracket.$$

 $R_{atom}-R_{it}$ reduce any sentence with $[!\sigma]$ to one without, by pushing $[!\sigma]$ of $[!\sigma]\varphi$ inward when φ is compound and eventually eliminating $[!\sigma]$ when φ is atomic. Hence $R_{atom}-R_{it}$ completely axiomatize **PAL** by reducing it to its "announcement-free", static fragment, which is just the modal logic given by M (along with any other axioms or rules on K_i that one chooses to add).

Theorem 1. Given a class \mathbb{M} of MN-models, let \mathbb{T} be the logic of \mathbb{M} in a propositional static language \mathcal{L} . Then the logic of \mathbb{M} in the **PAL** language extending \mathcal{L} is obtained by adding R_{atom} - R_{it} to \mathbb{T} .

It needs explaining why we should use neighborhood semantics, instead of the standard Kripke approach, to interpret **PAL**. One merit is that neighborhood semantics facilitates what may be called a *verifiability interpretation* of K_i operators. Think of $\mathcal{B}^i(w)$ as the family of propositions that the agent *i* can directly observe at the world *w*; then (\Box_{Prop}) means that $K_i\varphi$ is true at *w* iff some proposition that *i* can observe entails φ , that is, iff *i* can verify that φ .⁶ For instance, let us represent infinite sequences of coin tosses with maps $w : \mathbb{N} \to 2$ in $W = 2^{\mathbb{N}}$, so that w(n) = 1 means that the (n + 1)st toss comes up heads in the world $w \in W$; then what we can directly observe is the outcome of each toss or finite combinations thereof, so that $B \in \mathcal{B}^i(w)$ iff $B = \{v \in 2^{\mathbb{N}} \mid v(n) = w(n) \text{ for all } n \in J\}$ for some finite $J \subseteq \mathbb{N}$.⁷ Then, for instance, even though it is not directly observable that some toss comes up heads, it is verifiable (if true). We should note that such interpretation is precluded in Kripke semantics, where each $\mathcal{B}^i(w)$ is a singleton or has the smallest element, which means that there is a single observation that verifies everything verifiable.

The combination of the verifiability interpretation with **PAL** is suitable for the purpose of expressing epistemic inquiries. In scientific or any kind of inquiries, increase in knowledge (which **PAL** expresses by the restriction update of models) can turn the unobservable observable; for instance, scientists used to only observe a segment of light in a certain glass case, where they now observe a trail of a charged particle in a cloud chamber. The restriction update of **PAL** captures such a phenomenon precisely, by enabling sentences φ with $[\![\varphi]\!] \notin \mathcal{B}^i(w)$ to nonetheless have $[\![\varphi]\!]_S \in \mathcal{B}^i_S(w)$.

2 Reviewing "Neighborhood-Sheaf" semantics for First-Order Modal Logic

Next we review sheaf semantics for first-order modal logic, and in particular, "neighborhood-sheaf" semantics (Kishida 2011) (which subsumes Kripke-sheaf semantics (Gabbay et al. 2009) and topological-sheaf semantics (Awodey and Kishida 2008), as well as neighborhood semantics with constant domains (Arló-Costa and Pacuit 2006)); we will extend it to public announcements in the next section. A major advantage of using sheaf semantics as opposed to constant-domain semantics (either Kripke semantics or neighborhood semantics (Arló-Costa and Pacuit 2006)) is that the former facilitates more flexible treatment of identity of first-order individuals; this is particularly important in applying first-order public announcement logic to epistemic

⁶Topologic (Dabrowski et al. 1996) and evidence logic (van Benthem and Pacuit 2011) are based on similar interpretations of K_i and of topology or neighborhoods.

⁷The operation $\operatorname{int}^{i} : \mathcal{P}(2^{\mathbb{N}}) \to \mathcal{P}(2^{\mathbb{N}})$ which this \mathcal{B}^{i} induces with (int) is exactly the interior operation of the Cantor space.

Kishida

contexts where identity of objects are subject of epistemic inquiries. We refer the reader to (Awodey and Kishida 2008) and (Kishida 2011) for more explanation of how the semantics works both conceptually and technically.

Given MN-frames (X, \mathcal{B}_X) and (Y, \mathcal{B}_Y) , we say that a map $f : X \to Y$ is *open* if $f^{-1}[\operatorname{int}_Y(A)] = \operatorname{int}_X(f^{-1}[A])$. Then, by a *local isomorphism*, or *neighborhood sheaf*, we mean an open map $\pi : D \to W$ between *MCN-frames* (and not just MN-frames) such that, for every $a \in D$ with $\mathcal{B}_D(a) \neq \emptyset$, there is $B \in \mathcal{B}_D(a)$ with $\pi \upharpoonright B : B \to W$ injective (or, equivalently, $\pi \upharpoonright B : B \to \pi[B]$ bijective).⁸ It is worth noting that the Kripke sheaves (Gabbay et al. 2009) are just the neighborhood sheaves with $\mathcal{B}_W(w)$ a singleton for each $w \in W$, and that the topological sheaves (Awodey and Kishida 2008) are just the neighborhood sheaves over a topological space (W, \mathcal{B}_W) .

Fixing any MCN-frame *W*, write **LI**/*W* for the category of MCN-frames and local isomorphisms sliced over *W*. This category provides structures for interpreting first-order logic, when we regard *W* as a set of worlds. Indeed, whereas in interpreting propositional modal logic we regard worlds as having no internal structure, in interpreting first-order modal logic we regard worlds as models of first-order logic. So, given any (surjective) local isomorphism $\pi : D \to W$, we regard the inverse image $D_w = \pi^{-1}[\{w\}]$ of each $w \in W$ as a domain of individuals of the model *w*; then $D = \sum_{w \in W} D_w$, the disjoint union of all D_w , is the domain of individuals that are from some model or other, and, for any individual $a \in D$, $\pi(a)$ is the (unique) model it is from.⁹ The category **LI**/*W* provides with each $w \in W$ more structures for first-order logic. For instance, the *n*-fold product of π in **LI**/*W* comes with the set

$$D^n = \sum_{w \in W} (D_w \times \cdots \times D_w) = \{ (a_1, \dots, a_n) \mid \pi(a_1) = \cdots = \pi(a_n) \}$$

of *n*-tuples from the same model, and with the map

$$\pi^n: D^n \to W :: (a_1, \ldots, a_n) \mapsto \pi(a_1) = \cdots = \pi(a_n);$$

in particular, $D^0 = W$ and π^0 is the identity on W. (We will later discuss what basicneighborhood maps the *n*-fold product of $\pi : D \to W$ has on D^n .)

Accordingly, we interpret the classical part of first-order modal logic within each world or model $w \in W$. Let \mathcal{L} be a first-order (modal) language (that may have the relation symbol =, any function symbols, etc.). Then, extending the interpretation $\llbracket \sigma \rrbracket \subseteq W = D^0$ of sentences σ , we interpret a formula φ of \mathcal{L} with no free variables other than $\bar{x} = (x_1, \ldots, x_n)$ with the set $\llbracket \bar{x} \mid \varphi \rrbracket \subseteq D^n$ of *n*-tuples $\bar{a} = (a_1, \ldots, a_n) \in D^n$ of which φ is true (with a_i in place of x_i) in $\pi^n(\bar{a})$. So, for instance, $\bar{a} \in \llbracket \bar{x} \mid \neg \varphi \rrbracket$ iff

⁸See footnote 11 for why we require MCN-frames and not just MN-frames.

⁹Thus $\pi : D \to W$ gives essentially the same ontological picture as David Lewis's counterpart theory (Lewis 1968), though our interpretation of \Box differs significantly from Lewis's.

 $\bar{a} \notin [[\bar{x} | \varphi]]$, because $\neg \varphi$ is true of \bar{a} in $\pi^n(\bar{a})$ iff φ is not true of \bar{a} in $\pi^n(\bar{a})$. Thus $[[\cdot]]$ must satisfy, for each n,

 $\llbracket \bar{x} \mid \neg \varphi \rrbracket = D^n \setminus \llbracket \bar{x} \mid \varphi \rrbracket,$

$$[\neg]] = D^n \setminus \cdot, \qquad (\neg_{\rm FO})$$

so that

 $\llbracket \land \rrbracket = \cap, \qquad (\land_{FO})$ so that $\llbracket \bar{x} \mid \varphi \land \psi \rrbracket = \llbracket \bar{x} \mid \varphi \rrbracket \cap \llbracket \bar{x} \mid \psi \rrbracket, \qquad (\rightarrow_{FO})$ so that $\llbracket \bar{x} \mid \varphi \rightarrow \psi \rrbracket = (D^n \setminus \cdot) \cup \cdot, \qquad (\rightarrow_{FO})$ so that $\llbracket \bar{x} \mid \varphi \rightarrow \psi \rrbracket = (D^n \setminus \llbracket \bar{x} \mid \varphi \rrbracket) \cup \llbracket \bar{x} \mid \psi \rrbracket,$

straightforwardly extending $(\neg_{Prop}) - (\rightarrow_{Prop})$, and moreover

$$\bar{a} \in \llbracket \bar{x} \mid \exists y \varphi \rrbracket \iff (\bar{a}, b) \in \llbracket \bar{x}, y \mid \varphi \rrbracket, \tag{3}$$

$$(\bar{a}, b) \in \llbracket \bar{x}, y | \varphi \rrbracket \iff \bar{a} \in \llbracket \bar{x} | \varphi \rrbracket$$
 when y is not free in φ . (vac)

From the parallelism between (\neg_{Prop}) – $(\rightarrow_{\text{Prop}})$ and (\neg_{FO}) – $(\rightarrow_{\text{FO}})$, it is useful to extract the following, "many-sorted" view of this semantics. That is, for the connectives \neg through \rightarrow , which give sentences when taking sentences as arguments and give *n*-ary formulas when taking *n*-ary formulas as arguments, *W* provides a possible-world model of worlds $w \in W$ and propositions $[[\sigma]] \subseteq W$; *D* provides a possible-world model of individuals $a \in D$ and properties $[[x | \varphi]] \subseteq D$; and each D^n provides a possible-world model of *n*-tuples $\bar{a} \in D^n$ and *n*-ary relations $[[\bar{x} | \varphi]]$.

To interpret \Box , we take advantage of the "many-sorted" view, and straightforwardly extend (\Box_{Prop}); that is, we interpret \Box working on a sentence with \mathbf{int}_W on W, \Box working on a unary formula with \mathbf{int}_D on D, and \Box working on an n-ary formula with \mathbf{int}_{D^n} on D^n .¹⁰ Each D^n is an MCN-frame equipped with a basic-neighborhood map $\mathcal{B}^n : D^n \to \mathcal{PP}(D^n)$ such that, for each $\bar{a} \in D^n$ and $B \subseteq D^n$,

$$B \in \mathcal{B}^n(\bar{a}) \iff B = (B_1 \times \cdots \times B_n) \cap D^n$$
 where
 $B_i \in \mathcal{B}(a_i)$ for some *i* and,
for each *i*, either $B_i \in \mathcal{B}(a_i)$ or $B_i = D$

¹⁰We can contrast this to the non-sheaf frameworks of Kripke's semantics (Kripke 1963) and of neighborhood semantics with constant domains by Arló-Costa and Pacuit (2006), in both of which \Box is always interpreted with structures on W. In this regard, our interpretation is conceptually closer to that in David Lewis's counterpart theory (Lewis 1968), which interprets \Box working on a unary formula with a counterpart relation among individuals. Notwithstanding this and other conceptual difference, however, models by Kripke and by Arló-Costa and Pacuit can be subsumed as constant sheaves, in which D consists of copies of W having \mathcal{B}_D strictly parallel to \mathcal{B}_W .

in other words, when we write $p_i : D^n \to D :: \bar{a} \mapsto a_i$ for each projection, $\mathcal{B}^n(\bar{a})$ consists of the inverse images $p_i^{-1}[B]$ of $B \in \mathcal{B}(a_i)$ (for all *i*) and their (nonempty) finite intersections.¹¹ So (\Box_{Prop}) extends to the constraint that, for each *n*,

$$\llbracket \Box \rrbracket = \operatorname{int}_{D^n}, \qquad (\Box_{\mathrm{FO}})$$

so that $\bar{a} \in \llbracket \bar{x} \mid \Box \varphi \rrbracket \iff B \subseteq \llbracket \bar{x} \mid \varphi \rrbracket$ for some $B \in \mathcal{B}^n(\bar{a}).$

To sum up, we enter

Definition 2.1. A *neighborhood-sheaf model* for a first-order language \mathcal{L} is a pair $(\pi, \llbracket \cdot \rrbracket)$ of a surjective local isomorphism $\pi : D \to W$ and and a map $\llbracket \cdot \rrbracket$ that interprets¹²

- each *n*-ary predicate R of \mathcal{L} with a subset [[R]] of D^n , and
- each *n*-ary function symbol f of \mathcal{L} with an arrow $\llbracket f \rrbracket : D^n \to D$ in \mathbf{LI}/W (the case of individual constants is covered with n = 0).¹³

 ${}^{11}\pi^n$ with $(D^n, \mathcal{B}^n(\bar{a}))$ is the *n*-fold product of π in the category **LI**/W of *MCN-frames* and local isomorphisms over W. We can loosen the definition of local isomorphism $\pi : D \to W$ so that W and D can be MN-frames that are not MCN-frames, and think of the category of *MN-frames* and local isomorphisms over W. Yet the *n*-fold product of π in this category is "coarser" than in **LI**/W, given as $\mathcal{B}^n(\bar{a}) = \bigcup_{1 \le i \le n} \{p_i^{-1}[B] \mid B \in \mathcal{B}(a_i)\}$ for the maps $p_i : D^n \to D :: \bar{a} \mapsto a_i$. This \mathcal{B}^n is too coarse for the purpose of modelling \Box , because together with (\Box_{FO}) it entails

$$\begin{split} \bar{a} \in \llbracket \bar{x} \mid \Box \varphi \rrbracket \iff \exists i \exists B \in \mathcal{B}_D(a_i) \cdot p_i^{-1} [B] \subseteq \llbracket \bar{x} \mid \varphi \rrbracket \\ \iff \exists i \exists B \in \mathcal{B}_D(a_i) \cdot B \subseteq \llbracket x_i \mid \forall x_1 \cdots \forall x_{i-1} \forall x_{i+1} \cdots \forall x_n \varphi \rrbracket \\ \iff \exists i \cdot a_i \in \llbracket x_i \mid \Box \forall x_1 \cdots \forall x_{i-1} \forall x_{i+1} \cdots \forall x_n \varphi \rrbracket. \end{split}$$

To avoid this, we need a basic-neighborhood map $B_n : D^n \to \mathcal{PPD}^n$ that is "finer" than \mathcal{B}^n . This means that, to provide a model, we need an infinite sequence $((W, \mathcal{B}_W), (D, \mathcal{B}_D), (D^2, \mathcal{B}_2), \ldots)$ of data, rather than just a pair $((W, \mathcal{B}_W), (D, \mathcal{B}_D))$. In this article we choose MCN-frames as opposed to MN-frames, so that a pair $((W, \mathcal{B}_W), (D, \mathcal{B}_D))$ suffices.

¹²Given (\square_{FO}) and (vac), the requirement that $\pi : D \to W$ is an open map amounts to the condition that the square



commutes. In other words, the requirement means that $\llbracket y \mid \Box \sigma \rrbracket$ is well-defined from $\llbracket \sigma \rrbracket$.

¹³That $\llbracket f \rrbracket : D^n \to D$ is an arrow in \mathbf{LI}/W means that $\pi \circ f = \pi^n$ and that f is a local isomorphism; but for these to hold it is enough that $\pi \circ f = \pi^n$ and that f is continuous, meaning that $f^{-1}[\mathbf{int}_D(A)] \subseteq \mathbf{int}_{D^n}(f^{-1}[A])$. Moreover, a *neighborhood-sheaf interpretation* over a neighborhood-sheaf model $(\pi, \llbracket \cdot \rrbracket)$ for \mathcal{L} is a map that extends $\llbracket \cdot \rrbracket$ to all the terms t and formulas φ of \mathcal{L} , interpreting t with $\llbracket \bar{x} | t \rrbracket : D^n \to D$ and φ with $\llbracket \bar{x} | \varphi \rrbracket \subseteq D^n$, following $(\neg_{FO}) - (\Box_{FO})$ along with the constraint that $\llbracket \bar{x} | R\bar{x} \rrbracket = \llbracket R \rrbracket$ for n-ary predicate R.¹⁴

To illustrate how a neighborhood-sheaf model $(\pi : D \to W, \llbracket \cdot \rrbracket)$ interprets terms, let us take an individual constant c. (See Figure 1.) Since each $w \in W$ is a model



Figure 1: A global section $\llbracket c \rrbracket$ of a sheaf $\pi : D \to W$

of first-order logic, *c* has its reference in *w*, written $[\![c]\!](w)$, which lives in *w*—that is, $[\![c]\!](w) \in D_w$ and $\pi([\![c]\!](w)) = w$. Thus, the interpretation of *c* in the entire $(\pi, [\![\cdot]\!])$ is a map $[\![c]\!]: W \to D$ such that $\pi \circ [\![c]\!]: W \to W$ is the identity on *W*. Moreover, $[\![c]\!]$ is not just any map picking $[\![c]\!](w)$ randomly from D_w , but has to be an arrow in \mathbf{LI}/W , that is, a continuous map, picking $[\![c]\!](w)$ continuously. In topological terms, $[\![c]\!]$ forms a "global section".

The class of neighborhood-sheaf interpretations forms neighborhood-sheaf semantics, and its logic **FOMC** is obtained by adding the rule M and the axiom C to classical first-order logic.¹⁵ Correspondence results are also available; for instance, **FOS4**, firstorder logic with **S4**, is the logic of the subclass of topological-sheaf models (Awodey and Kishida 2008).

¹⁴There are constraints on interpreting terms and interpreting substitution of terms, but we omit them in this article. See (Awodey and Kishida 2008) and (Kishida 2011) instead.

¹⁵**FOMC** proves the converse Barcan formula $\Box \forall x \varphi \rightarrow \forall x \Box \varphi$ and also $\exists x \Box \varphi \rightarrow \Box \exists x \varphi$, but neither the Barcan $\forall x \Box \varphi \rightarrow \Box \forall x \varphi$ nor $\Box \exists x \varphi \vdash \exists x \Box \varphi$. Also, **FOMC** fails to prove $x = y \rightarrow \Box (x = y)$ and $x \neq y \rightarrow \Box (x \neq y)$.

3 First-Order Public-Announcement Logic

We are finally ready to combine the two semantics in the previous sections to provide a neighborhood-sheaf semantics for **FOPAL**, first-order public-announcement logic.

We define a language \mathcal{L} of **FOPAL** as a first-order language (possibly with the relation symbol =, function symbols, etc.) with operators K_i for $i \in I$ and $[!\sigma]$ for all sentences σ of \mathcal{L} . In this article, however, we settle for a pair of restrictions (leaving a formulation without them for future work): We assume that the set I of agents is a singleton, and simply write \Box for K_i ; we also require σ in $[!\sigma]$ be a sentence and with no free variables.

We define the restriction updating of sheaf models by taking advantage of the "many-sorted" view yet again. $[!\sigma]$ gives a sentence when taking a sentence as an argument and gives an *n*-ary formula when taking an *n*-ary formula as argument. So we interpret $[!\sigma]$ working on a sentence with a restriction update of W, $[!\sigma]$ working on a unary formula with a restriction update of D, and $[!\sigma]$ working on an *n*-ary formula with a restriction update of D^n . And these restriction updates are parallel to one another, and in particular to the propositional case $(\llbracket \sigma \rrbracket, \mathcal{B}_{\llbracket \sigma \rrbracket}, \llbracket \cdot \rrbracket_{\llbracket \sigma}]$.

The only question is to what subsets we should restrict D and D^n . In the propositional case, we eliminate worlds $w \notin \llbracket \sigma \rrbracket$ since they are no longer possibilities once σ is announced. So, similarly, we eliminate individuals a such that $\pi(a) \notin \llbracket \sigma \rrbracket$, since such a, not being from worlds in $\llbracket \sigma \rrbracket$, are no longer "possible individuals" once σ is announced; therefore we restrict D to $\pi^{-1}[\llbracket \sigma \rrbracket] = \sum_{w \in \llbracket \sigma \rrbracket} D_w$, the set of individuals from $\llbracket \sigma \rrbracket$. More generally, given $S \subseteq W$, the restriction of D^n from over W to over S is

$$D_S^n = (\pi^n)^{-1}[S] = \sum_{w \in S} D_w^n \subseteq D^n,$$

the set of *n*-tuples from S; in particular, $D_S^0 = S$. Thus, extending Definition 1.3 straightforwardly, we have the following.

Definition 3.1. Given a neighborhood-sheaf model $M = (\pi : D \to W, \llbracket \cdot \rrbracket)$ for \mathcal{L} and nonempty $S \subseteq W$, let $M_S = (\pi_S, \llbracket \cdot \rrbracket_S)$ be the neighborhood-sheaf model for \mathcal{L} as follows.

- $\pi_S = \pi \upharpoonright D_S : D_S \to S :: a \mapsto \pi(a);$
- $\mathcal{B}_S(w) = \{ B \cap S \mid B \in \mathcal{B}_W(w) \};$
- $\mathcal{B}_{D_S}(a) = \{ B \cap D_S \mid B \in \mathcal{B}_D(a) \};$
- $\llbracket R \rrbracket_S = \llbracket R \rrbracket \cap D_S^n$ for every *n*-ary predicate *R* of \mathcal{L} ; and

• $\llbracket f \rrbracket_S = \llbracket f \rrbracket \upharpoonright D_S^n : D_S^n \to D_S :: \bar{a} \mapsto \llbracket f \rrbracket(\bar{a})$ for every *n*-ary function symbol *f* of \mathcal{L} .

As the definition presupposes, and as is straightforward to check, π_S is a local isomorphism over *S* (with \mathcal{B}_S), and so M_S is in fact a neighborhood-sheaf model.¹⁶

So, for the model-update interpretation of $[!\sigma]$, let us enter

Definition 3.2. Given a language \mathcal{L} of **FOPAL**, a neighborhood-sheaf interpretation over a neighborhood-sheaf model $M = (\pi : D \to X, \llbracket \cdot \rrbracket)$ is called a *neighborhood-sheaf PAL-interpretation* over M if it satisfies the following, straightforwardly extending $(\llbracket \cdot \rrbracket)_{\text{Prop}}$. (It may be worth noting that $D_{\parallel \sigma \parallel}^n = \llbracket \bar{x} \mid \sigma \rrbracket$.)

$$\llbracket \bar{x} \mid \llbracket !\sigma]\varphi \rrbracket = \llbracket \bar{x} \mid \sigma \rrbracket \llbracket \to \rrbracket \llbracket \bar{x} \mid \varphi \rrbracket_{\llbracket \sigma \rrbracket}, \qquad (\llbracket !\cdot]_{FO})$$

so that $\bar{a} \in \llbracket \bar{x} \mid \llbracket !\sigma]\varphi \rrbracket \iff$ either $\bar{a} \in D^n \setminus D^n_{\llbracket \sigma \rrbracket}$
or $\bar{a} \in \llbracket \bar{x} \mid \varphi \rrbracket_{\llbracket \sigma \rrbracket} \subseteq D^n_{\llbracket \sigma \rrbracket}.$

The class of such interpretations forms neighborhood-sheaf semantics for **FOPAL**. Also, the subclasses of interpretations over Kripke sheaves and over topological sheaves respectively form Kripke-sheaf semantics and topological-sheaf semantics for **FOPAL**.

The semantics for **FOPAL** validates $R_{atom}-R_{it}$, just as in the case for propositional **PAL**. A language of **FOPAL** has one more operator, namely, \exists ; the semantics validates the reduction axiom

$$[!\sigma] \exists y \varphi \equiv \exists y [!\sigma] \varphi, \qquad \mathbf{R}_{\exists}$$

¹⁶It is also worth noting that the restriction preserves (finite) products, in the sense that, not only $(D^n)_S = (D_S)^n = D_S^n$, but also $(\mathcal{B}^n)_{D_S^n} = (\mathcal{B}_{D_S})^n$. To sum up in category-theoretic terms, the restriction is given by the "change of base" functor from **LI**/*W* to **LI**/*S*, which preserves products.

because for each $\bar{a} \in D^n$ we have

$$\begin{split} \bar{a} \in \llbracket \bar{x} \mid [!\sigma] \exists y \varphi \rrbracket \\ \iff \text{ either } \bar{a} \in D^n \setminus D_{\llbracket \sigma \rrbracket}^n \text{ or } \bar{a} \in \llbracket \bar{x} \mid \exists y \varphi \rrbracket_{\llbracket \sigma \rrbracket} \\ \iff \text{ either } (\bar{a}, b) \in D^{n+1} \setminus D_{\llbracket \sigma \rrbracket}^{n+1} \text{ for some } b \in D \setminus D_{\llbracket \sigma \rrbracket} \\ \text{ or } (\bar{a}, b) \in [\llbracket \bar{x}, y \mid \varphi \rrbracket_{\llbracket \sigma \rrbracket} \text{ for some } b \in D_{\llbracket \sigma \rrbracket} \\ \iff \text{ either } (\bar{a}, b) \in D^{n+1} \setminus D_{\llbracket \sigma \rrbracket}^{n+1} \text{ for some } b \in D_{\llbracket \sigma \rrbracket} \\ \iff \text{ either } (\bar{a}, b) \in D^{n+1} \setminus D_{\llbracket \sigma \rrbracket}^{n+1} \text{ for some } b \in D \\ \text{ or } (\bar{a}, b) \in [\llbracket \bar{x}, y \mid \varphi \rrbracket_{\llbracket \sigma \rrbracket} \text{ for some } b \in D \\ \iff \text{ there is } b \in D \text{ such that either } (\bar{a}, b) \in D^{n+1} \setminus D_{\llbracket \sigma \rrbracket}^{n+1} \\ \text{ or } (\bar{a}, b) \in \llbracket \bar{x}, y \mid \varphi \rrbracket_{\llbracket \sigma \rrbracket} \text{ for some } b \in D \\ \iff \text{ there is } b \in D \text{ such that } (\bar{a}, b) \in [\llbracket \bar{x}, y \mid \varphi \rrbracket_{\llbracket \sigma \rrbracket} \\ \iff \tilde{a} \in \llbracket \bar{x} \mid \exists y [!\sigma] \varphi \rrbracket. \end{split}$$

Therefore, again, R_{atom} - R_{it} and R_{\exists} completely axiomatize **FOPAL** by reducing it to its static fragment, that is, **FOMC** (or any stronger logic one chooses; in particular, **FOK** for Kripke-sheaf semantics and **FOS4** for topological-sheaf semantics).

Theorem 2. Given a class \mathbb{M} of neighborhood-sheaf models, let \mathbb{T} be the logic of \mathbb{M} in a first-order modal (but static) language \mathcal{L} . Then the logic of \mathbb{M} in the language of **FOPAL** extending \mathcal{L} is obtained by adding $R_{atom}-R_{it}$ and R_{\exists} to \mathbb{T} .

In this way, the sheaf semantics extends **PAL** to the first-order in a natural, straightforward, and modular fashion. Nonetheless, its flexibility in treating individuals and identity thereof facilitates interesting dynamic behaviors of models. To see this, let us take an example from Awodey and Kishida (2008), namely, a local homeomorphism (that is, local isomorphism with topology) $\pi : \mathbb{R}^+ \to S^1 :: a \mapsto (\cos 2\pi a, \sin 2\pi a)$ with $[\![\leq]\!] = \{(a, b) \in (\mathbb{R}^+)^2 \mid a \leq b\}.^{17}$ (See Figure 2.)

 \mathbb{R}^+ draws a spiral over the circle S^1 . In this spiral, the intervals (0, 1] and (0, 1) are $[x | \forall y.x \leq y]$ and $[x | \Box \forall y.x \leq y]$, the sets of elements that are "actually the least" and that are "necessarily the least" (in their own worlds); therefore, with φ short for $\forall y.x \leq y \rightarrow \Box \forall y.x \leq y$, $1 \notin [x | \varphi]$ and so φ is not valid in $(\pi, [\cdot])$. Now consider the image of $(0, 1) = [x | \Box \forall y.x \leq y]$ projected down to the circle, $S^1 \setminus \{(1, 0)\}$; it is $[\exists x \Box \forall y.x \leq y]$. Write σ for $\exists x \Box \forall y.x \leq y$. The restriction to $[\sigma]$ eliminates $1 \notin [x | \varphi]$, and so φ becomes valid in $(\pi_{[\sigma]}, [\cdot]_{[\sigma]})$; thus, $(\pi, [\cdot])$ validates $[!\sigma]\varphi$.

Now, recall that any neighborhood-sheaf model $(\pi : D \to W, \llbracket \cdot \rrbracket)$ has to interpret an individual constant *c* with a continuous, global section $\llbracket c \rrbracket : W \to D$. This means that

¹⁷Awodey and Kishida (2008), 157–159.



Figure 2: Restriction update of a sheaf

the spiral model above cannot interpret a language with any individual constant, since $\pi : \mathbb{R}^+ \to S^1$ has no global section. The public announcement of σ , however, "severs" the spiral at 1, 2, 3, ..., updating the model into a constant-domain model consisting of N-many copies of $[\sigma]$, and hence the new model accommodates a language with any constants for natural numbers. This example points to an application of **FOPAL** to dynamic processes of naming things in epistemic inquiries, in which the increase in knowledge makes our ontology better structured and thereby enables us to name more objects.

4 Conclusion

In this article, we laid out three elements of extension for the standard **PAL**, namely, the (monotone) neighborhood setting, the first-order version **FOPAL**, and its neighborhood-sheaf semantics. The first facilitates the verifiability interpretation of K_i operators. The second brings into the **PAL** analysis of dynamic and epistemic processes a first-order vocabulary, and in particular names for individuals. And, lastly, the sheaf semantics gives us a flexible way of modelling how a first-order ontology as well as a first-order vocabulary can evolve in dynamic and epistemic processes.

Obvious extensions of this work include the following. In this article we only dealt with a single-agent setting for **FOPAL**; so a multi-agent setting is naturally expected, and it is a natural step to discuss distributed knowledge and common knowledge as well. Also, we only dealt with public announcement of closed sentences, but public announcement of open formulas is also an informative event, so the fully general semantics of **FOPAL** should be able to deal with it. Yet the principal thesis of this article

is that the sheaf semantics enables us to lift up whatever structure we have at the propositional level to the first order straightforwardly. Therefore the semantics is expected to help us to lift up dynamic epistemic logics (van Ditmarsch et al. 2008) or other formalisms of logical dynamics (van Benthem 2011) such as belief revision (Baltag et al. 2008) to the first order.

Above all, bringing logical dynamics to the first order adds an entirely new aspect to the scope of logical dynamics, namely, dynamics in first-order ontology. The sheaf semantics enables us to flexibly model the interaction among the evolutions of knowledge, of ontology and of language; and it should be an interesting future project to introduce modal operators and other expressions for explicitly expressing and analyzing this interaction.

Acknowledgements This research was funded by the VIDI grant 639.072.904 of the Netherlands Organization for Scientific Research. The author is grateful to anonymous referees for and the audience at the workshop *Logic and Engineering of Natural Language Semantics 9*, and in particular to Hans Kamp and to Katsuhiko Sano, for their insightful comments and helpful suggestions. An extended version of this paper was presented at the *Logic and Interactive Rationality* seminar at the University of Amsterdam, and the author thanks the audience there, and in particular Alexandru Baltag and Johan van Benthem, for inspiring discussions and instructive comments. A grateful acknowledgment also goes to Lorenz Demey and Barteld Kooi for valuable discussions of this research as well as of their researches in relation to this.

References

H. Arló-Costa and E. Pacuit. First-order classical modal logic. *Studia Logica*, 84: 171–210, November 2006.

S. Awodey and K. Kishida. Topology and modality: the topological interpretation of first-order modal logic. *Review of Symbolic Logic*, 1:146–166, August 2008.

A. Baltag, H. P. van Ditmarsch, and L. S. Moss. Epistemic logic and information update. In P. Adriaans and J. van Benthem, editors, *Philosophy of Information*. Amsterdam: Elsevier, 2008.

J. van Benthem. *Logical Dynamics of Information and Interaction*. Cambridge: Cambridge University Press, 2011.

J. van Benthem and E. Pacuit. Dynamic logics of evidence-based beliefs. *Studia Logica*, 99:61–92, October 2011.

B. F. Chellas. *Modal Logic: An Introduction*. Cambridge: Cambridge University Press, 1980.

A. Dabrowski, L. S. Moss, and R. Parikh. Topological reasoning and the logic of knowledge. *Annals of Pure and Applied Logic*, 78, 1996.

L. Demey. Towards a dynamics of realistic knowledge: neighborhood semantics for public announcement logic. In *Fourth Conference of the Dutch-Flemish Association for Analytic Philosophy*, Leuven, Belgium, January 2010.

H. van Ditmarsch, W. van der Hoek, and B. Kooi. *Dynamic-Epistemic Logic*. Dordrecht: Springer, 2008.

D. M. Gabbay, V. Shehtman, and D. Skvortsov. *Quantification in Nonclassical Logic*, volume 1. Burlington: Elsevier, 2009.

H. H. Hansen. Monotonic modal logics. Master's thesis, University of Amsterdam, Amsterdam, the Netherlands, October 2003.

D. Harel. First-Order Dynamic Logic. Berlin: Springer, 1979.

D. Harel, D. Kozen, and J. Tiuryn. Dynamic Logic. Cambridge: MIT Press, 2000.

K. Kishida. Neighborhood-sheaf semantics for first-order modal logic. *Electronic Notes in Theoretical Computer Science*, 278:129–143, 2011.

B. Kooi. Dynamic term-modal logic. In J. van Benthem, S. Ju, and F. Veltman, editors, *A Meeting of the Minds: Proceedings of the Workshop on Logic, Rationality and Interaction, Beijing, 2007*, pages 173–185. London: College Publications, 2007.

S. Kripke. Semantical considerations on modal logic. *Acta Philosophica Fennica*, 16, 1963.

D. Lewis. Counterpart theory and quantified modal logic. *Journal of Philosophy*, 65, 1968.

M. Ma. Mathematics of public announcements. In H. P. van Ditmarsch, J. Lang, and S. Ju, editors, *Logic, Rationality and Interaction: Third International Workshop, LORI 2011, Guangzhou, China, October 10–13, 2011, Proceedings*, pages 193–205. Heidelberg: Springer, 2011.

J. Zvesper. *Playing with Information*. PhD thesis, University of Amsterdam, Amsterdam, the Netherlands, March 2010. ILLC Dissertation Series DS-2010-02.

Transition Semantics. The Dynamics of Dependence Logic

Pietro Galliani

University of Helsinki pgallian@gmail.com

Abstract

We examine the relationship between Dependence Logic and game logics. A variant of Dynamic Game Logic, called *Transition Logic*, is developed, and we show that its relationship with Dependence Logic is comparable to the one between First-Order Logic and Dynamic Game Logic discussed by van Benthem.

This suggests a new perspective on the interpretation of Dependence Logic formulas, in terms of assertions about *reachability* in games of imperfect information against Nature. We then capitalize on this intuition by developing expressively equivalent variants of Dependence Logic in which this interpretation is taken to the foreground.

1 Introduction

1.1 Dependence Logic

Dependence Logic (Väänänen 2007a) is an extension of First-Order Logic which adds *dependence atoms* of the form $=(t_1, \ldots, t_n)$ to it, with the intended interpretation of "the value of the term t_n is a function of the values of the terms $t_1 \ldots t_{n-1}$." In recent years, this logic has attracted a remarkable amount of interest as a formalism for the formal analysis of the properties of *dependence itself* in a first-order setting; and some recent papers (Grädel and Väänänen 2013, Engström 2012, Galliani 2012) explore the effects of replacing dependence atoms with other similar primitives such as *indepen-*

dence atoms (Grädel and Väänänen 2013), *multivalued dependence atoms* (Engström 2012), or *inclusion* or *exclusion* atoms (Galliani 2011; 2012).

In this section we will recall the definition of the semantics of this logic; then, in the rest of this work, we will examine its relationship with (an imperfect-information variant of) Dynamic Game Logic, and discuss how it (or variants thereof) may be used as a formalism for reasoning about games of imperfect information.

Definition 1.1 (Assignments and substitutions). Let M be a first order model and let V be a finite set of variables. Then an *assignment* over M with *domain* V is a function s from V to the set Dom(M) of all elements of M.

Furthermore, for any assignment s over M with domain V, any element $m \in Dom(M)$ and any variable v (not necessarily in V), we write s[m/v] for the assignment with domain $V \cup \{v\}$ such that

$$s[m/v](w) = \begin{cases} m & \text{if } w = v;\\ s(w) & \text{if } w \in V \setminus \{v\} \end{cases}$$

for all $w \in V \cup \{v\}$.

Definition 1.2 (Team). Let *M* be a first-order model and let *V* be a finite set of variables. A *team X* over *M* with *domain* Dom(X) = V is a set of assignments from *V* to *M*.

Definition 1.3 (Relations corresponding to teams). Let *X* be a team over *M*, and let *V* be a finite set of variables. and let \vec{v} be a finite tuple of variables in its domain. Then $X(\vec{v})$ is the relation $\{s(\vec{v}) : s \in X\}$. Furthermore, we write Rel(X) for X(Dom(X)).

As is often the case for Dependence Logic, we will assume that all our formulas are in Negation Normal Form:

Definition 1.4 (Dependence Logic, Syntax). Let Σ be a first-order signature. Then the set of all dependence logic formula with signature Σ is given by

$$\varphi ::= R\vec{t} \mid \neg R\vec{t} \mid = (t_1, \dots, t_n) \mid \varphi \lor \varphi \mid \varphi \land \varphi \mid \exists v\varphi \mid \forall v\varphi$$

where *R* ranges over all relation symbols, \vec{t} ranges over all tuples of terms of the appropriate arities, $t_1 \dots t_n$ range over all terms and *v* ranges over the set Var of all variables.

The set $Free(\varphi)$ of all *free variables* of a formula φ is defined precisely as in First Order Logic, with the additional condition that all variables occurring in a dependence atom are free with respect to it.

Definition 1.5 (Dependence Logic, Semantics). Let *M* be a first-order model, let *X* be a team over it, and let φ be a Dependence Logic formula with the same signature of *M* and with free variables in Dom(*X*). Then we say that *X* satisfies φ in *M*, and we write $M \models_X \varphi$, if and only if

- **TS-lit:** φ is a first-order literal and $M \models {}_{s}\varphi$ for all $s \in X$;
- **TS-dep:** φ is a dependence atom = (t_1, \ldots, t_n) and any two assignments $s, s' \in X$ which assign the same values to $t_1 \ldots t_{n-1}$ also assign the same value to t_n ;
- **TS-V:** φ is of the form $\psi_1 \lor \psi_2$ and there exist two teams Y_1 and Y_2 such that $X = Y_1 \cup Y_2$, $M \models_{Y_1} \psi_1$ and $M \models_{Y_2} \psi_2$;
- **TS-** \wedge : φ is of the form $\psi_1 \land \psi_2$, $M \models_X \psi_1$ and $M \models_X \psi_2$;
- **TS-∃:** φ is of the form $\exists v\psi$ and there exists a function $F: X \to \text{Dom}(M)$ such that $M \models_{X[F/v]} \psi$, where

$$X[F/v] = \{s[F(s)/v] : s \in X\};$$

TS- \forall : φ is of the form $\forall v\psi$ and $M \models_{X[M/v]} \psi$, where

$$X[M/v] = \{s[m/v] : s \in X, m \in Dom(M)\}.$$

The disjunction of Dependence Logic does not behave like the classical disjunction: for example, it is easy to see that $=(x) \lor =(x)$ is not equivalent to =(x), as the former holds for the team $X = \{\{(x, 0)\}, \{(x, 1)\}\}$ and the latter does not. However, it is possible to define the classical disjunction in terms of the other connectives:

Definition 1.6 (Classical Disjunction). Let ψ_1 and ψ_2 be two Dependence Logic formulas, and let u_1 and u_2 be two variables not occurring in them. Then we write $\psi_1 \sqcup \psi_2$ as a shorthand for

$$\exists u_1 \exists u_2 (=(u_1) \land =(u_2) \land ((u_1 = u_2 \land \psi_1) \lor (u_1 \neq u_2 \land \psi_2))).$$

Proposition 1. For all formulas ψ_1 and ψ_2 , all models M with at least two elements¹ whose signature contains that of ψ_1 and ψ_2 and all teams X whose domain contains the free variables of ψ_1 and ψ_2

$$M \models_X \psi_1 \sqcup \psi_2 \Leftrightarrow M \models_X \psi_1 \text{ or } M \models_X \psi_2.$$

The following four proportions are from (Väänänen 2007a):

Proposition 2. For all models M and Dependence Logic formulas φ , $M \models_{\emptyset} \varphi$.

Proposition 3 (Downwards Closure). If $M \models_X \varphi$ and $Y \subseteq X$ then $M \models_Y \psi$.

¹In general, we will assume through this whole work that all first-order models which we are considering have at least two elements. As one-element models are trivial, this is not a very onerous restriction.

Proposition 4 (Locality). If $M \models_X \varphi$ and $X(Free(\varphi)) = Y(Free(\varphi))$ then $M \models_Y \varphi$.

Proposition 5 (From Dependence Logic to Σ_1^1). Let $\varphi(\vec{v})$ be a Dependence Logic formula with free variables in \vec{v} . Then there exists a Σ_1^1 sentence $\Phi(R)$ such that

$$M \models_X \varphi \Leftrightarrow M \models \Phi(X(\vec{v}))$$

for all suitable models M and for all nonempty teams X. Furthermore, in $\Phi(R)$ the symbol R occurs only negatively.

As proved in (Kontinen and Väänänen 2009), there is also a converse for the last proposition:

Theorem 1 (From Σ_1^1 to Dependence Logic). Let $\Phi(R)$ be a Σ_1^1 sentence in which R occurs only negatively. Then there exists a Dependence Logic formula $\varphi(\vec{v})$, where $|\vec{v}|$ is the arity of R, such that

$$M \models_X \varphi \Leftrightarrow M \models \Phi(X(\vec{v}))$$

for all suitable models M and for all nonempty teams X whose domain contains \vec{v} .

Because of this correspondence between Dependence Logic and Existential Second Order Logic, it is easy to see that Dependence Logic is closed under existential quantification: for all Dependence Logic formulas $\varphi(\vec{v}, P)$ over the signature $\Sigma \cup \{P\}$ there exists a Dependence Logic formula $\exists P\varphi(\vec{v}, P)$ over the signature Σ such that

$$M \models_X \exists P \varphi(\vec{v}, P) \Leftrightarrow \exists P \text{ s.t. } M \models_X \varphi(\vec{v}, P)$$

for all models M with domain Σ and for all teams X over the free variables of φ . Therefore, in the rest of this work we will add second-order existential quantifiers to the language of Dependence Logic, and we will write $\exists P\varphi(\vec{v}, P)$ as a shorthand for the corresponding Dependence Logic expression.

1.2 Dynamic Game Logic

Game logics are logical formalisms for reasoning about games and their properties in a very general setting. Whereas the Game Theoretic Semantics approach attempts to use game-theoretic techniques to *interpret* logical systems, game logics attempt to put logic to the service of game theory, by providing a high-level language for the study of games.

They generally contain two different kinds of expressions:

- 1. *Game terms*, which are descriptions of games in terms of compositions of certain primitive *atomic games*, whose interpretation is presumed fixed for any given game model;
- 2. *Formulas*, which, in general, correspond to assertions about the abilities of players in games.

In this subsection, we are going to summarize the definition of a variant of Dynamic Game Logic (Parikh 1985).² Then, in the next subsection, we will discuss a remarkable connection between First-Order Logic and Dynamic Game Logic discovered by Johan van Benthem in (van Benthem 2003).

One of the fundamental semantic concepts of Dynamic Game Logic is the notion of *forcing relation:*

Definition 1.7 (Forcing Relation). Let *S* be a nonempty set of *states*. A *forcing relation* over *S* is a set $\rho \subseteq S \times \text{Parts}(S)$, where Parts(S) is the powerset of *S*.

In brief, a forcing relation specifies the abilities of a player in a perfect-information game: $(s, X) \in \rho$ if and only if the player has a strategy that guarantees that, whenever the initial position of the game is *s*, the terminal position of the game will be in *X*.

A (two-player) *game* is then defined as a pair of forcing relations satisfying some axioms:

Definition 1.8 (Game). Let *S* be a nonempty set of states. A *game* over *S* is a pair (ρ^E, ρ^A) of forcing relations over *S* satisfying the following conditions for all $i \in \{E, A\}$, all $s \in S$ and all $X, Y \subseteq S$:

Monotonicity: If $(s, X) \in \rho^i$ and $X \subseteq Y$ then $(s, Y) \in \rho^i$;

Consistency: If $(s, X) \in \rho^E$ and $(s, Y) \in \rho^A$ then $X \cap Y \neq \emptyset$;

Non-triviality: $(s, \emptyset) \notin \rho^i$.

Determinacy: If $(s, X) \notin \rho^i$ then $(s, S \setminus X) \in \rho^j$, where $j \in \{E, A\} \setminus \{i\}$.³

²The main difference between this version and the one of Parikh's original paper lies in the absence of the iteration operator γ^* from our formalism. In this, we follow (van Benthem 2003, van Benthem et al. 2008).

³This requirement is nothing but a formal version of Zermelo's Theorem: if one of the players cannot force the outcome of the game to belong to a set of "winning outcomes" X, this implies that the other player can force it to belong to the complement of X.

Definition 1.9 (Game Model). Let *S* be a nonempty set of states, let Φ be a nonempty set of *atomic propositions* and let Γ be a nonempty set of *atomic game symbols*. Then a *game model* over *S*, Φ and Γ is a triple $(S, \{(\rho_g^E, \rho_g^A) : g \in \Gamma\}, V)$, where (ρ_g^E, ρ_g^A) is a game over *S* for all $g \in \Gamma$ and where *V* is a valutation function associating each $p \in \Phi$ to a subset $V(p) \subseteq S$.

The language of Dynamic Game Logic, as we already mentioned, consists of *game terms*, built up from atomic games, and of *formulas*, built up from atomic proposition. The connection between these two parts of the language is given by the *test* operation φ ?, which turns any formula φ into a test game, and the *diamond* operation, which combines a game term γ and a formula φ into a new formula $\langle \gamma, i \rangle \varphi$ which asserts that agent *i* can guarantee that the game γ will end in a state satisfying φ .

Definition 1.10 (Dynamic Game Logic - Syntax). Let Φ be a nonempty set of *atomic propositions* and let Γ be a nonempty set of *atomic game formulas*. Then the sets of all game terms γ and formulas φ are defined as

$$\begin{array}{lll} \gamma & ::= & g \mid \varphi ? \mid \gamma; \gamma \mid \gamma \cup \gamma \mid \gamma^{d} \\ \varphi & ::= & \perp \mid p \mid \neg \varphi \mid \varphi \lor \varphi \mid \langle \gamma, i \rangle \varphi \end{array}$$

for p ranging over Φ , g ranging over Γ , and i ranging over $\{E, A\}$.

We already mentioned the intended interpretations of the test connective φ ? and of the diamond connective $\langle \gamma, i \rangle \varphi$. The interpretations of the other game connectives should be clear: γ^d is obtained by swapping the roles of the players in γ , $\gamma_1 \cup \gamma_2$ is a game in which the existential player *E* chooses whether to play γ_1 or γ_2 , and γ_1 ; γ_2 is the *concatenation* of the two games corresponding to γ_1 and γ_2 respectively.

Definition 1.11 (Dynamic Game Logic - Semantics). Let $G = (S, \{(\rho_g^E, \rho_g^A) : g \in \Gamma\}, V)$ be a game model over S, Γ and Φ . Then for all game terms γ and all formulas φ of Dynamic Game Logic over Γ and Φ we define a game $|\gamma|_G$ and a set $|\varphi|_G \subseteq S$ as follows:

DGL-atomic-game: For all $g \in \Gamma$, $|g|_G = (\rho_g^E, \rho_g^A)$;

DGL-test: For all formulas φ , $|\varphi|_G = (\rho^E, \rho^A)$, where

- $s\rho^E X$ iff $s \in |\varphi|_G$ and $s \in X$;
- $s\rho^A X$ iff $s \notin |\varphi|_G$ or $s \in X$

for all $s \in S$ and all X with $\emptyset \neq X \subseteq S$;

- **DGL-concat:** For all game terms γ_1 and γ_2 , $|\gamma_1; \gamma_2|_G = (\rho^E, \rho^A)$, where, for all $i \in \{E, A\}$ and for $|\gamma_1|_G = (\rho_1^E, \rho_1^A), |\gamma_2|_G = (\rho_2^E, \rho_2^A),$
 - $s\rho^i X$ if and only if there exists a Z such that $s\rho_1^i Z$ and for each $z \in Z$ there exists a set X_z satisfying $z\rho_2^i X_z$ such that

$$X = \bigcup_{z \in Z} X_z;$$

DGL-: For all game terms γ_1 and γ_2 , $|\gamma_1 \cup \gamma_2|_G = (\rho^E, \rho^A)$, where

- $s\rho^E X$ if and only if $s\rho_1^E X$ or $s\rho_2^E X$, and
- $s\rho^A X$ if and only if $s\rho_1^A X$ and $s\rho_2^A X$

where, as before, $|\gamma_1|_G = (\rho_1^E, \rho_1^A)$ and $|\gamma_2|_G = (\rho_2^E, \rho_2^A)$;⁴

DGL-dual: If $|\gamma|_G = (\rho^E, \rho^A)$ then $|\gamma^d|_G = (\rho^A, \rho^E)$;

DGL- \perp : $|\perp|_G = \emptyset$;

DGL-atomic-pr: $|p|_G = V(p);$

- **DGL-**¬: $|\neg \varphi|_G = S \setminus |\varphi|_G;$
- **DGL-**V: $|\varphi_1 \vee \varphi_2|_G = |\varphi_1|_G \cup |\varphi_2|_G;$

DGL- \diamond : If $|\gamma|_G = (\rho^E, \rho^A)$ then for all φ ,

 $|\langle \gamma, i \rangle \varphi|_G = \{ s \in S : \exists X_s \subseteq |\varphi|_G \text{ s.t. } s \rho^i X_s \}.$

If $s \in |\varphi|_G$, we say that φ is *satisfied* by s in G and we write $M \models {}_s \varphi$.

We will not discuss here the properties of this logic, or the vast amount of variants and extensions of it which have been developed and studied. It is worth pointing out, however, that van Benthem et al. (2008) introduced a *Concurrent Dynamic Game Logic* that can be considered one of the main sources of inspiration for the Transition Logic that we will develop in Subsection 3.2.

 $^{^{4}}$ Van Benthem et al. (2008) give the following alternative condition for the powers of the universal player:

[•] $s\rho^A X$ if and only if $X = Z_1 \cup Z_2$ for two Z_1 and Z_2 such that $s\rho_1^A Z_1$ and $s\rho_2^A Z_2$.

It is trivial to see that, if our games satisfy the monotonicity condition, this rules is equivalent to the one we presented.

1.3 Dynamic Game Logic and First Order Logic

In this subsection, we will briefly recall a remarkable result from (van Benthem 2003) which establishes a connection between Dynamic Game Logic and First-Order Logic.

In brief, as the following two theorems demonstrate, either of these logics can be seen as a special case of the other, in the sense that models and formulas of the one can be uniformly translated into models of the other in a way which preserves satisfiability and truth:

Theorem 2. Let $G = (S, \{(\rho_g^E, \rho_g^A) : g \in \Gamma\}, V)$ be any game model, let φ be any game formula for the same language, and let $s \in S$. Then it is possible to uniformly construct a first-order model G^{FO} , a first-order formula φ^{FO} and an assignment s^{FO} of G^{FO} such that

$$G \models {}_{s}\varphi \Leftrightarrow G^{FO} \models {}_{s^{FO}}\varphi^{FO}.$$

Theorem 3. Let *M* be any first order model, let φ be any first-order formula for the signature of *M*, and let *s* be an assignment of *M*. Then it is possible to uniformly construct a game model G^{DGL} , a game formula φ^{DGL} and a state s^{DGL} such that

$$M \models {}_{s}\varphi \Leftrightarrow G^{DGL} \models {}_{s^{DGL}}\varphi^{DGL}.$$

We will not discuss here the proofs of these two results. Their *significance*, however, is something about which is necessary to spend a few words. In brief, what this back-and-forth representation between First Order Logic and Dynamic Game Logic tells us is that it is possible to understand First Order Logic as a *logic for reasoning about determined games*!

In the next sections, we will attempt to develop a similar result for the case of Dependence Logic.

2 Transition Logic

2.1 A logic for imperfect information games against nature

We will now define a variant of Dynamic Game Logic, which we will call *Transition Logic*. It deviates from the basic framework of Dynamic Game Logic in two fundamental ways:

- It considers *one-player* games against Nature, instead of *two-player games* as is usual in Dynamic Game Logic;
- 2. It allows for *uncertainty* about the initial position of the game.

Hence, Transition Logic can be seen as a *decision-theoretic logic*, rather than a *game-theoretic* one: Transition Logic formulas, as we will see, correspond to assertions about the abilities of a single agent acting under uncertainty, instead of assertions about the abilities of agents interacting with each other.

In principle, it is certainly possible to generalize the approach discussed here to multiple agents acting in situations of imperfect information, and doing so might cause interesting phenomena to surface; but for the time being, we will content ourselves with developing this formalism and discussing its connection with Dependence Logic.

Our first definition is a fairly straightforward generalization of the concept of forcing relation:

Definition 2.1 (Transition system). Let *S* be a nonempty set of *states*. A *transition system* over *S* is a nonempty relation $\theta \subseteq \text{Parts}(S) \times \text{Parts}(S)$ satisfying the following requirements:

Downwards Closure: If $(X, Y) \in \theta$ and $X' \subseteq X$ then $(X', Y) \in \theta$;

Monotonicity: If $(X, Y) \in \theta$ and $Y \subseteq Y'$ then $(X, Y') \in \theta$;

Non-creation: $(\emptyset, Y) \in \theta$ for all $Y \subseteq S$;

Non-triviality: If $X \neq \emptyset$ then $(X, \emptyset) \notin \theta$.

Informally speaking, a transition system specifies the abilities of an agent: for all $X, Y \subseteq S$ such that $(X, Y) \in \theta$, the agent has a strategy which guarantees that the output of the transition will be in Y whenever the input of the transition is in X. The four axioms which we gave capture precisely this intended meaning, as we will see.

Definition 2.2 (Decision Game). A *decision game* is a triple $\Gamma = (S, E, O)$, where S is a nonempty set of *states*, E is a nonempty set of possible *decisions* for our agent and O is an *outcome function* from $S \times E$ to Parts(S).

If $s' \in O(s, e)$, we say that s' is a *possible outcome* of s under e; if $O(s, e) = \emptyset$, we say that *e* fails on input s.

Definition 2.3 (Abilities in a decision game). Let $\Gamma = (S, E, O)$ be a decision game, and let $X, Y \subseteq S$. Then we say that Γ allows the transition $X \to Y$, and we write $\Gamma : X \to Y$, if and only if there exists a $e \in E$ such that $\emptyset \neq O(s, e) \subseteq Y$ for all $s \in X$ (that is, if and only if our agent can make a decision which guarantees that the outcome will be in Y whenever the input is in X).

Theorem 4 (Transition Systems and abilities). A set $\theta \subseteq Parts(S) \times Parts(S)$ is a transition system if and only if there exists a decision game $\Gamma = (S, E, O)$ such that

 $(X, Y) \in \theta \Leftrightarrow \Gamma : X \to Y.$

What this theorem tells us is that our notion of transition system is the correct one: it captures precisely the abilities of an agent making choices under imperfect information and attempting to guarantee that, if the initial state is in a set X, the outcome will be in a set Y.

Definition 2.4 (Trump). Let S be a nonempty set of states. A *trump* over S is a nonempty, downwards closed family of subsets of S.

Whereas a transition system describes the abilities of an agent to transition from a set of possible initial states to a set of possible terminal states, a trump describes the agent's abilities to reach *some* terminal state from a set of possible initial states:⁵

Proposition 6. Let θ be a transition system and let $Y \subseteq S \neq \emptyset$. Then $\operatorname{reach}(\theta, Y) = \{X \mid (X, Y) \in \theta\}$ forms a trump. Conversely, for any trump X over S there exists a transition system θ such that $X = \operatorname{reach}(\theta, Y)$ for any nonempty $Y \subseteq S$.

We can now define the syntax and semantics of Transition Logic:

Definition 2.5 (Transition Model). Let Φ be a set of *atomic propositional symbols* and let Θ be a set of *atomic transition symbols*. Then a *transition model* is a tuple $T = (S, \{\theta_t : t \in \Theta\}, V)$, where S is a nonempty set of states, θ_t is a transition system over S for any $t \in \Theta$, and V is a function sending each $p \in \Phi$ into a trump of S.

Definition 2.6 (Transition Logic - Syntax). Let Φ be a set of atomic propositions and let Θ be a set of atomic transitions. Then the *transition terms* and *formulas* of our language are defined respectively as

$$\begin{aligned} \tau & ::= t \mid \varphi ? \mid \tau \otimes \tau \mid \tau \cap \tau \mid \tau ; \tau \\ \varphi & ::= \tau \mid p \mid \varphi \lor \varphi \mid \varphi \land \varphi \mid \langle \tau \rangle \varphi \end{aligned}$$

where t ranges over Θ and p ranges over Φ .

Definition 2.7 (Transition Logic - Semantics). Let $T = (S, \{\theta_t : t \in \Theta\}, V)$ be a transition model, let τ be a transition term, and let $X, Y \subseteq S$. Then we say that τ allows the transition from X to Y, and we write $T \models_{X \to Y} \tau$, if and only if

TL-atomic-tr: $\tau = t$ for some $t \in \Theta$ and $(X, Y) \in \theta_t$;

TL-test: $\tau = \varphi$? for some transition formula φ such that $T \models_X \varphi$ in the sense described later in this definition, and $X \subseteq Y$;

⁵The term "trump" is taken from (Hodges 1997), who used it to describe the set of all teams which satisfy a given formula.

TL- \otimes : $\tau = \tau_1 \otimes \tau_2$, and $X = X_1 \cup X_2$ for two X_1 and X_2 such that $T \models_{X_1 \to Y} \tau_1$ and $T \models_{X_2 \to Y} \tau_2$;

TL- \cap : $\tau = \tau_1 \cap \tau_2$, $T \models_{X \to Y} \tau_1$ and $T \models_{X \to Y} \tau_2$;

TL-concat: $\tau = \tau_1; \tau_2$ and there exists a $Z \subseteq S$ such that $T \models_{X \to Z} \tau_1$ and $T \models_{Z \to Y} \tau_2$.

Analogously, let φ be a transition formula, and let $X \subseteq S$. Then we say that *X* satisfies φ , and we write $T \models_X \varphi$, if and only if

TL- \top : $\varphi = \top$;

TL-atomic-pr: $\varphi = p$ for some $p \in \Phi$ and $X \in V(p)$;

TL- \lor : $\varphi = \psi_1 \lor \psi_2$ and $T \models_X \psi_1$ or $T \models_X \psi_2$;

TL- \land : $\varphi = \psi_1 \land \psi_2, T \models_X \psi_1 \text{ and } T \models_X \psi_2;$

TL- \diamond : $\varphi = \langle \tau \rangle \psi$ and there exists a *Y* such that $T \models_{X \to Y} \tau$ and $T \models_Y \psi$.

Proposition 7. For any transition model T, transition term τ and transition formula φ , the set

$$|\tau|_T = \{(X, Y) : T \models_{X \to Y} \tau\}$$

is a transition system and the set

$$|\varphi|_T = \{X : T \models_X \varphi\}$$

is a trump.

Proof. By induction.

We end this subsection with a few simple observations about this logic.

First of all, we did not take the negation as one of the primitive connectives. Indeed, Transition Logic, much like Dependence Logic, has an intrinsically *existential* character: it can be used to reason about which sets of possible states an agent *may* reach, but not to reason about which ones such an agent *must* reach. There is of course no reason, in principle, why a negation could not be added to the language, just as there is no reason why a negation cannot be added to Dependence Logic, thus obtaining the far more powerful *Team Logic* (Väänänen 2007b, Kontinen and Nurmi 2009): however, this possible extension will not be studied in this work.

The connectives of Transition Logic are, for the most part, very similar to those of Dynamic Game Logic, and their interpretation should pose no difficulties. The exception is the *tensor operator* $\tau_1 \otimes \tau_2$, which substitutes the game union operator $\gamma_1 \cup \gamma_2$ and

which, while sharing roughly the same informal meaning, behaves in a very different way from the semantic point of view (for example, it is not in general idempotent!)

The decision game corresponding to $\tau_1 \otimes \tau_2$ can be described as follows: first the agent chooses an index $i \in \{1, 2\}$, then he or she picks a strategy for τ_i and plays accordingly. However, the choice of *i* may be a function of the initial state: hence, the agent can guarantee that the output state will be in *Y* whenever the input state is in *X* only if he or she can split *X* into two subsets X_1 and X_2 and guarantee that the state in *Y* will be reached from any state in X_1 when τ_1 is played, and from any state in X_2 when τ_2 is played.

It is also of course possible to introduce a "true" choice operator $\tau_1 \cup \tau_2$, with semantical condition

TL-U:
$$T \models_{X \to Y} \tau_1 \cup \tau_2$$
 iff $T \models_{X \to Y} \tau_1$ or $T \models_{X \to Y} \tau_2$;

but we will not explore this possibility any further in this work, nor we will consider any other possible connectives such as, for example, the iteration operator

TL-*: $T \models_{X \to Y} \tau^*$ iff there exist $n \in \mathbb{N}$ and $Z_0 \dots Z_n$ such that $Z_0 = X$, $Z_n = Y$ and $T \models_{Z_i \to Z_{i+1}} \tau$ for all $i \in 1 \dots n-1$.

2.2 Transition Logic and Dependence Logic

This subsection contains the central result of this work, that is, the analogues of Theorems 2 and 3 for Dependence Logic and Transition Logic.

Representing Dependence Logic models and formulas in Transition Logic is fairly simple:

Definition 2.8 (M^{TL}). Let M be a first-order model. Then M^{TL} is the transition model (S, Θ, V) such that

- *S* is the set of all teams over *M*;
- The set of all atomic transition symbols is $\{\exists v, \forall v : v \in Var\}$, and hence Θ is $\{\theta_{\exists v}, \theta_{\forall v} : v \in Var\}$;
- For any variable v, $\theta_{\exists v} = \{(X, Y) : \exists F \text{ s.t. } X[F/v] \subseteq Y\}$ and $\theta_{\forall v} = \{(X, Y) : X[M/v] \subseteq Y\}$;
- For any first-order literal or dependence atom α , $V(\alpha) = \{X : M \models_X \varphi\}$.

Definition 2.9 (φ^{TL}). Let φ be a Dependence Logic formula. Then φ^{TL} is the transition term defined as follows:

- 1. If φ is a literal or a dependence atom, $\varphi^{TL} = \varphi$?;
- 2. If $\varphi = \psi_1 \lor \psi_2$, $\varphi^{TL} = (\psi_1)^{TL} \otimes (\psi_2)^{TL}$;
- 3. If $\varphi = \psi_1 \wedge \psi_2$, $\varphi^{TL} = (\psi_1)^{TL} \wedge (\psi_2)^{TL}$;
- 4. If $\varphi = \exists v \psi, \varphi^{TL} = \exists v; (\psi)^{TL};$
- 5. If $\varphi = \forall v \psi, \varphi^{TL} = \forall v; (\psi)^{TL}$.

Theorem 5. For all first-order models M, teams X and formulas φ , the following are equivalent:

- $M \models_X \varphi;$
- $\exists Y \ s.t. \ M^{TL} \models_{X \to Y} \varphi^{TL};$
- $M^{TL} \models_X \langle \varphi^{TL} \rangle \top;$
- $M^{TL} \models_{X \to S} \varphi^{TL}$.

Proof. By structural induction on φ .

One interesting aspect of this representation result is that Dependence Logic *formulas* correspond to Transition Logic *transitions*, not to Transition Logic *formulas*. This can be thought of as one first hint of the fact that Dependence Logic can be thought of as a logic of transitions: and in the later sections, we will explore this idea more in depth.

Representing Transition Models, game terms and formulas in Dependence Logic is somewhat more complex:

Definition 2.10 (T^{DL}) . Let $T = (S, (\theta_t : t \in \Theta), V)$ be a transition model. Furthermore, for any $t \in \Theta$, let $\theta_t = \{(X_i, Y_i) : i \in I_t\}$, and, for any $p \in \Phi$, let $V(p) = \{X_j : j \in J_p\}$. Then T^{DL} is the first-order model with domain⁶ $S \uplus \biguplus \{I_t : t \in \Theta\} \uplus \biguplus \{J_p : p \in \Phi\}$ whose signature contains

- For every $t \in \Theta$, a ternary relation R_t whose interpretation is $\{(i, x, y) : i \in I_t, x \in X_i, y \in Y_i\}$;
- For every $p \in \Phi$, a binary relation V_p whose interpretation is $\{(j, x) : j \in J_p, x \in X_j\}$.

⁶Here we write $A \uplus B$ for the *disjoint union* of the sets A and B.

Definition 2.11 (φ_x^{DL} and τ_x^{DL}). For any transition formula φ and variable *x*, the Dependence Logic formula φ_x^{DL} is defined as

- 1. \top_{x}^{DL} is \top ;
- 2. For all $p \in \Phi$, p_x^{DL} is $\exists j (=(j) \land V_p(j, x))$;
- 3. $(\psi_1 \lor \psi_2)_x^{DL}$ is $(\psi_1)_x^{DL} \sqcup (\psi_2)^{DL}$, where \sqcup is the classical disjunction introduced in Definition 1.6;
- 4. $(\psi_1 \wedge \psi_2)_r^{DL}$ is $(\psi_1)_r^{DL} \wedge (\psi_2)_r^{DL}$;

5.
$$(\langle \tau \rangle \psi)_x^{DL}$$
 is $\exists P((\tau)_x^{DL}(P) \land \forall y(\neg Py \lor (\psi)_y^{DL})),$

where for any transition term τ , variable x and unary relation symbol P, $\tau_x^{DL}(P)$ is defined as

- 6. For all $t \in \Theta$, $t_x^{DL}(P)$ is $\exists i (=(i) \land \exists y (R_t(i, x, y)) \land \forall y (\neg R_t(i, x, y) \lor Py));$
- 7. For all formulas φ , $(\varphi?)_x^{DL}(P)$ is $\varphi_x^{DL} \wedge Px$;
- 8. $(\tau_1 \otimes \tau_2)^{DL}_x(P) = (\tau_1)^{DL}_x(P) \lor (\tau_2)^{DL}_x(P);$

9.
$$(\tau_1 \cap \tau_2)_x^{DL}(P) = (\tau_1)_x^{DL}(P) \land (\tau_2)_x^{DL}(P);$$

10. $(\tau_1; \tau_2)_x^{DL}(P) = \exists Q((\tau_1)_x^{DL}(Q) \land \forall y(\neg Qy \lor (\tau_2)_y^{DL}(P)))$ for a new and unused variable y.

Theorem 6. For all transition models $T = (S, (\theta_t : t \in \Theta), V)$, transition terms τ , transition formulas φ , variables x, sets $P \subseteq S$ and teams X over T^{DL} with $X(x) \subseteq S$,⁷

$$T^{DL} \models_X \varphi_x^{DL} \Leftrightarrow T \models_{X(x)} \varphi$$

and

$$T^{DL} \models_X \tau_x^{DL}(P) \Leftrightarrow T \models_{X(x) \to P} \tau.$$

Proof. The proof is by structural induction on terms and formulas. We show the two most interesting cases:

⁷That is, such that X(x) is a set of states of the transition model.

5. $T^{DL} \models_X(\langle \tau \rangle \psi)_x^{DL}$ if and only if there exists a *P* such that $T^{DL} \models_X(\tau)_x^{DL}(P)$ and $T^{DL} \models_{X[T^{DL}/y]} \neg Py \lor (\psi)_y^{DL}$. By induction hypothesis, the first condition holds if and only if $T \models_{X(x) \to P} \tau$. As for the second one, it holds if and only if $X[T^{DL}/y] = Y_1 \cup Y_2$ for two Y_1, Y_2 such that $T^{DL} \models_{Y_1} \neg Py$ and $T^{DL} \models_{Y_2} \tau_y(\psi)$. But then we must have that $T \models_{Y_2(y)} \psi$ and that $P \subseteq Y_2(y)$; therefore, by downwards closure, $T \models_P \psi$ and finally $T \models_{X(x)} \langle \tau \rangle \psi$.

Conversely, suppose that there exists a *P* such that $T \models_{X(x)\to P} \tau$ and $T \models_{P} \psi$; then by induction hypothesis we have that $T^{DL} \models_{X}(\tau)_{x}^{DL}(P)$ and that $T^{DL} \models_{X[T^{DL}/y]} \neg Py \lor (\psi)_{x}^{DL}$, and hence $T^{DL} \models_{X}(\langle \tau \rangle \psi)_{x}^{DL}$.

10. $T^{DL} \models_X \exists Q((\tau_1)_x^{DL}(Q) \land \forall y(\neg Qy \lor (\tau_2)_y^{DL}(P)))$ if and only if there exists a Q such that $T \models_{X(x) \to Q} \tau_1$ and there exists a $Q' \supseteq Q$ such that $T \models_{Q' \to P} \tau_2$. By downwards closure, if this is the case then $T \models_{Q \to P} \tau_2$ too, and hence $T \models_{X(x) \to P} \tau_1; \tau_2$, as required.

Conversely, suppose that there exists a Q such that $T \models_{X(x) \to Q} \tau_1$ and $T \models_{Q \to P} \tau_2$. Then, by induction hypothesis $T^{DL} \models_X(\tau_1)_x^{DL}(Q)$; and furthermore, $X[T^{DL}/y]$ can be split into

$$Z_1 = \{s \in X[T^{DL}/y] : s(y) \notin Q\}$$

and

$$Z_2 = \{s \in X[T^{DL}/y] : s(y) \in Q\}$$

It is trivial to see that $T^{DL} \models_{Z_1} \neg Qy$; and furthermore, since $Z_2(y) = Q$ and $T \models_{Q \rightarrow P} \tau_2$, by induction hypothesis we have that $T^{DL} \models_{Z_2}(\tau_2)_y^{DL}$. Thus $T^{DL} \models_{X[T^{DL}/y]} \forall y(\neg Qy \lor (\tau_2)_y^{DL}(P))$ and finally $T^{DL} \models_X(\tau_1;\tau_2)_x^{DL}(P)$, and this concludes the proof.

Hence, the relationship between Transition Logic and Dependence Logic is analogous to the one between Dynamic Game Logic and First-Order Logic. In the next sections, we will develop variants of Dependence Logic which are syntactically closer to Transition Logic, while still being first-order: as we will see, the resulting frameworks are expressively equivalent to Dependence Logic on the level of satisfiability, but can be used to represent finer-grained phenomena of *transitions* between sets of assignments.

3 Dynamic variants of Dependence Logic

3.1 Dependence Logic and transitions between teams

Now that we have established a connection between Dependence Logic and a variant of Dynamic Game Logic, it is time to explore what this might imply for the further development of logics of imperfect information. If, as Theorems 5 and 6 suggest, Dependence Logic can be thought of as a logic of imperfect-information decision problems, perhaps it could be possible to develop variants of Dependence Logic in which expressions can be interpreted directly as transition systems?

In what follows, we will do exactly that, first with *Transition Dependence Logic* – a variant of Dependence Logic, expressively equivalent to it, which is also a quantified version of Transition Logic – and then with *Dynamic Dependence Logic*, in which *all* expressions are interpreted as transitions!

But why would we interested in such variants of Dependence Logic? One possible answer, which we will discuss in this subsection, is that transitions between teams are *already* a central object of study in the field of Dependence Logic, albeit in a nonexplicit manner: after all, the semantics of Dependence Logic interprets quantifiers in terms of transformations of teams, and disjunctions in terms of decompositions of teams into subteams. This intuition is central to the study of issues of interdefinability in Dependence Logic and its variants, like for example the ones discussed in (Galliani 2012). As a simple example, let us recall Definition 1.6:

$$\psi_1 \sqcup \psi_2 := \exists u_1 \exists u_2 (=(u_1) \land =(u_2) \land ((u_1 = u_2 \land \psi_1) \lor (u_1 \neq u_2 \land \psi_2))),$$

where u_1 and u_2 are new variables.

As we said in Proposition 1, $M \models_X \psi_1 \sqcup \psi_2$ if and only if $M \models_X \psi_1$ or $M \models_X \psi_2$. We will now sketch the proof of this result, and – as we will see – this proof will hinge on the fact that the above expression can be read as a specification of the following algorithm:

- 1. Choose an element $a \in Dom(M)$ and extend the team X by assigning a as the value of u_1 for all assignments;
- 2. Choose an element $b \in Dom(M)$ and further extend the team by assigning b as the value of u_2 for all assignments;
- 3. Split the resulting team into two subteams Y_1 and Y_2 such that
 - (a) ψ_1 holds in Y_1 , and the values of u_1 and u_2 coincide for all assignments in it;
(b) ψ_2 holds in Y_2 , and the values of u_1 and u_2 differ for all assignments in it.

Since the values of u_1 and u_2 are chosen to always be respectively a and b, one of Y_1 and Y_2 is empty and the other is of the form $X[ab/u_1u_2]$, and since u_1 and u_2 do not occur in ψ_1 or ψ_2 the above algorithm can succeed (for some choice of a and b) only if $M \models_X \psi_1$ or $M \models_X \psi_2$.

3.2 Transition Dependence Logic

As stated, we will now define a variant of Dependence Logic which can also be seen as a quantified variant of Transition Logic. We will then prove that the resulting Transition Dependence Logic is expressively equivalent to Dependence Logic, in the sense that any Dependence Logic formula is equivalent to some Transition Dependence Logic formula and vice versa.

Definition 3.1 (Transition Dependence Logic - Syntax). Let Σ be a first-order signature. Then the sets of all *transition terms* and of all *formulas* of Dependence Transition Logic are given by the rules

$$\begin{aligned} \tau & ::= \quad \exists v \mid \forall v \mid \varphi? \mid \tau \otimes \tau \mid \tau \cap \tau \mid \tau; \tau \\ \varphi & ::= \quad R\vec{t} \mid \neg R\vec{t} \mid = (t_1, \dots, t_n) \mid \varphi \lor \varphi \mid \varphi \land \varphi \mid \langle \tau \rangle \varphi. \end{aligned}$$

where *v* ranges over all variables in Var, *R* ranges over all relation symbols of the signature, \vec{t} ranges over all tuples of terms of the required arities, *n* ranges over \mathbb{N} and $t_1 \dots t_n$ range over the terms of our signature.

Definition 3.2 (Transition Dependence Logic - Semantics). Let *M* be a first-order model, let τ be a first-order transition term of the same signature, and let *X* and *Y* be teams over *M*. Then we say that the transition $X \to Y$ is *allowed* by τ in *M*, and we write $M \models_{X \to Y} \tau$, if and only if

- **TDL-**: τ is of the form $\exists v$ for some $v \in Var$ and there exists a *F* such that $X[F/v] \subseteq Y$;
- **TDL-** \forall : τ is of the form $\forall v$ for some $v \in Var$ and $X[M/v] \subseteq Y$;
- **TDL-test:** τ is of the form φ ?, $M \models_X \varphi$ in the sense given later in this definition, and $X \subseteq Y$;
- **TDL-** \otimes : τ is of the form $\tau_1 \otimes \tau_2$ and $X = X_1 \cup X_2$ for some X_1 and X_2 such that $M \models_{X_1 \to Y} \tau_1$ and $M \models_{X_2 \to Y} \tau_2$;

- **TDL-** \cap : τ is of the form $\tau_1 \cap \tau_2$, $M \models_{X \to Y} \tau_1$ and $M \models_{X \to Y} \tau_2$;
- **TDL-concat:** τ is of the form τ_1 ; τ_2 and there exists a team Z such that $M \models_{X \to Z} \tau_1$ and $M \models_{Z \to Y} \tau_2$.

Similarly, if φ is a formula and X is a team with domain Var. Then we say that X satisfies φ in M, and we write $M \models_X \varphi$, if and only if

- **TDL-lit:** φ is a first-order literal and $M \models {}_{s}\varphi$ in the usual first-order sense for all $s \in X$;
- **TDL-dep:** φ is a dependence atom = (t_1, \ldots, t_n) and any two $s, s' \in X$ which assign the same values to $t_1 \ldots t_{n-1}$ also assign the same value to t_n ;
- **TDL-**V: φ is of the form $\varphi_1 \lor \varphi_2$ and $M \models_X \varphi_1$ or $M \models_X \varphi_2$;

TDL- \wedge : φ is of the form $\varphi_1 \land \varphi_2$, $M \models_X \varphi_1$ and $M \models_X \varphi_2$;

TDL-: φ is of the form $\langle \tau \rangle \psi$ and there exists a *Y* such that $M \models_{X \to Y} \tau$ and $M \models_Y \psi$.

As the next theorem shows, in this semantics formulas and transitions are interpreted in terms of trumps and transition systems:

Theorem 7. For all Transition Dependence Logic formulas φ , all models M and all teams X and Y, we have that

Downwards Closure: If $M \models_X \varphi$ and $Y \subseteq X$ then $M \models_Y \varphi$;

Empty Team Property: $M \models_{\emptyset} \varphi$.

Furthermore, for all Transition Dependence Logic transition terms τ , all models M and all teams X, Y and Z,

Downwards Closure: If $M \models_{X \to Y} \tau$ and $Z \subseteq X$ then $M \models_{Z \to Y} \tau$;

Monotonicity: If $M \models_{X \to Y} \tau$ and $Y \subseteq Z$ then $M \models_{X \to Z} \tau$;

Non-creation: For all *Y*, $M \models_{\emptyset \to Y} \tau$;

Non-triviality: If $X \neq \emptyset$ then it is not the case that $M \models_{X \to \emptyset} \tau$.

Proof. The proof is by structural induction over φ and τ .

Also, it is not difficult to see, on the basis of the results of the previous section, that this new variant of Dependence Logic is equivalent to the usual one:

Theorem 8. For every Dependence Logic formula φ there exists a Transition Dependence Logic transition term τ_{φ} such that

$$M \models_X \varphi \Leftrightarrow \exists Y \text{ s.t. } M \models_{X \to Y} \tau_{\varphi} \Leftrightarrow M \models_X \langle \tau_{\varphi} \rangle \top$$

for all first-order models M and teams X.

Proof. τ_{φ} is defined by structural induction on φ , as follows:

- 1. If φ is a first-order literal or a dependence atom then $\tau_{\varphi} = \varphi$?;
- 2. If φ is $\varphi_1 \lor \varphi_2$ then $\tau_{\varphi} = \tau_{\varphi_1} \otimes \tau_{\varphi_2}$;
- 3. If φ is $\varphi_1 \land \varphi_2$ then $\tau_{\varphi} = \tau_{\varphi_1} \cap \tau_{\varphi_2}$;
- 4. If φ is $\exists v \psi$ then $\tau_{\varphi} = \exists v; \tau_{\psi};$
- 5. If φ is $\forall v \psi$ then $\tau_{\varphi} = \forall v; \tau_{\psi}$.

It is then trivial to verify, again by induction on φ , that $M \models {}_X\varphi$ if and only if $M \models {}_X\langle \tau_{\varphi} \rangle \top$, as required.

This representation result associates Dependence Logic *formulas* to Transition Dependence Logic *transition terms*. This fact highlights the dynamical nature of Dependence Logic operators, which we discussed in the previous subsection: in this framework, quantifiers describe *transformations* of teams, the Dependence Logic connectives are operations over games, and the literals are interpreted as tests. In fact, one might wonder what is the purpose of Transition Dependence Logic formulas: could we do away with them altogether, and develop a variant of Transition Dependence Logic in which *all* formulas are transitions?

Later, we will explore this idea further; but first, let us verify that Transition Dependence Logic is no more expressive than Dependence Logic.

Theorem 9. For every Transition Dependence Logic formula φ there exists a Dependence Logic formula $T(\varphi)$ such that

$$M \models_X \varphi \Leftrightarrow M \models_X T(\varphi)$$

for all first-order models M and teams X. Furthermore, for every Transition Dependence Logic transition term τ and Dependence Logic formula θ there is a Dependence Logic formula $U(\tau, \psi)$ such that

 $M \models_X U(\tau, \theta) \Leftrightarrow \exists Y \text{ s.t. } M \models_{X \to Y} \tau \text{ and } M \models_Y \theta$,

again for all first-order models M and teams X.

Proof. By structural induction over φ and τ .

However, in a sense, Transition Dependence Logic allows one to consider subtler distinctions than Dependence Logic does. The formula $\forall x \exists y (= (y, f(x)) \land Pxy)$, for example, could be translated as any of

- $\langle \forall x; \exists y \rangle (=(y, f(x)) \land Pxy);$
- $\langle \forall x; \exists y \rangle \langle =(y, f(x))? \rangle Pxy;$
- $\langle \forall x; \exists y \rangle \langle Pxy? \rangle = (y, f(x));$
- $\langle \forall x; \exists y \rangle \langle (Pxy?) \cap (=(y, f(x))?) \rangle \top$.

The intended interpretations of these formulas are rather different, even though they happen to be satisfied by the same teams: and for this reason, Transition Dependence Logic may be thought of as a proper refinement of Dependence Logic even though it has exactly the same expressive power.

3.3 Dynamic Predicate Logic

Dynamic Semantics is the name given to a family of semantical frameworks which subscribe to the following principle (Groenendijk and Stokhof 1991):

The meaning of a sentence does not lie in its truth conditions, but rather in the way it changes (the representation of) the information of the interpreter.

In various forms, this intuition can be found prefigured in some of the later work of Ludwig Wittgenstein, as well as in the research of philosophers of language such as Austin, Grice, Searle, Strawson and others (Dekker 2008); but its formal development can be traced back to the work of Groenendijk and Stokhof about the proper treatment of pronouns in formal linguistics (Groenendijk and Stokhof 1991).

We will not discuss here the formal definition of Groenendijk and Stockhof's formalism or its linguistic applications. All that is relevant for our purposes is that, according to it, formulas are interpreted as *transitions* from assignments to assignments: for example, the satisfaction conditions for conjunctions are given by

DPL- \land : φ is of the form $\psi_1 \land \psi_2$ and there exists an *h* such that $M \models_{s \to h} \psi_1$ and $M \models_{h \to s'} \psi_2$;

The similarity between this semantics and our semantics for transition terms should be evident. Hence, it seems natural to ask whether we can adopt, for a suitable variant of Dependence Logic, the following variant of Groenendijk and Stokhof's motto:

The meaning of a formula does not lie in its satisfaction conditions, but rather in the team transitions it allows.

In the next section, we will make use of this intuition to develop another, terser version of Dependence Logic; and finally, we will discuss some implications of this new version for the further developments and for the possible applications of this interesting logical formalism.

3.4 Dynamic Dependence Logic

We will now develop a formula-free variant of Transition Dependence Logic, along the lines of Groenendijk and Stockhof's Dynamic Predicate Logic.

Definition 3.3 (Dynamic Dependence Logic - Syntax). Let Σ be a first-order signature. The set of all formulas of Dynamic Dependence Logic over Σ is given by the rules

 $\tau ::= R\vec{t} \mid \neg R\vec{t} \mid = (t_1, \dots, t_n) \mid \exists v \mid \forall v \mid \tau \otimes \tau \mid \tau \cap \tau \mid \tau; \tau$

where, as usual, *R* ranges over all relation symbols of our signature, \vec{t} ranges over all tuples of terms of the required lengths, *n* ranges over \mathbb{N} , $t_1 \dots t_n$ range over all terms, and *v* ranges over \mathbb{Var} .

The semantical rules associated to this language are precisely as one would expect:

Definition 3.4 (Dynamic Dependence Logic - Semantics). Let M be a first-order model, let τ be a Dynamic Dependence Logic formula over the signature of M, and let X and Y be two teams over M with domain Var. Then we say that τ allows the transition $X \to Y$ in M, and we write $M \models_{X \to Y} \tau$, if and only if

- **DDL-lit:** τ is a first-order literal, $M \models {}_{s}\tau$ in the usual first-order sense for all $s \in X$, and $X \subseteq Y$;
- **DDL-dep:** τ is a dependence atom = (t_1, \ldots, t_n) , $X \subseteq Y$, and any two assignments $s, s' \in X$ which coincide over $t_1 \ldots t_{n-1}$ also coincide over t_n ;
- **DDL-** \exists : τ is of the form $\exists v$ for some $v \in Var$, and $X[F/v] \subseteq Y$ for some $F : X \to Dom(M)$;

DDL- \forall : τ is of the form $\forall v$ for some $v \in Var$, and $X[M/v] \subseteq Y$;

- **DDL-** \otimes : τ is of the form $\tau_1 \otimes \tau_2$ and $X = X_1 \cup X_2$ for two teams X_1 and X_2 such that $M \models_{X_1 \to Y} \tau_1$ and $M \models_{X_2 \to Y} \tau_2$;
- **DDL-** \cap : τ is of the form $\tau_1 \cap \tau_2$, $M \models_{X \to Y} \tau_1$ and $M \models_{X \to Y} \tau_2$;
- **DDL-concat:** τ is of the form τ_1 ; τ_2 , and there exists a Z such that $M \models_{X \to Z} \tau_1$ and $M \models_{Z \to Y} \tau_2$.

A formula τ is said to be *satisfied* by a team X in a model M if and only if there exists a Y such that $M \models_{X \to Y} \tau$; and if this is the case, we will write $M \models_X \tau$.

It is not difficult to see that Dynamic Dependence Logic is equivalent to Transition Dependence Logic (and, therefore, to Dependence Logic).

Proposition 8. Let φ be a Dependence Logic formula. Then there exists a Dynamic Dependence Logic formula φ' which is equivalent to it, in the sense that

$$M \models_X \varphi \Leftrightarrow M \models_X \varphi' \Leftrightarrow \exists Y \text{ s.t. } M \models_{X \to Y} \varphi'$$

for all suitable teams X and models M

Proposition 9. Let τ be a Dynamic Dependence Logic formula. Then there exists a Transition Dependence Logic transition term τ' such that

$$M \models_{X \to Y} \tau \Leftrightarrow M \models_{X \to Y} \tau'$$

for all suitable X, Y and M, and such that hence

$$M \models_X \tau \Leftrightarrow M \models_X \langle \tau' \rangle \top$$

Corollary 1. Dynamic Dependence Logic is equivalent to Transition Dependence Logic and to Dependence Logic

Proof. Follows from the two previous results and from the equivalence between Dependence Logic and Transition Dependence Logic.

4 Further work

In this work, we established a connection between a variant of Dynamic Game Logic and Dependence Logic, and we used it as the basis for the development of variants of Dependence Logic in which it is possible to talk directly about transitions from teams to teams. This suggests a new perspective on Dependence Logic and Team Semantics, one which allow us to study them as a special kind of *algebras of nondeterministic transitions between relations*. One of the main problems that is now open is whether it is possible to axiomatize these algebras, in the same sense in which Allen Mann (2009) offers an axiomatization of the algebra of trumps corresponding to IF Logic (or, equivalently, to Dependence Logic).

Furthermore, we might want to consider different choices of connectives, like for example ones related to the theory of database transactions. The investigation of the relationships between the resulting formalisms is a natural continuation of the currently ongoing work on the study of the relationship between various extensions of Dependence Logic, and promises of being of great utility for the further development of this fascinating line of research.

Acknowledgements The author wishes to thank Johan van Benthem and Jouko Väänänen for a number of useful suggestions and insights. Furthermore, he wishes to thank the reviewers for a number of highly useful suggestions and comments.

References

J. van Benthem. Logic games are complete for game logics. *Studia Logica*, 75: 183–203, 2003.

J. van Benthem, S. Ghosh, and F. Liu. Modelling simultaneous games in dynamic logic. *Synthese*, 165:247–268, 2008.

P. Dekker. A guide to dynamic semantics. *ILLC Prepublication Series*, (PP-2008-42), 2008.

F. Engström. Generalized quantifiers in dependence logic. *Journal of Logic, Language and Information*, 21(3):299–324, 2012.

P. Galliani. Multivalued dependence logic and independence logic. In *Non-classical Modal and Predicate Logics*, 2011.

P. Galliani. Inclusion and exclusion dependencies in team semantics: On some logics of imperfect information. *Annals of Pure and Applied Logic*, 163(1):68 – 84, 2012.

E. Grädel and J. Väänänen. Dependence and independence. *Studia Logica*, 101(2): 399–410, 2013.

J. Groenendijk and M. Stokhof. Dynamic Predicate Logic. *Linguistics and Philosophy*, 14(1):39–100, 1991.

W. Hodges. Compositional Semantics for a Language of Imperfect Information. *Journal of the Interest Group in Pure and Applied Logics*, 5 (4):539–563, 1997.

J. Kontinen and V. Nurmi. Team logic and second-order logic. In H. Ono, M. Kanazawa, and R. de Queiroz, editors, *Logic, Language, Information and Computation*, volume 5514 of *Lecture Notes in Computer Science*, pages 230–241. Springer Berlin / Heidelberg, 2009.

J. Kontinen and J. Väänänen. On definability in dependence logic. *Journal of Logic, Language and Information*, 3(18):317–332, 2009.

A. L. Mann. Independence-friendly cylindric set algebras. *Logic Journal of IGPL*, 17 (6):719–754, 2009.

R. Parikh. The logic of games and its applications. In *Selected papers of the international conference on "foundations of computation theory" on Topics in the theory of computation*, pages 111–139, New York, NY, USA, 1985. Elsevier North-Holland, Inc.

J. Väänänen. Dependence Logic. Cambridge University Press, 2007a.

J. Väänänen. Team Logic. In J. van Benthem, D. Gabbay, and B. Löwe, editors, *Interactive Logic. Selected Papers from the 7th Augustus de Morgan Workshop*, pages 281–302. Amsterdam University Press, 2007b.

Dynamic Measure Logic

Tamar Lando

Columbia University tal2108@columbia.edu

Abstract

This paper brings together Dana Scott's measure-based semantics for the propositional modal logic S4, and recent work in Dynamic Topological Logic. In a series of recent talks, Scott showed that the language of S4 can be interpreted in the Lebesgue measure algebra, \mathcal{M} , or algebra of Borel subsets of the real interval, [0, 1], modulo sets of measure zero. Conjunctions, disjunctions and negations are interpreted via the Boolean structure of the algebra, and we add an interior operator on \mathcal{M} that interprets the \Box -modality. In this paper we show how to extend this measure-based semantics to the bimodal logic of S4C. S4C is a dynamic topological logic that is interpreted in 'dynamic topological systems,' or topological spaces together with a continuous function acting on the space. We extend Scott's measure based semantics to this bimodal logic by defining a class of operators on the algebra \mathcal{M} , which we call O-operators and which take the place of continuous functions in the topological semantics for S4C. The main result of the paper is that S4C is complete for the Lebesgue measure algebra. A strengthening of this result, also proved here, is that there is a *single* measure-based model in which all non-theorems of S4C are refuted.

1 Introduction

Kripke models for normal modal logics, consisting of a set of possible worlds together with a binary accessibility relation, are, by now, widely familiar. But long before Kripke semantics became standard, Tarski showed that the propositional modal logic S4 can be interpreted in topological spaces. In the topological semantics for S4, a

topological space is fixed, and each propositional variable, p, is assigned an arbitrary subset of the space: the set of points where p is *true*. Conjunctions, disjunctions and negations are interpreted as set-theoretic intersections, unions and complements (thus, e.g., $\varphi \land \psi$ is *true* at all points in the intersection of the set of points where φ is true and the set of points where ψ is true. The \Box -modality of *S*4 is interpreted via the topological interior: $\Box \varphi$ is true at any point in the topological interior of the set of points at which φ is true.

In this semantics, the logic S4 can be seen as describing topological spaces. Indeed, with the topological semantics it became possible to ask not just whether S4is complete for the set of topological validities—formulas valid in *every* topological space—but also whether S4 is complete for any *given* topological space. The culmination of Tarski's work in this area was a very strong completeness result. In 1944, Tarski and McKinsey proved that S4 is complete for any dense-in-itself metric space. One particularly important case was the real line, \mathbb{R} , and as the topological semantics received renewed interest in recent years, more streamlined proofs of Tarski's result for this special case emerged (in, e.g., Aiello 2003, Bezhanishvili and Gehrke 2005, Mints and Zhang 2005).

The real line, however, can be investigated not just from a topological point of view, but from a measure-theoretic point of view. Here, the probability measure we have in mind is the usual Lebesgue measure on the reals. In the last several years Dana Scott introduced a new *probabilistic* or *measure-based* semantics for *S*4 that is built around Lebesgue measure on the reals and is in some ways closely related to Tarski's older topological semantics (see Scott 2009).

Scott's semantics is essentially algebraic: formulas are interpreted in the Lebesgue measure algebra, or the σ -algebra of Borel subsets of the real interval [0,1], modulo sets of measure zero (henceforth, "null sets"). We denote this algebra by \mathcal{M} . Thus elements of \mathcal{M} are equivalence classes of Borel sets. In Scott's semantics, each propositional variable is assigned to some element of \mathcal{M} . Conjunctions, disjunctions and negations are assigned to meets, joins and complements in the algebra, respectively. In order to interpret the $S4 \square$ -modality, we add to the algebra an "interior" operator (defined below), which we construct from the collection of *open* elements in the algebra, or elements that have an open representative. Unlike the Kripke or topological semantics, there is no notion here of truth *at a point* (or at a "world"). In (Fernandez-Duque 2010) and (Lando 2012) it was shown that S4 is complete for the L ebesgue measure algebra.

The introduction of a measure-based semantics for S4 raises a host of questions that are, at this point, entirely unexplored. Among them: What about natural extensions of S4? Can we give a measure-based semantics not just for S4 but for some of its extensions that have well-known topological interpretations?

Lando

This paper focuses on a family of logics called *dynamic topological logics*. These logics were investigated over the last fifteen years, in an attempt to describe "dynamic topological systems" by means of modal logic. A dynamic topological system is a pair $\langle X, f \rangle$, where X is a topological space and f is a continuous function on X. We can think of f as moving points in X in discrete units of time. Thus in the first moment, x is mapped to f(x), then to f(f(x)), etc. The most basic dynamic topological logic is S4C. In addition to the $S4 \square$ -modality, it has a temporal modality, which we denote by \bigcirc . Intuitively, we understand the formula $\bigcirc p$ as saying that at the "next moment in time," p will be true. Thus we put: $x \in V(\bigcirc p)$ iff $f(x) \in V(p)$. In (Kremer and Mints 2005) and (Slavnov 2003) it was shown that S4C is complete for the real line, \mathbb{R} . However, in (Slavnov 2005) it was shown that S4C is complete for Euclidean spaces of arbitrarily large finite dimension, and in (Fernandez-Duque 2005) it was shown that S4C is complete for \mathbb{R}^2 .

The aim of this paper is to give a measure-based semantics for the logic S4C, along the lines of Scott's semantics for S4. Again, formulas will be assigned to some element of the Lebesgue measure algebra, \mathcal{M} . But what about the *dynamical* aspect i.e., the interpretation of the \bigcirc -modality? We show that there is a very natural way of interpreting the \bigcirc -modality via operators on the algebra \mathcal{M} that take the place of continuous functions in the topological semantics. These operators can be viewed as transforming the algebra in discrete units of time. Thus one element is sent to another in the first instance, then to another in the second instance, and so on. The operators we use to interpret S4C are O-operators: ones that take "open" elements in the algebra to open elements (defined below). But there are obvious extensions of this idea: for example, to interpret the logic of *homeomorphisms* on topological spaces, one need only look at *automorphisms* of the algebra \mathcal{M} .

Adopting a measure-based semantics for S4C brings with it certain advantages. Not only do we reap the probabilistic features that come with Scott's semtantics for S4, but the curious *dimensional asymmetry* that appears in the topological semantics (where S4C is *in*complete for \mathbb{R} but complete for \mathbb{R}^2) disappears in the measure-based semantics. Our main result is that the logic S4C is complete for the Lebesgue-measure algebra. A strengthening of this result, also proved here, is that S4C is complete for a *single* model of the Lebesgue measure algebra. Due to well-known results by Oxtoby, this algebra is isomorphic to the algebra generated by Euclidean space of arbitrary dimension. In other words, S4C is complete for the reduced measure algebra generated by *any* Euclidean space.

2 Topological semantics for *S*4*C*

Let the language $L_{\Box,\bigcirc}$ consist of a countable set, $PV = \{p_n | n \in \mathbb{N}\}$, of propositional variables, and be closed under the binary connectives $\land, \lor, \rightarrow, \leftrightarrow$, unary operators, \neg, \Box, \diamond , and a unary modal operator \bigcirc (thus, $L_{\Box,\bigcirc}$ is the language of propositional *S*4 enriched with a new modality, \bigcirc).

Definition 2.1. A **dynamic topological space** is a pair $\langle X, f \rangle$, where X is a topological space and $f : X \to X$ is a continuous function on X. A **dynamic topological model** is a triple, $\langle X, f, V \rangle$, where X is a topological space, $f : X \to X$ is a continuous function, and $V : PV \to \mathcal{P}(X)$ is a valuation assigning to each propositional variable a subset of X. We say that $\langle X, f, V \rangle$ is a model over X.

We extend V to the set of all formulas in $L_{\Box,\bigcirc}$ by means of the following recursive clauses:

 $V(\varphi \lor \psi) = V(\varphi) \cup V(\psi)$ $V(\neg \varphi) = X - V(\varphi)$ $V(\Box \varphi) = Int (V(\varphi))$ $V(\bigcirc \varphi) = f^{-1}(V(\varphi))$

where 'Int' denotes the topological interior.

Let $N = \langle X, f, V \rangle$ be a dynamic topological model. We say that a formula φ is *satisfied* at a point $x \in X$ if $x \in V(\varphi)$, and we write $N, x \models \varphi$. We say φ is true in N $(N \models \varphi)$ if $N, x \models \varphi$ for each $x \in X$. We say φ is *valid* in $X (\models x\varphi)$, if for any model N over X, we have $N \models \varphi$. Finally, we say φ is *topologically valid* if it is valid in every topological space.

Definition 2.2. The logic S4C in the language $L_{\Box,\bigcirc}$ is given by the following axioms:

- the classical tautologies,
- S4 axioms for \Box .
- (A1) $\bigcirc (\varphi \lor \psi) \leftrightarrow (\bigcirc \varphi \lor \bigcirc \psi)$,
- (A2) $(\bigcirc \neg \varphi) \leftrightarrow (\neg \bigcirc \varphi)$,
- (A3) $\bigcirc \Box \varphi \rightarrow \Box \bigcirc \varphi$ (the axiom of continuity),

and the rules of modus ponens and necessitation for both \Box and \bigcirc . Following (Kremer and Mints 2005), we use S4C both for this axiomatization and for the set of all formulas derivable from the axioms by the inference rules.

We close this section by listing the known completeness results for S4C in the topological semantics.

Theorem 2.3. (*Completeness*) For any formula $\varphi \in L_{\Box,\bigcirc}$, the following are equivalent: (i) $S 4C \vdash \varphi$; (ii) φ is topologically valid; (iii) φ is true in any finite topological space;

(iv) φ is valid in \mathbb{R}^n for $n \ge 2$.

Proof. The equivalence of (i)-(iii) was proved in (Artemov et al. 1997). The equivalence of (i) and (iv) was proved in (Fernandez-Duque 2005). This was a strengthening of a result proved in (Slavnov 2005).

Theorem 2.4. (Incompleteness for \mathbb{R}) There exists $\varphi \in L_{\Box,\bigcirc}$ such that φ is valid in \mathbb{R} , but φ is not topologically valid.

Proof. See (Kremer and Mints 2005) and (Slavnov 2003).

3 Kripke semantics for *S*4*C*

In this section we show that the logic S4C can also be interpreted in the more familiar setting of Kripke frames. It is well known that the logic S4 (which does not include the 'temporal' modality, \bigcirc) is interpreted in transitive, reflexive Kripke frames, and that such frames *just are* topological spaces of a certain kind. It follows that the Kripke semantics for S4 is just a special case of the topological semantics for S4. In this section, we show that the logic S4C can be interpreted in transitive, reflexive Kripke frames with some additional 'dynamic' structure, and, again, that Kripke semantics for S4C is a special case of the more general topological semantics for S4C. Henceforth, we assume that Kripke frames are both transitive and reflexive.

Definition 3.1. A dynamic Kripke frame is a triple $\langle W, R, G \rangle$ where W is a set, R is a reflexive, transitive relation on W, and $G : W \rightarrow W$ is a function that is R-monotone in the following sense: for any $u, v \in W$, if uRv, then G(u) R G(v).

Definition 3.2. A dynamic Kripke model is a pair $\langle F, V \rangle$ where $F = \langle W, R, G \rangle$ is a dynamic Kripke frame and $V : PV \rightarrow \mathcal{P}(W)$ is a valuation assigning to each propositional variable an arbitrary subset of W. We extend V to the set of all formulas in $L_{\Box, \bigcirc}$ by the following recursive clauses:

 $V(\varphi \lor \psi) = V(\varphi) \cup V(\psi)$

 $V(\neg \varphi) = W - V(\varphi)$ $V(\bigcirc \varphi) = G^{-1}(V(\varphi))$ $V(\Box \varphi) = \{w \in W \mid v \in V(\varphi) \text{ for all } v \in W \text{ such that } w R v\}.$

Given a dynamic Kripke frame $K = \langle W, R, G \rangle$, we can impose a topology on W via the accessibility relation R. We define the *open* subsets of W as those subsets that are upward closed under R:

(*) $O \subseteq W$ is open iff $x \in O$ and xRy implies $y \in O$.

Recall that an *Alexandroff topology* is a topological space in which arbitrary intersections of open sets are open. The reader can verify that the collection of open subsets of *W* includes the entire space, the empty set, and is closed under arbitrary intersections and unions. Hence, viewing $\langle W, R \rangle$ as a topological space, the space is Alexandroff.

Going in the other direction, if X is an Alexandroff topology, we can define a relation R on X by:

(@) xRy iff x is a point of closure of $\{y\}$.

(Equivalently, y belongs to every open set containing x.) Clearly R is reflexive. To see that R is transitive, suppose that xRy and yRz. Let O be an open set containing x. Then since x is a point of closure for $\{y\}$, $y \in O$. But since y is a point of closure for $\{z\}$, $z \in O$. So x is a point of closure for $\{z\}$ and xRz. So far, we have shown that *static* Kripke frames, $\langle W, R \rangle$ correspond to Alexandroff topologies. But what about the dynamical aspect? Here we invite the reader to verify that *R*-monotonicity of the function G is equivalent to continuity of G in the topological setting. It follows that *dynamic Kripke frames are just dynamic Alexandroff topologies*.

In view of the fact that every finite topology is Alexandroff (if X is finite, then there are only finitely many open subsets of X), we have shown that finite topologies are just finite Kripke frames. This result, together with Theorem 2.3 (*iii*), gives the following completeness theorem for Kripke semantics:

Lemma 3.3. For any formula $\varphi \in L_{\square,\bigcirc}$, the following are equivalent: (i) $S 4C \vdash \varphi$; (ii) φ is true in any finite Kripke frame (= finite topological space).

In what follows, it will be useful to consider not just arbitrary finite Kripke frames, but frames that carry some additional structure. The notion we are after is that of a *stratified dynamic Kripke frame*, introduced by Slavnov (2005). We recall his definitions below.

Definition 3.4. Let $K = \langle W, R, G \rangle$ be a dynamic Kripke frame. A **cone** in K is any set $U_v = \{w \in W \mid vRw\}$ for some $v \in W$. We say that v is a **root** of U_v .

Note in particular that any cone, U_v , in K is an open subset of W—indeed, the smallest open subset containing v.

Definition 3.5. Let $K = \langle U, R, G \rangle$ be a finite dynamic Kripke frame. We say that K is **stratified** if there is a sequence $\langle U_1, \ldots, U_n \rangle$ of pairwise disjoint cones in K with roots u_1, \ldots, u_n respectively, such that $U = \bigcup_k U_k$; $G(u_k) = u_{k+1}$ for k < n, and G is injective. We say the stratified Kripke frame has depth n and (with slight abuse of notation) we call u_1 the **root** of the stratified frame.

Note that it follows from *R*-monotonicity of *G* that $G(U_k) \subseteq U_{k+1}$, for k < n.

Definition 3.6. Define the function CD ("circle depth") on the set of all formulas in $L_{\Box, \bigcirc}$ inductively, as follows.

$$\begin{split} CD(p) &= 0 \text{ for any propositional variable } p;\\ CD(\varphi \lor \psi) &= max \{CD(\varphi), CD(\psi)\};\\ CD(\neg \varphi) &= CD(\varphi);\\ CD(\Box \varphi) &= CD(\varphi);\\ CD(\bigcirc \varphi) &= 1 + CD(\varphi). \end{split}$$

We also refer to $CD(\varphi)$ *as the* \bigcirc *-depth of* φ *.*

Lemma 3.7. Suppose the formula φ is not a theorem of S4C, and $CD(\varphi) = n$. Then there is a stratified finite dynamic Kripke frame K with depth n + 1 such that φ is refuted at the root of K.

Proof. The proof is by Lemma 3.3 and by a method of 'disjointizing' finite Kripke frames (for the details, see Slavnov 2005).

4 Algebraic semantics for S4C

We saw that the topological semantics for S4C is a generalization of the Kripke semantics. Can we generalize further? Just as classical propositional logic is interpreted in Boolean algebras, we would like to interpret modal logics algebraically. Tarski and McKinsey showed that this can be done for the logic S4, interpreting the \Box -modality as an interior operator on a Boolean algebra. In this section we show that the same can be done for the logic S4C, interpreting the \bigcirc -modality via O-operators on a Boolean algebra.

We denote the top and bottom elements of a Boolean algebra by 1 and 0, respectively. **Definition 4.1.** A topological Boolean algebra is a Boolean algebra, \mathcal{A} , together with an interior operator I on \mathcal{A} that satisfies:

 $\begin{array}{l} (I_1) \ I1 = 1; \\ (I_2) \ Ia \leq a; \\ (I_3) \ IIa = Ia; \\ (I_4) \ I(a \wedge b) = Ia \wedge Ib. \end{array}$

Example 4.2. The set of all subsets $\mathcal{P}(X)$ of a topological space X with set-theoretic meets, joins and complements and where the operator I is just the topological interior operator (for $A \subseteq X$, I(A) = Int(A)) is a topological Boolean algebra. More generally, any collection of subsets of X that is closed under finite intersections, unions, complements and topological interiors is a topological Boolean algebra. We call any such algebra a topological field of sets.

Suppose *A* is a topological Boolean algebra with interior operator *I*. We define the *open* elements in *A* as those elements for which

$$Ia = a. (1)$$

Definition 4.3. Let A_1 and A_2 be topological Boolean algebras. We say $h : A_1 \rightarrow A_2$ is a **Boolean homomorphism** if h preserves Boolean operations. We say h is a **Boolean embedding** if h is an injective Boolean homomorphism. We say h is a **homomorphism** if h preserves Boolean operations and the interior operator. We say h is an **embedding** if h is an injective homomorphism. Finally, we say A_1 and A_2 are **isomorphic** if there is an embedding from A_1 onto A_2 .

Definition 4.4. Let A_1 and A_2 be topological Boolean algebras, and let $h : A_1 \to A_2$. We say h is an **O-map** if

- (*i*) *h* is a Boolean homomorphism.
- (ii) For any c open in A_1 , h(c) is open in A_2 .

An **O-operator** is an O-map from a topological Boolean algebra to itself.

Lemma 4.5. Let A_1 and A_2 be topological Boolean algebras, with interior operators I_1 and I_2 respectively. Suppose that $h : A_1 \to A_2$ is a Boolean homomorphism. Then h is an O-map iff for every $a \in A_1$,

$$h(I_1a) \le I_2(h(a)). \tag{2}$$

Proof. We let \mathcal{G}_1 and \mathcal{G}_2 denote the collection of open elements in A_1 and A_2 respectively. (\Rightarrow) Suppose *h* is an O-map. Then $h(I_1a) \in \mathcal{G}_2$ by Definition 4.4 (*ii*). Also, $I_1a \leq a$, so $h(I_1a) \leq h(a)$ (*h* is a Boolean homomorphism, hence preserves order). Taking interiors on both sides, we have $h(I_1a) = I_2(h(I_1a)) \leq I_2(ha)$. (\Leftarrow) Suppose that for every $a \in A_1$, $h(I_1a) \leq I_2(h(a))$. Let $c \in \mathcal{G}_1$. Then $c = I_1c$, so $h(c) = h(I_1c) \leq I_2(h(c))$. But also, $I_2(h(c)) \leq h(c)$. So $h(c) = I_2(h(c))$ and $h(c) \in \mathcal{G}_2$.

We are now in a position to state the algebraic semantics for the language $L_{\Box,\bigcirc}$.

Definition 4.6. A dynamic algebra is a pair $\langle A, h \rangle$, where A is a topological Boolean algebra and h is an O-operator on A. A dynamic algebraic model is an ordered triple, $\langle A, h, V \rangle$, where A is a topological Boolean algebra, h is an O-operator on A, and V : $PV \rightarrow A$ is a valuation, assigning to each propositional variable $p \in PV$ an element of A. We say $\langle A, h, V \rangle$ is a model over A. We can extend V to the set of all formulas in $L_{\Box, \Box}$ by the following recursive clauses:

$$V(\varphi \lor \psi) = V(\varphi) \lor V(\psi)$$
$$V(\neg \varphi) = -V(\varphi)$$
$$V(\Box \varphi) = IV(\varphi)$$
$$V(\bigcirc \varphi) = hV(\varphi).$$

(The remaining binary connectives, \rightarrow and \leftrightarrow , and unary operator, \diamond , are defined in terms of the above in the usual way.)

We define standard validity relations. Let $N = \langle A, h, V \rangle$ be a dynamic algebraic model. We say φ is *true* in $N(N \models \varphi)$ iff $V(\varphi) = 1$. Otherwise, we say φ is *refuted* in N. We say φ is *valid* in $A (\models_A \varphi)$ if for any algebraic model N over $A, N \models \varphi$. Finally, we let $DML_A = \{\varphi \mid \models_A \varphi\}$ (i.e., the set of validities in A). In our terminology, soundness of S4C for A is the claim: $S4C \subseteq DML_A$. Completeness of S4C for A is the claim: $DML_A \subseteq S4C$.

Proposition 4.7. (Soundness) Let A be a topological Boolean algebra. Then $S4C \subseteq DML_A$.

Proof. We have to show that the S4C axioms are valid in A and that the rules of inference preserve truth. To see that (A1) is valid, note that:

$$V(\bigcirc(\varphi \lor \psi)) = h(V(\varphi) \lor V(\psi))$$

= $h(V(\varphi)) \lor h(V(\psi))$ (*h* a Boolean homomorphism)
= $V(\bigcirc\varphi \lor \bigcirc\psi)$

Thus $V(\bigcirc(\varphi \lor \psi) \leftrightarrow (\bigcirc \varphi \lor \bigcirc \psi)) = 1$. Validity of (A2) is proved similarly. For (A3), note that:

$$V(\bigcirc \Box \varphi) = h(IV(\varphi))$$

$$\leq Ih(V(\varphi)) \quad \text{(by Lemma 4.5)}$$

$$= V(\Box \bigcirc \varphi)$$

So $V(\bigcirc \Box \varphi) \leq V(\Box \bigcirc \varphi)$ and $V(\bigcirc \Box \varphi \rightarrow \Box \bigcirc \varphi) = 1$. This takes care of the special \bigcirc -modality axioms. The remaining axioms are valid by soundness of *S*4 for any topological Boolean algebra (see e.g. Rasiowa and Sikorski 1963). To see that necessitation for \bigcirc preserves validity, suppose that φ is valid in *A* (i.e., for every algebraic model $N = \langle A, h, V \rangle$, we have $V(\varphi) = 1$). Then $V(\bigcirc \varphi) = h(V(\varphi)) = h(1) = 1$, and $\bigcirc \varphi$ is valid in *A*.

5 Reduced measure algebras

We would like to interpret S4C not just in arbitrary topological Boolean algebras, but in algebras carrying a probability measure—or 'measure algebras.' In this section we show how to construct such algebras from separable metric spaces together with a σ finite Borel measure (defined below).

Definition 5.1. Let A be a Boolean σ -algebra, and let μ be a non-negative function on A. We say μ is a **measure** on A if for any countable collection $\{a_n\}$ of disjoint elements in A, $\mu(\bigvee_n a_n) = \sum_n \mu(a_n)$.

If μ is a measure on A, we say μ is *positive* if 0 is the only element at which μ takes the value 0. We say μ is σ -*finite* if 1 is the countable join of elements in A with finite measure.¹ Finally, we say μ is *normalized* if $\mu(1) = 1$.

Definition 5.2. A measure algebra is a Boolean σ -algebra A together with a positive, σ -finite measure μ on A.

Lemma 5.3. Let A be a Boolean σ -algebra and let μ be a σ -finite measure on A. Then there is a normalized measure ν on A such that for all $a \in A$, $\mu(a) = 0$ iff $\nu(a) = 0$.

Proof. Since μ is σ -finite, there exists a countable collection $\{s_n | n \ge 1\} \subseteq A$ such that $\bigvee_{n\ge 1} s_n = 1$ and $\mu(s_n) < \infty$ for each $n \ge 1$. WLOG we can assume the s_n 's are pairwise

¹I.e., there is a countable collection of elements A_n in A such that $\bigvee_n A_n = 1$ and $\mu(A_n) < \infty$ for each $n \in \mathbb{N}$.

disjoint (i.e., $s_n \wedge s_m = 0$ for $m \neq n$). For any $a \in A$, let

$$\nu(a) = \sum_{n>1} 2^{-n} \frac{\mu(a \wedge s_n)}{\mu(s_n)}.$$

The reader can verify that ν has the desired properties.

In what follows, we show how to construct measure algebras from a topological space, *X*, together with a Borel measure on *X*. The relevant definition is given below.

Definition 5.4. Let X be a topological space. We say that μ is a **Borel measure** on X if μ is a measure defined on the σ -algebra of Borel subsets of X.²

Let X be a topological space, and let μ be a σ -finte Borel measure on X. We let *Borel*(X) denote the collection of Borel subsets of X and let $Null_{\mu}$ denote the collection of measure-zero Borel sets in X. Then *Borel*(X) is a Boolean σ -algebra, and $Null_{\mu}$ is a σ -ideal in *Borel*(X). We form the quotient algebra

$$\mathcal{M}_{X}^{\mu} = Borel(X)/Null_{\mu}$$

(Equivalently, we can define the equivalence relation ~ on Borel sets in X by $A \sim B$ iff $\mu(A \triangle B) = 0$, where \triangle denotes symmetric difference. Then \mathcal{M}_X^{μ} is the algebra of equivalence classes under ~.) Boolean operations in \mathcal{M}_X^{μ} are defined in the usual way in terms of underlying sets:

$$|A| \lor |B| = |A \cup B|$$
$$|A| \land |B| = |A \cap B|$$
$$-|A| = |X - A|$$

Lemma 5.5. There is a unique measure v on \mathcal{M}_X^{μ} such that $v|A| = \mu(A)$ for all A in Borel(X). Moreover, the measure v is σ -finite and positive.

Proof. See (Halmos 1959, pg. 79).

It follows from Lemma 5.5 that \mathcal{M}_X^{μ} is a measure algebra. We follow Halmos (1959) in referring to any algebra of the form \mathcal{M}_X^{μ} as a *reduced measure algebra*.³

²I.e., on the smallest σ -algebra containing all open subsets of *X*.

³In fact, Halmos allows as 'measure algebras' only algebras with a normalized measure. We relax this constraint here, in order to allow for the 'reduced measure algebra' generated by the entire real line together with the usual Lebesgue measure. This algebra is, of course, isomorphic to \mathcal{M}_X^{μ} , where X is the real interval [0, 1], and μ is the usual Lebesgue measure on X. This amendment was suggested by the anonymous referee.

Lemma 5.6. Let X be a topological space and let μ be a σ -finite Borel measure on X. Then for any $|A|, |B| \in \mathcal{M}_{X}^{\mu}, |A| \leq |B|$ iff $A \subseteq B \cup N$ for some $N \in Null_{\mu}$.

Proof. (\Rightarrow) If $|A| \leq |B|$, then $|A| \wedge |B| = |A|$, or equivalently $|A \cap B| = |A|$. This means that $(A \cap B) \triangle A \in Null_{\mu}$, so $A - B \in Null_{\mu}$. But $A \subseteq B \cup (A - B)$. (\Leftarrow) Suppose $A \subseteq B \cup N$ for some $N \in Null_{\mu}$. Then $A \cap (B \cup N) = A$, and $|A| \wedge |B \cup N| = |A|$. But $|B \cup N| = |B|$, so $|A| \wedge |B| = |A|$, and $|A| \leq |B|$.

For the remainder of this section, let *X* be a separable metric space, and let μ be a σ -finite Borel measure on *X*. Where the intended measure is obvious, we will drop superscripts, writing \mathcal{M}_X for \mathcal{M}_X^{μ} .

So far we have seen only that \mathcal{M}_{X}^{μ} is a Boolean algebra. In order to interpret the \Box modality of S4C in \mathcal{M}_{X}^{μ} , we need to construct an interior operator on this algebra (thus transforming \mathcal{M}_{X}^{μ} into a topological Boolean algebra). We do this via the topological structure of the underlying space, X. Let us say that an element $a \in \mathcal{M}_{X}^{\mu}$ is *open* if a = |U| for some open set $U \subseteq X$. We denote the collection of open elements in \mathcal{M}_{X}^{μ} by \mathcal{G}_{X}^{μ} (or, dropping superscripts, \mathcal{G}_{X}).

Proposition 5.7. \mathcal{G}_{x}^{μ} is closed under (i) finite meets and (ii) arbitrary joins.

Proof. (i) This follows from the fact that open sets in *X* are closed under finite intersections. (ii) Let $\{a_i \mid i \in I\}$ be a collection of elements in G_X^{μ} . We need to show that $\sup \{a_i \mid i \in I\}$ exists and is equal to some element in \mathcal{G}_X^{μ} . Since *X* is separable, there exists a countable dense set *D* in *X*. Let \mathcal{B} be the collection of open balls in *X* centered at points in *D* with rational radius. Then any open set in *X* can be written as a union of elements in \mathcal{B} . Let *S* be the collection of elements $B \in \mathcal{B}$ such that $|B| \leq a_i$ for some $i \in I$. We claim that

$$\sup \{a_i \mid i \in I\} = |\bigcup S|.$$

First, we need to show that $|\bigcup S|$ is an upper bound on $\{a_i | i \in I\}$. For each $i \in I$, $a_i = |U_i|$ for some open set $U_i \subseteq X$. Since U_i is open, it can be written as a union of elements in \mathcal{B} . Moreover, each of these elements is a member of S (if $B \in \mathcal{B}$ and $B \subseteq U_i$, then $|B| \leq |U_i| = a_i$). So $U_i \subseteq \bigcup S$ and $a_i = |U_i| \leq |\bigcup S|$.

For the reverse inequality (\geq) we need to show that if *m* is an upper bound on $\{a_i \mid i \in I\}$, then $|\bigcup S| \leq m$. Let m = |M|. Note that *S* is countable (since $S \subseteq \mathcal{B}$ and \mathcal{B} is countable). We can write $S = \{B_n \mid n \in \mathbb{N}\}$. Then for each $n \in \mathbb{N}$, there exists $i \in I$ such that $|B_n| \leq a_i \leq m$. By Lemma 5.5, $B_n \subseteq M \cup N_n$ for some $N_n \in Null_\mu$. Taking unions, $\bigcup_n B_n \subseteq M \cup \bigcup_n N_n$, and $\bigcup_n N_n \in Null_\mu$. By Lemma 5.5, $|S| = |\bigcup_n B_n| \leq m$.

We can now define an interior operator, I_X^{μ} , on \mathcal{M}_X^{μ} via the collection of open elements, \mathcal{G}_X^{μ} . For any $a \in \mathcal{M}_X^{\mu}$, let

$$I_X^{\mu}a = \sup \{ c \in \mathcal{G}_X^{\mu} \, | \, c \le a \}.$$

Lemma 5.8. I_X^{μ} is an interior operator.

Proof. For simplicity of notation, we let I denote I_X^{μ} and let \mathcal{G} denote \mathcal{G}_X^{μ} . Then (I_1) follows from the fact that $1 \in \mathcal{G}$. (I_2) follows from the fact that a is an upper bound on $\{c \in \mathcal{G} \mid c \leq a\}$. For (I_3) note that by (I_2) , we have $IIa \leq Ia$. Moreover, if $c \in \mathcal{G}$ with $c \leq a$, then $c \leq Ia$ (since Ia is supremum of all such c). Thus $\bigvee \{c \in \mathcal{G} \mid c \leq a\} \in \bigvee \{c \in \mathcal{G} \mid c \leq Ia\}$, and $Ia \leq IIa$. For (I_4) note that since $a \wedge b \leq a$, we have $I(a \wedge b) \leq Ia$. Similarly, $I(a \wedge b) \leq Ib$, so $I(a \wedge b) \leq Ia \wedge Ib$. For the reverse inequality, note that $Ia \wedge Ib \leq a$ (since $Ia \leq a$), and similarly $Ia \wedge Ib \leq b$. So $Ia \wedge Ib \leq a \wedge b$. Moreover, $Ia \wedge Ib \in \mathcal{G}$. It follows that $Ia \wedge Ib \leq I(a \wedge b)$.

Remark 5.9. Is the interior operator I_X^{μ} non-trivial? (That is, does there exist $a \in \mathcal{M}_X^{\mu}$ such that $Ia \neq a$?) This depends on the space, X, and the measure, μ . If we let X be the real interval, [0, 1], and let μ be the Lebesgue measure on Borel subsets of X, then the interior operator is non-trivial (for the proof, see Lando 2012). But suppose μ is a non-standard measure on the real interval, [0, 1], defined by:

$$\mu(A) = \begin{cases} 1 & \text{if } \frac{1}{2} \in A \\ 0 & \text{otherwise.} \end{cases}$$

Then $Borel([0, 1])/Null_{\mu}$ is the algebra 2, and both elements of this algebra are 'open.' So Ia = a for each element a in the algebra.

Remark 5.10. The operator I_X^{μ} does not coincide with taking topological interiors on underlying sets. More precisely, it is in general not the case that for $A \subseteq X$, $I_X^{\mu}(|A|) =$ |Int (A)|, where 'Int(A)' denotes the topological interior of A. Let X be the real interval [0, 1] with the usual topology, and let μ be Lebesgue measure restricted to measurable subsets of X. Consider the set $X - \mathbb{Q}$ and note that $|X - \mathbb{Q}| = |X|$ (\mathbb{Q} is countable, hence has measure zero). We have: $I_X^{\mu}(|X - \mathbb{Q}|) = I_X^{\mu}(|X|) = I_X^{\mu}(1) = 1$. However, $|Int (X - \mathbb{Q})| = |\emptyset| = 0$.

Remark 5.11. Note that an element $a \in \mathcal{M}_X^{\mu}$ is open just in case $I_X^{\mu}a = a$. Indeed, if a is open, then $a \in \{c \in \mathcal{G}_X^{\mu} | c \le a\}$. So $a = \sup\{c \in \mathcal{G}_X^{\mu} | c \le a\} = I_X^{\mu}a$. Also, if $I_X^{\mu}a = a$, then a is the join of a collection of elements in \mathcal{G}_X^{μ} , and so $a \in \mathcal{G}_X^{\mu}$. This shows that the definition of 'open' elements given above fits with the definition in (1).

In what follows, it will sometimes be convenient to express the interior operator I_X^{μ} in terms of underlying open sets, as in the following Lemma:

Lemma 5.12. Let $A \subseteq X$. Then $I_X^{\mu}(|A|) = |\bigcup \{O \text{ open } | |O| \le |A|\}|$.

Proof. By definition of I_X^{μ} , $I_X^{\mu}(|A|) = \sup\{c \in \mathcal{G}_X^{\mu} | c \le |A|\}$. Let \mathcal{B} and D be as in the proof of Proposition 5.7, and let S be the collection of elements $B \in \mathcal{B}$ such that $|B| \le |A|$. Then by the proof of Proposition 5.7, $I_X^{\mu}(|A|) = |\bigcup S|$. But now $\bigcup S = \bigcup \{O \text{ open } ||O| \le |A|\}$. (This follows from the fact that any open set $O \subseteq X$ can be written as a union of elements in \mathcal{B} .) Thus, $I_X^{\mu}(|A|) = |\bigcup S| = |\bigcup \{O \text{ open } ||O| \le |A|\}$.

We have shown that \mathcal{M}_X^{μ} together with the operator I_X^{μ} is a topological Boolean algebra. Of course, for purposes of our semantics, we are interested in O-operators on \mathcal{M}_X^{μ} . How do such maps arise? Unsurprisingly, a rich source of examples comes from continuous functions on the underlying topological space *X*. Let us spell this out more carefully.

Definition 5.13. Let X and Y be topological spaces and let μ and ν be Borel measures on X and Y respectively. We say $f : X \to Y$ is **measure-zero preserving** (MZP) if for any $A \subseteq Y$, $\nu(A) = 0$ implies $\mu(f^{-1}(A)) = 0$.

Lemma 5.14. Let X and Y be separable metric spaces, and let μ and ν be σ -finite Borel measures on X and Y respectively. Suppose B is a Borel subset of X with $\mu(B) = \mu(X)$, and $f: B \to Y$ is measure-zero preserving and continuous. Define $h_f^{[\cdot]}: \mathcal{M}_Y^{\nu} \to \mathcal{M}_X^{\mu}$ by

$$h_{f}^{[\cdot]}(|A|) = |f^{-1}(A)|.$$

Then $h_f^{[\cdot]}$ is an O-map. In particular, if X = Y, then $h_f^{[\cdot]}$ is an O-operator.

Proof. First, we must show that $h_f^{[\cdot]}$ is well-defined.⁴ Indeed, if |A| = |B|, then $v(A \triangle B) = 0$. And since f is MZP, $\mu(f^{-1}(A) \triangle f^{-1}(B)) = \mu(f^{-1}(A \triangle B)) = 0$. So $f^{-1}(A) \sim f^{-1}(B)$. This shows that $h_f^{[\cdot]}|A|$ is independent of the choice of representative, A. Furthermore, it is clear that $h_f^{[\cdot]}$ is a Boolean homomorphism. To see that it is an O-map, we need only show that if $c \in \mathcal{G}_Y^{\nu}$, $h_f^{[\cdot]}(c) \in \mathcal{G}_X^{\mu}$. But if $c \in \mathcal{G}_Y^{\nu}$ then c = |U| for some open set $U \subseteq Y$. By continuity of f, $f^{-1}(U)$ is open in B. So $f^{-1}(U) = O \cap B$ for some O open in X. So $h_f^{[\cdot]}(c) = |f^{-1}(U)| = |O| \in \mathcal{G}_X^{\mu}$.

By the results of the previous section, we can now interpret the language of S4C in reduced measure algebras. In particular, we say an algebraic model $\langle A, h, V \rangle$ is a *dynamic measure model* if $A = \mathcal{M}_X^{\mu}$ for some separable metric space X and a σ -finite Borel measure μ on X.

⁴Note that by continuity of f, $f^{-1}(A)$ is a Borel set in B, hence also a Borel set in X.

We are particularly interested in the reduced measure algebra generated by the real interval, [0, 1], together with the usual Lebesgue measure.

Definition 5.15. Let I be the real interval [0, 1] and let λ denote Lebesgue measure restricted to the Borel subsets of I. The Lebesgue measure algebra is the algebra $\mathcal{M}_{I}^{\lambda}$.

Because of its central importance, we denote the Lebesgue measure algebra without subscripts or superscripts, by \mathcal{M} . Furthermore, we denote the collection of open elements in \mathcal{M} by \mathcal{G} and the interior operator on \mathcal{M} by I.

As in Definition 4.6, we let $DML_{\mathcal{M}} = \{\varphi \mid \models_{\mathcal{M}} \varphi\}$ (i.e., the set of validities in \mathcal{M}). In our terminology, soundness of S4C for \mathcal{M} is the claim: $S4C \subseteq DML_{\mathcal{M}}$. Completeness of S4C for \mathcal{M} is the claim: $DML_{\mathcal{M}} \subseteq S4C$.

Proposition 5.16. (Soundness) $S4C \subseteq DML_{\mathcal{M}}$.

Proof. Immediate from Proposition 4.7.

Remark 5.17. The algebra \mathcal{M} is isomorphic to the algebra $Leb([0, 1])/Null_{\mu}$ where Leb([0, 1]) is the σ -algebra of Lebesgue-measureable subsets of the real interval [0, 1], and $Null_{\mu}$ is the σ -ideal of Lebesgue measure-zero sets. This follows from the fact that every Lebesgue-measureable set in [0, 1] differs from some Borel set by a set of measure zero.

6 Isomorphism between reduced measure algebras

In this section we use a well-known result of Oxtoby's to show that any reduced measure algebra generated by a separable metric space with a σ -finite, nonatomic Borel measure is isomorphic to \mathcal{M} . By Oxtoby's result, we can think of \mathcal{M} as the *canonical* separable measure algebra.

In the remainder of this section, let \mathcal{J} denote the space $[0,1] - \mathbb{Q}$ (with the usual metric topology), and let δ denote Lebesgue measure restricted to the Borel subsets of \mathcal{J} .

Definition 6.1. A topological space X is **topologically complete** if X is homeomorphic to a complete metric space.

Definition 6.2. Let X be a topological space. A Borel measure μ on X is **nonatomic** if $\mu(\{x\}) = 0$ for each $x \in X$.

Theorem 6.3. (Oxtoby, 1970) Let X be a topologically complete, separable metric space, and let μ be a normalized, nonatomic Borel measure on X. Then there exists a Borel set $B \subseteq X$ and a function $f : B \to \mathcal{J}$ such that $\mu(X - B) = 0$ and f is a measure-preserving homeomorphism (where the measure on \mathcal{J} is δ).

Proof. See (Oxtoby 1970).

Lemma 6.4. Suppose X and Y are separable metric spaces, and μ and v are normalized Borel measures on X and Y respectively. If $f : X \to Y$ is a measure preserving homoemorphism, then \mathcal{M}_X^{μ} is isomorphic to $\mathcal{M}_Y^{\nu, 5}$

Proof. For simplicity of notation, we drop superscripts, writing simply \mathcal{M}_X , \mathcal{G}_X , and I_X , etc. Let $h_f^{[\cdot]} : \mathcal{M}_Y \to \mathcal{M}_X$ be defined by $h_f^{[\cdot]}(|A|) = |f^{-1}(A)|$. This function is well-defined because f is MZP and continuous. (The first property ensures that $h_f^{[\cdot]}(|A|)$ is independent of representative A; the second ensures that $f^{-1}(A)$ is Borel.) Clearly $h_f^{[\cdot]}$ is a Boolean homomorphism. We can define the mapping $h_{f^{-1}}^{[\cdot]} : \mathcal{M}_X \to \mathcal{M}_Y$ by $h_f^{[\cdot]}(|A|) = |f(A)|$. Then $h_f^{[\cdot]}$ and $h_{f^{-1}}^{[\cdot]}$ are inverses, so $h_f^{[\cdot]}$ is bijective. We need to show that $h_f^{[\cdot]}$ preserves interiors—i.e., $h_f^{[\cdot]}(I_Ya) = I_X h_f^{[\cdot]}(a)$. The inequality (\leq) follows from the fact that $h_f^{[\cdot]}$ is an O-map (see Lemma 5.14). For the reverse inequality, we need to see that $h_f^{[\cdot]}(A)$ is an upper bound on $\{c \in \mathcal{G}_X \mid c \leq h_f^{[\cdot]}(a)\}$. If $c \in \mathcal{G}_X$, then $h_{f^{-1}}^{[\cdot]}(c) \in \mathcal{G}_Y$ and if $c \leq h_f^{[\cdot]}(a)$, then $h_{f^{-1}}^{[\cdot]}(c) \leq h_{f^{-1}}^{[\cdot]}(h_f^{[\cdot]}(a)) = a$. Thus $h_{f^{-1}}^{[\cdot]}(c) \leq I_Ya$, and $c = h_f^{[\cdot]}(h_f^{[\cdot]}(c)) \leq h_f^{[\cdot]}(I_Ya)$.

Corollary 6.5. Let X be a separable metric space, and let μ be a nonatomic σ -finite Borel measure on X with $\mu(X) > 0$. Then,

$$\mathcal{M}^{\mu}_{X}\cong\mathcal{M}.$$

Proof. By Lemma 5.3, we can assume that μ is normalized.⁶ Let X_{comp} be the completion of the metric space X. Clearly X_{comp} is separable. We can extend the Borel measure μ on X to a Borel measure μ^* on X_{comp} by letting $\mu^*(S) = \mu(S \cap X)$ for any Borel set S

⁵We can relax the conditions of the lemma, so that instead of requiring that *f* is measure-preserving, we require only that $\nu(f(S)) = 0$ iff $\mu(A) = 0$. In fact, we can further relax these conditions so that $f : B \to C$, where $B \subseteq X$, $C \subseteq Y$, $\mu(B \triangle X) = 0$, and $\nu(C \triangle Y) = 0$. We prove the lemma as stated because only this weaker claim is needed for the proof of Corollary 6.5.

⁶More explicitly: If μ is σ -finite, then by Lemma 5.3 there is a normalized Borel measure μ^* on X such that $\mu^*(S) = 0$ iff $\mu(S) = 0$ for each $S \subseteq X$. It follows that $\mathcal{M}_X^{\mu} \cong \mathcal{M}_X^{\mu^*}$ (where the isomorphism is not, in general, measure-preserving).

in X_{comp} . The reader can convince himself that μ^* is a normalized, nonatomic, σ -finite Borel measure on X_{comp} , and that $\mathcal{M}_{X_{comp}}^{\mu^*} \cong \mathcal{M}_X^{\mu}$. By Theorem 6.3, there exists a set $B \subseteq X_{comp}$ and a function $f: B \to \mathcal{J}$ such that $\mu^*(B) = 1$ and f is a measure-preserving homeomorphism. By Lemma 6.4, $\mathcal{M}_{\mathcal{J}} \cong \mathcal{M}_B$. We have:

$$\mathcal{M}\cong\mathcal{M}_{\mathcal{J}}\cong\mathcal{M}_{B}\cong M_{X_{comp}}^{\mu^*}\cong\mathcal{M}_{X}^{\mu}.$$

7 Invariance maps

At this point, we have at our disposal two key results: completeness of S4C for finite stratified Kripke frames, and the isomorphism between \mathcal{M}_X^{μ} and \mathcal{M} for any separable metric space X and σ -finite, nonatomic Borel measure μ . Our aim in what follows will be to transfer completeness from finite stratified Kripke frames to the Lebesgue measure algebra, \mathcal{M} . But how to do this?

We can view any topological space as a topological Boolean algebra—indeed, as the topological field of all subsets of the space (see Example 4.2). Viewing the finite stratified Kripke frames in this way, what we need is 'truth-preserving' maps between the algebras generated by Kripke frames and \mathcal{M}_X^{μ} , for appropriately chosen X and μ . The key notion here is that of a "dynamic embedding" (defined below) of one dynamic algebra into another. Although our specific aim is to transfer truth from Kripke algebras to reduced measure algebras, the results we present here are more general and concern truth preserving maps between arbitrary dynamic algebras.

Recall that a *dynamic algebra* is a pair $\langle A, h \rangle$, where A is a topological Boolean algebra, and h is an O-operator on A.

Definition 7.1. Let $M_1 = \langle A_1, h_1 \rangle$ and $M_2 = \langle A_2, h_2 \rangle$ be two dynamic algebras. We say a function $h : M_1 \to M_2$ is a **dynamic embedding** if

(i) h is an embedding of A_1 into A_2 ;

$$(ii) h \circ h_1 = h_2 \circ h.$$

Lemma 7.2. Let $M_1 = \langle A_1, h_1, V_1 \rangle$ and $M_2 = \langle A_2, h_2, V_2 \rangle$ be two dynamic algebraic models. Suppose that $h : \langle A_1, h_1 \rangle \rightarrow \langle A_2, h_2 \rangle$ is a dynamic embedding, and for every propositional variable p,

$$V_2(p) = h \circ V_1(p)$$

Then for any $\varphi \in L_{\Box,\bigcirc}$ *,*

$$V_2(\varphi) = h \circ V_1(\varphi).$$

Proof. By induction on the complexity of φ .

Corollary 7.3. Let $M_1 = \langle A_1, h_1, V_1 \rangle$ and $M_2 = \langle A_2, h_2, V_2 \rangle$ be two dynamic algebraic models. Suppose that $h : \langle A_1, h_1 \rangle \rightarrow \langle A_2, h_2 \rangle$ is a dynamic embedding, and for every propositional variable p,

$$V_2(p) = h \circ V_1(p).$$

Then for any $\varphi \in L_{\Box,\bigcirc}$ *,*

$$M_1 \models \varphi \quad iff \quad M_2 \models \varphi.$$

Proof. $M_2 \models \varphi$ iff $V_2(\varphi) = 1$ iff $h \circ V_1(\varphi) = 1$ (by Lemma 7.2) iff $V_1 = 1$ (since *h* is an embedding)

Let $\langle X, F \rangle$ be a dynamic topological space and let A_X be the topological field of all subsets of *X* (see Example 4.2). We define the function h_F on A_X by

$$h_F(S) = F^{-1}(S).$$

It is not difficult to see that h_F is an O-operator. We say that $\langle A_X, h_F \rangle$ is the dynamic algebra *generated by* (or *corresponding to*) to the dynamic topological space $\langle X, F \rangle$.

Our goal is to embed the dynamic algebras generated by finite dynamic Kripke frames into a dynamic measure algebra, $\langle \mathcal{M}_X^{\mu}, h \rangle$, where *X* is some appropriately chosen separable metric space and μ is a nonatomic, σ -finite Borel measure on *X*. In view of Corollary 7.3 and completeness for finite dynamic Kripke frames, this will give us completeness for the measure semantics. The basic idea is to construct such embeddings via 'nice' maps on the underlying topological spaces. To this end, we introduce the following new definition:

Definition 7.4. Suppose X and Y are a topological spaces, and μ is a Borel measure on X. Let $\gamma : X \to Y$. We say γ has the **M-property** with respect to μ if for any subset $S \subseteq Y$:

- (i) $\gamma^{-1}(S)$ is Borel;
- (ii) for any open set $O \subseteq X$, if $\gamma^{-1}(S) \cap O \neq \emptyset$ then $\mu(\gamma^{-1}(S) \cap O) > 0$.

Lemma 7.5. Suppose $\langle X, F \rangle$ is a dynamic topological space, where X is a separable metric space, F is measure-zero preserving, and let μ be a σ -finite Borel measure on X with $\mu(X) > 0$. Suppose $\langle Y, G \rangle$ is a dynamic topological space, and $\langle A_Y, h_G \rangle$ is the corresponding dynamic algebra. Let B be a subset of X with $\mu(B) = \mu(X)$, and suppose we have a map $\gamma : B \to Y$ that satisfies:

- (i) γ is continuous, open and surjective;
- (*ii*) $\gamma \circ F = G \circ \gamma$;
- (iii) γ has the M-property with respect to μ .

Then the map $\Phi: \langle A_Y, h_G \rangle \to \langle \mathcal{M}_Y^{\mu}, h_F^{|\cdot|} \rangle$ defined by

$$\Phi(S) = |\gamma^{-1}(S)|$$

is a dynamic embedding.

Proof. By the fact that \mathcal{M}_{X}^{μ} is isomorphic to \mathcal{M}_{B}^{μ} , we can view Φ as a map from $\langle A_{Y}, h_{G} \rangle$ into $\langle \mathcal{M}_{B}^{\mu}, h_{F}^{|\cdot|} \rangle$, where h_{F}^{μ} is viewed as an operator on \mathcal{M}_{B}^{μ} . Note that Φ is well-defined by the fact that γ satisfies clause (*i*) of the M-property. We need to show that (*i*) Φ is an embedding of $\langle A_{Y}, h_{G} \rangle$ into $\langle \mathcal{M}_{B}^{\mu}, h_{F}^{|\cdot|} \rangle$, and (*ii*) $\Phi \circ h_{G} = h_{F}^{|\cdot|} \circ \Phi$.

- (*i*) Clearly Φ is a Boolean homomorphism. We prove that Φ is injective and preserves interiors.
 - (Injectivity) Suppose $\Phi(S_1) = \Phi(S_2)$ and $S_1 \neq S_2$. Then $\gamma^{-1}(S_1) \sim \gamma^{-1}(S_2)$, and $S_1 \triangle S_2 \neq \emptyset$. Let $y \in S_1 \triangle S_2$. By surjectivity of γ , we have $\gamma^{-1}(y) \neq \emptyset$. Moreover, $\mu(\gamma^{-1}(y)) > 0$ (since γ has the M-property w.r.t. μ , and the entire space *B* is open). So $\mu(\gamma^{-1}(S_1) \triangle \gamma^{-1}(S_2)) = \mu(\gamma^{-1}(S_1 \triangle S_2)) \ge \mu(\gamma^{-1}(y)) > 0$. And $\gamma^{-1}(S_1) \nsim \gamma^{-1}(S_2)$. \perp .
 - (Preservation of Interiors) For clarity, we will denote the topological interior in the spaces Y and B by Int_Y and Int_B respectively, and the interior operator on \mathcal{M}_B^{μ} by *I*. Let $S \subseteq Y$. It follows from continuity and openness of $\gamma : B \to Y$, that

$$\gamma^{-1}(Int_Y(S)) = Int_B(\gamma^{-1}(S)).$$

Note that,

-
$$\Phi(Int_Y(S)) = |\gamma^{-1}(Int_Y(S))|$$

$$= |Int_B(\gamma^{-1}(S))|$$

$$= |\bigcup \{O \text{ open in } B \mid O \subseteq \gamma^{-1}(S)\}|$$
- $I(\Phi(S)) = I |\gamma^{-1}(S)|$

$$= |\bigcup \{O \text{ open in } B \mid |O| \le |\gamma^{-1}(S)|\} | \quad \text{(by Lemma 5.12)}$$

Thus it is sufficient to show that for any open set $O \subseteq B$,

$$O \subseteq \gamma^{-1}(S)$$
 iff $|O| \leq |\gamma^{-1}(S)|$.

The left-to-right direction is obvious. For the right-to-left direction, suppose (toward contradiction) that $|O| \leq |\gamma^{-1}(S)|$ but that $O \not\subseteq \gamma^{-1}(S)$. Then $O \subseteq \gamma^{-1}(S) \cup N$ for some $N \subseteq B$ with $\mu(N) = 0$. Moreover, since $O \not\subseteq \gamma^{-1}(S)$, there exists $x \in O$ such that $x \notin \gamma^{-1}(S)$. Let $y = \gamma(x)$. Then $\gamma^{-1}(y) \cap O \neq \emptyset$. Since γ has the M-property with respect to μ , it follows that $\mu(\gamma^{-1}(y) \cap O) > 0$. But $\gamma^{-1}(y) \cap O \subseteq N$ (since $\gamma^{-1}(y) \cap O \subseteq O \subseteq \gamma^{-1}(S) \cup N$, and $\gamma^{-1}(y) \cap \gamma^{-1}(S) = \emptyset$). \bot .

We've shown that Φ is an embedding of $\langle A_Y, h_G \rangle$ into $\langle \mathcal{M}_B^{\mu}, h_F^{|\cdot|} \rangle$. In view of the isomorphism between \mathcal{M}_X^{μ} and \mathcal{M}_B^{μ} , we have shown that Φ is an embedding of $\langle A_Y, h_G \rangle$ into \mathcal{M}_X^{μ} .

(*ii*) We know that $\gamma \circ F = G \circ \gamma$. Taking inverses, we have $F^{-1} \circ \gamma^{-1} = \gamma^{-1} \circ G^{-1}$. Now let $S \subseteq Y$. Then:

$$\Phi \circ h_G(S) = |\gamma^{-1}(G^{-1}(S))|$$
$$= |F^{-1}(\gamma^{-1}(S))|$$
$$= h_F^{|\cdot|} \circ \Phi(S).$$

8 Completeness of *S*4*C* for the Lebesgue measure algebra

In this section we prove the main result of the paper: Completeness of S4C for the Lebesgue measure algebra, \mathcal{M} . Recall that completeness is the claim that $DML_{\mathcal{M}} \subseteq S4C$. In fact, we prove the contrapositive: For any formula $\varphi \in L_{\Box,\bigcirc}$, if $\varphi \notin S4C$, then $\varphi \notin DML_{\mathcal{M}}$. Our strategy is as follows. If φ is a non-theorem of S4C, then by Lemma 3.7, φ is refuted in some finite stratified Kripke frame $K = \langle W, R, G \rangle$. Viewing the frame algebraically (i.e., as a topological field of sets), we must construct a dynamic embedding $\Phi : \langle A_W, h_G \rangle \rightarrow \langle \mathcal{M}, h \rangle$, where $\langle A_W, h_G \rangle$ is the dynamic Kripke algebra generated by the dynamic Kripke frame K, and h is some O-operator on \mathcal{M} . In view of the isomorphism between \mathcal{M} and \mathcal{M}_X^{μ} for any separable metric space, X, and nonatomic, σ -finite Borel measure μ on X with $\mu(X) > 0$, it is enough to construct a dynamic embedding of the Kripke algebra into \mathcal{M}_X^{μ} , for appropriately chosen X and μ .

The constructions in this section are a modification of the constructions introduced in (Slavnov 2005), where it is proved that S4C is complete for topological models in

Euclidiean spaces of arbitrarily large finite dimension. The modifications we make are measure-theoretic, and are needed to accommodate the new 'probabilistic' setting. We are very much indebted to Slavnov for his pioneering work in (Slavnov 2005).⁷

8.1 Outline of the proof

Let us spell out the plan for the proof a little more carefully. The needed ingredients are all set out in Lemma 7.5. Our first step will be to construct the dynamic topological space $\langle X, F \rangle$, where X is a separable metric space, and F is a measure-zero preserving, continuous function on X. We must also construct a measure μ on the Borel sets of X that is nonatomic and σ -finite, such that $\mu(X) > 0$. We want to embed the Kripke algebra $\langle A_W, h_G \rangle$ into $\langle \mathcal{M}_X^{\mu}, h_F^{|} \rangle$, and to do this, we must construct a topological map $\gamma : B \to W$, where $B \subseteq X$ and $\mu(B) = 1$, and γ satisfies the requirements of Lemma 7.5. In particular, we must ensure that (*i*) γ is open, continuous and surjective, (*ii*) $\gamma \circ F = G \circ \gamma$ and (*iii*) γ has the M-property with respect to μ .

In Section 8.2, we show how to construct the dynamic space $\langle X, F \rangle$, and the Borel measure μ on X. In Section 8.3, we construct the map $\gamma : X \to W$, and show that it has the desired properties.

8.2 The topological carrier of the countermodel

Let

$$X_n = I^1 \sqcup \cdots \sqcup I^n$$

where I^k is the *k*-th dimensional unit cube and \Box denotes disjoint union. We would like X_n to be a metric space, so we think of the cubes I^k as embedded in the space \mathbb{R}^n , and as lying at a certain fixed distance from one another. For simplicity of notation, we denote points in I^k by (x_1, \ldots, x_k) , and do not worry about how exactly these points are positioned in \mathbb{R}^n .

For each k < n, define the map $F_k : I^k \to I^{k+1}$ by $(x_1, \ldots, x_k) \mapsto (x_1, \ldots, x_k, \frac{1}{2})$. We let

$$F(x) = \begin{cases} F_k(x) & \text{if } x \in I_k, k < n \\ x & \text{if } x \in I_n \end{cases}$$

Clearly F is injective. For each $k \ge 2$ we choose a privileged "midsection" $D_k = [0, 1]^{k-1} \times \{\frac{1}{2}\}$ of I_k . Thus, $f(I_k) = D_{k+1}$ for k < n.

⁷Where possible, we have preserved Slavnov's original notation in (Slavnov 2005).



Figure 1: The space $X_3 = I^1 \sqcup I^2 \sqcup I^3$. Note that $\mu(I^1) = 1$, $\mu(I^2) = 2$, and $\mu(I^3) = 3$. The shaded regions in I^2 and I^3 denote the midsections, D_2 and D_3 , respectively.

The space X_n will be the carrier of our countermodels (we will choose *n* according to the \bigcirc -depth of the formula which we are refuting, as explained in the next section). We define a non-standard measure, μ , on X_n . This somewhat unusual measure will allow us to transfer countermodels on Kripke frames back to the measure algebra, \mathcal{M}_Y^{μ} .

Let μ on I_1 be Lebesgue measure on \mathbb{R} restricted to Borel subsets of I_1 . Suppose we have defined μ on I^1, \ldots, I^k . For any Borel set B in I^{k+1} , let $B_1 = B \cap D_{k+1}$, and $B_2 = B \setminus D_{k+1}$. Then $B = B_1 \sqcup B_2$. We define

$$\mu(B) = \mu(F^{-1}(B_1)) + \lambda(B_2)$$

where λ is the usual Lebesgue measure in \mathbb{R}^{k+1} . Finally, for any Borel set $B \subseteq X_n$, we let $\mu(B) = \sum_{k=1}^{n} \mu(B \cap I^k)$

Note that $\mu(I^1) = 1$, and in general $\mu(I^{k+1}) = \mu(I^k) + 1$. Thus $\mu(X_n) = \mu(I^1 \sqcup \cdots \sqcup I^n) = \sum_{1}^{n} k = \frac{1}{2}(n^2 + n)$.

Lemma 8.1. μ is a nonatomic, σ -finite Borel measure on X_n .

Proof. Clearly μ is nonatomic. Moreover, since $\mu(X_n) < \infty$, μ is σ -finite. The only thing left to show is that μ is countably additive. Suppose that $\{B_m\}_{m \in \mathbb{N}}$ is a collection of pairwise disjoint subsets of X_n .

Claim 8.2. For any $k \le n$,

$$\mu\left(\bigcup_{m}(B_m\cap I^k)\right)=\sum_{m}\mu(B_m\cap I^k).$$

Proof of Claim: By induction on k.⁸ But now we have:

$$\mu(\bigcup_{m} B_{m}) = \sum_{k} \mu[(\bigcup_{m} B_{m}) \cap I^{k}] \quad \text{(by definition of } \mu)$$
$$= \sum_{k} \mu[\bigcup_{m} (B_{m} \cap I^{k})]$$
$$= \sum_{k} \sum_{m} \mu(B_{m} \cap I^{k}) \quad \text{(by Claim 8.2)}$$
$$= \sum_{m} \sum_{k} \mu(B_{m} \cap I^{k})$$
$$= \sum_{m} \mu(B_{m}) \quad \text{(by definition of } \mu)$$

Lemma 8.3. *X* is a separable metric space and $F : X_n \to X_n$ is measure-preserving and continuous.

Proof. The set of rational points in I^k is dense in k ($k \le n$), so X_n is separable. Continuity of F follows from the fact that F is a translation in \mathbb{R}^n ; F is measure-preserving by the construction of μ .

8.3 Completeness

Assume we are given a formula $\varphi \in L_{\Box,\bigcirc}$ such that φ is not a theorem of *S*4*C* and let $n = CD(\varphi) + 1$. By Lemma 3.7, there is a finite stratified, dynamic Kripke model $K = \langle W, R, G, V_1 \rangle$ of depth *n* such that φ is refuted at the root of *K*. In other words, there is a collection of pairwise disjoint cones W_1, \ldots, W_n with roots w_0^1, \ldots, w_0^n respectively, such that $W = \bigcup_{k \le n} W_k$; *G* is injective; and $G(w_k) = w_{k+1}$ for each k < n; and $K, w_0^1 \not\models$

⁸The base case is by countable additivity of Lebesgue measure on the unit interval, [0, 1]. For the induction step, suppose the claim is true for k - 1. Then we have:

 $\mu(\bigcup_{m} (B_{m} \cap I^{k})) = \mu \left[F^{-1}(\bigcup_{m} (B_{m} \cap I^{k} \cap D^{k}))\right] + \lambda \left[\bigcup_{m} (B_{m} \cap I^{k}) \setminus D^{k}\right] \quad \text{(by definition of } \mu)$ $= \mu \left[\bigcup_{m} F^{-1}(B_{m} \cap I^{k} \cap D^{k})\right] + \sum_{m} \lambda((B_{m} \cap I^{k}) \setminus D^{k}) \quad \text{(by countable additivity of } \lambda)$ $= \sum_{m} \mu \left[F^{-1}(B_{m} \cap I^{k} \cap D^{k})\right] + \sum_{m} \lambda((B_{m} \cap I^{k}) \setminus D^{k}) \quad \text{(by induction hypothesis)}$ $= \sum_{m} \mu \left[F^{-1}(B_{m} \cap I^{k} \cap D^{k})\right] + \lambda((B_{m} \cap I^{k}) \setminus D^{k})$ $= \sum_{m} (B_{m} \cap I^{k}) \quad \text{(by definition of } \mu)$

 φ . Let the space $X = X_n = I^1 \sqcup \cdots \sqcup I^n$ and the measure μ be as defined in the previous section. We construct a map $\tilde{\gamma} : X \to W$ in a countable number of stages. To do this we will make crucial use of the notion of ϵ -nets, defined below:

Definition 8.4. Given a metric space S and $\epsilon > 0$, a subset Ω of S is an ϵ -net for S if for any $y \in S$, there exists $x \in \Omega$ such that $d(x, y) < \epsilon$ (where d denotes the distance function in S).

Observe that if S is compact, then for any $\epsilon > 0$ there is a *finite* ϵ -net for S.

Basic construction. Let $w_{\text{root}}^1 = w_0^1$, and let w_1, \ldots, w_{r_1} be the *R*-successors of w_{root}^1 . At the first stage, we select r_1 pairwise disjoint closed cubes T_1, \ldots, T_{r_1} in I^1 , making sure that their total measure adds up to no more than $(\frac{1}{2})^{0+2}$ —that is, $\sum_{k \le r_1} \mu(T_k) < \frac{1}{4}$. For each *x* in the interior of T_k we let $\tilde{\gamma}(x) = w_k$ ($k \le r_1$). With slight abuse of notation we put $\tilde{\gamma}(T_k) = w_k$. We refer to T_1, \ldots, T_{r_1} as *terminal cubes*, and we let $I_1^1 = I^1 - \bigcup_{k=1}^{r_1} \text{Int}(T_k)$.

At any subsequent stage, we assume we are given a set I_i^1 that is equal to I^1 with a finite number of open cubes removed from it. Thus I_i^1 is a compact set. We find a $\frac{1}{2^i}$ -net Ω_i for I_i^1 and for each point $y \in \Omega_i$, we choose r_1 pairwise disjoint closed cubes, $T_1^y, \ldots, T_{r_1}^y$ in the $\frac{1}{2^i}$ -neighborhood of y, putting $\tilde{\gamma}(T_k^y) = w_k$ (for $k \le r_1$, with the same meaning as above). Again, we refer to the T_k 's as terminal cubes. Since Ω_i is finite, we create only a finite number of new terminal cubes at this stage, and we make sure to do this in such a way as to remove a total measure of no more than $(\frac{1}{2})^{i+2}$. We let I_{i+1}^1 be the set I_i^1 minus the interiors of the new terminal cubes.

After doing this countably many times, we are left with some points in I^1 that do not belong to the interior of any terminal cube. We call such points *exceptional points* and we put $\tilde{\gamma}(x) = w_{\text{root}}^1$ for each exceptional point $x \in I^1$. This completes the definition of $\tilde{\gamma}$ on I^1 .

Now assume that we have already defined $\tilde{\gamma}$ on I^{j} . We let $w_{\text{root}}^{j+1} = w_{0}^{j+1}$ and let $w_{1}, \ldots, w_{r_{j+1}}$ be the *R*-successors of w_{root}^{j+1} . We define $\tilde{\gamma}$ on I^{j+1} as follows. At first we choose r_{j+1} closed cubes $T_{1}, \ldots, T_{r_{j+1}}$ in I^{j+1} , putting $\tilde{\gamma}(T_{k}) = w_{k}$ (for $k \leq r_{j+1}$). In choosing $T_{1}, \ldots, T_{r_{j+1}}$, we make sure that these cubes are not only pairwise disjoint (as before) but also disjoint from the midsection D_{j+1} . Again, we also make sure to remove a total measure of no more than $(\frac{1}{2})^{0+2} \mu(I^{j+1})$. We let $I_{1}^{j+1} = I^{j+1} - \bigcup_{k=1}^{r_{j+1}} \text{Int}(T_{k})$.

At stage *i*, we assume we are given a set I_i^{j+1} equal to I^{j+1} minus the interiors of a finite number of closed cubes. Thus I_i^{j+1} is compact, and we choose a finite $\frac{1}{2^i}$ -net Ω_i for I_i^{j+1} . For each $y \in \Omega_i$ we choose r_{j+1} closed terminal cubes $T_1, \ldots, T_{r_{j+1}}$ in the $\frac{1}{2^i}$ -neighborhood of y. We make sure that these cubes are not only pairwise disjoint, but disjoint from the midsection D_{j+1} . Since Ω_i is finite, we add only finitely many

new terminal cubes in this way. It follows that there is an ϵ -neighborhood of D_{j+1} that is disjoint from all the terminal cubes added up to this stage. Moreover, for each terminal cube T of I^j defined at the *i*th stage, $F(T) \subseteq D_{j+1}$, and we let T' be some closed cube in I^{j+1} containing F(T) and of height at most ϵ . To ensure that the equality $\tilde{\gamma} \circ F(x) = G \circ \tilde{\gamma}(x)$ hold s for all points x belonging to the interior of terminal cubes of I^j , we put:

$$\tilde{\gamma}(T') = G \circ \tilde{\gamma}(T).$$

Finally, we have added only finitely many terminal cubes at this stage, and we do so in such a way as to make sure that the total measure of these cubes is no more than $(\frac{1}{2})^{i+2} \mu(I^{j+1})$. We let I_{i+1}^{j+1} be the set I_i^{j+1} minus the new terminal cubes added at this stage.

We iterate this process countably many times, removing a countable number of terminal cubes from I^{j+1} . For all exceptional points x in I^{j+1} (i.e., points that do not belong to the interior of any terminal cube defined at any stage) we put $\tilde{\gamma}(x) = w_{\text{root}}^{j+1}$. Noting that exceptional points of I^j are pushed forward under F to exceptional points in I^{j+1} , we see that the equality $\tilde{\gamma} \circ F(x) = G \circ \tilde{\gamma}(x)$ holds for exceptional points as well.

This completes the construction of $\tilde{\gamma}$ on X. We pause now to prove two facts about the map $\tilde{\gamma}$ that will be of crucial importance in what follows.

Lemma 8.5. Let $E(I^j)$ be the collection of all exceptional points in I^j for some $j \le n$. Then $\mu(E(I^j)) \ge \frac{1}{2} \mu(I^j)$.

Proof. At stage *i* of construction of $\tilde{\gamma}$ on I^j , we remove from I^j terminal cubes of total measure no more than $(\frac{1}{2})^{i+2} \mu(I^j)$. Thus over countably many stages we remove a total measure of no more than $\mu(I^j) \sum_{i\geq 0} (\frac{1}{2})^{i+2} = \frac{1}{2} \mu(I^j)$. The remaining points in I^j are all exceptional, so $\mu(E(I^j)) \ge \mu(I^j) - \frac{1}{2} \mu(I^j) = \frac{1}{2} \mu(I^j)$.

Lemma 8.6. Let $x \in I^j$ be an exceptional point for some $j \le n$. Then $\tilde{\gamma}(x) = w_0^j$, and for any $\epsilon > 0$ and any $w_k \in W_j$ there is a terminal cube T contained in the ϵ -neighborhood of x with $\tilde{\gamma}(T) = w_k$.

Proof. Since $x \in I^j$ is exceptional, it belongs to I_i^j for each $i \in \mathbb{N}$. We can pick *i* large enough so that $\frac{1}{2^i} < \frac{\epsilon}{2}$. But then in the notations above, there exists a point $y \in \Omega_i$ such that $d(x, y) < \frac{\epsilon}{2}$. The statement now follows from the Basic Construction, since for each $w_k \in W_j$ there is a terminal cube T_k in the $\frac{1}{2^i}$ -neighborhood of *y* (and so also in the $\frac{\epsilon}{2}$ -neighborhood of *y*) with $\tilde{\gamma}(T_k) = w_k$.

Construction of the maps, γ_l . In the basic construction we defined a map $\tilde{\gamma}$: $X \to W$ that we will use in order to construct a sequence of 'approximation' maps, $\gamma_1, \gamma_2, \gamma_3, \ldots$, where $\gamma_1 = \tilde{\gamma}$. In the end, we will construct the needed map, γ , as

the limit (appropriately defined) of these approximation maps. We begin by putting $\gamma_1 = \tilde{\gamma}$. The *terminal cubes of* γ_1 and the *exceptional points of* γ_1 are the terminal cubes and exceptional points of the Basic Construction. Note that each of I^1, \ldots, I^n contains countably many terminal cubes of γ_1 together with exceptional points that don't belong to any terminal cube.

Assume that γ_l is defined and that for each terminal cube *T* of γ_l , all points in the interior of *T* are mapped by γ_l to a single element in *W*, which we denote by $\gamma_l(T)$. Moreover, assume that:

- (*i*) $\gamma_l \circ F = G \circ \gamma_l$;
- (*ii*) for any terminal cube T of γ_l in I^j , F maps T into some terminal cube T' of γ_l in I^{j+1} , for j < n.

where *F* is again the embedding $(x_1, \ldots, x_j) \mapsto (x_1, \ldots, x_j, \frac{1}{2})$.

We now define γ_{l+1} on the interiors of the terminal cubes of γ_l . In particular, for any terminal cube T of γ_l in I^1 , let $T^1 = T$ and let T^{j+1} be the terminal cube of I^{j+1} containing $F(T^j)$, for j < n. Then we have a system T^1, \ldots, T^n exactly like the system I^1, \ldots, I^n in the Basic Construction. We define γ_{l+1} on the interiors of T^1, \ldots, T^n in the same way as we defined $\tilde{\gamma}$ on I^1, \ldots, I^n , letting $w_{\text{root}}^j = \gamma_l(T^j)$ and letting w_1, \ldots, w_{r^j} be the *R*-successors of w_{root}^j . The only modification we need to make is a measure-theoretic one. In particular, in each of the terminal cubes T^{j} , we want to end up with a set of exceptional points that carries non-zero measure (this will be important for proving that the limit map we define, γ , has the M-property with respect to μ). To do this, assume γ_{l+1} has been defined on T^1, \ldots, T^j , and that for $k \le j, \mu(E(T^k)) \ge \frac{1}{2}\mu(T^k)$, where $E(T^k)$ is the set of exceptional points in T^k . When we define γ_{l+1} on T^{j+1} , we make sure that at the first stage we remove terminal cubes with a total measure of no more than $\frac{1}{2}^{0+2}\mu(T^{j+1})$. At stage *i* where we are given T_i^{j+1} we remove terminal cubes with a total measure of no more than $(\frac{1}{2})^{i+2} \mu(T^{j+1})$. Again, this can be done because at each stage i we remove only a finite number of terminal cubes, so we can make the size of these cubes small enough to ensure we do not exceed the allocated measure. Thus, over countably many stages we remove from T^{j+1} a total measure of no more than $\mu(T^{j+1}) \sum_{i \ge 0} (\frac{1}{2})^{i+2} = \frac{1}{2} \mu(T^{j+1})$. Letting $E(T^{j+1})$ be the set of exceptional points in T^{j+1} , we have $\mu(E(T^{j+1})) \ge \frac{1}{2}\mu(T^{j+1})$.

We do this for each terminal cube T of γ_l in I^1 . Next we do the same for all the remaining terminal cubes T of γ_l in I^2 (i.e. those terminal cubes in I^2 that are disjoint from D_2), and again, for all the remaining terminal cubes T of γ_l in I^3 (the terminal cubes in I^3 that are disjoint from D_3), etc. At the end of this process we have defined γ_{l+1} on the interior of each terminal cube of γ_l . For any point $x \in X$ that does not belong

to the interior of any terminal cube of γ_l , we put $\gamma_{l+1}(x) = \gamma_l(x)$. The terminal cubes of γ_{l+1} are the terminal cubes of the Basic Construction applied to each of the terminal cubes of γ_l . The points in the interior of terminal cubes of γ_l that do not belong to the interior of any terminal cube of γ_{l+1} are the exceptional points of γ_{l+1} .

In view of the measure-theoretic modifications we made above, we have the following analog of Lemma 8.5:

Lemma 8.7. Let $l \in \mathbb{N}$ and let T be any terminal cube of γ_l and E(T) be the set of exceptional points of γ_{l+1} in T. Then

$$\mu(E(T)) \ge \frac{1}{2}\,\mu(T).$$

Furthermore, the reader can convince himself that we have the following analog of Lemma 8.6 for the maps γ_l :

Lemma 8.8. Let x be an exceptional point of γ_l and let $\gamma_l(x) = w$. Then for any $\epsilon > 0$ and any v such that wRv, there is a terminal cube T of γ_l contained in the ϵ -neighborhood of x with $\gamma_l(T) = v$.

Finally, note that if x is an exceptional point of γ_l for some l, then $\gamma_l(x) = \gamma_{l+k}(x)$ for any $k \in \mathbb{N}$. We let B denote the set of points that are exceptional for some γ_l , and define the map $\gamma : B \to W$ as follows:

$$\gamma(x) = \lim_{l \to \infty} \gamma_l(x).$$

Lemma 8.9. $\mu(B) = \mu(X)$.

Proof. Let T_l be the set of all points that belong to some terminal cube of γ_l . Note that $T_l \supseteq T_{l+1}$ for $l \in \mathbb{N}$, and $\mu(T_1)$ is finite. Thus $\mu(\bigcap_l T_l) = \lim_{l \to \infty} \mu(T_l) = 0$. (The limit value follows from Lemma 8.7.) Finally, note that $B = X - \bigcap_l T_l$. So *B* is Borel, and $\mu(B) = \mu(X) - \mu(\bigcap_l T_l) = \mu(X)$.

We have constructed a map $\gamma : B \to W$ where $\mu(B) = \mu(X)$. Moreover, by the Basic Construction, we have $\gamma_l \circ F(x) = G \circ \gamma_l(x)$ for each $l \in \mathbb{N}$. It follows that $\gamma \circ F(x) = G \circ \gamma(x)$ for $x \in B$. All that is left to show is that (*i*) γ is continuous, open, and surjective; and (*ii*) γ has the M-property with respect to μ .

Lemma 8.10. γ has the *M*-property with respect to μ .

Proof. We show that for any subset $S \subseteq W$, (i) $\gamma^{-1}(S)$ is Borel; and (ii) for any open set $O \subseteq X$, if $\gamma^{-1}(S) \cap O \neq \emptyset$ then $\mu(\gamma^{-1}(S) \cap O) \neq 0$. Note that since W is finite, it is sufficient to prove this for the case where $S = \{w\}$ for some $w \in W$.

- (*i*) Note that $x \in \gamma^{-1}(w)$ iff *x* is exceptional for some γ_l and *x* belongs to some terminal cube *T* of γ_{l-1} , with $\gamma_{l-1}(T) = w$. There are only countably many such cubes, and the set of exceptional points in each such cube is closed. So $\gamma^{-1}(w)$ is a countable union of closed sets, hence Borel.
- (*ii*) Suppose that *O* is open in *X* with $\gamma^{-1}(w) \cap O \neq \emptyset$. Let $x \in \gamma^{-1}(w) \cap O$. Again, *x* is exceptional for some γ_l . Pick $\epsilon > 0$ such that the ϵ -neighborhood of *x* is contained in *O*. By Lemma 8.8, there is a terminal cube *T* of γ_l contained in the ϵ -neighborhood of *x* such that $\gamma_l(T) = w$ (since *wRw*). Letting E(T) be the set of exceptional points of γ_{l+1} in *T*, we know that $E(T) \subseteq \gamma^{-1}(w)$. But by Lemma 8.7, $\mu(E(T)) \ge \frac{1}{2}\mu(T) > 0$. So E(T) is a subset of $\gamma^{-1}(w) \cap O$ of non-zero measure, and $\mu(\gamma^{-1}(w) \cap O) > 0$.

In what follows, for any $w \in W$, let $U_w = \{v \in W | wRv\}$ (i.e., U_w is the smallest open set in W containing w).

Lemma 8.11. γ *is continuous.*

Proof. Let *U* be an open set in *W* and suppose that $x \in \gamma^{-1}(U)$. Let $\gamma(x) = w \in U$. Then *x* is exceptional for some γ_l . So *x* belongs to an (open) terminal cube *T* of γ_{l-1} with $\gamma_{l-1}(T) = w$. By *R*-monotonicity of $\langle \gamma_l(y) \rangle$ for all $y \in B$, we know that for any $y \in T$, $\gamma(y) \in U_w$ —i.e., $T \subseteq \gamma^{-1}(U_w)$. Moreover, since $w \in U$ and *U* is open, we have $U_w \subseteq U$. Thus $x \in T \subseteq \gamma^{-1}(U)$. This shows that $\gamma^{-1}(U)$ is open in *X*.

Lemma 8.12. γ is open.

Proof. Let *O* be open in *B* and let $w \in \gamma(O)$. We show that $U_w \subseteq \gamma(O)$. We know that there exists $x \in O$ such that $\gamma(x) = w$. Moreover, *x* is exceptional for some γ_l . Pick $\epsilon > 0$ small enough so that the ϵ -neighborhood of *x* is contained in *O*. By Lemma 8.8, for each $v \in U_w$ there is a terminal cube T_v of γ_l contained in the ϵ -neighborhood of *x* such that $\gamma_l(T_v) = v$. But then for any exceptional point y_v of γ_{l+1} that lies in T_v , we have $\gamma(y_v) = \gamma_{l+1}(y_v) = v$, and $y_v \in O$. We have shown that for all $v \in U_w$, $v \in \gamma(O)$. It follows that $\gamma(O)$ is open.

Lemma 8.13. γ is surjective.

Proof. This follows immediately from the fact that γ 'hits' each of the roots, w_0^1, \ldots, w_0^{n+1} , of K and γ is open.

Corollary 8.14. φ is refuted in \mathcal{M} .
Proof. We stipulated that φ is refuted in the dynamic Kripke model $K = \langle W, R, G, V_1 \rangle$. Equivalently, letting $M_1 = \langle A_K, h_G, V_1 \rangle$ be the dynamic algebraic model corresponding to K, φ is refuted in M_1 . By Lemma 8.11, Lemma 8.12, Lemma 8.13, and Lemma 8.10, we showed that $\gamma : X \to W$ is (i) continuous, open and surjective; (ii) $\gamma \circ f = G \circ \gamma$; and (iii) γ has the M-property with respect to μ . Thus by Lemma 7.5, the map $\Phi : \langle A_K, h_G \rangle \to \langle M_X^{\mu}, h_F^{|} \rangle$ defined by

$$\Phi(S) = |\gamma^{-1}(S)|$$

is a dynamic embedding. We now define the valuation $V_2: PV \to \mathcal{M}_X^{\mu}$ by:

$$V_2(p) = \Phi \circ V_1(p)$$

and we let $M_2 = \langle \mathcal{M}_X^{\mu}, h_F^{[\cdot]}, V_2 \rangle$. By Corollary 7.3, $M_2 \not\models \varphi$. In view of the isomorphism $M_X^{\mu} \cong \mathcal{M}$, we have shown that φ is refuted in \mathcal{M} .

We have shown that for *any* formula $\varphi \notin S4C$, φ is refuted in \mathcal{M} . We conclude the section by stating this completeness result more formally as follows:

Theorem 8.15. $DML_{\mathcal{M}} \subseteq S4C$.

9 Completeness for a single measure model

In this section we prove a strengthening of the completeness result of the previous section, showing that there is a single dynamic measure model $\langle \mathcal{M}, h, V \rangle$ in which every non-theorem of *S*4*C* is refuted.

Definition 9.1. Denote by \mathcal{M}^{ω} the product $\mathcal{M} \times \mathcal{M} \times \mathcal{M} \dots$ This is a Boolean algebra, where Boolean operations are defined component-wise:

$$(a_1, a_2, a_3, \dots) \lor (b_1, b_2, b_3, \dots) = (a_1 \lor b_1, a_2 \lor b_2, a_3 \lor b_3, \dots)$$
$$(a_1, a_2, a_3, \dots) \land (b_1, b_2, b_3, \dots) = (a_1 \land b_1, a_2 \land b_2, a_3 \land b_3, \dots)$$
$$-(a_1, a_2, a_3, \dots) = (-a_1, -a_2, -a_3, \dots)$$

Definition 9.2. We say $(a_1, a_2, a_3, ...)$ is an **open** element in \mathcal{M}^{ω} if a_k is open in \mathcal{M} for each $k \in \mathbb{N}$.

The collection of open elements in \mathcal{M}^{ω} is closed under finite meets, arbitrary joins and contains the top and bottom element (since operations in \mathcal{M}^{ω} are componentwise). We define the operator I_{ω} on \mathcal{M}^{ω} by:

$$I_{\omega}(a_1, a_2, a_3, \dots) = (Ia_1, Ia_2, Ia_3, \dots).$$

Then I_{ω} is an interior operator on \mathcal{M}^{ω} (the proof is the same as the proof of Lemma 5.8). So the algebra \mathcal{M}^{ω} together with the interior operator I_{ω} is a topological Boolean algebra.

Lemma 9.3. There is a dynamic algebraic model $M = \langle \mathcal{M}^{\omega}, h, V \rangle$ such that for any formula $\varphi \in L_{\Box, \bigcirc}$, the following are equivalent:

- (i) $S4C \vdash \varphi$;
- (*ii*) $M \models \varphi$.

Proof. Let $\langle \varphi_k \rangle$ be an enumeration of all non-theorems of S4C (there are only countably many formulas, so only countably many non-theorems). By completeness of S4C for \mathcal{M} , for each $k \in \mathbb{N}$, there is a model $M_k = \langle \mathcal{M}, h_k, V_k \rangle$ such that $M_k \not\models \varphi_k$. We construct a model $M = \langle \mathcal{M}^{\omega}, h, V \rangle$, where h and V are defined as follows. For any $\langle a_k \rangle_{k \in \mathbb{N}} = (a_1, a_2, a_3, \ldots) \in \mathcal{M}^{\omega}$, and for any propositional variable p:

$$h((a_1, a_2, a_2, \dots)) = \langle h_k(a_k) \rangle_{k \in \mathbb{N}}$$
$$V(p) = \langle V_k(p) \rangle_{k \in \mathbb{N}}.$$

(The fact that *h* is an O-operator follows from the fact that *h* is computed componentwise according to the h_k 's, and each h_k is an O-operator).

We can now prove the lemma. The direction $(i) \Rightarrow (ii)$ follows from Proposition 4.7. We show $(ii) \Rightarrow (i)$, by proving the contrapositive. Suppose that $S4C \not\models \varphi$. Then $\varphi = \varphi_k$ for some $k \in \mathbb{N}$. We claim that

$$\pi_k V(\varphi) = V_k(\varphi)$$

where π_k is the projection onto the *k*th coordinate. (Proof: By induction on complexity of φ , and the fact that π_k is a topological homomorphism.) In particular, $\pi_k V(\varphi_k) = V_k(\varphi_k) \neq 1$. So $V(\varphi_k) \neq 1$, and $M \nvDash \varphi_k$.

Lemma 9.4. \mathcal{M}^{ω} is isomorphic to \mathcal{M} .

Proof. We need to construct an isomorphism from \mathcal{M}^{ω} onto \mathcal{M} . Let $(a_1, a_2, a_3, ...)$ be an arbitrary element in \mathcal{M}^{ω} . Then for each $k \in \mathbb{N}$, we can choose a set $A_k \subseteq [0, 1]$ such that $a_k = |A_k|$ and $1 \notin A_k$. We define a sequence of points s_k in the real interval [0, 1] as follows:

$$s_0 = 0$$

 $s_1 = 1/2$
 $s_2 = 3/4$

In general, $s_k = \frac{2^k - 1}{2^k}$ $(k \ge 1)$. We now define a sequence of intervals I_k having the a_k 's as endpoints:

$$I_0 = [0, \frac{1}{2})$$
$$I_1 = [\frac{1}{2}, \frac{3}{4})$$
$$I_2 = [\frac{3}{4}, \frac{7}{8})$$

and in general $I_k = [s_k, s_{k+1})$. Our idea is to map each set A_k into the interval I_k . We do this by letting $B_k = \{l_k x + s_k | x \in A_k\}$ where l_k is the length of I_k . Clearly $B_k \subseteq I_k$ and $B_k \cap B_j = \emptyset$ for all $k \neq j$. We can now define the map $h : \mathcal{M}^{\omega} \to \mathcal{M}$ by:

$$h(a_1, a_2, a_3, \dots) = |\bigcup_{k \in \mathbb{N}} B_k|$$

where B_k is defined as above. The reader can now verify that h is an isomorphism. \Box

Corollary 9.5. There is a dynamic measure model $M = \langle \mathcal{M}, h, V \rangle$ such that for any formula $\varphi \in L_{\Box, \bigcirc}$, the following are equivalent:

- (*i*) $S4C \vdash \varphi$;
- (*ii*) $M \models \varphi$.

Proof. Immediate from Lemma 9.3 and Lemma 9.4.

Acknowledgements I would like to thank Grigori Mints and Dana Scott for valuable comments, and Sergei Slavnov for his pioneering work in this area.

References

G. B. M. Aiello, J. van Benthem. Reasoning about space: The modal way. *Journal of Logic and Computation*, 13(6):889–920, 2003.

S. Artemov, J. Davoren, and A. Narode. Modal logics and topological semantics for hybrid systems. MSI 97-05, Cornell University, 1997.

G. Bezhanishvili and M. Gehrke. Completeness of s4 with respect to the real line: Revisited. *Annals of Pure and Applied Logic*, 131(1-3):287–301, 2005.

D. Fernandez-Duque. Dynamic topological completeness for r^2 . Logic Journal of *IGPL*, 15:77–107, 2005.

D. Fernandez-Duque. Absolute completeness of s4u for its measure-theoretic semantics. *Advances in Modal Logic*, 8, 2010.

P. Halmos. Boolean algebras. New York: Stevens & Co., 1959.

P. Kremer and G. Mints. Dynamic topological logic. *Annals of Pure and Applied Logic*, 131(1-3):133–158, 2005.

T. Lando. Completeness of s4 for the lebesgue measure algebra. *Journal of Philosophical Logic*, 41(2):287–316, 2012.

G. Mints and T. Zhang. A proof of topological completeness of s4 in (0,1). *Annals of Pure and Applied Logic*, 133(1-3):231–235, 2005.

J. Oxtoby. Homeomorphic measures in metric spaces. *Proceedings of the American Mathematical Society*, 24(3):419–423, 1970.

H. Rasiowa and R. Sikorski. *The mathematics of metamathematics*. Państwowe Wydawnictowo Naukowe, Warsaw, Poland, 1963.

D. Scott. Mixing modality and probability. Lecture notes, 2009.

S. Slavnov. Two counterexamples in the logic of dynamic topological systems. TR 2003015, Cornell University, 2003.

S. Slavnov. On completeness of dynamic topological logic. *Moscow Mathematical Journal*, 5:477–492, 2005.

Surprise in Probabilistic Dynamic Epistemic Logic

Lorenz Demey

Center for Logic and Analytic Philosophy, KU Leuven lorenz.demey@hiw.kuleuven.be

Abstract

This paper presents a new analysis of surprise in the framework of probabilistic dynamic epistemic logic. This analysis is based on current psychological theories, and as a result, several experimentally observed aspects of surprise can be derived as theorems within the logical system. Most importantly, however, I will argue that unlike other formal accounts of surprise, the current analysis is able to capture the essentially dynamic nature of surprise. This conceptual elucidation also yields additional empirical benefits: the new analysis can capture important aspects of surprise that are not covered by earlier frameworks, such as its transitory nature.

1 Introduction

The phenomenon of surprise is ubiquitous in everyday life. People get surprised all the time; for example, by an unexpected flash of light, or—more 'down to earth' about the fact that their local grocery store has run out of milk (after all, the store is usually well-stocked!). The role of surprise in human life has been intensively studied in psychology from cognitive, social, developmental and educational perspectives. Furthermore, computer scientists have implemented the psychological findings about human surprise in artificial agents, and used logical models to describe these agent architectures. Surprise even crops up in various philosophical debates, such as those concerning the role of surprising evidence in Bayesian epistemology, or concerning the so-called surprise examination paradox.¹

The overarching goal of this paper is to provide a new analysis of the phenomenon of surprise in the framework of probabilistic dynamic epistemic logic. This account is based on the vast amount of experimental work on surprise in psychology, which should benefit its empirical adequacy. The paper's main thesis, however, is of a more conceptual nature: surprise is an essentially *dynamic* phenomenon, and any good formal analysis should represent this dynamics explicitly. I will argue that all current formalizations of surprise in artificial intelligence and logic fail to fully capture this dynamics, and show that the framework developed in this paper *is* able to capture it. As an additional benefit, this new framework can be used to analyze some aspects of surprise that could not be analyzed before.

This enterprise is motivated by a variety of interrelated issues. In the first place, a logical perspective on surprise can help to elucidate the basic properties of this notion. Starting from the concrete empirical results about surprise, a complete axiomatization is proposed in which the observed behavioral patterns can be derived as theorems. In other words, the fundamental laws of surprise can be 'reverse engineered' out of the concrete behavior that they generate. Secondly, the resulting logical system serves as a highly expressive language to formally specify agent architectures; it belongs to the general framework of (dynamic) epistemic logic, which is becoming a contemporary 'lingua franca' in multi-agent systems. Thirdly, and most importantly, this project constitutes a concrete illustration of the so-called dynamic turn in logic (van Benthem 1996; 2011, Demey 2013), which maintains that many theorems, phenomena, etc. which are usually expressed or analyzed in an entirely statical way, actually have a lot of dynamics going on, and could benefit significantly from analyses which explicitly represent this underlying dynamics. Considering the main conceptual thesis of this paper (as stated above), it should be clear that the paper offers a new illustration of the dynamic turn in logic.

The remainder of the paper is organized as follows. Section 2 briefly reviews the literature on surprise in cognitive science, multi-agent systems and logic. In Section 3 I argue that two earlier formalizations do not adequately represent the dynamic nature of surprise, and make some suggestions on how this can be achieved. In Section 2.2, then, I show how these suggestions can be developed into a full-fledged dynamic logic of surprise, which can capture several key aspects of surprise, such as its transitory (short-lived) nature and its role in belief revision. Finally, Section 5 wraps things up.

¹These philosophical debates will not be directly addressed in this paper; for overviews, the reader can consult (Talbott 2008) and (Chow 1998), respectively.

Demey

2 Three perspectives on surprise

This section provides an overview of the literature on surprise in cognitive science, multi-agent systems, and logic, focusing on those topics and debates that are most relevant for our current purposes. For more comprehensive overviews, the reader can consult (Macedo et al. 2009; 2012, Reisenzein and Meyer 2009), and the longer version of this paper (Demey forthcoming b).

2.1 Cognitive Science

The emotion of surprise is probably of old phylogenetic origin (Reisenzein et al. 1996). This short-lived state of mind is caused in an agent when she encounters an event that she did not expect. Surprise comes in degrees of intensity, which depend monotonically on the degree of unexpectedness of the surprise-causing event (Stiensmeier-Pelster et al. 1995).

The cognitive-psychoevolutionary theory of surprise (Meyer et al. 1997) claims that typically, an unexpected event elicits a sequence of four processes. First, the event is appraised as unexpected, i.e. as conflicting with a previously held belief. Secondly, if the degree of unexpectedness is sufficiently large, then ongoing processes are interrupted and attention is shifted to the unexpected event. Thirdly, the unexpected event is analyzed and evaluated, which can lead to the fourth process, viz. revision of the relevant beliefs.

The fact that this sequence ends in belief revision helps to explain the transitory (short-lived) character of surprise. When a surprising event occurs again and again, subjects tend to 'get used' to it, and after a few occurrences they do not find it surprising at all anymore (Experiment II Charlesworth 1964). Initially, the surprising event is unexpected: it conflicts with a previously held belief *B*. This leads to a process of belief revision, which removes *B* from the agent's stock of beliefs (and perhaps replaces it with another belief). When the same event happens again, it is no longer surprising, because it no longer conflicts with a previously held belief.

2.2 Multi-agent Systems

Since surprise typically leads to processes of learning and belief revision in humans, it is a natural move to endow *artificial agents* with the capability of feeling surprise, which can guide them in their actions. In a recent series of papers, Macedo and Cardoso have done exactly this (Macedo and Cardoso 2001a;b; 2004, Macedo et al. 2004; 2006). This work is based on the cognitive theories of surprise described above, and

can thus also be seen as a simulation of the human surprise mechanism (with various simplifications, obviously).

In the simplest model (Macedo and Cardoso 2001b), the anticipated intensity of surprise elicited by a piece of information φ is calculated as follows:²

$$S(\varphi) := 1 - P(\varphi). \tag{1}$$

The unexpectedness of φ is represented by $1 - P(\varphi)$. Here, $P(\varphi)$ denotes the subjective probability of φ , which is computed based on frequencies stored in the agent's memory. Thus (1) clearly shows that the intensity of surprise about φ is a monotone increasing function of the unexpectedness of φ .

2.3 Logic

Lorini (2008) has argued that researchers attempting to incorporate surprise and other emotions into multi-agent systems can benefit from the accuracy of logical frameworks for the formal specification of emotions. Therefore, Lorini and Castelfranchi (2006; 2007) have developed a logical framework for surprise. Just like Macedo and Cardoso's, this framework is based on the cognitive theories of surprise described in Subsection 2.1, and can thus be seen as a formal-logical model of human surprise.

I will now discuss some of the main features of this framework. The base logic is a system of probabilistic epistemic logic with a belief operator *B* and formulas about (linear combinations of) probabilities, such as $P(\varphi) \ge 0.5$ and $P(\varphi) + 2P(\psi) \ge 0.7$. This system is extended with PDL-style dynamic operators, and two unary operators *Test* and *Datum*. The formulas $Test(\varphi)$ and $Datum(\varphi)$ are to be read as "the agent is currently scrutinizing φ " and "the agent has perceptual datum φ ", respectively. Furthermore, there are actions $observe(\varphi)$ and $retrieve(\varphi)$, which represent observing that φ is the case and retrieving (from memory) that φ . Each of these actions gives rise to a PDL-style dynamic operator. The two most important axioms are:

$$[observe(\varphi)]Datum(\varphi), \tag{2}$$

$$[retrieve(\varphi)]Test(\varphi). \tag{3}$$

Axiom (2) says that after the agent observes that φ , this becomes a perceptual datum; analogously, axiom (3) says that after the agent has retrieved φ , this becomes an item under scrutiny.

²There exist more complex (and realistic) proposals for defining surprise in terms of unexpectedness (probability) (Macedo et al. 2004). However, the experimental data do not seem to single out one of these complex definitions over the other ones. Furthermore, the main conceptual points of this paper (regarding the dynamic nature of surprise) can perfectly be made using (1). Therefore, I will stick to the simpler definition.

With these resources, the notion of *mismatch-based surprise* can be defined. This emotion arises when there is a conflict between a perceptual datum ψ and a currently scrutinized belief φ ; 'conflict' here means that the agent believes that φ and ψ cannot be jointly true. Furthermore, the *intensity* of a mismatch-based surprise is defined as the probability that the agent assigns to the scrutinized belief φ . Hence, the more confident the agent is in her belief that φ , the more intensely she will be surprised upon receiving a perceptual datum that conflicts with φ (this captures exactly the idea that the intensity of surprise is a monotone function of the degree of unexpectedness). Formally:

$$MismatchS(\psi,\varphi) := Datum(\psi) \wedge Test(\varphi) \wedge B(\psi \to \neg \varphi), \tag{4}$$

Intensity
$$S(\psi, \varphi) = c := Mismatch S(\psi, \varphi) \land P(\varphi) = c.$$
 (5)

3 Surprise as a dynamic phenomenon

In this section I will argue that neither Macedo and Cardoso's computational nor Lorini and Castelfranchi's logical models of surprise adequately capture the dynamic nature of surprise. Afterwards I will suggest how the dynamics of surprise *can* adequately be formalized.

3.1 Quasi-static analyses of surprise

Let's first fix some terminology. Surprise is caused by an unexpected event. Any mental state (beliefs, desires, emotions, etc.) that the agent had (just) *before* perceiving the unexpected event will be called '*prior*'; any such state that she has (just) *after* perceiving the event will be called '*posterior*'.³ A statement that involves only prior notions or only posterior notions will be called 'temporally coherent'; a statement that involves both prior and posterior notions will be called 'temporally incoherent'.

Consider Macedo and Cardoso's analysis of surprise, and recall their Definition (1) of surprise intensity as unexpectedness:

$$S(\varphi) = 1 - P(\varphi).$$

The left side contains a posterior notion: the intensity of the surprise felt by the agent after the unexpected event. The right side, however, contains a prior notion: the agent's subjective probability before the unexpected event. Hence, Definition (1) is a temporally incoherent statement.

³ This terminology is analogous to the use of 'priors' and 'posteriors' in Bayesian frameworks. However, it should be emphasized that in this paper, 'prior' and 'posterior' are defined in terms of (being before or after) *perceiving* the unexpected event, while in Bayesian frameworks they are defined in terms of (being before or after) the *probabilistic update* triggered by that event.

To see this more clearly, note that there are two ways of reading (1) as a temporally coherent statement: (i) by considering both *S* and *P* to be prior notions, and (ii) by considering both *S* and *P* to be posterior notions. For interpretation (i), consider a case where the agent assigns a low (prior) probability to φ ; Definition (1) then says that she should experience a highly intensive surprise about φ . Under interpretation (i), this surprise is prior; in other words, the agent is highly surprised about an event *before* she has even perceived it—which is clearly absurd. For interpretation (ii), consider a case where the agent is highly surprised after perceiving an occurrence of φ ; Definition (1) then says that she assigns a low probability to φ . Under interpretation (ii), this probability is posterior; in other words, even after the agent has observed an occurrence of φ , she still assigns a low probability to it—which clearly contradicts the common assumption that agents process new information via Bayesian updating.⁴

I now turn to Lorini and Castelfranchi's analysis of surprise. Let's first consider the qualitative notion of mismatch-based surprise—ignoring, for the moment, surprise intensity. Recall their Definition (4):

$$MismatchS(\psi,\varphi) \equiv Datum(\psi) \land Test(\varphi) \land B(\psi \to \neg \varphi).$$

The left side contains a posterior notion: the agent's mismatch-based surprise after the unexpected event. The right side is more complicated. The first conjunct is posterior: ψ is only a perceptual datum after it has been observed by the agent; this dynamics was explicitly represented in (2). The second conjunct is both prior *and* posterior: φ was under scrutiny before the observation of the unexpected event, and remains so afterwards. The third and final conjunct is prior: the agent believed that ψ and φ cannot be jointly true before the unexpected event; typically, she will drop this belief as a result of her surprise (recall from Subsection 2.1 that surprise typically leads to a process of belief revision). Thus, in total, Definition (4) is temporally incoherent.⁵

Finally, let's consider the quantitative aspects of Lorini and Castelfranchi's system. Recall their Definition (5) of surprise intensity:

Intensity
$$S(\psi, \varphi) = c \equiv MismatchS(\psi, \varphi) \land P(\varphi) = c.$$

The left side contains a posterior notion: the intensity of the agent's mismatch-based surprise after she has perceived the unexpected event. The right side is, again, more complicated. The first conjunct—which was also the left side of (4)—is posterior: the

⁴And $P(\varphi|\varphi) = 1$, so after the occurrence of φ , the agent should assign probability 1 to it.

⁵Again, there are two ways of reading (4) as a temporally coherent statement: by considering all notions that appear in it to be prior, or by considering all those notions to be posterior. It is easy to see, however, that both interpretations quickly lead to counterintuitive consequences. Similar remarks apply to (5), which will be discussed next.

agent experiences mismatch-based surprise only after perceiving the unexpected event. The second conjunct, however, involves a prior notion, viz. the probability that the agent assigns to the scrutinized item φ before perceiving the unexpected event. Hence, Definition (5) is temporally incoherent as well.

An intuitively right principle about surprise should look somewhat like this: if the agent has a (prior) belief that ψ and φ are incompatible, and assigns (prior) probability c to φ , then after retrieving φ and observing an occurrence of ψ , she will experience a (posterior) mismatch-based surprise with intensity c. Formally, this looks as follows:

$$(B(\psi \to \neg \varphi) \land P(\varphi) = c) \longrightarrow [retrieve(\varphi); observe(\psi)] Intensity S(\psi, \varphi) = c.$$
 (6)

However, to derive (6) in Castelfranchi and Lorini's system, one needs principles such as (7) and (8), which link the agent's prior and posterior states by claiming that her observation of the occurrence of ψ does not change her relevant beliefs and probabilities in any way. This is highly counterintuitive: both (7) and (8) run entirely against the idea that surprise triggers a process of belief revision; additionally, (8) clearly contradicts the common assumption that agents process new information via Bayesian conditionalization.

$$B(\psi \to \neg \varphi) \to [observe(\psi)]B(\psi \to \neg \varphi),$$
 (7)

$$P(\varphi) = c \rightarrow [observe(\psi)]P(\varphi) = c.$$
 (8)

3.2 Towards a fully dynamic analysis of surprise

I have shown that both Macedo and Cardoso's definition of surprise intensity (1) and Lorini and Castelfranchi's definitions of mismatch-based surprise and its intensity (4–5) are temporally incoherent. There is a uniform explanation for this: surprise is an essentially dynamic phenomenon, but none of these authors explicitly represent this dynamics, so they have to 'smuggle' it into their systems—which thus end up being temporally incoherent.⁶

To obtain a temporally coherent definition of surprise, which respects the different 'stages' (before vs. after perceiving the unexpected event), the dynamics of surprise needs to be represented explicitly. I will use a public announcement operator $[!\varphi]$ for this purpose (technical details will be discussed in the next section). Whether a certain notion is to be interpreted as prior or as posterior, is now encoded directly in the syntax of the language: if the notion is within the scope of a dynamic operator, it is posterior, otherwise it is prior. For example, $P(\varphi) = 0.2$ means that the agent's *prior* probability of φ is 0.2, while $[!\varphi]P(\varphi) = 0.2$ means that her *posterior* probability of φ is 0.2.

⁶A similar story can be told about the role of dynamics in Aumann's celebrated 'agreeing to disagree' theorem in game theory (Demey 2010; forthcoming c).

We will work with a simple measure of surprise intensity S, based on Macedo and Cardoso's (1).⁷ When the surprise dynamics is explicitly represented, (1) is transformed into the following:

$$[!\varphi]S(\varphi) = c \longleftrightarrow P(\varphi) = 1 - c.$$
(9)

This principle says that the agent will be surprised about φ with intensity *c* after the unexpected event iff she assigns probability 1 - c to φ before the unexpected event. It thus says exactly the same as (1), but now in a temporally coherent way: both sides of (9) are prior statements.⁸ Furthermore, note that the right-to-left direction of (9) is similar in spirit to (6), which was very intuitive, but which was only derivable using additional implausible principles such as (7–8).

4 Modeling surprise in Probabilistic DEL

In the previous section, I made some suggestions on how the dynamics of surprise can be represented explicitly. In this section, these suggestions will be developed into a full-fledged logical system. I will also show how this system can naturally capture several important properties of surprise.

4.1 Semantic setup

Given the dynamic nature of surprise, and its connection with epistemic states and processes (beliefs, unexpectedness, belief revision, etc.), it is natural to work in the general framework of dynamic epistemic logic. This framework is rapidly becoming a 'lingua franca' or 'universal toolbox', which has been applied to problems in game theory, philosophy, artificial intelligence, etc. (Demey 2013).

Fix a countable set *Prop* of proposition letters. In this paper I will only work with a single agent, so it is not necessary to introduce agent indices. The formal language \mathcal{L} is given by the following Backus-Naur form:

$$\varphi ::= p \mid \neg \varphi \mid (\varphi \land \varphi) \mid K\varphi \mid \sum c_i P(\varphi) \ge c \mid S(\varphi) \ge c \mid S(\varphi) \le c \mid [!\varphi]\varphi,$$

where $p \in Prop$ and $c_i, c \in \mathbb{Q}$. As usual, $K\varphi$ means that the agent knows that φ . Similarly, $P(\varphi) \ge c$ means that the agent assigns probability (degree of belief) at least *c* to

⁷Recall Footnote 2.

⁸The left formula *as a whole* is prior; the *subformula* $S(\varphi) = c$ occurs inside the scope of the $[!\varphi]$ -operator, and is thus posterior. In other words, principle (9) is able to express a connection between the agent's *prior* probability and her *posterior* surprise intensity in a temporally coherent way, exactly by making use of the dynamic $[!\varphi]$ -operator.

 φ . Arbitrary linear combinations of probability terms are allowed mainly for technical reasons that need not concern us here (Fagin and Halpern 1994). Because of this generality, any type of (in)equality of probabilities is expressible (Def. 2 Kooi 2003). The formula $S(\varphi) \ge c$ says that the agent is surprised about φ with intensity at least *c*. Here, full expressivity is not allowed, and so the \ge - and \le -forms are both taken as primitive. One can then define $S(\varphi) < c$ as $\neg(S(\varphi) \ge c)$, etc.

Finally, $[!\varphi]\psi$ should be read as: "after a public announcement of φ , it will be the case that ψ ". Its dual is $\langle !\varphi \rangle \psi := \neg [!\varphi] \neg \psi$. Public announcement is usually explicated in terms of rational communication, but actually, almost any public event can be modeled using public announcements; for example, a strike of lightning can be modeled as a public announcement of the proposition 'lightning occurs (at time *t* and location ℓ)'.⁹ It thus makes perfect sense to represent an unexpected event (whatever its exact nature) as a public announcement.¹⁰

We now turn to the models on which this language will be interpreted:

Definition 4.1. A *surprise model* is a tuple $\mathbb{M} := \langle W, R, \mu, \sigma, V \rangle$, where *W* is a nonempty and finite set of states, *R* is an equivalence relation on *W*, and *V*: $Prop \rightarrow \mathscr{O}(W)$ is a valuation function. Furthermore, μ assigns to every state $w \in W$ a probability mass function $\mu(w): W \rightarrow [0, 1]$ that satisfies two conditions: (i) if $(w, v) \notin R$ then $\mu(w)(v) = 0$, and (ii) $\mu(w)(w) > 0$. Finally, σ assigns to every state $w \in W$ a surprise measure, i.e. a partial function $\sigma(w): \mathscr{O}(W) \rightarrow [0, 1]$.

Definition 4.2. The class of all surprise models will be denoted C_S . Furthermore, C_S^* is the class of all surprise models whose surprise measures are entirely undefined, i.e. such that $\sigma(w)(X)$ is undefined for all $w \in W$ and $X \subseteq W$.

A surprise model is thus just an ordinary finite¹¹ Kripke model $\langle W, R, V \rangle$ with additional components μ and σ . First of all, $\mu(w)(v) = c$ means that at state w, the agent assigns probability c to v being the actual state. Similarly, $\sigma(w)(X) = c$ means that at

⁹Van Benthem et al. (2009, p. 71) make a similar comment: "While much of the theory has been developed with conversation and communication in mind, it is important [...] to stress that we are not doing some sort of formal linguistics. The formal systems we will be dealing with apply just as well to observation, experimentation, learning, or any sort of information-carrying scenario".

¹⁰This also resolves a terminological tension in the literature on surprise. Agents are surprised *about* some *propositional* content (a piece of information), but their surprise is *caused by* some (non-propositional) *event*. In the new system, the propositional content of the surprise is formalized as the proposition φ , while its cause is formalized as the *public announcement of* that proposition. In short: φ is a proposition, but $!\varphi$ is an event.

¹¹The assumption that surprise models are finite ensures that probabilities can be represented using simple probability mass functions. This assumption can be dropped; the general case uses σ -algebras to represent probabilities (Sack 2009, Demey and Sack forthcoming). However, the main points of this paper are of a more conceptual nature, and can perfectly be made using the less sophisticated setup.

state *w*, the agent experiences surprise with intensity *c* about *X* (i.e. about one of the states in *X* being the actual state). Note the following differences between $\mu(w)$ and $\sigma(w)$ (for any state $w \in W$):

- μ(w) is a total function, so μ(w)(v) is defined for every state v ∈ W; on the other hand, σ(w) is a partial function, so it is allowed that σ(w)(X) is undefined for some sets X ⊆ W,
- $\mu(w)$ is required to satisfy conditions (i) and (ii), whose motivation is discussed extensively in (Demey and Sack forthcoming); on the other hand, $\sigma(w)$ is not required to satisfy any additional conditions whatsoever,
- $\mu(w)$ is defined on individual states, and can additively be lifted to sets of states: $\mu(w)(X) = \sum_{x \in X} \mu(w)(x)$ (this essentially reflects the finite additivity of probabilities); on the other hand, $\sigma(w)$ is defined directly on sets of states, so it might happen that $\sigma(w)(\{x, y\}) \neq \sigma(w)(\{x\}) + \sigma(w)(\{y\})$.

These differences show that unlike the well-behaved epistemological notion of probability (degree of belief), the psychological notion of (degree of) surprise satisfies no static regularities whatsoever. This is a clear manifestion of the essentially dynamic nature of surprise in the definition of surprise models.¹²

I now turn to the logic's semantics. This is entirely as expected; the formal clauses are stated in Definition 2.9. Given a formula $\varphi \in \mathcal{L}$ and a surprise model \mathbb{M} , I use $\llbracket \varphi \rrbracket^{\mathbb{M}}$ to denote the set $\{w \in W \mid \mathbb{M}, w \models \varphi\}$. The clause for surprise formulas holds for $\geq \in \{\geq, \leq\}$; I will return to it later (see Lemma 2). Note that to interpret a formula of the form $\llbracket \varphi \rrbracket \psi$ at a surprise model \mathbb{M} , the subformula ψ has to be interpreted at the updated model $\mathbb{M} \upharpoonright \varphi$, which is well-defined because of Definition 4.4 and Lemma 1. Finally, Definition 4.5 states the usual definition of semantic validity.

Definition 4.3. Consider a surprise model M and state w in M. Then:

¹²One might consider adding the requirements that if $X \subseteq Y \subseteq W$, then $\sigma(w)(X) \ge \sigma(w)(Y)$ and $\sigma(w)(W-X) = 1 - \sigma(w)(X)$, in analogy to the well-known Kolmogorov axioms for probability. However, the only motivation for such requirements seems to be the observation that "surprise is inversely correlated with probability", which is only plausible if 'surprise' is read as posterior and 'probability' as prior. I will return to this suggestion after the dynamics has been formally introduced (cf. Lemma 4).

 $\mathbb{M}, w \models p$ $w \in V(p)$ (for $p \in Prop$), iff $\mathbb{M}, w \models \neg \varphi$ iff $\mathbb{M}, w \neq \varphi,$ $\mathbb{M}, w \models \varphi \land \psi$ $\mathbb{M}, w \models \varphi \text{ and } \mathbb{M}, w \models \psi,$ iff $\mathbb{M}, w \models K\varphi$ for all $v \in W$: if wRv then $\mathbb{M}, v \models \varphi$, iff $\sum c_i \mu(w)(\llbracket \varphi_i \rrbracket^{\mathbb{M}}) \ge c,$ $\mathbb{M}, w \models \sum c_i P(\varphi_i) \ge c$ iff $\begin{cases} \sigma(w)(\llbracket \varphi \rrbracket^{\mathbb{M}}) \ge c & \text{if } \sigma(w)(\llbracket \varphi \rrbracket^{\mathbb{M}}) \text{ is defined} \\ c = 0 & \text{otherwise,} \end{cases}$ $\mathbb{M}, w \models S(\varphi) \ge c$ iff if $\mathbb{M}, w \models \varphi$ then $\mathbb{M} \upharpoonright \varphi, w \models \psi$. $\mathbb{M}, w \models [!\varphi]\psi$ iff

Definition 4.4. Consider an arbitrary surprise model $\mathbb{M} = \langle W, R, \mu, \sigma, V \rangle$ and formula $\varphi \in \mathcal{L}$, and suppose that $\mathbb{M}, w \models \varphi$ for some $w \in W$. Then the *updated model* $\mathbb{M} \upharpoonright \varphi := \langle W^{\varphi}, R^{\varphi}, \mu^{\varphi}, \sigma^{\varphi}, V^{\varphi} \rangle$ is defined as follows:

- $W^{\varphi} := \llbracket \varphi \rrbracket^{\mathbb{M}} = \{ w \in W \mid \mathbb{M}, w \models \varphi \},\$
- $R^{\varphi} := R \cap (\llbracket \varphi \rrbracket^{\mathbb{M}} \times \llbracket \varphi \rrbracket^{\mathbb{M}}),$
- $\mu^{\varphi}(w)(v) := \frac{\mu(w)(v)}{\mu(w)(\llbracket\varphi\rrbracket^{\Bbbk})}$ for all $w, v \in W^{\varphi}$,
- $\sigma^{\varphi}(w)(X) := 1 \mu(w)(X)$ for all $w \in W^{\varphi}, X \subseteq W^{\varphi}$,
- $V^{\varphi}(p) := V(p) \cap \llbracket \varphi \rrbracket^{\mathbb{M}}$ for every $p \in Prop$.

Definition 4.5. For any formula $\varphi \in \mathcal{L}$ and class of models *C*, we say that $C \models \varphi$ iff $\mathbb{M}, w \models \varphi$ for all models $\mathbb{M} \in C_S$ and states *w* in \mathbb{M} .

Lemma 1. The class C_S is closed under public announcements, i.e. if $\mathbb{M} \in C_S$, then also $\mathbb{M} \upharpoonright \varphi \in C_S$ (for any formula $\varphi \in \mathcal{L}$). This does not hold for C_S^* .

Proof. The C_S case is trivial: for the non-surprise components, (see Lemma 9 Demey 2010), and since Definition 4.1 does not require the surprise measures to satisfy any additional requirements, there is nothing else to prove. For C_S^* , note that by Definition 4.4, the updated surprise measures are total functions, even if the original surprise measures were entirely undefined.

The public announcement of φ in a model \mathbb{M} deletes all $\neg \varphi$ -states from that model; this is a standard idea (van Ditmarsch et al. 2007). The probability functions are changed by Bayesian conditionalization on the announced proposition φ (Kooi 2003). To see this more clearly, note that the definition of the updated probability function can be rewritten using conditional probabilities: $\mu^{\varphi}(w)(v) = \mu(w)(v | \llbracket \varphi \rrbracket^{\mathbb{M}})$. Most importantly, the updated surprise measure $\sigma^{\varphi}(w)$ is defined in terms of the original probability function $\mu(w)$. This is the only substantial property of surprise that is assumed in the logic's semantic setup; it is clearly of a dynamic nature (linking the original and the updated model).

Even though the surprise measures $\sigma(w)$ are allowed to be partial, Lemma 2 below shows that this does not lead to any truth value gaps in the semantics. When we are modeling concrete scenarios, we typically want to assume that the agent initially (i.e. before any unexpected events have taken place) experiences no surprise. Lemma 2 therefore justifies the following heuristic rule (HEUR) :

When modeling a scenario, it can be assumed that the 'initial' model \mathbb{M} (which represents the situation before any unexpected events have taken place) leaves all surprise measures undefined, i.e. that $\mathbb{M} \in C_S^*$.

Lemma 2. Consider an arbitrary surprise model $\mathbb{M} = \langle W, R, \mu, \sigma, V \rangle$ and formula $\varphi \in \mathcal{L}$, and suppose that $\sigma(w)(\llbracket \varphi \rrbracket^{\mathbb{M}})$ is undefined. Then $\mathbb{M}, w \models S(\varphi) = 0$.

Proof. Since $\sigma(w)(\llbracket \varphi \rrbracket^{\mathbb{M}})$ is undefined, it follows by the semantic clause for $S(\varphi) \ge c$ that $\mathbb{M}, w \models S(\varphi) \ge 0$ (and $\mathbb{M}, w \not\models S(\varphi) \ge c$ for all $c \ne 0$). Entirely analogously, $\mathbb{M}, w \models S(\varphi) \le 0$ (and $\mathbb{M}, w \not\models S(\varphi) \le c$ for all $c \ne 0$).

The following lemma states that the language \mathcal{L} contains no redundancies. In particular, the surprise operator cannot be defined in terms of the other available operators.

Lemma 3. There exists no formula $\varphi \in \mathcal{L} - \{S\}$ such that $\models \varphi \leftrightarrow S(p) \ge 0.5$.

Proof. Consider the surprise models M_1 and M_2 , defined as follows:

- $\mathbb{M}_1 = \langle W_1, R_1, \mu_1, \sigma_1, V_1 \rangle, W_1 = \{w_1\}, R_1 = \{(w_1, w_1)\}, \mu(w_1)(w_1) = 1, \sigma_1(w_1)(X) = 0.6 \text{ for all } X \subseteq W_1, \text{ and } V_1(p) = W_1,$
- $\mathbb{M}_2 = \langle W_2, R_2, \mu_2, \sigma_2, V_2 \rangle, W_2 = \{w_2\}, R_2 = \{(w_2, w_2)\}, \mu_2(w_2)(w_2) = 1, \sigma_2(w_2)(X) = 0.4 \text{ for all } X \subseteq W_2, \text{ and } V_2(p) = W_2.$

One can show by induction on the complexity of φ that

for all
$$\varphi \in \mathcal{L} - \{S\}$$
: $\mathbb{M}_1, w_1 \models \varphi$ iff $\mathbb{M}_2, w_2 \models \varphi$.

But it also holds that $\mathbb{M}_1, w_1 \models S(p) \ge 0.5$, while $\mathbb{M}_2, w_2 \not\models S(p) \ge 0.5$.

The distinction between the original and the updated model corresponds exactly to the distinction between prior and posterior notions that was introduced in the previous section. Hence, the definition $\sigma^{\varphi}(w)(X) = 1 - \mu(w)(X)$ defines *posterior* surprise in terms of *prior* probability. As a consequence, all the properties of probability are manifested in the posterior surprise measure (recall Footnote 12):

Lemma 4. Consider an arbitrary surprise model $\mathbb{M} = \langle W, R, \mu, \sigma, V \rangle$ and formula $\varphi \in \mathcal{L}$, and suppose that $\mathbb{M}, w \models \varphi$ for some $w \in W$. For all $w \in W^{\varphi}$ and $X \subseteq Y \subseteq W^{\varphi}$, it holds that $\sigma^{\varphi}(w)(X) \ge \sigma^{\varphi}(w)(Y)$ and that $\sigma^{\varphi}(w)(W - X) = 1 - \sigma^{\varphi}(w)(X)$.

Proof. Both items follow immediately from the definition of σ^{φ} and the fact that $\mu(w)$ is a probability mass function. For example, if $X \subseteq Y$, then $\mu(w)(X) \le \mu(w)(Y)$, and hence $\sigma^{\varphi}(w)(X) = 1 - \mu(w)(X) \ge 1 - \mu(w)(Y) = \sigma^{\varphi}(w)(Y)$.

Before moving to the logic's proof theory, I will illustrate and justify its semantics by discussing a simple example in full detail.

Example 1. Consider the following scenario. Mary does not know whether it is currently snowing. In fact, it is indeed currently snowing, but since Mary does not yet know about this, she experiences no surprise about it whatsoever. Furthermore, since it is July and Mary knows that snow in July is very rare at her current location, she considers it very unlikely that it is currently snowing. This example can be formalized using the following surprise model: $\mathbb{M} = \langle W, R, \mu, \sigma, V \rangle$, $W = \{w, v\}$, $R = W \times W$, $\mu(w)(w) = \mu(v)(w) = 0.05$, $\mu(w)(v) = \mu(v)(v) = 0.95$, $V(p) = \{w\}$, and $\sigma(w)(X)$ and $\sigma(v)(X)$ undefined for all $X \subseteq W$. (Note that we have followed the heuristic rule HEUR discussed above.) The proposition letter *p* represents 'it is snowing'; the state *w* represents the actual world. This model is a faithful representation of the scenario described above; for example:

$$\mathbb{M}, w \models \neg Kp \land \neg K \neg p \land P(p) = 0.05 \land P(\neg p) = 0.95 \land S(p) = 0.$$

Now suppose that Mary goes outside and sees that it is actually snowing. This can be modeled as a public announcement of *p* (recall Footnote 9). Applying Definition 4.4, we obtain the updated model $\mathbb{M} \upharpoonright p$, with $W^p = \{w\}, R = \{(w, w)\},$

$$\mu^{p}(w)(\llbracket p \rrbracket^{\mathbb{M} \restriction p}) = \mu^{p}(w)(w) = \frac{\mu(w)(w)}{\mu(w)(\llbracket p \rrbracket^{\mathbb{M}})} = \frac{\mu(w)(w)}{\mu(w)(w)} = 1,$$

$$\sigma^{p}(w)(\llbracket p \rrbracket^{\mathbb{M} \restriction p}) = \sigma^{p}(w)(\{w\}) = 1 - \mu(w)(\{w\}) = 1 - 0.05 = 0.95.$$

Using this updated model $\mathbb{M} \upharpoonright p$, we find that

$$\mathbb{M}, w \models [!p](Kp \land P(p) = 1 \land P(\neg p) = 0 \land S(p) = 0.95).$$

So after going outside, Mary comes to know that it is in fact snowing. She also adjusts her probabilities: she is now certain that it is snowing; i.e. she assigns probability 1 to p being true and probability 0 to p being false. These are the main *cognitive* effects of Mary's observation that it is snowing. However, on the *emotional* side, she is also

highly surprised to find out that it is snowing, because she initially considered this highly unlikely. These are the results that one would intuitively expect, so the semantic setup introduced above seems to yield an adequate representation of (the interactions between) the cognitive (epistemic and probabilistic) and emotional (surprise) effects of a public announcement.

4.2 Axiomatization

I now turn to the logic's proof theory. *Reduction axioms* are equivalences which allow us to push the public announcement operator through any of the other connectives, thus yielding an effective procedure to rewrite any formula as an equivalent formula that does not contain any dynamic operators. The reduction axioms for all operators of $\mathcal{L} - \{S\}$ are well-known (Kooi 2003, van Ditmarsch et al. 2007), cf. items 1–5 of Definition 4.6 below. What about reduction axioms for *S*? Recall that in Subsection 3.2 I suggested a dynamified (and temporally coherent!) version (9) of Macedo and Cardoso's original (1). With only minor modifications,¹³ this suggestion can be turned into reduction axioms for *S*; cf. items 6–7 below.

Definition 4.6. The reduction axioms for public announcement:

| 1. $[!\varphi]p \qquad \longleftrightarrow \qquad \varphi \rightarrow p (\text{for}$ | or $p \in Prop$) |
|---|---|
| 2. $[!\varphi] \neg \psi \qquad \longleftrightarrow \qquad \varphi \rightarrow \neg [!\varphi] \psi$ | |
| 3. $[!\varphi](\psi_1 \land \psi_2) \longleftrightarrow [!\varphi]\psi_1 \land [!\varphi]$ | $]\psi_2$ |
| 4. $[!\varphi]K\psi \qquad \longleftrightarrow \qquad \varphi \to K[!\varphi]\psi$ | |
| 5. $[!\varphi] \sum c_i P(\psi_i) \ge c \longleftrightarrow \varphi \to \sum c_i(\langle !\varphi \rangle \varphi)$ | $\langle \varphi \rangle \psi) \ge c P(\varphi)$ |
| 6. $[!\varphi]S(\psi) \ge c \qquad \longleftrightarrow \qquad \varphi \to P(\langle !\varphi \rangle \psi)$ | $(b) \leq 1 - c$ |
| 7. $[!\varphi]S(\psi) \le c \qquad \longleftrightarrow \qquad \varphi \to P(\langle !\varphi \rangle \psi)$ | $(b) \ge 1 - c$ |

We are now ready to axiomatize the logic of surprise.

Definition 4.7. SURPRISE is the logic axiomatized as follows:

- all of propositional logic,
- S5 for the knowledge operator K,
- three sets of axioms which are discussed in detail in (Demey and Sack forthcoming):

¹³Trivial modifications are that the statement about = needs to be 'split out' into statements about \leq and \geq , and that in the reduction axioms the argument of *S* should be an arbitrary formula ψ , and not just φ itself. A more serious modification is that the right sides of the reduction axioms should not contain simply $P(\psi)$, but rather $P(\langle !\varphi \rangle \psi)$, to 'pre-encode' the effect of the public announcement.

- the Kolmogorov axioms for the probability operator P,
- auxiliary axioms for linear inequalities,
- the axioms $K\varphi \to P(\varphi) = 1$ and $\varphi \to P(\varphi) > 0$ (these correspond to conditions (i) and (ii) in Definition 4.1),
- auxiliary axioms and rules for the surprise operator S:

$$\begin{aligned} &-S(\varphi) \ge 0, \\ &-S(\varphi) > 0 \to S(\varphi) \le 1, \\ &-\neg(S(\varphi) \le k \land S(\varphi) \ge k') \text{ for all } k < k', \\ &-S(\varphi) > 0 \to (S(\varphi) \ge k \lor S(\varphi) \le k), \\ &-\text{ if } \vdash \varphi \leftrightarrow \psi \text{ then } \vdash S(\varphi) \gtrless c \leftrightarrow S(\psi) \gtrless c \text{ (for } \gtrless \in \{\ge, \le\}), \end{aligned}$$

- necessitation for public announcement: if $\vdash \psi$, then $\vdash [!\varphi]\psi$,
- the reduction axioms for public announcement described in Definition 4.6.

This logic is sound and complete with respect to C_S (see Demey forthcoming b, for a proof). Note that the static axioms for surprise are all concerned with the technical details of this particular formalization of surprise (such as the totality of \geq), rather than with any substantial properties of surprise itself. The only substantial axioms for surprise are thus its reduction axioms (items 6–7 of Definition 4.6), which together constitute a dynamified version of Macedo and Cardoso's original definition (1). I take this to be a clear manifestion of the essentially dynamic nature of surprise in the axiomatization of the logic.

4.3 Some interesting modeling results

I will now show that the logical system developed in the previous subsections is able to capture several properties of surprise. However, there is one technical caveat. Recall that φ can only be publicly announced if it is true *before* the announcement. It is natural to assume that φ will still be true *after* the announcement. However, because public announcements take into account higher-order information, it might happen that φ , simply by being announced, becomes false. A typical example is $\varphi = p \land \neg Kp$. If no such 'self-falsifying' effects occur, φ is called successful:

Definition 4.8. A formula $\varphi \in \mathcal{L}$ is called *successful* iff $\models [!\varphi]\varphi$.

When modeling 'real-life' scenarios in a single-agent setting, formulas typically do not involve higher-order information,¹⁴ so at least from this modeling perspective, the assumption of successfulness in many of the propositions below is quite harmless. I now turn to the first concrete result.

Proposition 1. The following formula is satisfiable:

| φ | Λ | $\neg K\varphi$ | \wedge | $P(\varphi) = 0.2$ | \wedge | $S(\varphi) = 0$ |
|-----------|----------|--|----------|--------------------|----------|---------------------|
| | \wedge | $\langle ! \varphi \rangle (K \varphi$ | \wedge | $P(\varphi) = 1$ | \wedge | $S(\varphi) = 0.8)$ |
| | \wedge | $\langle ! \varphi \rangle \langle ! \varphi \rangle (K \varphi$ | \wedge | $P(\varphi) = 1$ | \wedge | $S(\varphi) = 0$). |

Proof. Consider $\mathbb{M} := \langle W, R, \sigma, \mu, V \rangle$, with $W = \{w, v\}, R = W \times W, V(p) = \{w\}, \mu(w)(w) = 0.2, \mu(w)(v) = 0.8$ and $\sigma(w)(X)$ and $\sigma(v)(X)$ undefined for all $X \subseteq w$ (all components which have not been mentioned are irrelevant, and can thus be assigned values at random). One can easily check that this is a surprise model, and that the formula stated above (with φ instantiated to p) is true at \mathbb{M}, w . Finally, note that $\mathbb{M} \in C_s^*$, i.e. we have followed the heuristic rule HEUR.

Proposition 1 shows that the logic is capable of doing what it was designed to do, viz. explicitly representing surprise dynamics. It describes the following scenario. Initially, φ is true, but the agent does not know this. Furthermore, she assigns rather low prior probability to it (and thus does not expect its announcement). However, because she does not yet know that φ is actually true, she experiences no surprise about it whatsoever. Next, the unexpected announcement of φ occurs, and three things happen: (i) the agent comes to know that φ , (ii) she processes this new information by Bayesian conditionalization and thus assigns probability 1 to it, and (iii) she experiences a very high degree of surprise about φ (inversely correlated to the low probability that she initially assigned to it). After another announcement of φ , the agent's knowledge and probabilities are not changed; however, because this second announcement was no longer unexpected (after all, in the meanwhile she has come to know that φ), her surprise about φ drops again to 0. The formula in Proposition 1 captures this scenario in a very natural way, using nested public announcement operators to explicitly represent the successive layers of surprise dynamics.

We now turn to Proposition 2 below. This says that an occurrence of φ can lead to surprise about φ itself, but also about all of its consequences. For example, it follows from items 1 and 2 that if an agent assigns probability 0.2 to $p \land q$, then after the announcement of this conjunction, she is surprised with intensity 0.8 about $p \land q$, but

¹⁴In a single-agent setting one is typically surprised about 'facts of nature', not about one's *epistemic attitudes about* such facts. In a multi-agent setting, however, it would be natural to have scenarios like "Alice was surprised when finding out that *Bob knows that* φ ".

also about p and q individually. Items 3 and 4 are trivial consequences of 1 and 2; they are mentioned to highlight the subtleties of unsuccessful formulas: if φ is not assumed to be successful, then 4 continues to hold, but 3 doesn't.

Proposition 2. Assume that $\varphi \in \mathcal{L}$ is successful, and that $\models \varphi \rightarrow \psi$. Then:

- 1. $\models P(\varphi) \ge c \rightarrow [!\varphi]S(\psi) \le 1 c,$
- 2. $\models P(\varphi) \le c \rightarrow [!\varphi]S(\psi) \ge 1 c,$
- 3. $\models P(\varphi) \ge c \rightarrow [!\varphi]S(\varphi) \le 1 c,$
- 4. $\models P(\varphi) \le c \rightarrow [!\varphi]S(\varphi) \ge 1 c.$

Proof. Straightforward applications of the semantics.

The fact that an occurrence of φ can lead to surprise about its consequences presupposes that the agent is actually able to *draw* those consequences (if the agent did not realize that ψ is a logical consequence of φ , then an unexpected occurrence of φ would cause her to be surprised about φ , but not about ψ). In other words, Proposition 2 shows that the logical system assumes the agent to be logically omniscient.¹⁵ An even clearer illustration of this assumption is provided by item 1 of Proposition 3 below, which says that the agent is never surprised about semantic validities. Similarly, items 2 and 3 say that if an agent already knows φ , or assigns probability 1 to it, then she will not be surprised about it. These principles are clearly false for actual human beings, which are not logically omniscient, and can e.g. be genuinely surprised upon learning (that some formula is actually) a tautology; rather, the main importance of item 1 is that it elucidates Wittgenstein's famous anti-psychologistic claim that "there can never be surprises in logic" (Proposition 6.1251 Wittgenstein 1922).

Proposition 3. Assume $\varphi \in \mathcal{L}$ is successful. Then:

- 1. if $\models \varphi$, then $\models [!\varphi]S(\varphi) = 0$,
- 2. $\models P(\varphi) = 1 \rightarrow [!\varphi]S(\varphi) = 0$,
- 3. $\models K\varphi \rightarrow [!\varphi]S(\varphi) = 0.$

Proof. Straightforward applications of the semantics.

¹⁵This also illustrates the thoroughly *epistemic* character of surprise: the problem of logical omniscience originally is a problem for epistemic logic, but it automatically carries over into the surprise component.

I will finish this subsection by proving two more substantial results, both of which illustrate how important empirical properties of surprise can be obtained as semantic validities of the logical system.

Proposition 4. Assume $\varphi \in \mathcal{L}$ is successful. Then for all $n \geq 2$, we have:¹⁶

$$\models [!\varphi]^n S(\varphi) = 0.$$

Proof. First of all, note that since φ is successful, it holds that $\models \varphi \leftrightarrow \langle !\varphi \rangle \varphi$; call this principle (†). Consider an arbitrary surprise model $\mathbb{M} = \langle W, R, \mu, \sigma, V \rangle$ and state *w*, and assume that $\mathbb{M}, w \models \varphi$. For any $n \ge 0$, we abbreviate

$$\langle W^n, R^n, \mu^n, \sigma^n, V^n \rangle = \mathbb{M} \upharpoonright n := (\cdots (\mathbb{M} \underbrace{\uparrow \varphi) \upharpoonright \varphi \cdots}_{n \text{ times}}) \cdot \varphi$$

Let's now show that $\mathbb{M}, w \models [!\varphi]^{n+1} P(\varphi) = 1$ for all $n \ge 0$. This follows directly from the following calculation:

$$\mu^{n+1}(w)(\llbracket \varphi \rrbracket^{\mathbb{M} \restriction n+1}) = \mu^{n+1}(w)(\llbracket \langle !\varphi \rangle \varphi \rrbracket^{\mathbb{M} \restriction n})$$

= $\mu^{n+1}(w)(\llbracket \varphi \rrbracket^{\mathbb{M} \restriction n})$ (†)
= $\mu^{n}(w)(\llbracket \varphi \rrbracket^{\mathbb{M} \restriction n}) = 1$. (Definition 4.4)

We now use this to justify the (‡)-labeled step in the following calculation:

$$\sigma^{n+2}(w)(\llbracket\varphi\rrbracket^{\mathbb{M}\lceil n+2}) = \sigma^{n+2}(w)(\llbracket\langle !\varphi\rangle\varphi\rrbracket^{\mathbb{M}\lceil n+1}) = \sigma^{n+2}(w)(\llbracket\varphi\rrbracket^{\mathbb{M}\lceil n+1}) \qquad (\dagger) = 1 - \mu^{n+1}(w)(\llbracket\varphi\rrbracket^{\mathbb{M}\lceil n+1}) \qquad (\text{Definition 4.4}) = 1 - 1 = 0. \qquad (\ddagger)$$

This shows that $\mathbb{M}, w \models [!\varphi]^{n+2}S(\varphi) = 0$ for all $n \ge 0$.

Informally speaking, Proposition 4 says that after two public announcements of φ , the agent is no longer surprised about φ . It thus nicely captures the transitory nature of surprise, which was discussed in Subsection 2.1. Furthermore, the proof closely resembles the informal explanation which was given there: the first announcement of φ causes the agent to update her probabilities and to assign probability 1 to φ , so that the second (and subsequent) announcement is no longer unexpected, and thus no longer surprising.¹⁷

| - | | |
|---|--|--|
| | | |
| | | |
| | | |

¹⁶[$!\varphi$]^{*n*} is defined inductively: [$!\varphi$]⁰ $\psi := \psi$, and [$!\varphi$]^{*n*+1} $\psi := [!\varphi][!\varphi]^n\psi$.

¹⁷The fact that surprise intensity drops to 0 after only two announcements is no problem for Proposition 4, even though for most real subjects this drop happens more gradually and requires several more repetitions (Charlesworth 1964). The more gradual decrease in surprise intensity is the consequence of personal and coincidental factors, such as intelligence and fatigue. Both the informal explanation in Subsection 2.1 and Proposition 4 make abstraction of such factors, and predict that the drop in surprise intensity will already happen after the second repetition.

Finally, Proposition 5 says that if an occurrence of (a public announcement of) φ leads an agent to change her probability of ψ from *a* to *b* in a non-trivial¹⁸ fashion, then she will experience at least *some* surprise about ψ . In other words: surprise is a *necessary condition* for belief revision (in the current framework: probability revision). This is perfectly in line with the cognitive-psychoevolutionary theory of surprise described in Subsection 2.1, which holds that surprise is part of a sequence of processes triggered by an unexpected event; the final stage of this sequence is typically a process of belief revision.

Proposition 5. Consider $\varphi, \psi \in \mathcal{L}$ and suppose that $\models \neg \psi \rightarrow [!\varphi] \neg \psi$. Then

$$\models (P(\psi) = a \land [!\varphi]P(\psi) = b \land a \neq b) \to [!\varphi]S(\psi) > 0.$$

Proof. Consider an arbitrary surprise model $\mathbb{M} = \langle W, R, \mu, \sigma, V \rangle$ and state *w*, and assume that the antecedent of the formula above is true at \mathbb{M} , *w*. For a reductio, assume that \mathbb{M} , *w* $\not\models [!\varphi]S(\psi) > 0$. Then it follows that

$$0 = \sigma^{\varphi}(w)(\llbracket \psi \rrbracket^{\mathbb{M} \restriction \varphi}) = \sigma^{\varphi}(w)(\llbracket \langle !\varphi \rangle \psi \rrbracket^{\mathbb{M}}) = 1 - \mu(w)(\llbracket \langle !\varphi \rangle \psi \rrbracket^{\mathbb{M}}),$$

and thus $\mu(w)(\llbracket \langle !\varphi \rangle \psi \rrbracket^{\mathbb{M}}) = 1$. From the assumption that $\models \neg \psi \rightarrow \llbracket !\varphi \rbrack \neg \psi$ in the statement of the proposition, it follows that $\llbracket \langle !\varphi \rangle \psi \rrbracket^{\mathbb{M}} \subseteq \llbracket \psi \rrbracket^{\mathbb{M}}$, and thus

$$1 = \mu(w)(\llbracket \langle !\varphi \rangle \psi \rrbracket^{\mathbb{M}}) \le \mu(w)(\llbracket \psi \rrbracket^{\mathbb{M}}) = a,$$

so a = 1. Since $\models \langle !\varphi \rangle \psi \to \varphi$, we similarly get that $\mu(w)(\llbracket \varphi \rrbracket^{\mathbb{M}}) = 1$, and hence

$$b = \mu^{\varphi}(w)(\llbracket \psi \rrbracket^{\mathbb{M} \upharpoonright \varphi}) = \mu^{\varphi}(w)(\llbracket \langle !\varphi \rangle \psi \rrbracket^{\mathbb{M}}) = \frac{\mu(w)(\llbracket \langle !\varphi \rangle \psi \rrbracket^{\mathbb{M}})}{\mu(w)(\llbracket \varphi \rrbracket^{\mathbb{M}})} = \frac{1}{1} = 1.$$

We thus have a = 1 = b, which contradicts the assumption that $a \neq b$.

Corollary 1. For any $\varphi \in \mathcal{L}$, it holds that

$$= (P(\varphi) = a \land [!\varphi]P(\varphi) = b \land a \neq b) \to [!\varphi]S(\varphi) > 0$$

Proof. It always holds that $\models \neg \varphi \rightarrow [!\varphi] \neg \varphi$, so by putting $\psi = \varphi$, the condition of Proposition 5 is always satisfied.

¹⁸This non-triviality requirement is captured by the condition that $\models \neg \psi \rightarrow [!\varphi]\neg \psi$, i.e. the public announcement of φ should not turn any $\neg \psi$ -states into ψ -states. In other words, the change of $P(\psi)$ from *a* to *b* is non-trivial if $[\![\psi]\!]^{\mathbb{M}}$ does not grow. (If $[\![\psi]\!]^{\mathbb{M}}$ grows, then it is trivial that the value of $P(\psi)$ might change: if $A \subseteq B$, then $P(A) \leq P(B)$.) Intuitively, exactly the same argument can be made about $[\![\psi]\!]^{\mathbb{M}}$ shrinking rather than growing (i.e. about the requirement that $\models \psi \rightarrow [!\varphi]\psi$), but it turns out that this second requirement is technically speaking not necessary for Proposition 5 to hold. This disanalogy is similar to the disanalogy between items 3 and 4 of Proposition 2.

4.4 A Lockean thesis for surprise

The current framework allows us to express statements such as 'the agent is surprised about φ with intensity 0.8'. In many natural cases, however, we might want to say that the agent is surprised about φ , without wishing to commit ourselves to some particular value for her surprise intensity. This is entirely analogous to the epistemic cases, where we might sometimes want to say that the agent believes that φ , without committing ourselves to some particular degree of belief.

A widespread proposal is to define 'belief' as 'sufficiently high degree of belief'; this proposal is called the *Lockean thesis* (Demey forthcoming a):

$$B\varphi :\equiv P(\varphi) \ge \tau \tag{10}$$

(here, $\tau \in (0.5, 1)$ is some threshold value).

Because of the high similarity between the epistemic case and the surprise case, it seems natural to apply the Lockean thesis also to surprise. In other words, we will introduce a 'qualitative' surprise operator by saying that the agent is surprised about φ iff she is surprised about φ with some sufficiently high intensity. I will argue below that the most natural choice for the value of the surprise intensity threshold is τ , i.e. the same value as the degree of belief threshold. Formally, the Lockean thesis for surprise thus looks as follows:

$$S\varphi :\equiv S(\varphi) \ge \tau.$$
 (11)

Principles (10) and (11) allow us to talk about an agent's 'qualitative' beliefs and surprises. Furthermore, since both principles make use of the same threshold value τ , there is a natural connection between the two operators they define. Proposition 6 says that after an announcement of φ , the agent will be surprised about ψ iff (assuming that φ is true) she initially believed that ψ would be false then. This qualitative observation is in line with the cognitive-psychoevolutationary theory of surprise described in Subsection 2.1, which holds that surprise stems from a conflict between unexpected data and a previously held belief.

Proposition 6. $\models [!\varphi]S\psi \iff \varphi \rightarrow B[!\varphi]\neg\psi.$

Proof. Consider the reduction axiom for surprise formulas:

$$[!\varphi]S(\psi) \ge \tau \iff \varphi \to P(\langle !\varphi \rangle \psi) \le 1 - \tau.$$
(12)

We have the following chain of SURPRISE-equivalences:

$$\begin{split} P(\langle !\varphi \rangle \psi) &\leq 1 - \tau \quad \leftrightarrow \quad 1 - P(\langle !\varphi \rangle \psi) \geq \tau \\ &\leftrightarrow \quad P(\neg \langle !\varphi \rangle \psi) \geq \tau \\ &\leftrightarrow \quad P([!\varphi] \neg \psi) \geq \tau \end{split}$$

and thus (12) can be rewritten as

 $[!\varphi]S(\psi) \ge \tau \iff \varphi \to P([!\varphi]\neg\psi) \ge \tau.$

Applying (11) and (10) to the left- and right-hand sides, respectively, yields the desired result. \Box

5 Conclusion

In this paper I have presented a new analysis of surprise in the framework of probabilistic dynamic epistemic logic. This analysis is based on current psychological theories, and as a result, several experimentally observed aspects of surprise can be derived as theorems within the logical system (recall, for example, Proposition 5 on the role of surprise in belief revision). Furthermore, being based on the contemporary 'lingua franca' of (dynamic) epistemic logic, it offers a natural, well-understood and highly expressive language for the formal description of agent architectures (cf. Proposition 1).

Most importantly, however, the analysis naturally captures the dynamic nature of surprise. This is clearly manifested in the logic's semantics (the surprise measures $\sigma(w)$ are not required to satisfy any static properties) as well as in its proof theory (the only substantial axioms for surprise are its reduction axioms). These reduction axioms jointly constitute a temporally coherent definition of surprise, in contrast to earlier, temporally incoherent formalizations such as Macedo and Cardoso's and Lorini and Castelfranchi's. This temporal coherence has several advantages. First and foremost, by explicitly distinguishing between prior and posterior notions, the proposed analysis is able to reach a high level of *conceptual hygiene*. This conceptual advantage also yields additional *empirical* benefits: the new analysis can capture important aspects of surprise that are not covered by earlier frameworks, such as its transitory nature (cf. Proposition 4).¹⁹

Acknowledgements A longer version of this paper will be published as (Demey forthcoming b); however, Subsection 4.4 is new. Earlier versions of this paper were presented at the LIRa seminar (ILLC, Amsterdam, October 2012), the Reasoning Club PhD conference (Brussels, September 2012), LOFT 10 (Sevilla, June 2012) and a workshop on modal logic (Brussels, May 2012). I would like to thank the audiences of these talks for their helpful remarks and suggestions. In particular, I would like to thank Alexandru Baltag, Johan van Benthem, Jan van Eijck, Jan Heylen, Emiliano

¹⁹Unsurprisingly, the aspect of transitoriness is itself of a highly dynamic character, involving repeated occurrences of the unexpected event.

Lorini, Alexandru Marcoci, Ahti-Veikko Pietarinen, Sonja Smets and Jean Paul Van Bendegem for their feedback on earlier versions of this paper. This research was supported by a PhD fellowship of the Research Foundation – Flanders (FWO).

References

J. van Benthem. *Exploring Logical Dynamics*. CSLI Publications, Stanford, CA, 1996.

J. van Benthem. *Logical Dynamics of Information and Interaction*. Cambridge University Press, Cambridge, 2011.

J. van Benthem, J. Gerbrandy, and B. P. Kooi. Dynamic update with probabilities. *Studia Logica*, 93:67–96, 2009.

W. R. Charlesworth. Instigation and maintenance of curiosity behavior as a function of surprise versus novel and familiar stimuli. *Child Development*, 35:1169–1186, 1964.

T. Y. Chow. The surprise examination or unexpected hanging paradox. *American Mathematical Monthly*, 105:41–51, 1998.

L. Demey. Agreeing to disagree in probabilistic dynamic epistemic logic. Master's thesis, ILLC, Universiteit van Amsterdam, Amsterdam, 2010.

L. Demey. Believing in Logic and Philosophy. PhD thesis, KU Leuven, 2013.

L. Demey. Contemporary epistemic logic and the Lockean thesis. *Foundations of Science*, forthcoming a.

L. Demey. The dynamics of surprise. Logique et Analyse, forthcoming b.

L. Demey. Agreeing to disagree in probabilistic dynamic epistemic logic. *Synthese*, forthcoming c.

L. Demey and J. Sack. Epistemic probabilistic logic. In H. van Ditmarsch, J. Y. Halpern, W. van der Hoek, and B. Kooi, editors, *Handbook of Logics for Knowledge and Belief*. College Publications, London, forthcoming.

H. van Ditmarsch, W. van der Hoek, and B. P. Kooi. *Dynamic Epistemic Logic*. Springer, Dordrecht, 2007.

R. Fagin and J. Halpern. Reasoning about knowledge and probability. *Journal of the ACM*, 41:340–367, 1994.

B. P. Kooi. Probabilistic dynamic epistemic logic. *Journal of Logic, Language and Information*, 12:381–408, 2003.

E. Lorini. Agents with emotions: A logical perspective. *Association for Logic Pro*gramming Newsletter, 21(2–3):1–9, 2008.

E. Lorini and C. Castelfranchi. The unexpected aspects of surprise. *Int. J. of Pattern Rec. and AI*, 20:817–833, 2006.

E. Lorini and C. Castelfranchi. The cognitive structure of surprise: Looking for basic principles. *Topoi*, 26:133–149, 2007.

L. Macedo and A. Cardoso. Creativity and surprise. In G. Wiggins, editor, *Proc. of the AISB '01 Symposium on Creativity in Arts and Science*, pages 84–92, York, 2001a. The Society for the Study of Artificial Intelligence and Simulation Behaviour.

L. Macedo and A. Cardoso. Modelling forms of surprise in an artificial agent. In J. Moore and K. Stenning, editors, *Proc. of the 23rd An. Conf. of the Cog. Sci. Soc.*, pages 588–593, Edinburgh, 2001b. Erlbaum.

L. Macedo and A. Cardoso. Exploration of unknown environments with motivational agents. In N. Jennings and M. Tambe, editors, *Proc. of the Third Int. Joint Conf. on Autonomous Agents and MAS*, pages 328–335, New York, NY, 2004. IEEE Computer Society.

L. Macedo, A. Cardoso, and R. Reisenzein. Modeling forms of surprise in artificial agents: Empirical and theoretical study of surprise functions. In K. Forbus, D. Gentner, and T. Regier, editors, *Proc. of the 26th An. Conf. of the Cog. Sci. Soc.*, pages 588–593, Mahwah, NJ, 2004. Erlbaum.

L. Macedo, A. Cardoso, and R. Reisenzein. A surprise-based agent architecture. In R. Trappl, editor, *Proc. of the 18th European Meeting on Cybernetics and Systems Research*, pages 583–588, Vienna, 2006. Austrian Society for Cybernetic Studies.

L. Macedo, A. Cardoso, R. Reisenzein, E. Lorini, and C. Castelfranchi. Artificial surprise. In J. Vallverdú and D. Casacuberta, editors, *Handbook of Research on Synthetic Emotions and Sociable Robotics: New Applications in Affective Computing and AI*, pages 267–291. IGI Global, Hershey, PA, 2009. L. Macedo, R. Reisenzein, and A. Cardoso. Surprise and anticipation in learning. In N. M. Seel, editor, *Encyclopedia of the Sciences of Learning*, pages 3250–3253. Springer, New York, NY, 2012.

W.-U. Meyer, R. Reisenzein, and A. Schützwohl. Towards a process analysis of emotions: The case of surprise. *Motivation and Emotion*, 21:251–274, 1997.

R. Reisenzein and W.-U. Meyer. Surprise. In D. Sander and K. R. Scherer, editors, *Oxford Companion to the Affective Sciences*, pages 386–387. Oxford University Press, Oxford, 2009.

R. Reisenzein, W.-U. Meyer, and A. Schützwohl. Reacting to surprising events: A paradigm for emotion research. In N. Frijda, editor, *Proc. of the 9th Conf. of the Int. Soc. for Research on Emotions*, pages 292–296, Toronto, 1996. ISRE.

J. Sack. Extending probabilistic dynamic epistemic logic. *Synthese*, 169:241–257, 2009.

J. Stiensmeier-Pelster, A. Martini, and R. Reisenzein. The role of surprise in the attribution process. *Cognition and Emotion*, 9:5–31, 1995.

W. Talbott. Bayesian epistemology. In E. N. Zalta, editor, *Stanford Encyclopedia of Philosophy*. CSLI, Stanford University, Stanford, CA, 2008.

L. Wittgenstein. *Tractatus Logico-Philosophicus*. Routledge and Kegan Paul, London, 1922.

A Geo-logical Solution to the Lottery Paradox, with Applications to Conditional Logic

Hanti Lin and Kevin T. Kelly

Australian National University, Carnegie Mellon University hanti.lin@anu.edu.au, kk3n@andrew.cmu.edu

Abstract

We defend a set of acceptance rules that avoids the lottery paradox, that is closed under classical entailment, and that accepts uncertain propositions without ad hoc restrictions. We show that the rules we recommend provide a semantics that validates exactly Adams' conditional logic and are exactly the rules that preserve a natural, logical structure over probabilistic credal states that we call *probalogic*. To motivate probalogic, we first expand classical logic to *geo-logic*, which fills the entire unit cube, and then we project the upper surfaces of the geo-logical cube onto the plane of probabilistic credal states by means of standard, linear perspective, which may be interpreted as an extension of the classical principle of indifference. Finally, we apply the geometrical/logical methods developed in the paper to prove a series of trivialization theorems against question-invariance as a constraint on acceptance rules and against rational monotonicity as an axiom of conditional logic in situations of uncertainty.

1 The Lottery Paradox

If Bayesians are right, one's credal state should be a probability measure p over propositions, where probabilities represent degrees of belief. It seems that one also *accepts* propositions in light of p. Acceptance of proposition A is sometimes portrayed as a momentous inference making A certain, in the sense that one would bet one's life against nothing that A is true (e.g., Levi 1967). But that extreme standard would eliminate almost all ordinary examples of accepted propositions. We therefore entertain a more modest view of acceptance, according to which the set of propositions accepted in light of p should, in some sense, aptly capture some characteristics of p to others or, in everyday cognition, to ourselves. That view is non-inferential in the sense that p is not conditioned on the propositions accepted, but it is inferential in another sense—the accepted propositions may serve as premises in arguments whose conclusions are also accepted in the same, weak sense.

It seems that high probability short of full certainty suffices for acceptance, a view now referred to as the *Lockean thesis*. But the Lockean rule licenses acceptance of inconsistent sets of propositions, however high the threshold r < 1 is set. For there exists a fair lottery with more than 1/(1-r) tickets. It is accepted that some ticket wins, since that proposition carries probability 1. But for each ticket, it is also accepted that the ticket loses, since that proposition has probability greater than r. So an inconsistent set of propositions is accepted. That is Henry Kyburg's *lottery paradox* (Kyburg 1961).

To elude the paradox, one must abandon either the full Lockean thesis or classical consistency. Kyburg pursued the second course by rejecting the classical inference rule that *A*, *B* jointly imply $A \land B$, so that the collection of propositions of form "ticket *i* does not win" does not entail "no ticket wins". Most responses side with classical logic and constrain the Lockean thesis in some manner to avoid contradictions. For example, Jeffrey (1970) recommended that the entire practice of acceptance be abandoned in favor of reporting probabilities. Levi (1967) rejected the idea that acceptance can be based on probability alone, since utilities should also be consulted. Or one may impose as a necessary condition that accepted propositions be certain (van Fraassen 1995, Arló-Costa and Parikh 2005) or to cases in which no logical contradiction happens to result (Pollock 1995, Ryan 1996, Douven 2002).

Our approach is different. Instead of restricting the Lockean thesis, we *revise* it. In particular, we defend an *unrestricted* rule of acceptance that is *contradiction-free* and yet capable of accepting *uncertain* propositions—even propositions of fairly low probability. Like the Lockean rule, the proposed rule has a parameter that controls its strictness. When the parameter is tuned toward 1, the proposed rule is almost indistinguishable from the classical logical closure of the Lockean rule; but as the parameter drops toward 0, the proposed rule's geometry shifts steadily away from that of the Lockean rule so as to avert the lottery paradox.

The rule we recommend was invented by (Levi 1996, p.286), who saw no justification for it except as a loose approximation to an alternative rule he took to be justified by decision-theoretic means (Levi 1967; 1969).¹ We provide a justification of

¹Levi writes: "I do not know how to derive it from a view of the cognitive aims of inquiry [i.e. seeking more information and avoiding error] that seems attractive." (Levi 1996, p.286) We rediscovered the rule as

Lin and Kelly

the rule in terms of *preservation of logical structure* implicit in the space of probabilistic credal states. The crux is to order probabilistic credal states according to relative logical strength, as Boolean algebra does for propositions. We do so in two steps. First, we start with a sigma algebra of propositions (closed under negation and countable disjunction) and then *extend* that sigma algebra to cover the entire unit cube by introducing a new connective \neg_d interpreted as negation to *degree d*, so that $\neg_0\varphi$ is equivalent to φ and $\neg_1\varphi$ is equivalent to φ . The resulting logical structure is called *geologic* (Section 4). Next, we *view* the geological structure *in perspective* through the *picture plane* of possible credal states to obtain a logical structure over credal states that we call *probalogic* (Sections 5 and 6). Then it is natural to require that every acceptance rule preserves probalogical structure when it maps probabilistic credal states to standard, Boolean propositions.

The requirement that acceptance rules preserve probalogical structure has appealing consequences for the theory of acceptance. First, we show that the rules we recommend are exactly the rules that preserve probalogical structure (Section 7). Moreover, no plausible logical structure on probability measures is preserved by the Lockean rule or its variants (Section 8).

Another justification of the proposed acceptance rules concerns the logic of conditionals and defeasible reasoning. Frank P. Ramsey proposed an influential, epistemic condition for acceptance of conditional statements, now commonly referred to as the *Ramsey test*:

If two people are arguing 'If A, then B?' and are both in doubt as to A, they are adding A hypothetically to their stock of knowledge and arguing on that basis about B; so that in a sense 'If A, B' and 'If A, $\neg B$ ' are contradictories. We can say that they are fixing their degrees of belief in B given A. (Ramsey 1929, footnote 1)²

Suppose that an agent is in a probabilistic credal state p and adopts an acceptance rule. We propose the following interpretation of the Ramsey test: the agent accepts the (flat) conditional 'if A then B' when, by the acceptance rule she adopts, she would accept B in the credal state $p(\cdot|A)$ that results from p by conditioning on A. Thus, conditional acceptance is reduced to Bayesian conditioning and acceptance of non-conditional propositions. This natural semantics allows one to characterize the axioms of conditional logic in terms of their geometrical constraints on acceptance rules, in

a consequence of our work on Ockham's razor. The problem was to extend the Ockham efficiency theorem (Kelly 2008) from methods that choose theories to methods that update probabilistic degrees of belief on theories. That required a concept of retraction of credal states, expounded in (Kelly 2010). We thank Teddy Seidenfeld for bringing the prior publication of the rule to our attention.

²We take the liberty of substituting A, B for p, q in Ramsey's text.

much the same way that axioms of modal logic are standardly characterized in terms of constraints on accessibility relations among worlds. Accordingly, for each of the axioms in Adams' conditional logic (Adams 1975), we solve for its geometrical constraint on acceptance rules (Section 9). These constraints are shown to be satisfied by the rules that preserve probalogical structure, so the probalogic-preserving rules validate Adams' logic with respect to the Ramsey test (Section 10). Conversely, Adams' logic is shown to be *complete* with respect to the Ramsey test when acceptance follows probalogic-preserving rules (Section 12). The result is a new probabilistic semantics: it defines validity simply as preservation of acceptance, which improves upon Adams' ϵ - δ semantics (Adams 1975); and it allows for accepting propositions of low probabilities, which improves upon Pearl's infinitesimal semantics (Pearl 1989). Thus, the recommended acceptance rules are vindicated both by probalogic and by conditional logic.

One might hope for validating a stronger logic of flat conditionals than Adams', e.g. system R (Lehmann and Magidor 1992) or, slightly stronger, the AGM axioms for belief revision (Harper 1975, Alchourrón et al. 1985). We close the door on that hope with a new trivialization theorem (Section 11). In light of that result, we propose that Adams' conditional logic reflects Bayesian ideals better than AGM belief revision does.

Finally, the acceptance rules we recommend are sensitive to framing effects determined by an underlying question. One might hope that the advantages of the proposed rules could be obtained without question-dependence. Again, we close the door on that hope with a series of trivialization theorems (Sections 13 and 14), employing the geometrical techniques described above. We conclude that, all things considered, the advantages of the recommended acceptance rules within questions justify their dependence on questions.

2 The geometry of the Lottery Paradox

Let $\mathcal{E}_{\kappa} = \{E_i : i \in I\}$ be a countable collection of mutually exclusive and jointly exhaustive propositions over some underlying set of possibilities, where κ (either ω or some finite *n*) is the cardinality of the index set I.³ We think of \mathcal{E}_{κ} as a *question* in context whose *potential answers* are the various E_i . Let \mathcal{A}_{κ} be the least collection of propositions containing \mathcal{E}_{κ} that is closed under negation and countable disjunction and conjunction, and let \mathcal{P}_{κ} denote the set of all (countably additive) probability measures defined on \mathcal{A}_{κ} . We think of \mathcal{P}_{κ} as the space of *probabilistic credal states* over answers

³Note that in the mathematics that follows, we never distinguish between questions of a given cardinality, so no confusion results from identifying questions in terms of cardinality.

Lin and Kelly

to question \mathcal{E}_{κ} . The subscripts are suppressed in the sequel unless we wish to emphasize cardinality.

We assume that acceptance rules produce sets of propositions that are closed under classical entailment so that, without loss of generality, each acceptance rule may be viewed as a map $\alpha : \mathcal{P} \to \mathcal{A}$, where proposition $\alpha(p)$ is understood as the strongest proposition accepted in light of probability measure *p*. Then proposition *A* is *accepted* by rule α at credal state *p*, written $p \Vdash_{\alpha} A$, if and only if $\alpha(p)$ entails *A*. The *acceptance zone* of *A* under α is defined as the set of all credal states at which *A* is accepted by α .

For example, the *Lockean* acceptance rule with threshold set to *r* in the unit interval is just the mapping:

$$\lambda_r(p) = \bigwedge \{A \in \mathcal{A} : p(A) \ge r\}.$$
 (1)

Each probability measure p in \mathcal{P} can be represented with respect to \mathcal{E} as the κ dimensional vector $(p(E_i) : i \in I)$ with components in the unit interval summing to one. In the context of question \mathcal{E} , we identify p with its vector, so that the *i*-th component p_i equals $p(E_i)$. When $\kappa = 3$, for example, \mathcal{P}_3 corresponds to the set of all such 3-vectors, which is the equilateral triangle in \mathbb{R}^3 whose corners have Cartesian coordinates $e_1 = (1, 0, 0), e_2 = (0, 1, 0)$ and $e_3 = (0, 0, 1)$ (Figure 1). To avoid ambiguity, we let $(e_i)_i$ pick out the *j*-th component of e_i . Reformulate the Lockean rule (1) as



Figure 1: The space \mathcal{P}_3 of probabilistic credal states

follows:4

$$\lambda_r(p) = \bigwedge \{\neg E_i : p(\neg E_i) \ge r \text{ and } i \in I\};$$
(2)

$$= \bigwedge \{\neg E_i : p_i \le 1 - r \text{ and } i \in I\}.$$
(3)

By this formulation, the acceptance zone of $\neg E_1$ under λ_r with respect to question \mathcal{E}_3 is depicted in Figure 2. The Lockean rule is now expressed *geometrically*—its



Figure 2: Acceptance zone for $E_2 \vee E_3$ under λ_r

acceptance zone for $\neg E_1$ has a definite, trapezoidal *shape* that results from truncating the triangular space \mathcal{P}_3 parallel to one side. As threshold *r* is dropped, the trapezoid becomes thicker. The acceptance zones of $\neg E_2$ and $\neg E_3$ are included in Figure 3.a. By closure under entailment, proposition E_1 is accepted exactly when both $\neg E_2$ and $\neg E_3$ are accepted, so the corner, diamond-shaped zones license acceptance of potential answers to \mathcal{E} . When $r \leq 2/3$, the propositions $\neg E_1, \neg E_2, \neg E_3$ are all accepted at the probability measures contained in the small, dark, central triangle (Figure 3.b). But that set of propositions is inconsistent so, by closure under entailment, the dark, central triangle is the acceptance zone of the inconsistent proposition \bot . That is just the lottery paradox for thresholds $r \leq 2/3$ (interpret E_i as the proposition "ticket *i* wins").

Geometrically, the lottery paradox arises because the Lockean rule's acceptance zones for the various propositions $\neg E_i$ crash clumsily into one another as the probability threshold *r* decreases. It is easy to design alternative acceptance zones that *bend*

⁴This is equivalent to the original formulation because, first, every proposition A is equivalent to the conjunction of all propositions of form $\neg E_i$ that are entailed by A and, second, propositions of form $\neg E_i$ that are entailed by A are at least as probable as A.



Figure 3: Acceptance zones under λ_r

progressively as they approach the center of the triangle so that they eventually meet without overlapping like the leaves of a camera shutter (Figure 4). The proposed ac-



Figure 4: Progressively bent zones that avert collision

ceptance zones are almost indistinguishable from those of the Lockean rule when r is close to 1. As r approaches 0, the bending becomes more pronounced and the lottery paradox is avoided.

A special, symmetric case of the proposed rule, which we call the *symmetric camera shutter rule*, modifies the Lockean rule as follows. Test whether answer E_i to \mathcal{E} should be rejected at credal state p by considering, not probability p_i itself, but the probability ratio:

$$\sigma(p)_i = \frac{p_i}{\max_j p_j},$$

resulting in the modified rule:

$$\nu_r(p) = \bigwedge \{\neg E_i : \sigma(p)_i \le 1 - r \text{ and } i \in I\}.$$
(4)

The symmetric camera shutter rule is algebraically the same as the Lockean rule (3) except that probability is divided by the probability distribution's mode. Say that acceptance rule α is *everywhere consistent* if and only if $p \nvDash_{\alpha} \perp$ for each p in \mathcal{P} , and say that α is *non-skeptical* if and only if for each E_i in \mathcal{E} there exists p in \mathcal{P} such that $p(E_i) < 1$ and $p \Vdash_{\alpha} E_i$. Then:

Proposition 1. Let \mathcal{E} contains at least two answers. The symmetric camera shutter rule v_r is everywhere consistent and non-skeptical, for each r such that 0 < r < 1.

Proof. For everywhere consistency, note that since $\sum_i p_i = 1$, so there exists $i \in I$ such that $p_i = \max_i p_j$. Then, since r > 0,

$$\tau(p)_i = 1 \nleq 1 - r,$$

(

so $p \nvDash_{v_r} \neg E_i$, by formula (4). It follows that $p \nvDash_{v_r} \bot$. For non-skepticism, let E_i be an arbitrary answer, and it suffices to show that E_i is accepted by v_r at some credal state p such that $p_i < 1$. Let $p_i = 1/(2 - r)$. Since \mathcal{E} contains at least two answers, choose j in I distinct from i and let $p_j = (1 - r)/(2 - r)$. Since a probability distribution is normalized, $p_k = 0$ for all $k \neq i, j$. Note that p_i is the mode of p, since r > 0. So for each $k \neq i$:

$$\sigma(p)_k \leq 1 - r,$$

$$\sigma(p)_i = 1 \leq 1 - r$$

since r > 0. Hence $p \Vdash_{v_r} E_i$, by formula (4), with $p_i = 1/(2 - r) < 1$, since r < 1. \Box

On the other hand:

Proposition 2. Suppose that \mathcal{E} is countably infinite. The Lockean rule λ_r is either skeptical or somewhere inconsistent, for each r such that $0 \le r \le 1$.

Proof. If Lockean rule is not skeptical, then r < 1, and thus there exists p in \mathcal{P}_{ω} such that $p_i \leq 1-r$, for each $i \in I$. So by formula (3), $\lambda_r(p) = \bot$, and hence λ_r is somewhere inconsistent.
3 Respect for logic

The range of acceptance rule $\alpha : \mathcal{P} \to \mathcal{A}$ has a natural, Boolean logical structure:

$$(\mathcal{A},\leq,\vee,\wedge,\perp,\top),$$

where the partial order \leq corresponds to classical entailment or relative strength of propositions and \lor and \land are the least upper bound and the greatest lower bound with respect to \leq , which correspond to the usual propositional operations of disjunction and conjunction.⁵ If there were also a motivated logical structure on the space:

$$(\mathcal{P},\leq,\vee,\wedge,\perp,\top)$$

of probabilistic credal states, in which \leq is intended, again, to reflect relative strength, then an obvious constraint on acceptance rules would be to *preserve* logical structure in the sense that:

$$p \le q \implies \alpha(p) \le \alpha(q);$$
 (5)

$$\alpha(p \lor q) = \alpha(p) \lor \alpha(q); \tag{6}$$

$$\alpha(p \wedge q) = \alpha(p) \wedge \alpha(q); \tag{7}$$

$$\alpha(e_i) = E_i; \tag{8}$$

$$\alpha(\top) = \top; \tag{9}$$

$$\alpha(\bot) = \bot. \tag{10}$$

Any plausible logical structure over \mathcal{P} should also satisfy the following constraint:

the unit vectors e_i , for $i \in I$, are exactly the strongest credal states in \mathcal{P} . (11)

Then we already have the following assurance against inconsistency:

Proposition 3 (No lottery paradox). Suppose that acceptance rule α and relative strength \leq over \mathcal{P} satisfy conditions (5), (8), and (11). Then α is everywhere consistent.

Proof. Suppose for reductio that for some credal state p, $\alpha(p) = \bot$. Then, by condition (11), there exists a strongest state e_i such that $e_i \le p$. So $\alpha(e_i) \le \alpha(p)$, by (5). Then by (8), we have that $E_i = \alpha(e_i) \le \alpha(p) = \bot$. So $E_i \le \bot$, which is false in the Boolean logical structure of \mathcal{A} .

⁵In algebraic logic, $A \le B$ means that A is at least as strong as B.

Therefore, the lottery paradox witnesses the failure of the Lockean rule to preserve logical structure. But the lottery paradox is only the most glaring consequence of the Lockean rule's disrespect for logical structure. It is plausible to suppose that with respect to question \mathcal{E} , if credal state *p* accords maximal probability to answer E_i , compared to all the alternative answers to \mathcal{E} , and if E_j is a distinct answer to \mathcal{E} , then credal state $p(\cdot|\neg E_j)$ is at least as strong as *p*:

$$p_i = \max_k p_k \text{ and } E_i \neq E_j \implies p(\cdot | \neg E_i) \le p.$$
 (12)

But then the Lockean rule again fails to preserve relative strength, i.e., it violates condition (5). Recall from Figure 3 that a consistent Lockean rule's acceptance zone for E_2 is a diamond. The diamond has the wrong shape—its sides meet at an angle that is too acute. For consider a credal state p very close to the inner apex of the diamond, as depicted in Figure 5. Let $q = p(\cdot|\neg E_3)$. By condition (12), we have that $q \le p$. But



Figure 5: Deeper trouble for the Lockean rule

point *q* lies on the side of the triangle opposite e_3 because $q_3 = 0$, and *q* lies on the ray from e_3 that passes through *p* because $q_1/q_2 = p_1/p_2$. So $\lambda(q) = \neg E_3 \nleq E_2 = \lambda(p)$. Therefore, $q \le p$ but $\lambda(q) \nleq \lambda(p)$, which violates (5). That is another, counterintuitive way to fail to preserve logical order even when the lottery paradox does not arise.

The preceding argument illustrates a further point: intuitions about relative strength of credal states are tied to conditioning. The boundaries of acceptance zones determined by the Lockean rule do not follow the geometrical rays that correspond to the trajectories of probabilistic credal states under conditioning. For that reason, the Lockean rule is a bad choice for trying to explicate the acceptance of conditionals in terms of conditional probabilities. Specifically, consider the following interpretation of the Ramsey test. Let *A*, *B* be arbitrary propositions in \mathcal{A} . Let the *flat conditional* with antecedent *A* and consequent *B* be expressed by $A \Rightarrow B$. (The arrow notation ' \Rightarrow ' does

not denote any binary operation in any algebra; so mathematically, $A \Rightarrow B$ is simply the ordered pair (A, B). Instead, $A \Rightarrow B$ is used to indicate that we are talking about the acceptance condition of a flat conditional with antecedent A and consequent B.) We propose that the Ramsey test be explicated by the following acceptance condition for flat conditionals:⁶

$$p \Vdash_{\alpha} A \Rightarrow B \iff p(\cdot|A) \Vdash_{\alpha} B \text{ or } p(A) = 0.$$
 (13)

So when the antecedent has nonzero probability, this semantic rule says that flat conditional $A \Rightarrow B$ is accepted at credal state p if and only if, when one "adds A hypothetically to one's stock of knowledge" and thereby hypothetically conditions p on A to obtain posterior credal state $p(\cdot|A)$, one accepts the consequent B in the posterior state. Consider again the consistent Lockean rule λ and credal state p in Figure 5. Then, as evident from the picture, we have:

$$p \Vdash_{\lambda} E_2,$$
$$p \nvDash_{\lambda} \neg E_3 \Rightarrow E_2.$$

Note that $\neg E_3$ is entailed by E_2 , so Lockean rule λ instructs one to retract her acceptance of E_2 when a logical consequence of E_2 is learned or supposed. That allows one to retract acceptance too easily and violates almost all logics of conditionals that have interested logicians, including Adams' conditional logic (Adams 1975).⁷ On the other hand, to preserve acceptance under logically entailed information, it suffices to require conditions (5) and (12). For by (12), credal state q would have been at least as strong as credal state p and hence, by (5), any proposition accepted in p remains accepted in q, e.g., E_2 .

The angles formed by the sides of the acceptance zones are crucial to the preservation of logical structure. The acceptance rules we recommend—the camera shutter rules—do have acceptance zones with the correct angles at their corners and, therefore, do not encounter any of the preceding logical difficulties. We will show that the camera shutter rules preserve a very natural logical structure on state space \mathcal{P} and, therefore, yield a soundness and completeness theorem for Adams' conditional logic that is simpler than more natural than Adams' original 1975 version.

⁶If $p(A) \neq 0$, $p(\cdot|A)$ is defined to be $p(\cdot \wedge A)/p(A)$; otherwise it is undefined.

⁷Specifically, the principle that acceptance be preserved under logically implied information can be shown to be equivalent to Cautious Monotonicity, given two other axioms in Adams' conditional logic (a.k.a system P: Right Weakening, And. That system will be discussed in detail latter.

4 Geologic

Consider classical, infinitary propositional logic, which allows for countable disjunction and conjunction.⁸ Start with propositional constants \bot , \top and propositional variables $V_{\kappa} = \{E_i : i \in I\}$, where the countable index set *I* has cardinality κ . Let $\bigvee_j \varphi_j$ and $\bigwedge_j \varphi_j$ be countable disjunction and conjunction, respectively. Let language L_{κ} be the least set containing the propositional constants in V_{κ} that is closed under negation, countable disjunction, and countable conjunction. We interpret the propositional variables to be mutually exclusive and exhaustive. Under that restriction, each assignment is an κ -dimensional basis vector e_i . Let \mathcal{B}_{κ} denote the set of all such vectors. The valuation function for classical logic is definable as follows. In the base case:

$$v_{e_i}(E_j) = e_i \cdot e_j; \quad v_{e_i}(\top) = 1; \quad v_{e_i}(\bot) = 0,$$

where \cdot denotes the vector *inner product* $x \cdot y = \sum_{i \in I} x_i y_i$. In the inductive case:

$$v_{e_i}(\neg \varphi) = 1 - v_{e_i}(\varphi); \quad v_{e_i}(\bigvee_j \varphi_j) = \max_j(v_{e_i}(\varphi_j)); \quad v_{e_i}(\bigwedge_j \varphi_j) = \min_j(v_{e_i}(\varphi_j))$$

Logical entailment is definable in terms of valuation as follows:

$$\varphi \models \psi \quad \iff \quad v_{e_i}(\varphi) \le v_{e_i}(\psi), \text{ for all } i \in I.$$

Let the *proposition* $[\![\varphi]\!]_{\kappa}$ expressed by φ in language L_{κ} denote the set of all assignments in \mathcal{B}_{κ} in which φ evaluates to 1. Each proposition $[\![\varphi]\!]_{\kappa}$ is represented uniquely by its *valuation vector*:

$$v_{\kappa}(\varphi) = (v_{e_i}(\varphi) : i \in I),$$

which belongs to 2^{κ} . Define the following relations and operations over 2^{κ} :

$$u \le v \iff u_i \le v_i$$
, for all *i* in *I*; (14)

$$(\neg v)_i = 1 - v_i; \tag{15}$$

$$(\bigvee_{j} v^{j})_{i} = \max_{j} v_{i}^{j};$$
(16)

$$(\bigwedge_{i} v^{j})_{i} = \min_{j} v_{i}^{j}.$$
(17)

Then the structure of classical, infinitary logic is captured⁹ by the mathematical structure:

$$\mathcal{L}_{\kappa} = (2^{\kappa}, \leq, \bigvee, \bigwedge, 1, 0).$$

⁸For classic studies concerning completeness infinitary logic, cf. (Karp 1964) and (Barwise 1969). Our applications make no reference to completeness or to proof systems for infinitary logic.

⁹I.e., the Lindenbaum-Tarski algebra of language L_{κ} is isomorphic to \mathcal{L}_{κ} .



Figure 6.a illustrates \mathcal{L}_3 , which bears a suggestive resemblance to the unit cube $[0, 1]^3$

Figure 6: Bead-and-string logic vs. geologic

(Figure 6.b), but it is really just a string-and-bead figure whose strings happen to be sized and stretched to outline a cube. However, one can *extend* classical propositional logic on L_{κ} to a *fuzzy* language L_{κ}^{*} that generates fuzzy propositions covering the entire κ -dimensional unit cube $[0, 1]^{\kappa}$.¹⁰ A *fuzzy proposition* is just a *fuzzy subset* (Zadeh 1965) of \mathcal{B}_{κ} , which is representable by a *fuzzy characteristic function* from \mathcal{B}_{κ} to [0, 1] and, hence, by a *fuzzy valuation vector* v in $[0, 1]^{\kappa}$. Formula (14) represents the fuzzy subset relation and formulas (15) through (17) correspond to fuzzy complement, intersection, and union over fuzzy propositions.

Here is one natural way to extend classical logic over L_{κ} to cover the κ -dimensional unit cube. For each real number *d* in the unit interval, let the *partial negation* $\neg_d \varphi$ be understood as the negation of φ to *degree d*, interpreted as follows:

$$v_{e_i}(\neg_d \varphi) = d v_{e_i}(\neg \varphi) + (1 - d) v_{e_i}(\varphi).$$

In particular, $\neg_0 \varphi$ is equivalent to φ , whereas $\neg_1 \varphi$ is equivalent to $\neg \varphi$. Between these extremes, $\neg_{1/2} \varphi$ hovers semantically midway between φ and $\neg \varphi$. Let L_{κ}^* be the result of expanding language L_{κ} with \neg_d . Otherwise, the preceding definitions of valuation

¹⁰The idea may sound similar to multi-valued logic, but it is quite different. In multi-valued logic, (discrete) logical formulas in L_{κ} are interpreted over an expanded, continuous space of assignments (Novak et al. 2000)—such logics generate a discrete, weakening of classical logic, rather than a continuous, conservative extension of classical logic.

function v_{e_i} and valuation vector $(v_{e_i}(\varphi) : i \in I)$ remain unaltered.¹¹ Partial negation never generates values outside of the unit interval, so all valuation vectors for L_{κ}^* are in the unit cube $[0, 1]^{\kappa}$. Conversely, every vector v in $[0, 1]^{\kappa}$ is the valuation vector of some formula in L_{κ}^* , namely:

$$\bigvee_{i\in I}(\neg_{1-\nu_i}E_i\wedge E_i).$$

So the propositions expressible by the fuzzy language \mathcal{L}_{κ}^* correspond to the vectors in the κ -dimensional unit cube $[0, 1]^{\kappa}$. Therefore, we refer to the logic just defined as *geologic*.

Formulas (14) to (17) still make sense for fuzzy valuation functions (because they correspond to the standard definitions of the fuzzy set theoretic operations). Therefore, the *structure* of geologic is:

$$\mathcal{L}^*_{\kappa} = ([0,1]^{\kappa}, \leq, \bigvee, \bigwedge, 1, 0).$$

Since the valuation definition for geologic is exactly the same as for classical logic over the fragment L_{κ} , it follows that \mathcal{L}_{κ}^* restricted to L_{κ} is just \mathcal{L}_{κ} —in other words, geologic is a conservative extension of classical, infinitary logic.

Since the operations in \mathcal{L}_{κ}^* correspond to fuzzy set theoretical operations on propositions, it is immediate that the geological operations satisfy associativity, commutativity, distributivity, and the De Morgan rules (Zadeh 1965). Excluded middle and disjunctive syllogism, on the other hand, can fail spectacularly for propositions in the unit cube's interior. For example, let *c* denote the center $(\frac{1}{2}, \ldots, \frac{1}{2}, \ldots)$ of the unit cube. Then:

$$\neg c = c;$$

$$c \lor \neg c = c;$$

$$(e_1 \lor c) \land \neg c = c.$$

In spite of that, we think of geologic as the natural *extension* of classical logic to fuzzy propositions. Associativity, commutativity, distributivity, and the De Morgan rules are all motivated by symmetries of the unit cube. Excluded middle is not motivated by symmetry—it is a mere artifact of an impoverished syntax. Furthermore, unlike modal logic, which is also a conservative extension of classical propositional logic, geologic arises from the addition of a *truth-functional* negation.

¹¹More directly, one can simply introduce a new unary connective $a\varphi$ called *scalar multiple* interpreted by $v_{e_i}(a\varphi) = av_{e_i}(\varphi)$. But we found it harder to motivate usage of such a connective.

Filling the interior of the Boolean algebra to make it a genuine cube provides an explanatory, geometrical perspective on classical logic. For example, given points v and u in the unit cube, find the smallest parallelepiped solid S(v, u) containing v and u whose sides are parallel to the sides of the cube. Then the uppermost vertex of S(v, u) is $v \lor u$ and the lowermost vertex of S(v, u) is $v \land u$ (Figure 7.a). The parallelepiped S(v, u) is like a sub-crystal within the cube, which is another reason for thinking of geologic as *geo*logical.¹² The geometry of full geological negation is just reflection through the



Figure 7: Geological operations

center *c* of the cube, which is a natural generalization of Boolean complementation. To construct the partial negation $\neg_d v$ of *v*, first reflect *v* through *c* to obtain the full negation $\neg v$. Now draw a straight line segment between *v* and $\neg v$. Then $\neg_d v$ is the point that lies proportion *d* of the way from *v* to $\neg v$ along the line segment (Figure 7.b). Consider the classical De Morgan rules. Since full negation involves projection through the center *c* of the cube, think of *c* as the aperture of a pinhole camera. It is a familiar fact that projection through an aperture inverts the image. But the disjunction $v \lor u$ is the top vertex of the parallelepiped spanning *v* and *u*. Projecting the parallelepiped through the aperture inverts it and turns the top vertex into the bottom vertex—the conjunction of the projection of *v* with the projection of *u* (Figure 8).

¹²Note that the same geometrical relationships would hold even if the unit cube were stretched along its various axes to form a prism. We will return to that theme in the last section of the paper.



Figure 8: Geometry of the De Morgan Rules

5 Logic from a probabilistic perspective

For our purposes, the point of geologic is that it affords a unified perspective on logic and probability.¹³ The set \mathcal{P}_3 of possible credal states is a horizontal, triangular plane through the unit cube of geological propositions (Figure 6.b). Thus, credal state space \mathcal{P}_3 has a natural embedding *within* geologic. That embedding generalizes to each countable cardinality κ .

Valuation and probability assignment can both be viewed as inner products within the geological cube:

$$v_{e_i}(u) = e_i \cdot u,$$

$$p(b) = p \cdot b,^{14}$$

where *u* is a vector in $[0, 1]^{\kappa}$ corresponding to an arbitrary, geological proposition, v_{e_i} is a valuation function corresponding to a classical assignment, *b* is a Boolean valuation vector in 2^{κ} corresponding to a classical proposition, and *p* is a probability measure/vector in \mathcal{P}_{κ} .

Say that probability measure p is *uniform* with respect to \mathcal{E} if and only if p assigns only zero or a fixed value to the answers in \mathcal{E} . The *support* of p is the disjunction of all elements of \mathcal{E} that p assigns non-zero probability to (recall that \mathcal{E} is countable).

¹³Due to the truth-functionality of conjunction in fuzzy logic, the fuzzy logic community tends to view fuzzy logic in isolation from probability theory, rather than as a tool for understanding probability theory, as we propose.

The *classical principle of indifference* is a mapping σ that associates each uniform probability distribution p with its support. For example, σ associates $(\frac{1}{2}, \frac{1}{2}, 0)$ with the classical proposition $\sigma(\frac{1}{2}, \frac{1}{2}, 0) = (1, 1, 0)$. Construct a ray from \perp through uniform distribution p and then $\sigma(p)$ is the (classical) proposition on the upper surface of the unit cube that the ray points to (Figure 9.a). Algebraically, $\sigma(p)$ is the (unique) scalar



Figure 9: Indifference as projection

multiple of p in the unit cube that has at least one component equal to 1, which amounts to the formula encountered earlier in the definition of the symmetric camera shutter acceptance rules:

$$\sigma(p)_i = \frac{p_i}{\max_i p_i}, \text{ for } i \in I.$$

Say that geological proposition *u* is *fully satisfiable* if and only if there exists e_i such that $v_{e_i}(u) = 1$, i.e. *u* has a component equal to 1. So $\sigma(p)$ is the (unique), fully satisfiable, geological valuation vector that is proportional to *p*. In classical logic, the mapping $\sigma(p)$ is defined only for uniform *p*, but it is defined for *all p* in geologic, since the (continuous) upper surface of the geological cube covers the entire triangle of probability measures (Figure 9.b). Now *every* probability measure *p* has a unique, geological proposition $\sigma(p)$ that stands to *p* in much the same way that the support of *p* stands to uniform *p*.

Mapping σ has a heuristic interpretation. Think of the unit cube as a room with tiled walls (Figure 10). Imagine that there is a digital camera embedded in the baseboard of

the room at corner \perp . Think of the triangle \mathcal{P}_3 as the picture plane corresponding to the 2-dimensional image received by the camera. Then the inverse σ^{-1} of σ is the classical perspective rendering of the room's interior on the picture plane. The perspective is extreme because the camera is literally embedded in the lower corner of the room, so the floor and adjacent walls are tangent to the camera's view and are rendered as the boundaries of the triangular picture.



Figure 10: A literal, probabilistic perspective on logic

The point of the preceding detour through geologic is that the picture plane is the space \mathcal{P}_3 of probability measures on \mathcal{A}_3 and the walls and ceiling of the office are the fully satisfiable propositions in geologic. So Figure 10 literally illustrates geologic from a *probabilistic perspective*. That perspective sheds new light on the lottery paradox and its associated conundrums. In particular, note the similarity between the acceptance zones of the proposed, paradox-avoiding rule (Figure 4) and the projected coordinate lines of the unit cube (Figure 10.b). The boundaries of the former always follow the latter.

6 Probalogic

We understand "logic" in the broad, pragmatic sense that logic is wherever logical structure is. If the logical structure pertains to relative strength of credal states, then there is a logic of such states, even though the states in question are not necessarily propositional and the logical relations among them are not plausibly interpreted as arguments. And if the structure happens to be relative to pragmatic factors such as a question that elevates the significance of certain propositions as relevant answers,

then logic, itself, is pragmatic—we do not insist that logic must in some sense be prior to or independent of such considerations. Our view accords with an ancient tradition according to which logic is a tool or organon for *inquiry*, which typically begins with some question and ends with an answer thereto. In this section, we introduce a logic of probabilistic credal states in the broad, pragmatic sense just outlined.

When credence is modeled as qualitative belief in a proposition, it is straightforward to judge the relative strength of credal states in terms of the classical, logical strength of the propositions believed:

$$B\varphi \le B\psi \quad \Longleftrightarrow \quad \varphi \le \psi.$$

We propose, in a similar spirit, that probabilistic credal states inherit their logical strength from their unique, geological images:

$$p \le q \quad \Longleftrightarrow \quad \sigma(p) \le \sigma(q)$$
 (18)

$$\iff \sigma(p)_i \le \sigma(q)_i \text{ for all } i \in I.$$
(19)

Disjunction \lor and conjunction \land are standardly defined, respectively, as the least upper bound and the greatest lower bound with respect to \leq . We call the resulting logical structure on probability measures *probalogic*:

$$(\mathcal{P}, \leq, \lor, \land).$$

Probalogic is just geologic from a probabilistic perspective.

Consider arbitrary credal state p in \mathcal{P}_3 . Which credal states are probalogically at least as weak as p? First, project p up to geological proposition $\sigma(p)$ on the upper surface of the geological cube. The geological consequences of $\sigma(p)$ consist of the parallelepiped containing \top whose sides are parallel to the sides of the unit cube and whose bottom-most corner is $\sigma(p)$ (Figure 11). Since $\sigma(p)$ is incident to an upper surface of the cube, the parallelepiped is, in this case, a rectangle lying entirely in one upper face of the unit cube (or, in degenerate cases, entirely within an upper edge of the unit cube). The probalogical consequences of p are contained within the linear perspective projection of that rectangle onto the picture plane \mathcal{P}_3 . Note that, according to the usual rules of linear perspective, parallel sides of the rectangles meet at *vanishing points*, which correspond to the corners of \mathcal{P}_3 that are not closest to p. Similarly, the geological propositions in the range of σ that are geologically at least as strong as $\sigma(p)$ are in the rectangle with sides parallel to the sides of the unit cube that has $\sigma(p)$ as its upper corner and the nearest unit vector e_i to $\sigma(p)$ as its lower corner. So the inverse image of that rectangle under σ is the set of the credal states that are probalogically at least as strong as $\sigma(p)$. We call the partial order \leq so defined *relative probalogical* strength.



Figure 11: Probalogical strength

Probalogical disjunction, conjunction, and negation can be defined similarly, as the projections of the corresponding, geological disjunction:

$$p \lor q = \sigma^{-1}(\sigma(p) \lor \sigma(q)); \tag{20}$$

$$p \wedge q = \sigma^{-1}(\sigma(p) \wedge \sigma(q));$$
 (21)

$$\neg p = \sigma^{-1}(\neg \sigma(p)). \tag{22}$$

Since the geological disjunction of two propositions on the upper surface of the unit cube is also on the upper surface of the unit cube, \mathcal{P}_3 is closed under probalogical disjunction. Geometrically, these logical operations can be constructed as perspective renderings of the corresponding geological operations on the cube (Figures 12, 13, and 14). Probalogical constants and operations are not necessarily defined. In finite questions, \top denotes the uniform distribution, but in countably infinite questions there is no such distribution. There is no interpretation of \bot . Letting $\bot = \top$ is obvious unappealing, but any choice of \bot that is off-center is equally implausible. Geological negation is closed over the upper faces of the unit cube, so probalogical negation is defined only over the lower edges of the unit cube. Furthermore, if $\sigma(p)$ and $\sigma(q)$ are on different upper faces of the unit cube, then the conjunction $\sigma(p) \land \sigma(q)$ lies below the upper faces of the unit cube, so $p \land q = \sigma^{-1}(\sigma(p) \land \sigma(q))$ is undefined.

Although we will not pursue the idea in this paper, there is a way to expand \mathcal{P} to a space over which probalogical conjunction and disjunction are closed. Some assumptions are so certain that one does not even conceive of their falsity—e.g., that a particle cannot have two distinct momenta at the same time but can have a definite momentum and position at the same time. But when experience gets strange, we may come to



Figure 12: Probalogical conjunction and disjunction within a face



Figure 13: Probalogical conjunction and disjunction across faces

doubt our basic assumptions without having thought yet of any concrete alternatives. In such cases, a natural response is to transfer probability mass to a non-descript "catchall hypothesis" absent from the original algebra \mathcal{A}_3 . Within \mathcal{A}_3 , the resulting credal state appears to be normalized to a value less than 1. Accordingly, let \mathcal{P}^* denote the set of all additive measures p on \mathcal{A} such that $0 \le p(\top) \le 1$. Then the problem of closure under negation and conjunction is solved by plausibly *extending* σ to a bijection between \mathcal{P}^*



Figure 14: Negation around the perimeter

and the entire unit cube as follows (Figure 15):

$$\sigma^*(p)_i = p(\top) \cdot \frac{p_i}{\max_i p_i}, \text{ for } i \in I.$$

So equations (18) to (22), with σ replaced by σ^* , induce a probalogical structure on \mathcal{P}^* that is closed under the probalogical operations of conjunction, disjunction, and negation.



Figure 15: σ extended to measures normalized to a value ≤ 1

Lin and Kelly

7 Acceptance that respects Probalogic

A *probalogical* acceptance rule v is an acceptance rule that preserves probalogical structure in the sense of morphism conditions (5) to (8).¹⁵ As described in the preceding section, condition (7) is understood to hold only when $p \land q$ is defined over \mathcal{P} .

Recall the camera-shutter-like acceptance rules introduced above as one geometrical strategy for solving the lottery paradox. The rules can be stated a bit more generally, by allowing the threshold *r* and the strictness of the inequality to vary with *i*. Say that acceptance rule *v* is a *camera shutter rule* for \mathcal{E} if and only if there exist thresholds $\{r_i : i \in I\}$ in the unit interval and inequalities $\{\triangleleft_i : i \in I\}$ that are either \leq or <, such that for each *p* in \mathcal{P} and $i \in I$:

1. $v(p) = \bigwedge \{\neg E_i : \sigma(p)_i \triangleleft_i 1 - r_i \text{ and } i \in I\};$

2. if
$$\triangleleft_i = \leq$$
 then $r_i > 0$;

3. if
$$\triangleleft_i = <$$
 then $r_i < 1$.

Note that 0 is omitted in the second condition to make it possible to not accept $\neg E_i$, and 1 is omitted in the third condition to make it possible to accept $\neg E_i$ —else morphism condition (8) would be violated trivially. The main result of this section is that, over countable dimensions, the camera shutter rules are precisely the rules that preserve probalogic.

Theorem 1 (representation of probalogical rules). Suppose that \mathcal{E} is countable. Then an arbitrary acceptance rule is probalogical if and only if it is a camera shutter rule.

The proof proceeds by a series of lemmas. Let p, q be in \mathcal{P} . Define:

$$q \leq_i p \iff \sigma(p)_i \leq \sigma(q)_i.$$

Lemma 1. Suppose that $q \leq_i p$. Then $p = (p \lor e_i) \land (p \lor q)$.

Proof. See Figure 16. By the definition of probalogic in terms of geologic, it suffices to show that

$$\sigma(p) = \sigma((p \lor e_i) \land (p \lor q)).$$

By geologic, the *j*-th component of the right hand side expands to:

 $\min(\max(\sigma(p)_i, \sigma(e_i)_i), \max(\sigma(p)_i, \sigma(q)_i)).$

¹⁵Note that (5) is redundant, for it is derivable from (6).



Figure 16: Proof of Lemma 1

Since $(e_i)_i = 1$, it follows that $\max(\sigma(p)_i, \sigma(e_i)_i) = 1$. Since $\sigma(q)_i \leq \sigma(p)_i$, it follows that $\max(\sigma(p)_i, \sigma(q)_i) = \sigma(p)_i$. So $\sigma((p \lor e_i) \land (p \lor q))_i = \sigma(p)_i$. Now let E_j be in \mathcal{E} for $j \neq i$. Then $(e_i)_j = 0$, so $\max(\sigma(p)_j, \sigma(e_i)_j) = \sigma(p)_j$. In general, $\min(x, \max(x, y)) = x$, so we have as well that $\sigma((p \lor e_i) \land (p \lor q))_j = \sigma(p)_j$.

Lemma 2. Let v satisfy morphism conditions (6), (7), and (8). Let $i \in I$. Then:

 $p \Vdash_{\nu} \neg E_i \text{ and } q \leq_i p \implies q \Vdash_{\nu} \neg E_i.$

Proof. Suppose that $p \Vdash \neg E_i$ and that $q \leq_i p$. Since $p \Vdash \neg E_i$, it follows that $p_i < \max_k p_k$. For otherwise, $e_i \leq p$, so by morphism condition 5, $e_i \Vdash \neg E_i$, contrary to morphism condition (8). Since it is also the case that $q \leq_i p$, Lemma 1 yields that $p = (p \lor e_i) \land (p \lor q)$. Suppose for reductio that v(q) is logically compatible with E_i . Then by morphism condition (6), $v(p \lor q)$ is compatible with E_i . By morphism condition (8), $v(e_i)$ is compatible with E_i . So again by morphism condition (6), $v(p \lor e_i)$ is compatible with E_i , by morphism condition (7) and by the fact that E_i is an atom in algebra \mathcal{A} . But $p \Vdash_v \neg E_i$. Contradiction. Hence, $q \Vdash_v \neg E$.

Proof of Theorem 1. For the only if side, let $i \in I$. Define:

$$1 - r_i = \sup\{\sigma(p)_i : p \in \mathcal{P} \text{ and } p \Vdash_{\mathcal{V}} \neg E_i\}.$$

Suppose that $\sigma(p)_i <_i 1 - r_i$. Then $p \Vdash_{\nu} \neg E_i$, by Lemma 2. Suppose that $\sigma(p)_i > 1 - r_i$. Then $p \nvDash_{\nu} \neg E_i$, by the definition of $1 - r_i$. Finally, suppose that $\sigma(p)_i = \sigma(q)_i = 1 - r_i$. Consider the case in which there exists r in \mathcal{P} such that $\sigma(r) = 1 - r_i$ and $r \Vdash_{\nu} \neg E_i$. Then $p \Vdash_{\nu} \neg E$ and $q \Vdash_{\nu} \neg E$, by Lemma 2. In the alternative case, it is immediate that $p \nvDash_{\nu} \neg E$ and $q \nvDash_{\nu} \neg E$. Thus, $p \Vdash_{\nu} \neg E_i$ if and only if $q \Vdash_{\nu} \neg E_i$. Set $\triangleleft_i = \leq$ in the former case and set $\triangleleft_i = <$ in the latter case. In the former case, suppose for reductio that $r_i = 0$. Then $\nu(e_i) \Vdash \neg E_i$, contradicting morphism condition (8), so $r_i > 0$, as required. In the latter case, suppose for reductio that $r_i = 1$. Then $\nu(e_i) \nvDash E_i$, contradicting morphism condition (8), so $r_i > 0$, as required. For the if side of the theorem, suppose that ν is a camera shutter rule for countable \mathcal{E} . For morphism condition (5), suppose that $p \leq q$. Then $\sigma(p)_i \leq \sigma(q)_i$, for each $i \in I$. Then $q \Vdash_{\nu} E_i$ implies $p \Vdash_{\nu} E_i$, so $\nu(p) \leq \nu(q)$. For morphism condition (6), let $\nu(p) = A$ and $\nu(q) = B$, so:

$$A = \bigwedge \{ \neg E_i : \sigma(p)_i \triangleleft_i 1 - r_i \};$$

$$B = \bigwedge \{ \neg E_i : \sigma(q)_i \triangleleft_i 1 - r_i \}.$$

Let $\mathcal{D} = \{\neg E_i : A \leq E_i \text{ and } B \leq E_i\}$ and note that $A \lor B = \bigwedge \mathcal{D}$. Suppose that $\neg E_i$ is in \mathcal{D} . Then $\sigma(p)_i \triangleleft_i 1 - r_i$ and $\sigma(q)_i \triangleleft_i 1 - r_i$. Hence, $\max(\sigma(p)_i, \sigma(q)_i) \triangleleft_i 1 - r_i$. Thus, $v(p \land q) \leq \neg E_i$. Suppose that $\neg E_i$ is not in \mathcal{D} . Then either $\sigma(p)_i \not\triangleleft_i 1 - r_i$ or $\sigma(q)_i \not\triangleleft_i 1 - r_i$, so $\max(\sigma(p)_i, \sigma(q)_i) \not\triangleleft_i 1 - r_i$ and, thus, $v(p \land q) \not\leq \neg E_i$. Hence, $\neg E_i$ is in \mathcal{D} if and only if $v(p \land q) \leq \neg E_i$. Therefore, $v(p \lor q) = \bigwedge \mathcal{D} = A \lor B$. The dual argument works for morphism condition (7).

Recall that the conditions (5)-(7) omit preservation of negation and of the infinitary versions of disjunction and conjunction. There are good reasons to drop those conditions.

Proposition 4. In finite dimensions, no probalogical acceptance rule preserves infinite conjunction and disjunction.

Proof. Consider probalogical acceptance rule v for question $\{E_i : i \in I\}$. By morphism condition (8), $v(e_1) = E_1$ and $v(e_2) = E_2$. Let *L* be the straight line connecting e_1 with e_2 . Note that no uniform distribution with infinite support is encountered along this line, so it is continuous. So by morphism condition (5), there is a boundary point *b* such that $q \Vdash_v E_1$, for all *q* closer to e_1 than *b*, and $q \nvDash_v E_1$, for all *q* farther from e_1 than *b*. Let *m* be the mid-point of *L*. Consider the case in which *p* is between *m* and e_1 . Consider the case in which $b \Vdash_v E_1$. Let $\{p_i : i \in \mathbb{N}\}$ be a discrete sequence of points in line segment $\overline{e_1 \ b}$ that converges to *b*. Then:

$$\bigvee_i p_i = b = \bigwedge_i q_i.$$

Suppose that $b \nvDash_{v} E_{1}$. Then $v(\bigvee_{i} p_{i}) \neq \bigvee_{i} v(p_{i})$. Alternatively, suppose that $b \Vdash_{v} E_{1}$. Then $v(\bigwedge_{i} q_{i}) \neq \bigwedge_{i} v(q_{i})$.

Proposition 5. In finite dimensions, no probalogical acceptance rule also preserves probalogical negation.

Proof. Let $p = (\frac{2}{3}, \frac{1}{3}, 0)$. Assume, for reductio, that acceptance rule v is probalogical and preserves probalogical negation as well. So by Proposition 1, v is a camera shutter rule. Suppose that v rejects E_2 in p. So $\sigma(e_2) = \frac{1}{2} \triangleleft_2 1 - r_2$. Note, in Figure 14, that $\neg p = (0, \frac{1}{3}, \frac{2}{3})$. So by preservation of negation, v does not reject E_2 at $\neg p$. Thus: $\sigma(e_2) = \frac{1}{2} \not\triangleleft_2 1 - r_2$, which is a contradiction. The case in which v does not reject E_2 in p is similar. The argument generalizes to arbitrary, finite dimensions.

On the other hand, setting each $r_i = \frac{1}{2}$ almost preserves negation, in the sense that negation is preserved at all points on the perimeter of the triangle *except* at the six probability assignments with range $\{0, \frac{1}{3}, \frac{2}{3}\}$. But even so, no other setting for the r_i other than $\frac{1}{2}$ has that property, so the demands imposed by negation preservation are unreasonably strict.

8 Acceptance that does not respect Probalogic

The acceptance rules we recommend, the camera shutter rules, are exactly the rules that preserve probalogical structure. Alternative acceptance rules proposed by Kyburg (1961) and by Pollock (1995) fail to preserve probalogical structure—actually, they fail to preserve any plausible logical structure.

Each Kyburgian acceptance rule χ_r is a Lockean rule without closure under conjunction:

$$\chi_r = \{A \in \mathcal{A} : p(A) \ge r\}.$$

Let question \mathcal{E} be ternary and set $r = \frac{2}{3}$. In Figure 17, the set $\chi_{\frac{2}{3}}(c)$ of propositions accepted at the center $c = (\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$ is indicated by a solid line and the set $\chi_{\frac{2}{3}}(e_3)$ is indicated by a dashed line. Rule $\chi_{\frac{2}{3}}$ does not preserve logical order in any plausible sense, for corner e_3 is at least as strong as center c, but $\chi_{\frac{2}{3}}(e_3)$ is, intuitively, not at least as strong as $\chi_{\frac{2}{3}}(c)$ due to the retraction of $e_1 \vee e_2$.

There is, therefore, a hidden dilemma in Kyburg's thesis that one should give up closure of accepted propositions under conjunction. On the one hand, if only \top is accepted at the uniform measure *c*, then there is no lottery paradox and, hence, there is no motivation for failing to close the accepted propositions under conjunction. On the other hand, if some proposition other than \top is accepted at *c*—say, a disjunction *D* that is incompatible with E_i —then, using the same argument as above, when one jumps from the center *c* to the stronger state e_i , one must accept E_i (which has probability



Figure 17: Kyburgian acceptance rule

one) and retract D (which has probability zero) and thus one must fail to expand the set of accepted propositions. In contrast, all camera shutter rules preserve probalogic.

Pollock (1995), Ryan (1996), and Douven (2002) all propose variants of the Lockean acceptance rule, which we will call *Pollockian*. The basic idea is to *restrict* the Lockean rule to cases in which it produces no paradox. The idea is illustrated, for ternary \mathcal{E} , in Figure 18. The basic difference between Pollockian and Lockean rules in



Figure 18: Pollockian acceptance rules

3-dimension is that the former return \top whenever the latter return \perp (compare to Figure 3). The choice of \top as a substitute for \perp is natural enough, on grounds of symmetry, but due to the *shape* of Pollockian acceptance zones, there still exists no single logical structure that all Pollockian rules preserve.

Proposition 6. Suppose that \mathcal{E} is ternary. Let \leq be an arbitrary partial order on \mathcal{P} whose binary least upper bound operation \vee is totally defined. Then there exists at least one Pollockian acceptance rule that is not a structure preserving map from $(\mathcal{P}_3, \leq, \vee)$ to $(\mathcal{A}_3, \leq, \vee)$.

Proof. Suppose the contrary for reductio. Let π_r be a Pollockian rule. When $r > \frac{2}{3}$, as in Figure 18.a, the rule π_r accepts $E_1 \lor E_2$ at $p = (\frac{1}{2}, \frac{1}{2}, 0)$ and $E_2 \lor E_3$ at $q = (0, \frac{2}{3}, \frac{1}{3})$, respectively, whose disjunction is $E_1 \lor E_2 \lor E_3 = \top$. So, to preserve disjunction, $p \lor q$ must lie within the white triangle, where \top is accepted. If we let *r* approach $\frac{2}{3}$ from above, as in Figure 18.b, the white triangle converges to the center point $c = (\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$, so $p \lor q = c$. Now consider the case in which $r < \frac{2}{3}$ (Figure 18.c). By preservation of disjunction, we have:

$$T = \pi_r(c)$$

$$= \pi_r(p \lor q)$$

$$= \pi_r(p) \lor \pi_r(q)$$

$$= (E_1 \lor E_2) \lor E_2$$

$$= E_1 \lor E_2.$$

Hence $\top = E_1 \lor E_2$, a contradiction.

A dilemma for Pollockian theorists is that, on the one hand, symmetry precludes accepting anything other than \top at the center point *c*, but that implies that there is no logical structure on \mathcal{P} that all Pollockian rules preserve. In contrast, all camera shutter rules preserve probalogic.

9 The geometry of Conditional Logic

As illustrated in Figure 5, acceptance zones with a wrong shape can invalidate plausible principles of nonmonotonic reasoning. In fact, each axiom of the logic for flat conditionals corresponds to a definite, geometrical constraint on acceptance zones. The correspondences are established in this section and are used below to demonstrate that each probalogical rule validates a plausible set of axioms for conditional logic due to Adams (1975).

The acceptance condition of a conditional is defined by (13) as an explication of Ramsey test:

$$p \Vdash_{\alpha} A \Rightarrow B \iff p(\cdot|A) \Vdash_{\alpha} B \text{ or } p(A) = 0.$$

The set of axioms known as *Adams' conditional logic* (Adams 1975) or *system* P have been widely recognized as central to conditional and nonmonotonic reasoning (Kraus et al. 1990). They state closure properties for a set of accepted conditionals. Here we rewrite them as closure properties for the set of conditionals accepted at a fixed credal state p under a fixed acceptance rule α (where the horizontal line means material implication):

 $(\mathbf{Reflexivity}) \ \hline p \quad \Vdash_{\alpha} \quad A \Rightarrow A$

(Left Equivalence) $\frac{p \Vdash_{\alpha} A \Rightarrow B}{p \Vdash_{\alpha} C \Rightarrow B}$ if A is classically equivalent to C.

(**Right Weakening**) $\frac{p}{p} \Vdash_{\alpha} A \Rightarrow B$ if *B* classically entails *C*.

(And)
$$\begin{array}{cc} p & \Vdash_{\alpha} & A \Rightarrow B \\ p & \Vdash_{\alpha} & A \Rightarrow C \\ \hline p & \Vdash_{\alpha} & A \Rightarrow (B \land C) \end{array}$$

$$(\mathbf{Or}) \begin{array}{c} p & \Vdash_{\alpha} & A \Rightarrow C \\ p & \Vdash_{\alpha} & B \Rightarrow C \\ \hline p & \Vdash_{\alpha} & (A \lor B) \Rightarrow C \end{array}$$

(Cautious Monotonicity)
$$\begin{array}{c} p \quad \Vdash_{\alpha} \quad A \Rightarrow B\\ p \quad \Vdash_{\alpha} \quad A \Rightarrow C\\ \hline p \quad \Vdash_{\alpha} \quad (A \land B) \Rightarrow C \end{array}$$

Say that acceptance rule α validates an axiom for conditional logic if and only if, for each credal states p, α together with p satisfies that axiom. Say that α validates a set of axioms if and only if α validates each axiom in that set. The some axioms in Adams' logic are validated trivially.

Proposition 7. Each acceptance rule validates And, Left Equivalence, and Right Weakening.

Proof. Immediate from the modeling assumption: with respect to each credal state and each acceptance rule, there is a strongest accepted proposition that entails all the other accepted propositions.

Proposition 8. Let α be an acceptance rule. Then, α validates Reflexivity if and only if α accepts every certain proposition in the following sense: $p \Vdash_{\alpha} A$ for each credal state p in \mathcal{P} and each proposition A in \mathcal{A} such that p(A) = 1.

Proof. For the only if side, suppose that α validates Reflexivity. Suppose further that p(A) = 1. Then we have that $p \Vdash_{\alpha} A \Rightarrow A$ (by Reflexivity), and thus that $p(\cdot|A) \Vdash_{\alpha} A$ (note that $p(\cdot|A)$ exists), and hence that $p \Vdash_{\alpha} A$ (because $p = p(\cdot|A)$). So α accepts every certain proposition. For the converse, let α be an acceptance rule and p a credal state. Either credal state $p(\cdot|A)$ is undefined, and thus we have that $p \Vdash_{\alpha} A \Rightarrow A$ by default. Or $p(\cdot|A)$ is defined, and thus p(A|A) = 1 and then $p(\cdot|A) \Vdash_{\alpha} A$ (by acceptance of every certain proposition) and hence we have that $p \Vdash_{\alpha} A \Rightarrow A$ (by definition). \Box

Axioms Cautious Monotonicity and Or impose substantial geometrical constraints on acceptance rules. Let A be a proposition in \mathcal{A} . Let $\mathcal{P}|A$ denote the set of all p in \mathcal{P} such that p(A) = 1, which we will call the *facet* of simplex \mathcal{P} for proposition A. The *line segment* with endpoints p, q in simplex \mathcal{P} is defined by convex combination:

$$\overline{pq} = \{ap + (1-a)q : a \in [0,1]\}.^{16}$$

Say that *q* is a *projection* of *p* from facet $\mathcal{P}|\neg A$ onto facet $\mathcal{P}|A$ if and only if (i) there exists a line segment *L* through *p* with endpoint *q* in $\mathcal{P}|A$ and the other endpoint in facet $\mathcal{P}|\neg A$ and (ii) *p* is not in the complementary facet $\mathcal{P}|\neg A$. Projection is equivalent to Bayesian conditioning:

Lemma 3. Credal state q is a projection of p from facet $\mathcal{P}|\neg A$ onto facet $\mathcal{P}|A$ if and only if $p(\cdot|A)$ is defined and $q = p(\cdot|A)$.

Proof. This lemma is trivially true when p is in $\mathcal{P}|A$ or in $\mathcal{P}|\neg A$, so suppose that p is neither in $\mathcal{P}|A$ nor in $\mathcal{P}|\neg A$ and, thus, that both $p(\cdot|A)$ and $p(\cdot|\neg A)$ are defined. For the if side, consider line segment $L = \overline{p(\cdot|A)} p(\cdot|\neg A)$, whose endpoints are in $\mathcal{P}|A$ and $p(\cdot|\neg A)$, respectively. Note that p lies on L, since for each B in \mathcal{A} ,

$$p(B) = p(B|A)p(A) + p(B|\neg A)p(\neg A) = a p(B|A) + (1 - a) p(B|\neg A),$$

where a = p(A). Therefore, $p(\cdot|A)$ is a projection of p from $\mathcal{P}|\neg A$ onto $\mathcal{P}|A$. For the only if side, suppose that q is a projection of p from facet $\mathcal{P}|\neg A$ onto facet $\mathcal{P}|A$. So q is in $\mathcal{P}|A$ and there exists credal state r in $\mathcal{P}|\neg A$ such that line segment \overline{qr} contains p. Then, p lies in the interior of \overline{qr} , since p is neither in $\mathcal{P}|A$ nor in $\mathcal{P}|\neg A$. So there exists a in the open interval (0, 1) such that p = aq + (1 - a)r. Then it suffices to show that $q = p(\cdot|A)$. Consider the case in which $E_i \nleq A$. Then $E_i \le \neg A$. Since q

¹⁶Addition is defined as vector addition; multiplication is defined as scalar multiplication.

is in facet $\mathcal{P}|A$, we have that $q(E_i) = 0 = p(E_i|A)$. Now consider the case in which $E_i \leq A$. Then since *r* is in facet $\mathcal{P}|\neg A$, we have that $r(E_i) = 0$, so $p(E_i) = aq(E_i)$. Similarly, we have that q(A) = 1 and r(A) = 0, so $p(A) = a \cdot 1 + 0 = a$. Hence, $q(E_i) = p(E_i)/a = p(E_i)/p(A) = p(E_i)p(A|E_i)/p(A) = p(E_i|A)$. So $q(\cdot)$ agrees with $p(\cdot|A)$ for all E_i in \mathcal{E} and, thus, for all B in \mathcal{A} , as required.

Proposition 9 (Geometry of Cautious Monotonicity). Let α be an acceptance rule. Then, α validates Cautious Monotonicity if and only if the following condition holds: for each credal state p and for each proposition A, if α accepts A at p, then α accepts A at the projection of p on the facet $\mathcal{P}|B$, for each logical consequence B of A (as long as the projection exists). In light of Lemma 3, the condition may be restated as:

$$p \Vdash_{\alpha} A, A \leq B, and p(\cdot|B) \text{ is defined} \implies p(\cdot|B) \Vdash_{\alpha} A.$$
 (23)

Proof. The proof of the only if side involves unpacking the definitions and checking that the projection condition (23) is simply an instance of Cautious Monotonicity. For the if side, assume that the projection condition (23) holds. Suppose that $p \Vdash_{\alpha} A \Rightarrow B$ and $p \Vdash_{\alpha} A \Rightarrow C$. It suffices to show that $p \Vdash_{\alpha} (A \land B) \Rightarrow C$. If $p(\cdot|A \land B)$ is undefined, then by default $p \Vdash_{\alpha} (A \land B) \Rightarrow C$. So suppose that $p(\cdot|A \land B)$ is defined and, thus, $p(\cdot|A)$ is defined. Then argue as follows:

Proposition 10 (Geometry of Or). Let α be an acceptance rule that validates Reflexivity. Then, α validates Or if and only if the following condition holds: for each line segment L connecting two complementary facets $\mathcal{P}|B$ and $\mathcal{P}|\neg B$, and for each proposition A in \mathcal{A} , if α accepts A at both endpoints of L, then α accepts A at each point on L; in light of Lemma 3, the condition may be restated as:

$$p(\cdot|B) \Vdash_{\alpha} A , \ p(\cdot|\neg B) \Vdash_{\alpha} A \implies p \Vdash_{\alpha} A.$$

$$(24)$$

Proof. For the only if side, argue as follows:

$$\begin{array}{rcl} p(\cdot|B) \Vdash_{\alpha} A, \ p(\cdot|\neg B) \Vdash_{\alpha} B \\ \Longrightarrow & p \Vdash_{\alpha} B \Rightarrow A, \ p \Vdash_{\alpha} \neg B \Rightarrow A \\ \Longrightarrow & p \Vdash_{\alpha} (B \lor \neg B) \Rightarrow A \\ \Longrightarrow & p(\cdot|B \lor \neg B) \Vdash_{\alpha} A \\ \Longrightarrow & p \Vdash_{\alpha} A. \end{array}$$
by axiom Or,

For the converse, suppose that $p \Vdash_{\alpha} A \Rightarrow C$ and $p \Vdash_{\alpha} B \Rightarrow C$. It suffices to show that $p \Vdash_{\alpha} (A \lor B) \Rightarrow C$. If both $p(\cdot|A)$ and $p(\cdot|B)$ are undefined, then $p(\cdot|A \lor B)$ is undefined and thus we have that $p \Vdash_{\alpha} (A \lor B) \Rightarrow C$ by default. If one is defined and the other is undefined—say, $p(\cdot|A)$ is defined and $p(\cdot|B)$ is undefined—then p(B) = 0 and thus $p(\cdot|A \lor B) = p(\cdot|A)$ is defined, so:

$$p \Vdash_{\alpha} A \Rightarrow C$$

$$\implies p(\cdot|A) \Vdash_{\alpha} C$$

$$\implies p(\cdot|A \lor B) \Vdash_{\alpha} C \qquad \text{by } p(\cdot|A \lor B) = p(\cdot|A),$$

$$\implies p \Vdash_{\alpha} (A \lor B) \Rightarrow C.$$

Last, suppose that both $p(\cdot|A)$ and $p(\cdot|B)$ are defined. So $p(\cdot|A \vee B)$ is defined. Then argue for Or as follows:

| | $p \Vdash_{\alpha} A \Rightarrow C, \ p \Vdash_{\alpha} B \Rightarrow C$ | |
|------------|--|--|
| \implies | $p(\cdot A) \Vdash_{\alpha} C, \ p(\cdot B) \Vdash_{\alpha} C$ | |
| \implies | $q(\cdot A) \Vdash_{\alpha} C, \ q(\cdot B) \Vdash_{\alpha} C$ | letting $q = p(\cdot A \vee B)$, |
| | | so $q(\cdot A) = p(\cdot A)$ and $q(\cdot B) = p(\cdot B)$, |
| \implies | $q \Vdash_{\alpha} C \lor \neg A, \ q \Vdash_{\alpha} C \lor \neg B$ | (*) see the explanation below, |
| \implies | $q \Vdash_{\alpha} C \lor \neg (A \lor B)$ | by classical entailment, |
| \implies | $q \Vdash_{\alpha} C \lor \neg (A \lor B), \ q \Vdash_{\alpha} A \lor B$ | since $q(A \lor B) = 1$ and Proposition 8 applies, |
| \implies | $q \Vdash_{\alpha} C$ | by classical entailment, |
| \implies | $p(\cdot A \lor B) \Vdash_{\alpha} C$ | |
| \implies | $p \Vdash_{\alpha} (A \lor B) \Rightarrow C.$ | |

It only remains to establish step (*). By the symmetric roles of *A* and *B*, it suffices to show that $q(\cdot|A) \Vdash_{\alpha} C$ implies that $q \Vdash_{\alpha} C \lor \neg A$. If $q(\cdot|\neg A)$ is undefined, then $q(A) = 1 - q(\neg A) = 1 - 0 = 1$ and thus $q = q(\cdot|A) \Vdash_{\alpha} C \leq C \lor \neg A$, so $q \Vdash_{\alpha} C \lor \neg A$. If $q(\cdot|\neg A)$ is defined, then we have both that $q(\cdot|A) \Vdash_{\alpha} C$ (by supposition) and that $q(\cdot|\neg A) \Vdash_{\alpha} \neg A$ (by Reflexivity and Proposition 8). So we have both that $q(\cdot|A) \Vdash_{\alpha} C \lor \neg A$, by the convexity condition (24).

10 The Geometry of System P

In this section we examine, for each axiom in conditional logic system P, its geometrical constraint on acceptance rules. It is an easy corollary of the geometrical characterizations in the preceding section that:

Theorem 2 (Lin 2011). Each probalogical rule validates system P.

Proof sketch. When $|\mathcal{E}| = 3$, one can easily verify that probalogical rules satisfy the geometric conditions given in propositions 7-10 when the consequence relations in question have antecedents of nonzero probability. The routine verification can be easily generalized to a proof for all countable dimensional cases.

We now proceed to establish a partial converse to Theorem 2. Recall that acceptance zones for answers have the following form under probalogical rules:

$$p \Vdash_{\nu} E_{i} \iff E_{i} = \nu(p)$$

$$\iff E_{i} = \bigwedge \{\neg E_{j} : \sigma(p)_{j} \triangleleft_{j} 1 - r_{j}\}$$

$$\iff \forall j \neq i, \ \sigma(p)_{j} \triangleleft_{j} 1 - r_{j}$$

$$\iff \forall j \neq i, \ \frac{p_{j}}{\max_{k} p_{k}} \triangleleft_{j} 1 - r_{j}$$

$$\iff \forall j \neq i, \ \frac{p_{j}}{p_{i}} \triangleleft_{j} 1 - r_{j}.^{17}$$

Namely, answer E_i is accepted if and only if each rival has a sufficiently low odds to E_i . To allow for more generalized rules entertained below (Section 15), we relax the conditions that the rejection threshold $1 - r_j$ is in the unit interval and that it is constant for all *i*. Accordingly, say that the acceptance zone of answer E_i under α is a *blunt diamond* (Figure 19.a) if and only if it takes the following form: there exist thresholds $\{t_{ij} : j \in I \setminus \{i\}\}$ in interval $[0, \infty]$ and inequalities $\{\triangleleft_{ij} : j \in I \setminus \{i\}\}$ that are either \leq or <, such that for each $p \in \mathcal{P}$:

- 1. $p \Vdash_{\alpha} E_i \iff \forall j \neq i, \ \frac{p_j}{p_i} \triangleleft_{ij} t_{ij};$
- 2. if $\triangleleft_{ij} = \leq$ then $t_{ij} < \infty$;
- 3. if $\triangleleft_{ij} = \langle \text{then } t_{ij} \rangle > 0$.

Say that acceptance rule α is *corner-monotone* if and only if (i) $\alpha(e_i) = E_i$ for each $i \in I$, and (ii) for each $p \in \mathcal{P}$ such that $\alpha(p) = E_i$, we have that $\alpha(q) = E_i$ for all q in line segment $\overline{p e_i}$. Corner-monotonicity is a very natural constraint on acceptance rules and it is by all the rules we have discussed. Our partial converse to Theorem 2 is as follows.



Figure 19: Acceptance zone of E_2

Theorem 3 (Blunt diamond, Lin 2011). Let α be an acceptance rule. If α is everywhere consistent, satisfies corner-monotonicity, and validates system P, then for each answer E_i to question \mathcal{E} , the acceptance zone of E_i under α is a blunt diamond.

Proof sketch. Here we present a geometric argument for case $|\mathcal{E}| = 3$, which can be easily generalized to each countable dimension. Solve for the acceptance zone of E_2 under α , as depicted in Figure 19.b. By corner-monotonicity, the credal states along side $\overline{e_2} e_1$ of the triangle at which α accepts E_2 form a continuous, unbroken line segment with e_2 as an endpoint, which is depicted as the heavy, grey line segment lying on $\overline{e_2} e_1$. The same is true for side $\overline{e_2} e_3$.¹⁸ Connect the endpoints of the grey line segments to the opposite corners by straight lines, which enclose the grey blunt diamond at the corner e_2 .

Argue as follows that $p \Vdash_{\alpha} E_2$, for each point p in the blunt diamond. Consider the projection p' of p to the facet $\mathcal{P}|(E_2 \lor E_3)$. Note that p' is in the heavy, grey line segment alone side $\overline{e_2 e_3}$. On line segment $\overline{e_1 p'}$ ray, acceptance rule α accepts E_1 at one endpoint (e_1) and accepts E_2 at the other endpoint (p'), so α accepts $E_1 \lor E_2$ at both endpoints. Then, by Proposition 10, we have that $p \Vdash_{\alpha} E_1 \lor E_2$. By applying the same argument to the projection of p to the facet for proposition $E_1 \lor E_3$, we have that $p \Vdash_{\alpha} E_3 \lor E_2$. Then $p \Vdash_{\alpha} E_2$, since E_2 is entailed by $E_1 \lor E_2$ plus $E_2 \lor E_3$.

Argue as follows that $q \nvDash_{\alpha} E_2$, for each point q outside of the blunt diamond. Since q lies outside of the blunt diamond, there exists at least one answer E_i other than

¹⁸There is an issue whether the line segments are open or closed at the endpoints distinct from e_2 , which would give rise to a possible mixture of strict and weak inequalities, as stated in the theorem. That detail is handled in the formal proof in (Lin 2011), but is ignored here.

 E_2 such that the projection q' of q to the facet $\mathcal{P}|(E_2 \vee E_i)$ does not touch the grey line segment along side $\overline{e_2 e_i}$. Figure 19.b. illustrates the case for i = 3. Suppose for reductio that $q \Vdash_{\alpha} E_2$. Then, by applying Proposition 9 to the projection q' of q, we have that $q' \Vdash_{\alpha} E_2$. But $q' \nvDash_{\alpha} E_2$, for q' lies outside of the grey line segment contradiction.

11 AGM geometry is trivial

A popular, stronger system for the logic of flat conditionals, R, is obtained from P by adding the following axiom (Lehmann and Magidor 1992):

(Rational Monotonicity)
$$\begin{array}{ccc} p & \mathbb{F}_{\alpha} & A \Rightarrow \neg B \\ p & \mathbb{F}_{\alpha} & A \Rightarrow C \\ \hline p & \mathbb{F}_{\alpha} & (A \land B) \Rightarrow C \end{array}$$

Recall the *probabilistic* Ramsey test assumed in the preceding sections of this paper:

$$p \Vdash_{\alpha} A \Rightarrow B \iff \alpha(p(\cdot|A)) \le B \text{ or } p(A) = 0.$$

Given this test, validation of system R trivializes uncertain acceptance in the sense defined as follows. Say that acceptance rule α is *skeptical* if and only if there is some answer to \mathcal{E} that is accepted by α over no open subset of \mathcal{P} . Say that α is *opinionated* if and only if there is no open subset of \mathcal{P} over which α accepts an incomplete, disjunctive answer to \mathcal{E} as strongest. Finally, B is *trivial* if and only if α is either skeptical or opinionated.

Theorem 4 (Skepticism or opinionation). Let question \mathcal{E} has cardinality ≥ 3 . Suppose that acceptance rule α is everywhere consistent, corner-monotone, and validates system R. Then α is trivial.

Since the probabilistic Ramsey test is based on probabilistic conditioning, acceptance rules must respect the geometry of conditioning in order to validate axioms of nonmonotonic reasoning. What Theorem 4 says is that these geometrical constraints become hopelessly severe when one adds rational monotonicity to system P. Of course, the situation is quite different if one drops probabilistic conditioning from the Ramsey test.¹⁹ A *conditional* acceptance rule is a mapping $\beta : \mathcal{P} \times \mathcal{A} \to \mathcal{A}$, where $\beta(p|A) = B$ is interpreted as saying that *B* is the strongest proposition accepted in *p* in light of new

¹⁹The approach that follows is due to Hannes Leitgeb, who presented his unpublished results at the Opening Celebration of the Center for Formal Epistemology at Carnegie Mellon University in the Summer of 2010. The discussion in this section is based on detailed slides he presented at that meeting and on personal communication with him at that time.

information A. Then one can state a new, *non-probabilistic* Ramsey test directly in terms of conditional acceptance:

$$p \Vdash_{\beta} A \Rightarrow B \iff \beta(p|A) \le B.$$
 (25)

Such a conditional acceptance rule is an abstract concept that can be filled out in various different ways. For example, say that conditional acceptance rule β is *Bayesian* if and only if there exists a (non-conditional) acceptance rule α such that:

$$\beta(p|A) = \begin{cases} \alpha(p(\cdot|A)) & \text{if } p(A) > 0; \\ \bot & \text{otherwise.} \end{cases}$$
(26)

When β is Bayesian, the new information A is used to condition the credal state p to obtain $p(\cdot|A)$ and then some new propositional belief state S' is accepted in light of $p(\cdot|A)$ (the upper path in Figure 20). If β is Bayesian, then the non-probabilistic



Figure 20: Two paths

Ramsey test for β is equivalent to the probabilistic Ramsey test for α , so Theorem 4 still applies to β . But β need not be Bayesian. For example, β may sidestep Bayesian conditioning entirely by using α to accept a propositional belief state $S = \alpha(p)$ in p and by subsequently applying a propositional belief revision operator $*_p$ (that may depend on p) to convert $\alpha(p)$ into a new propsitional belief state $S' = \alpha(p) *_p A$ (the lower path in Figure 20).

$$\beta(p|A) = \alpha(p) *_p A. \tag{27}$$

In that case, the validation of system R depends entirely on the propositional revision operator $*_p$ —probabilistic conditioning and α are both irrelevant, so the geometrical proof of Theorem 4 is also sidestepped. To validate Rational Monotonicity, it is sufficient to require that each $*_p$ be an *AGM belief revision operator* (Harper 1975, Al-

Lin and Kelly

chourrón et al. 1985), thanks to the translation between nonomonotonic logic and belief revision due to Makinson and Gärdenfors (1991).²⁰

The escape route just described does not really vindicate or explain Rational Monotonicity from a Bayesian perspective, since Bayesian conditioning is bypassed and Rational Monotonicity is simply imposed on the propositional belief revision operator $*_p$. Moreover, as explained above, non-Bayesian conditional acceptance rules validate system R. On the other hand, it is an immediate corollary of Theorem 2 that the Bayesian rules of form:

$$\beta(p|A) = \nu(p(\cdot|A)) \tag{28}$$

all validate system P with respect to the non-probabilistic Ramsey test. We propose, therefore, that system P reflects Bayesian ideals better than system R.

The proof of Theorem 4 proceeds by establishing a slightly stronger result. The following two properties of an acceptance rule are derivable from validation of system R: Say that α satisfies *preservation* if and only if, for each credal state p and each proposition E, if (new information) E is consistent with (the prior belief state) $\alpha(p)$, then (the posterior belief state) $\alpha(p(\cdot|E))$ entails $\alpha(p) \wedge E$. Say that α satisfies *inclusion* if and only if, for each credal state p and each proposition E, if or each credal state p and each proposition E, $\alpha(p(\cdot|E))$ is entailed by $\alpha(p) \wedge E$.²¹ So, to prove Theorem 4, it suffices to prove the following theorem:

Theorem 5. Let question \mathcal{E} has cardinality ≥ 3 . Suppose that acceptance rule α is everywhere consistent, corner-monotone, and satisfies preservation and inclusion. Then α is trivial.

The proof of Theorem 5 proceeds by a sequence of lemmas and occupies the balance of this section. Suppose that rule α is everywhere consistent, corner-monotone, and satisfies preservation and inclusion. Suppose further that α is not skeptical. It suffices to show that α is opinionated. Let E_i, E_j be distinct answers to \mathcal{E} . Choose an arbitrary, third answer E_m to \mathcal{E} (since \mathcal{E} is assumed to have at least three answers). Let $\Delta e_i e_j e_m$ denote the two dimensional facet $\mathcal{P}|(E_i \vee E_j \vee E_m)$ (Figure 21.a). Let $\overline{e_i e_m}$ denote the one-dimensional facet $\mathcal{P}|(E_i \vee E_m)$, and similarly for $\overline{e_i e_j}$ and $\overline{e_j e_m}$. Let L_{im} be the set of the credal states on line segment $\overline{e_i e_m}$ at which E_i is accepted by α as strongest; namely:

$$L_{im} = \{ p \in \overline{e_i e_m} : \alpha(p) = E_i \}.$$

²⁰It is not necessary, though, because to validate system R one does not have to require that $*_p$ satisfies the consistency axiom in AGM—but all the other axioms have to be satisfied. We thank David Etlin for pointing this out.

²¹Inclusion is derivable from system P alone. Preservation is derivable from a special case Rational Monotonicity alone in which A is replaced by the tautology \top . They are so named because of their correspondence to the AGM axioms K^*3 and K^*4 .



Figure 21: Why system R is trivial

Lemma 4. L_{im} is a connected line segment of nonzero length that contains e_i but does not contain e_m .

Proof. By non-skepticism, there exists open subset O of \mathcal{P} over which α accepts E_i as strongest. Let O' be the image $\{o|_{E_i \lor E_m} : o \in O\}$ of O under conditioning on $E_i \lor E_m$. Since O is open, O' is an open subset of $\overline{e_i e_m}$. Note that the conditioning proposition $E_i \vee E_m$ is consistent with the prior belief state E_i , so preservation applies. Since α satisfies preservation, α accepts old belief E_i over O'. It follows that α accepts E_i as strongest over O', because α is consistent and the only proposition strictly strongest than E_i in the algebra is the inconsistent proposition \perp . So L_{im} is nonempty. Then, since α is corner-monotone, L_{im} is a nonempty, connected line segment that contains e_i . It remains to show that L_{im} does not contain e_m . Suppose for reductio that L_{im} contains e_m , then L_{im} must be so large that it is identical to $\overline{e_i e_m}$, by corner-monotonicity. By the same argument for showing that there is an open subset O' of $\overline{e_i e_m}$ over which α accepts E_i , we have that there is an open subset O'' of $\overline{e_i e_m}$ over which α accepts E_m . So α accepts both E_m and E_i over O'', and hence by closure under conjunction, α accepts their conjunction, which is an inconsistent proposition. So α is not consistent contradiction.

Let *a* be the endpoint of L_{im} that is closest to e_m ; namely, probability measure *a* is such that:

$$a \in \overline{e_i e_m},$$

$$a(E_m) = \sup\{p(E_m) : p \in L_{im}\}.$$

By the lemma we just proved, point *a* lies in the interior of side $\overline{e_i e_m}$. Applying the above argument for pair (i, m) to pair (j, m), we have that the set L_{im} , defined by

$$L_{jm} = \{ p \in \overline{e_j e_m} : \alpha(p) = E_j \},\$$

is a connected line segment of nonzero length that contains e_j but does not contain e_m , with endpoint *b* that lies in the interior of side $\overline{e_j e_m}$. Since both points *a*, *b* lie in the interiors of their respective sides, we have the following constructions. Let *A* be the line that connects *a* to e_j , *B* be the line that connects *b* to e_i , and *C* be the line that connects e_m through the intersection *d* of *A* and *B*, to point *c* on side $\overline{e_i e_j}$.

Lemma 5. α accepts E_i as strongest over the interior of $\triangle ade_i$.

Proof. Consider an arbitrary point p in the interior of $\triangle ade_i$ (Figure 21.b). Argue as follows that α accepts $E_i \lor E_j$ at p. Take p as a prior state and consider $\neg E_j$ as the conditioning information. Note that credal state $p(\cdot|\neg E_j)$ falls inside L_{im} , so α accepts E_i as strongest at the posterior credal state $p(\cdot|\neg E_j)$. Then, since α satisfies inclusion, we have that:

$$\alpha(p) \wedge \neg E_i \models E_i$$

(namely the posterior belief state E_i is entailed by the conjunction of the the prior belief state and the conditioning information). Then, by the consistency of α and the mutual exclusion among the answers, we have only three possibilities for $\alpha(p)$:

$$\alpha(p)$$
 is either E_i , or E_j , or $E_i \lor E_j$.

Rule out the last two alternatives as follows. Suppose for reductio that the prior belief state $\alpha(p)$ is E_j or $E_i \vee E_j$. Consider $\neg E_i$ as the conditioning information, which is consistent with the prior belief state and thus makes preservation applicable. Then, since α satisfies preservation, the posterior belief state $\alpha(p(\cdot|\neg E_i))$ must entail $\alpha(p) \land \neg E_i$ (i.e. the conjunction of the prior belief state and the information). But the latter proposition $\alpha(p) \land \neg E_i$ equals E_j , by the reductio hypothesis. So $\alpha(p(\cdot|\neg E_i)) = E_j$, by the consistency of α . Hence $p(\cdot|\neg E_i)$ lies on line segment L_{jm} by the construction of L_{jm} —but that is impossible (Figure 21.b). Ruling out the last two alternatives for $\alpha(p)$, we conclude that $\alpha(p) = E_i$.

Lemma 6. α accepts E_i as strongest over the interior of $\overline{e_i c}$.

Proof. Let *p* be an arbitrary interior point of $\triangle ade_i$. So $\alpha(p) = E_i$. Consider proposition $E_i \lor E_i$ as the conditioning information. Then, since α satisfies preservation, the

posterior belief state $\alpha(p(\cdot|E_i \vee E_j))$ entails $\alpha(p) \wedge (E_i \vee E_j)$ (i.e. the conjunction of the prior belief state and the information), which equals E_i . Then, by consistency, the posterior belief state is determined:

$$\alpha(p(\cdot|E_i \vee E_i)) = E_i$$

Let q be an arbitrary point in the interior of $\overline{e_i c}$. Then q can be expressed as $q = p(\cdot|E_i \vee E_j)$ for some point p in the interior of $\triangle ade_i$ (Figure 21.c). So, by the formula we just proved, $\alpha(q) = \alpha(p(\cdot|E_i \vee E_j)) = E_i$, as required.

Lemma 7. There is no open subset of $\overline{e_i e_j}$ over which α accepts $E_i \lor E_j$ as strongest.

Proof. We have established in the last lemma that α accepts E_i as strongest over the interior of $\overline{e_i c}$. By the same argument, α accepts E_j as strongest over the interior of $\overline{e_j c}$ (Figure 21.c). So if α accepts disjunction $E_i \lor E_j$ as strongest somewhere on $\overline{e_i e_j}$, α does so at some of the three points: e_i , e_j , and c. (We can rule out the first two alternatives; but the for the sake of the lemma, this result suffices.)

Since the choice of E_i and E_j is arbitrary, the last lemma generalizes to the following:

Lemma 8. For each pair of distinct answers E_i , E_j to \mathcal{E} , there is no open subset of $\overline{e_i e_j}$ over which α accepts $E_i \lor E_j$ as strongest.

The last lemma establishes opinionation for all edges of the simplex. The next step is to extend opinionation to the whole simplex.

Lemma 9. For each disjunction D of at least two distinct answers to the question, there is no open subset of \mathcal{P} over which α accepts D as strongest.

Proof. Suppose for reductio that some disjunction $E_i \vee E_j \vee X$ of at least two distinct answers is accepted by α as strongest over some open subset O of \mathcal{P} . Take $E_i \vee E_j \vee X$ as the prior belief state at each point in O and consider $E_i \vee E_j$ as the conditioning information. So the image $O' (= \{p(\cdot|E_i \vee E_j) : p \in O\})$ of O under conditioning on $E_i \vee E_j$ is an open subset of 1-dimensional space $\overline{e_i e_j}$. Let p' be an arbitrary point in O'. Since α satisfies inclusion, posterior belief state $\alpha(p')$ is entailed by $(E_i \vee E_j \vee X) \wedge$ $(E_i \vee E_j)$ (i.e. the conjunction of the prior state and the new information), which equals $E_i \vee E_j$. But $\alpha(p')$ also entails $E_i \vee E_j$, for otherwise the process of conditioning p'on $\neg E_j$ to obtain e_i would violate the fact that α satisfies inclusion and accepts E_i at e_i . So $\alpha(p') = E_i \vee E_j$. Hence α accepts $E_i \vee E_j$ as strongest over open subset O' of $\overline{e_i e_j}$, which contradicts the last lemma.

Proof of Theorem 5. Since the last lemma states that α is opinionated, we are done. \Box

Proof of Theorem 4. Immediate from Theorem 5.

12 A new probabilistic semantics for flat conditionals

Axiom system P is characteristic of Adams' logic of flat conditionals, so it is not surprising that the probalogical rules yield a new probabilistic semantics for which Adams' logic is sound. In fact, Adams' logic is both sound and complete for the new semantics.

Let \mathcal{L} be a set of sentences of a propositional language that is closed under conjunction, disjunction, and negation. Let \Rightarrow be a sentential connective standing for "if ... then ...". The language for the logic of *flat conditionals*, written $\mathcal{L}_{\Rightarrow}$, is the set of all sentences $\varphi \Rightarrow \psi$ with $\varphi, \psi \in \mathcal{L}$. Adams' logic of flat conditionals for language $\mathcal{L}_{\Rightarrow}$ is just the system P that we have stated, except that now it is construed as a system of rules of inference (with the symbol " $p \Vdash_{\alpha}$ " deleted). Say that γ is *derivable* from Γ in Adams' logic of flat conditionals, written $\Gamma \vdash_{\text{Adams}} \gamma$, if and only if γ is derivable from Γ in a finite number of steps using the rules of inference in system P.

A probabilistic model of acceptance for language $\mathcal{L}_{\Rightarrow}$ is a triple:

$$M = (\alpha, p, \llbracket \cdot \rrbracket),$$

where $\alpha : \mathcal{P} \to \mathcal{A}$ is an acceptance rule, p is a probability measure in the domain \mathcal{P} of α , and $\llbracket \cdot \rrbracket$ is a classical interpretation of \mathcal{L} to the codomain \mathcal{A} of α . When $M = (\alpha, p, \llbracket \cdot \rrbracket)$, say that α is the *underlying* acceptance rule of M. Let $\varphi \Rightarrow \psi$ be a flat conditional in $\mathcal{L}_{\Rightarrow}$. Acceptance of flat conditional $\varphi \Rightarrow \psi$ in model $M = (\alpha, p, \llbracket \cdot \rrbracket)$, written $M \Vdash \varphi \Rightarrow \psi$, is defined by the probabilistic Ramsey test:

$$\begin{split} M \Vdash \varphi \Rightarrow \psi & \longleftrightarrow \quad p \Vdash_{\alpha} \llbracket \varphi \rrbracket \Rightarrow \llbracket \psi \rrbracket, \\ & \longleftrightarrow \quad p(\cdot | \llbracket \varphi \rrbracket) \Vdash_{\alpha} \llbracket \psi \rrbracket \text{ or } p(\llbracket \varphi \rrbracket) = 0. \end{split}$$

Let Γ be a set of flat conditionals in $\mathcal{L}_{\Rightarrow}$. Acceptance of Γ in model M is defined by: $M \Vdash \Gamma$ if and only if $M \Vdash \gamma$ for all $\gamma \in \Gamma$. Validity is defined straightforwardly, as preservation of acceptance. Let C be a class of acceptance rules. Say that C validates the inference from Γ to γ , written $\Gamma \Vdash_c \gamma$, if and only if for each probabilistic model Mwhose underlying acceptance rule is in C, if $M \Vdash \Gamma$, then $M \Vdash \gamma$.

The proposed probabilistic semantics has the following attractive properties: (i) it is based on the probabilistic Ramsey test for accepting conditionals; (ii) it defines validity simply as preservation of acceptance, which improves upon Adams' ϵ - δ semantics (Adams 1975); and (iii) it allows for accepting propositions of probabilities

significantly less than 1, which improves upon Pearl's infinitesimal semantics (Pearl 1989). To establish the soundness and completeness result for Adams' logic of flat conditionals, it suffices to assume that the underlying acceptance rule is probalogical, or equivalently, a camera shutter rule:

Theorem 6 (soundness and completeness, Lin 2011). Let N be the class of the camera shutter rules. Then, for each finite sentence set Γ and each sentence γ in the language $\mathcal{L}_{\Rightarrow}$ of flat conditionals, $\Gamma \vdash_{\text{Adams}} \gamma$ if and only if $\Gamma \Vdash_{N} \gamma$.

13 Question-invariance

To this point, we have considered acceptance only within a fixed question \mathcal{E} . But one can and should consider the behavior of acceptance rules *across* questions. Let Ω denote some infinite collection of possibilities. A *question* $\mathcal{E} = \{E_i : i \in I\}$ is a countable partition of Ω such that each answer/cell E_i is infinite—the requirement of infinite answers rules out the artificial possibility of a maximally informative question whose answers cannot be strengthened. Let $\mathcal{A}_{\mathcal{E}}$ denote the least collection of propositions containing \mathcal{E} and closed under negation and countable disjunction and conjunction. Let \mathbb{E} denote the set of all such questions over Ω , and let \mathbb{P} denote the set of all countably additive probability measures p such that p is defined on $\mathcal{A}_{\mathcal{E}}$ for some question \mathcal{E} in \mathbb{E} . If p is in \mathbb{P} , let \mathcal{A}_p denote the domain of p and let \mathcal{E}_p denote the (unique) question that generates \mathcal{A}_p . A (*cross-question*) *acceptance rule* is a map β defined on \mathbb{P} such that β always maps p to a proposition in \mathcal{A}_p . Then the rules discussed earlier in this paper can be defined explicitly across questions as follows:

$$\lambda_r(p) = \bigwedge \{\neg E_i : p_i \le 1 - r \text{ and } i \in I_p\};$$

$$\lambda(p) = \lambda_{s(p)}(p), \text{ where } s(p) = 1 - \frac{1}{2|I_p|};$$

$$\nu_r(p) = \bigwedge \{\neg E : \sigma(p)_i \triangleleft_i 1 - r_i \text{ and } i \in I_p\};$$

$$\pi_r(p) = \begin{cases} \top & \text{if } \lambda_r(p) = \bot;\\ \lambda_r(p) & \text{otherwise.} \end{cases}$$

Rule λ_r is the Lockean rule with a fixed threshold across all questions in \mathbb{E} . Rule ν_r is the probalogical rule. Rule λ is the ad hoc Lockean rule whose threshold is adjusted to avoid lottery paradoxes in finite questions. Rule π_r is the Pollockian rule that substitutes \top for \bot whenever the latter is produced by λ_r .

Say that cross-question acceptance rule β is *question-invariant* if and only if:

$$p(A) = q(A) \implies (p \Vdash_{\beta} A \iff q \Vdash_{\beta} A),$$

for each p, q in \mathbb{P} and for each A that is in both \mathcal{A}_p and \mathcal{A}_q . Question-invariance is appealing. First, question-invariance makes it easier to compute whether to accept Ain light of p, since all of the detailed structure of p aside from p(A) can be ignored. Second, question-invariance allows for the accumulation of accepted propositions as one's question is refined by new concepts and theories. Third, question-invariance allows individual scientists pursuing distinct questions to *pool* their accepted conclusions. Probalogical rules, however, are not even remotely question-invariant. For example, in a four ticket lottery, the probalogical rule $v_{2/3}$ licenses acceptance of "ticket 1 will lose" when the question is "will ticket 1 lose or not?", but not when the question is "which ticket will win?". That makes one wonder whether the question-dependence of probalogical rules is a design defect that could have been avoided. We now proceed to demonstrate that *no* question-invariant rule has the three crucial virtues of the probalogical rules: consistency, logical closure, and non-skeptical acceptance of uncertain propositions.

Here is the first sign of trouble. Say that acceptance rule β is *non-skeptical* about answer *E* in question \mathcal{E} if and only if β accepts *E* at some probability measure *p* defined on $\mathcal{A}_{\mathcal{E}}$ such that p(E) < 1. Say that acceptance rule β is *gullible* about *E* in \mathcal{E} if and only if β accepts *E* at some *p* defined on $\mathcal{A}_{\mathcal{E}}$ such that p(E) = 0. Then:

Proposition 11. Suppose that β is question-invariant. If β is non-skeptical about answer *E* in ternary question \mathcal{E} , then β is gullible about *E* in \mathcal{E} .

Proof. Consider the equilateral triangle $\triangle q u v$ depicted in Figure 22.a. Note that p lies



Figure 22: Triangles preserve acceptance

on a line parallel to $\overline{e_2 e_3}$ extending the base \overline{uv} of the triangle $\triangle q uv$ and q is at the

apex. Suppose that $p \Vdash_{\beta} E_1$. Then $u, v \Vdash_{\beta} E_1$, by question-invariance. So $u \Vdash_{\beta} \neg E_2$ and $v \Vdash_{\beta} \neg E_3$. Then by question-invariance again, $q \Vdash_{\beta} \neg E_2$ and $q \Vdash_{\beta} \neg E_3$. So $q \Vdash_{\beta} \neg E_2 \land \neg E_3 = E_1$. Therefore, if β accepts E_1 at p, then β also accepts E_1 at q. Now we can chain such triangles all the way to the bottom of \mathcal{P}_3 to obtain s such that $s \Vdash E_1$ and $s(E_1) = 0$. Note that if $p(E_1) < 1$, there is room in \mathcal{P}_3 for such a chain.

It gets worse. Say that β is *dogmatic* about answer *E* in question \mathcal{E} if and only if β accepts *E* at each probability measure defined on $\mathcal{R}_{\mathcal{E}}$.

Proposition 12. Suppose that β is question-invariant. If β is non-skeptical about answer *E* in ternary question \mathcal{E} , then β is dogmatic about *E* in \mathcal{E} .

Proof. Consider the situation depicted in Figure 22.b, in which question-invariant rule β accepts E_1 at s, with $s(E_1) = 0$, and let q be an arbitrary credal state in \mathcal{P}_3 . Then there exists an equilateral triangle with s on its base and with q at its apex, so β also accepts E_1 at the arbitrarily chosen state q.

Here is the *coup de grâce*. Say that β is *everywhere inconsistent* if and only if $\beta(p) = \bot$, for all p in \mathbb{P} . Nothing could be more useless than an acceptance rule that accepts the contradiction in every possible credal state and every possible question.

Theorem 7. Suppose that β is question-invariant. If β is non-skeptical about at least two distinct answers in some ternary question, then β is everywhere inconsistent.

Proof. Suppose that β is non-skeptical about at least two distinct answers E_i, E_j in ternary question \mathcal{E} . Then, by Proposition 12, β accepts $E_i \wedge E_j$ and, thus, \perp at every state in question \mathcal{E} . But \perp has the same probability, namely 0, at every state in every question. So, by question-invariance, \perp is accepted at every state in every question. \Box

It follows from the preceding propositions that *none* of the rules listed above is question-invariant. That fact is obvious for probalogical rules and the ad hoc rules, all of which base acceptance explicitly on the underlying question. However, even the logically closed Lockean rule with fixed threshold is question-dependent whenever the threshold is strictly between 0 and 1—for then the rule is neither skeptical nor everywhere inconsistent (at threshold 0 it is everywhere inconsistent and at threshold 1 it is skeptical). If closure under conjunction is dropped, the Lockean rule with a fixed threshold is question-invariant and is non-skeptical, but is also consistent, so it escapes Theorem 7 (recall that set-valued rules are not covered by that proposition).

We are inclined to view Theorem 7 as a reductio argument against questioninvariance. That conclusion fits naturally with a minimalist, pragmatic interpretation of accepted proposition A as a more or less apt *proxy* for one's underlying credal state p,
rather than as new "information" that alters p (e.g., by conditioning p on A). Questioninvariance would be *nice*, but it is not rationally mandated under the minimalist conception of acceptance, and its price in terms of logical virtues *within* questions is too high.

14 Refinement-monotonicity

Invariance across all questions is a strong requirement. In this section, we consider the consequences of requiring invariance only over questions that refine or coarsen the given question. Say that \mathcal{E} refines \mathcal{F} (or that \mathcal{F} coarsens \mathcal{E}) if and only if each answer to \mathcal{E} entails some answer to \mathcal{F} . When \mathcal{E} refines \mathcal{F} , write $\mathcal{E} \leq \mathcal{F}$. By extension, say that *p* refines *q* (written $p \leq q$) when *q* is the restriction of *p* to \mathcal{A}_q , which implies that $\mathcal{E}_p \leq \mathcal{E}_q$. Say that cross-question acceptance rule β is refinement-invariant if and only if:

$$p \leq q \Longrightarrow (p \Vdash_{\beta} A \iff q \Vdash_{\beta} A),$$

for each p, q in \mathbb{P} and for each proposition A in \mathcal{A}_q . However:

Proposition 13. Refinement-invariance is equivalent to question-invariance.

Proof. Suppose that p(A) = q(A). Let r = (p(A), 1 - p(A)) over question $\{A, \neg A\}$. Then $p \leq r \geq q$. By refinement-invariance, if follows that $p \Vdash_{\beta} A \iff q \Vdash_{\beta} A$. The converse is immediate.

Refinement-invariance demands that acceptance be preserved under both refinement and coarsening. Since questions tend to become more precise as inquiry proceeds, perhaps it suffices merely to preserve acceptance under refinement. Accordingly, say that β is *refinement-monotone* if and only if:

$$p \leq q \implies \beta(p) \leq \beta(q),$$

for all p, q in \mathbb{P} . Refinement-monotonicity suffices for the accumulation of accepted conclusions as the question is refined and for the pooling of propositions accepted across diverse questions. With respect to the latter, let p, q, r be in \mathbb{P} . Say that r is a *conjunction* of p, q if and only if r is a maximally coarse common refinement of p, q. Then say that β *preserves conjunction* if and only if $\beta(r) \leq \beta(p) \land \beta(q)$, for each p and q in \mathbb{P} and for each conjunction r of p and q. Then it is easy to show that:

Proposition 14. Conjunction-preservation is equivalent to refinement-monotonicity.

Alas, probalogical rules also violate refinement-monotonicity—as witnessed by the simple lottery example in the preceding section of this paper. Again, the failure is not a defect but an ineluctable consequence of the logical virtues of probalogical rules.

Theorem 8. Suppose that β is refinement-preserving, validates system P in each question, and is non-skeptical about both answers in some binary question. Then there exists a facet of at least two dimensions over which β accepts \perp everywhere.

The alternative rules listed above also violate refinement-monotonicity, even though they all fail to validate system P. Choosing a probalogical rule at least yields the net advantage of validating P.

The proof of Theorem 8 proceeds by a sequence of lemmas that rely heavily on the geometrical characterizations of the axioms of P established in Section 9. Consider the binary question $\{E_0, F_0\}$, whose space of credal states is depicted in Figure 23.a as the line next to the triangle. Assume that β is non-skeptical about answers E_0 and



Figure 23: Acceptance snakes up the triangle

 F_0 , so that β accepts E_0 at p_0 and F_0 at q. Since E_0 is infinite, split E_0 into infinite answers F_1 and E_1 to produce the refined, ternary question $\{F_0, F_1, E_1\}$ (Figure 23.b). Suppose that β is refinement-monotone. Then proposition F_0 is accepted throughout the line segment L depicted in Figure 23.b, which is defined to be the set of all credal states that refine q. Similarly, proposition $E_0 = E_1 \vee F_1$ is accepted throughout the line segment M, which is the set of all credal states that refine p_0 . Let line segment Nconnect the right endpoint of L in Figure 23.b to the opposite corner e_1 , intersecting Mat credal state u; then project u to the (one-dimensional) facet for proposition $E_1 \vee F_0$ to obtain credal state p_1 . The following lemma concerns p_1 .

Lemma 10. Suppose β is refinement-monotone and validates system P. Then $p_1 \Vdash_{\beta} E_1$.

Lin and Kelly

Proof. Proposition E_1 is accepted by β at e_1 , by the geometry of Reflexivity (Proposition 8); and F_0 is accepted at each point on L, by construction. So the disjunction $E_1 \vee F_0$ is accepted by β at both endpoints of N. Then, since u lies on N, we have that $u \Vdash_{\beta} E_1 \vee F_0$, by the geometry of Or (Proposition 10). We have noted that $u \Vdash_{\beta} E_1 \vee F_1$. So $u \Vdash_{\beta} E_1$, because $E_1 = (E_1 \vee F_0) \wedge (E_1 \vee F_1)$. Then, since p_1 is the projection of u onto the facet for a logical consequence of E_1 , the geometry of Cautious Monotonicity (Proposition 9) yields that $p_1 \Vdash_{\beta} E_1$, as required.

The result is that E_1 is accepted by β with a lower probability than E_0 . Split E_1 into two infinite, exclusive propositions E_2 and F_2 and, thus, obtain the finer, quaternary question $\{F_0, F_1, F_2, E_2\}$. Restrict attention to the two-dimensional, triangular facet for proposition $F_0 \lor F_2 \lor E_2$, as depicted in Figure 23.c. Construct credal state p_2 as we did for p_1 , and argue similarly that E_2 is accepted at p_2 , with an even lower probability than the probability at which E_1 is accepted at p_1 . This construction can be repeated until we obtain a refined, finite question $\{F_0, F_1, \ldots, F_n, E_n\}$ such that E_n is accepted at p_n with low probability (Figure 23.d)—so low that p_n is far away from corner e_n and lies on or above the line L. Therefore:

Lemma 11. Continuing from the preceding lemma, $p_n \Vdash_{\beta} E_n$.

Then inconsistency arises:

Lemma 12. Continuing from the preceding lemma, let line segment $p_n f_n$ intersect L at v. Then $v \Vdash_{\beta} \bot$.

Proof. Proposition E_n is accepted by β at e_n , by construction; and F_0 is accepted at f_n , by the geometry of Reflexivity (Proposition 8). So the disjunction $E_n \vee F_n$ is accepted by β at both endpoints of line segment $\overline{p_n f_n}$. Then, by the geometry of Or (Proposition 10), $v \Vdash_{\beta} E_n \vee F_n$. But $v \Vdash_{\beta} F_0$, because v lines on L and thus refines q. Since $\bot = F_0 \wedge (E_n \vee F_n)$, we have that $v \Vdash_{\beta} \bot$, as required.

Here is the *coup de grâce*, of which Theorem 8 is an immediate corollary.

Lemma 13. Continuing from the preceding lemma, let \mathcal{P}_{n+2} be the set of probability measures defined on $\mathcal{A}_{\mathcal{E}_{n+2}}$, where \mathcal{E}_{n+2} is the question $\{F_0, \ldots, F_n, E_n\}$. Then β accepts \perp at each credal state p in facet $\mathcal{P}_{n+2}|(F_0 \vee F_n \vee E_n)$.

Proof. Let \triangle denote the two-dimensional facet $\mathcal{P}_{n+2}|(F_0 \lor F_n \lor E_n)$. Suppose that v lies in the interior, but not the sides, of \triangle . Since \bot is accepted at v, we have that \bot is accepted at the three corners f_0, f_n, e_n of \triangle , by projecting v to the three corners and by the geometry of Cautious Monotonicity (Proposition 9). Then, since each side of \triangle has endpoints that are corners, we have that \bot is accepted on the three sides of \triangle , by the geometry of Or (Proposition 10). Then, since each point on \triangle is on a line segment with endpoints on the sides of \triangle , we have that \bot is accepted at each credal state on \triangle , as required. When *v* is *not* in the interior of \triangle , *v* lies on side $\overline{e_n f_0}$ of \triangle and, thus, cannot be projected to the opposite corner f_n . But in that case we can apply the geometry of Or (Proposition 10) to line segment $\overline{v f_n}$ to show that F_n is accepted at every credal state on $\overline{v f_n}$. Similarly, $F_0 \lor E_n$ is accepted at every credal state on $\overline{w e_n}$, where *w* is defined to be the intersection of line *L* and $\overline{f_0 f_n}$. So \bot is accepted at the intersection of $\overline{v f_n}$ and $\overline{w e_n}$, which is in the interior of \triangle —the second case is thus reduced to the first case. \Box

15 Probalogic generalized

We close with a natural generalization of the probalogical framework. The uniform probability measure over \mathcal{E} is the center of the simplex \mathcal{P} of probability measures on the least algebra over question \mathcal{E} and served as the probalogically weakest credal state in \mathcal{P} in the presentation to this point. But, as Levi (1967) has emphasized, the answers to question \mathcal{E} typically have different *contents* (e.g., "quantum mechanics is true" has a great deal of content but "quantum mechanics is false" has very little). Therefore, a credal state that assigns less probability to a cell that has more content could sensibly be understood as weaker than a uniform state that accords the same probability to all cells. In that case, probalogic should be relative not only to question \mathcal{E} , but to an assignment of contents to the answers to \mathcal{E} . The result is a family of probalogics sensitive both to the cardinality of question \mathcal{E} and to the relative contents of the answers to \mathcal{E} .

We approach the issue as follows. If the answers E_i differ in content, it is natural to weight answers by *weakness* and to think of the neutral credal state as the center of mass of the answers. As a result, the weakest credal state is biased toward answers of low content. In particular, the center of \mathcal{P} is stronger than a state closer to a very weak answer. Recall that probalogic is just the geological cube in perspective. The sides of the cube have equal length. To represent differences in content, deform the cube into an oblong box whose side lengths are inversely proportional to the strengths of the corresponding answers. (Figure 24). Just like the cube, the oblong box may be viewed as a *generalized geologic* (recall that geological structure does not uniquely determine the metric). Project the generalized geologic from the box to the triangle credal state space, just as before, to induce a *generalized probalogic* on it. Then the credal states stronger than p are those in the grey region of Figure 24.d. Disjunction and conjunction are defined as before.

The weakest proposition in the generalized geologic is (m_1, m_2, m_3) (i.e. the vertex of the box that is most distant from the origin), so its rectilinear projection *w* to the triangle is the weakest credal state in the corresponding probalogic. Projection pre-



Figure 24: Deformation of geologic and corresponding deformation of probalogic

serves ratios between the rectangular coordinates, so we have: $w = \left(\frac{m_1}{M}, \frac{m_2}{M}, \frac{m_3}{M}\right)$, where $M = \sum_{i \in I} m_i$. The coordinates of *w* uniquely determine the generalized probalogic that has *w* as the weakest state. Intuitively, the result is like viewing a phone booth, rather than a cubical office, from the origin (Figure 24.d).²² Acceptance rules are still defined as maps that preserve probalogical structure and they look like Figure 25. Although the generalized probalogical acceptance rules appear "oblique", the boundaries of acceptance zones still follow rays from the corners—so they still validate exactly Adams'

²²In terms of projective geometry, the geological transformation is a non-rotational, non-reflective linear transformation and, thus, the induced probalogical transformation is a projective transformation that fixes all the corners.



Figure 25: Generalized probalogical acceptance rule

conditional logic. Algebraically, the generalized rules take the following form:

$$\alpha(p) = \bigwedge \left\{ \neg E_i : \frac{p(E_i)/m_i}{\max_j p(E_j)/m_j} \triangleleft_i 1 - r_i \text{ and } i \in I \right\}.$$

The acceptance rules introduced in (Levi 1996, p.286) are the same, except that we allow different thresholds r_i for different answers E_i while Levi does not. As we mentioned at the outset, Levi sees no justification for these rules, relative to his momentous understanding of acceptance as an explicit decision to condition one's credal state on the accepted proposition and, therefore, to bet one's life on it against nothing. Our own justification for the rules, grounded in a weaker conception of acceptance as apt description of one's credal state relative to a question, is again, that they preserve naturally defined logical structures over credal states relative to a question and that they validate exactly Adams' logic of conditionals.

Acknowledgements The authors are indebted to David Etlin for recommending us to explicitly state a stronger negative result for rational monotonicity. We are indebted to Teddy Seidenfeld for referring us to the passage in which Levi defines the camera shutter rules. We are indebted to Hannes Leitgeb for sharing his alternative approach based on conditional acceptance rules with us. We are also indebted to Clark Glymour and Greg Wheeler for useful questions.

References

E. W. Adams. The Logic of Conditionals. D. Reidel, Dordrecht, 1975.

C. E. Alchourrón, P. Gärdenfors, and D. Makinson. On the logic of theory change: Partial meet contraction and revision functions. *The Journal of Symbolic Logic*, 50: 510–530, 1985.

H. Arló-Costa and R. Parikh. Conditional probability and defeasible inference. *Journal of Philosophical Logic*, 34:97–119, 2005.

K. Barwise. Infinitary logic and admissible sets. *Journal of Symbolic Logic*, 34: 226–252, 1969.

I. Douven. A new solution to the paradoxes of rational acceptability. *British Journal for the Philosophy of Science*, 53:391–410, 2002.

B. van Fraassen. Fine-grained opinion, probability and the logic of full belief. *Journal of Philosophical Logic*, 24:349–377, 1995.

W. Harper. Rational belief change, popper functions and counterfactuals. *Synthese*, 30(1-2):221–262, 1975.

R. Jeffrey. Dracula meets wolfman: Acceptance vs. partial belief. In M. Swain, editor, *Induction, Acceptance, and Rational Belief.* 1970.

C. Karp. *Languages with Expressions of Infinite Length*. North Holland, Dordrecht, 1964.

K. Kelly. Ockham's razor, truth, and information. In J. van Benthem and P. Adriaans, editors, *Handbook of the Philosophy of Information*. Elsevier, Dordrecht, 2008.

K. Kelly. Ockham's razor, truth, and probability. In P. Bandyopadhyay and M. Forster, editors, *Handbook on the Philosophy of Statistics*. Elsevier, Dordrecht, 2010.

S. Kraus, D. Lehmann, and M. Magidor. Nonmonotonic reasoning, preferential models and cumulative logics. *Artificial Intelligence*, 44:167–207, 1990.

H. Kyburg. *Probability and the Logic of Rational Belief.* Wesleyan University Press, Middletown, 1961.

D. Lehmann and M. Magidor. What does a conditional base entails? *Artificial Intelligence*, 55:1–60, 1992.

I. Levi. *Gambling With Truth: An Essay on Induction and the Aims of Science*. MIT Press, Cambridge, 2nd edition, 1967.

I. Levi. Information and inference. Synthese, 19:369-91, 1969.

I. Levi. For the Sake of the Argument: Ramsey Test Conditionals, Inductive Inference and Non-monotonic Reasoning. Cambridge University Press, Cambridge, 1996.

H. Lin. A New Theory of Acceptance that Solves the Lottery Paradox and Provides a Simplified Probabilistic Semantics for Adams' Logic of Conditionals. Master's thesis, Carnegie Mellon University, Pittsburgh, PA, 2011.

D. Makinson and P. Gärdenfors. Relations between the logic of theory change and nonmonotonic logic. In A. Fuhrmann and M. Morreau, editors, *The Logic of Theory Change (Verlag Lecture Notes in Computer Science 465)*, pages 183–205, Springer, 1991.

V. Novak, I. Perfilieva, and J. Mockor. *Mathematical principles of fuzzy logic*. Kluwer, Dordrecht, 2000.

J. Pearl. Probabilistic semantics for nonmonotonic reasoning: A survey. In *Proc. First International Conference on Principles of Knowledge Representation and Reasoning*, pages 505–516. 1989.

J. Pollock. Cognitive Carpentry. MIT Press, Cambridge, MA, 1995.

F. P. Ramsey. General propositions and causality. In A. Mellor, editor, *F. Ramsey and Philosophical Papers*. Cambridge University Press, 1990, Cambridge, 1929.

S. Ryan. The epistemic virtues of consistency. Synthese, 109:121-41, 1996.

L. Zadeh. Fuzzy sets. Information and Control, 8:338-353, 1965.

Changing Types: Information Dynamics for Qualitative Type Spaces

Dominik Klein and Eric Pacuit

TiLPS Tilburg University of Maryland d.klein@uvt.nl, e.j.pacuit@uvt.nl

Abstract

Many different approaches to describing the players' knowledge and beliefs can be found in the literature on the epistemic foundations of game theory. We focus here on non-probabilistic approaches. The two most prominent are the so-called Kripke- or Aumann- structures and knowledge structures (non-probabilistic variants of Harsanyi type spaces). Much of the recent work on Kripke structures has focused on dynamic extensions and simple ways of incorporating these. We argue that many of these ideas can be applied to knowledge structures as well. Our main result characterizes precisely when one *type* can be transformed into another *type* by a specific type of information update. Our work in this paper suggest that it would be interesting to pursue a theory of "information dynamics" for knowledge structures (and eventually Harsanyi type spaces).

1 Introduction

The central thesis of the epistemic program in game theory is that the basic mathematical model of a game situation should include an explicit parameter describing the players' *informational attitudes*¹, see (Brandenburger 2008) for the relevant references

¹This is, of course, something of a truism regarding games of *incomplete* or *imperfect* information. But, the thesis is intended to apply to *all* game situations (see Section 5 of Brandenburger 2010, for a precise description about the crucial differences between an epistemic model of a game and a *Bayesian game*).

and a discussion of the key results, and (Perea 2012) for an introduction to this literature. Games are played in specific *informational contexts*, in which players have specific knowledge and beliefs about each other.² Many different formal models have been used to represent such informational contexts of a game (see Bonanno and Battigalli 1999, van der Hoek and Pauly 2006, van Benthem et al. 2011, and references therein, for a discussion). In this paper, we are not only interested in structures that describe the informational context of a game, but how these structures can *change* in response to the players' observations, communicatory acts or other dynamic operations of information change (cf. van Benthem 2011).

We focus our attention on the players' hard information about the game (which we refer to as *knowledge* following standard terminology in the game theory and epistemic logic literature) and its dynamics. Broadly speaking, there are two different types of models that have been used to describe the players' knowledge (and beliefs) in a game situation. Both types of models include a nonempty set S of states of nature (elements of S are intended to represent possible outcomes of a game situation).³ The first type of models are the so-called Aumann- or Kripke-structures (Aumann 1999, Fagin et al. 1995). These structures describe the players' knowledge in terms of an epistemic indistinguishability relation over a (finite) set of states W. The second type of models are the knowledge structures of Fagin et al. (1991; 1999), which are non-probabilistic variants of *Harsanyi type spaces* (Harsanyi 1967).⁴ The key concept here is a **type** which describes the players' infinite hierarchy of knowledge (i.e., what the players know about the ground facts, what the players know about each others knowledge of the ground facts, what players know about what each other know about each others knowledge of the ground facts, and so on.) The precise relationship between these two types of models was clarified in (Fagin et al. 1991; 1999).

Our goal in this paper is to show how to adapt recent work modeling information change on Kripke structures as a product update with an *event model* (van Ditmarsch et al. 2007) to the more general setting where the players' knowledge is represented using knowledge structures. To the best of our knowledge, this is the first attempt to develop a theory of information change for knowledge structures in the style of recent

²This is nicely explained by Adam Brandenburger and Amanda Friedenberg (2010, pg. 801): "In any particular structure, certain beliefs, beliefs about beliefs, ..., will be present and others won't be. So, there is an important implicit assumption behind the choice of a structure. This is that it is "transparent" to the players that the beliefs in the [type] structure — and only those beliefs — are possible....The idea is that there is a "context" to the strategic situation (eg., history, conventions, etc.) and this "context" causes the players to rule out certain beliefs."

³Often, it is assumed that the elements of S can be described by some logical language (for example, propositional logic), but this is not crucial for us in this paper.

⁴See (Siniscalchi 2008) for a modern introduction to type spaces as models of beliefs and (Myerson 2004) for a discussion of Harsanyi's classic paper.

Klein and Pacuit

work on dynamic epistemic logic. Our main result (Theorem 2) characterizes precisely when a type in a fixed knowledge structure can be transformed into another type in that structure using the product update operation.

There are two main motivations for this technical study. The first is to explore generalizations of the product update operation. This is done in Section 3.1 where we also generalize a result of (van Ditmarsch and French 2009) characterizing when a Kripke structure can be transformed into another Kripke structure by a product update. The second motivation for this work is to initiate a study of information dynamics for epistemic models of games. The players' information in a game can change in two ways. First, the players' knowledge and beliefs change during the play of a sequential game (for example, they learn about the choices of the other players as the game is played). The second way that the players' information can change is in response to some exogenous event. For example, during a poker game, a player may accidentally drop his cards or a gust of wind may allow a subset of the players to see certain cards. Of course, one may argue that game-theoretic models should abstract away from these payoff irrelevant events. We agree that the type of events we have in mind here are irrelevant to a game-theoretic analysis. But, these events do change the *context*⁵ of a game by revealing or hiding important information to all or some of the players. This paper is a first step towards a more general project that uses the dynamic epistemic logic framework to represent changes in the informational context of a game.

Our paper is organized as follows. Section 2 provides the necessary background on (dynamic) epistemic logic and knowledge structures. Note that this Section was written for a reader already familiar with the key concepts and definitions. Consult (van Benthem 2011) and (Fagin et al. 1999) for motivations and a broader discussion of the literature. Our main result is in Section 3.2 with the technical preliminaries found in Section 3.1. We conclude in Section 4 with a discussion of topics for future research.

2 Background

2.1 A primer on Dynamic Epistemic Logic

We assume the reader is familiar with the basics of (dynamic) epistemic logic, and so, we only give the key definitions here (see the textbooks Fagin et al. 1995, van Benthem 2011, for an introduction to the subsequent definitions). Let *I* be the finite set of players and At a (finite or infinite) set of atomic propositions.⁶

⁵Here, we take the "context" of a game to be *all* events that influence the players' beliefs in the game situation.

⁶Atomic propositions are intended to represent properties of states of nature.

Definition 2.1. The **epistemic language**, denoted \mathcal{L}_{EL} , is the smallest set of formulas generated by the following grammar:

$$\varphi ::= p | \neg \varphi | \varphi \land \varphi | K_i \varphi$$

where $p \in At$ and $i \in I$. Define $L_i \varphi$ as the dual of K_i (i.e., $L_i \varphi := \neg K_i \neg \varphi$) and the other boolean connectives (e.g., \lor, \rightarrow) as usual.

The intended interpretation of $K_i \varphi$ is "agent *i* knows that φ (is true)". The standard semantics for \mathcal{L}_{EL} are Kripke structures.

Definition 2.2. A **Kripke structure** (for a set of atomic propositions At) is a tuple $\langle W, \{R_i\}_{i \in I}, V \rangle$ where *W* is a set of states, $R_i \subseteq W \times W$ is an equivalence relation⁷, and $V : At \rightarrow \wp(W)$ is a valuation function. To simplify notation, we may write $w \in \mathcal{M}$ when $w \in W$.

Formulas of \mathcal{L}_{EL} are interpreted at states in a Kripke model in the standard way, we only remind the reader of the definition for the knowledge modality:

 $\mathcal{M}, w \models K_i \varphi$ iff for all $v \in W$ if $wR_i v$ then $\mathcal{M}, v \models \varphi$

The central idea of *dynamic* epistemic logic is to describe events that change a situation and the (uncertain) perceptions of these events by the agents' as a so-called event model.

Definition 2.3. An event model is a tuple $\langle E, \{Q_i\}_{i \in I}, pre \rangle$ where *E* is a set of basic events, $Q_i \subseteq E \times E$ is an equivalence relation⁸ and $pre : E \to \mathcal{L}_{EL}$ assigns to each primitive event a formula that serves as a **precondition** for that event. We write $e \in \mathcal{E}$ if *e* is an event in \mathcal{E} .

The primitive events represent the basic observations available to the agents in a dynamic situation. Similar to Kripke structures, uncertainty about which events are taking place is represented by relations Q_i . Given our assumptions that each Q_i is an equivalence relation, the intended interpretation of eQ_if is that agent *i* cannot distinguish between events *e* and *f*. The key operation of *product update* describes how to incorporate into a Kripke structure \mathcal{M} (describing an epistemic situation) the epistemic event described by an event model \mathcal{E} .

⁷In this paper, we restrict attention structures where the epistemic relations are equivalence relations. These are known in the literature as **S5**-structures or Aumann structures.

⁸To keep things manageable for this initial study, we restrict attention to event models with equivalence relations. For much of what follows, this assumption is not crucial.

Definition 2.4. The **product update** of a Kripke model $\mathcal{M} = \langle W, \{R_i\}_{i \in I}, V \rangle$ and an event model $\mathcal{E} = \langle E, \{Q_i\}_{i \in I}, \mathsf{pre} \rangle$ is a Kripke model $\mathcal{M} \oplus \mathcal{E} = \langle W', \{R'_i\}_{i \in I}, V' \rangle$ defined as follows:

- $W' = \{(w, e) \in W \times E \mid \mathcal{M}, w \models \mathsf{pr}e(e)\}$
- $(w, e)R'_i(w', e')$ iff wR_iw' and eQ_ie'
- $(w, e) \in V'(p)$ iff $w \in V(p)$

This operation (together with variants appropriate for modeling belief and preference change) has been extensively studied in the literature. We do not provide an overview of this literature here (see van Ditmarsch et al. 2007, van Benthem 2011, for an extensive analysis). Rather, the focus is on how to understand this theory of information dynamics in the context of models of knowledge (and beliefs) typically found in the game theory literature. We need one additional notion from the general theory of modal logic.

Definition 2.5. Suppose that $\mathcal{M}_1 = \langle W_1, R_1, V_1 \rangle$ and $\mathcal{M}_2 = \langle W_2, R_2, V_2 \rangle$ are Kripke structures. A nonempty relation $Z \subseteq W_1 \times W_2$ is a **bisimulation** provided for all $w_1 \in W_1$ and $w_2 \in W_2$, if $w_1 Z w_2$ then:

(atomic harmony) For all $p \in At$, $w_1 \in V_1(p)$ iff $w_2 \in V_2(p)$.

(**zig**) If $w_1R_1v_1$ then there is a $v_2 \in W_2$ such that $w_2R_2v_2$ and v_1Zv_2 .

(zag) If $w_2R_2v_2$ then there is a $v_1 \in W_1$ such that $w_1R_1v_1$ and v_1Zv_2 .

We write $\mathcal{M}_1, w_1 \rightleftharpoons \mathcal{M}_2, w_2$ if there is a bisimulation relating w_1 with w_2 . We write $\mathcal{M}_1 \rightleftharpoons \mathcal{M}_2$ if there is a bitotal bisimulation between \mathcal{M}_1 and \mathcal{M}_2 , that is a bisimulation Z such that for every $v \in \mathcal{M}_1$ there is some $W \in \mathcal{M}_2$ with vZw and vice versa. The relation Z is called a **simulation from** \mathcal{M}_1 to \mathcal{M}_2 , denoted $\mathcal{M}_1, w_1 \rightharpoonup \mathcal{M}_2, w_2$, if Z satisfies the **atomic harmony** and **zig** properties. Z is called **total** provided for each $w_1 \in W_1$ there is a $w_2 \in W_2$ such that w_1Zw_2 . Finally, Z is called **functional** if it is total and a function from \mathcal{M}_1 to \mathcal{M}_2 (i.e. for every $w_1 \in W_1$ and $w_2, \tilde{w}_2 \in W_2$ it is the case that w_1Zw_2 and $w_1Z\tilde{w}_2$ implies $w_2 = \tilde{w}_2$).

2.2 Knowledge structures

Knowledge structures were introduced in (Fagin et al. 1999) as an alternative semantics for the basic epistemic language \mathcal{L}_{EL} .⁹ They are non-probabilistic versions of Harsanyi

⊲

⁹See (Fagin et al. 1999) for an extended discussion of knowledge structures aimed at game theorists. Fagin (1994) and Fagin and Vardi (1985) show how variants of knowledge structures can provide an elegant semantics for many modal logics.

type spaces which are the predominant model of knowledge and beliefs in the literature on the epistemic foundations of game theory (Brandenburger (2010) offers some explanation about why this is the case).

The key concept is a κ -world (also called a **type** in the game theory literature) describing the players' infinite hierarchy of knowledge (belief) of a given state of affairs.

Definition 2.6. Let *S* be a (finite or infinite) nonempty set (whose elements are called states). A κ -world is a vector of functions $\vec{f} = \langle f_0, f_1, f_2 \dots \rangle$ of length κ (a possibly infinite ordinal) defined inductively as follows.

- A 1-world is a vector $\langle f_0 \rangle$ where f_0 is a state of nature (i.e., $f_0 \in S$).¹⁰
- For $\kappa > 1$ of the form $\kappa = \lambda + 1$ (i.e. κ is a successor ordinal) a κ -world is a vector $\langle f_0 \dots f_\lambda \rangle$ such that $\langle f_i | i < \lambda \rangle$ is a λ -world and f_λ is a function from the set of agents *I* to the power set of the set of λ -worlds over *S* (i.e., $f_\lambda : I \to \wp(\mathcal{F}_\lambda(S))$, where $\mathcal{F}_\lambda(S)$ denotes the set of all λ -worlds over *S*) that satisfies the following conditions. Let $\vec{f}_{<\beta}$ denote the initial segment of \vec{f} of length β .
 - **Extendability** If $0 < \alpha < \lambda$, then $\vec{g} \in f_{\alpha}(i)$ iff there is some $\vec{h} \in f_{\lambda}(i)$ such that $\vec{g} = \vec{h}_{<\alpha}$ (i.e., higher-order worlds are extensions of lower-order worlds and every lower-order world has at least one higher-order extension).

In addition, since we intend κ -worlds to represent the **knowledge** of the players, we impose two additional conditions:

- **Correctness** For each agent $i \in I$, $\vec{f}_{<\lambda} \in f_{\lambda}(i)$ (i.e., every agent must consider the actual state of the world possible).
- **Introspection** For all $i \in I$, if $\langle g_0, g_1, \ldots \rangle \in f_{\kappa}(i)$, then $g_{\lambda}(i) = f_{\lambda}(i)$, for all λ with $0 < \lambda < \kappa$ (i.e., players cannot consider states possible that differ in their description of their own lower-order beliefs).
- Finally, for κ a limit ordinal a κ-world is a vector of functions (f_i | i < κ) such that for every λ < κ the vector (f_i | i < λ) is a λ-world.

We denote the set of all κ -worlds over *S* by $\mathcal{F}_{\kappa}(S)$.

¹⁰For the comparison with epistemic logic, it is useful to think of the set of states *S* as the set of propositional valuations on a set At of atomic propositions. In this case f_0 would be a propositional valuation function.

Klein and Pacuit

The intended interpretation is that $f_{\kappa}(i) \subseteq \mathcal{F}_{\kappa}(S)$ is the set of all κ -worlds player *i* considers possible. Then, κ -worlds \vec{f} are descriptions of the state of affairs and the players higher-order knowledge (up to level κ). Thus, we can interpret the basic epistemic language at κ -worlds. For simplicity, we assume that each atomic proposition *E* corresponds to a subsset of the set of states *S*. This language is interpreted as follows:

$$\vec{f} \models E \Leftrightarrow f_0 \in E$$

$$\vec{f} \models \neg \varphi \Leftrightarrow \vec{f} \not\models \varphi$$

$$\vec{f} \models \varphi \land \psi \Leftrightarrow \vec{f} \models \varphi \text{ and } \vec{f} \models \psi$$

$$\vec{f} \models K_i \varphi \Leftrightarrow \text{ for each } \vec{g} \in f_l(i) : \vec{g} \models \varphi$$

where *l* is the quantifier depth¹¹ of φ .

There is an alternative way of defining truth of the knowledge modality by defining an accessibility relation on $\mathcal{F}_{\kappa}(S)$, which transforms $\mathcal{F}_{\kappa}(S)$ into a Kripke model. We can then use the standard definition of a modal operator. For a κ -world $\vec{f} = \langle f_0, f_1, \ldots \rangle$, let $\vec{f^i} = \langle f_1(i), f_2(i), \ldots \rangle$ (note that the state of nature is not part of $\vec{f^i}$) and define a relation \sim_i on the $\mathcal{F}_{\kappa}(S)$ as follows: $\vec{f} \sim_i \vec{g}$ iff $\vec{f^i} = \vec{g^i}$ (equality is defined component-wise). If $\vec{f} \sim_i \vec{g}$ then we say \vec{f} and \vec{g} are equivalent according to agent *i*. It is easy to see that these relations are equivalence relations. They turn $\mathcal{F}_{\kappa}(S)$ into a Kripke structure (with At = $\wp(S)$ and the valuation function *V* defined by $w \in V(E)$ iff $w \in E$). Fagin et al. (1991) show (Theorem 2.4.) that the interpretation of the epistemic language obtained by interpreting $\langle \mathcal{F}_{\kappa}(S), \{\sim_i\}_{i\in I}, V\rangle$ as a Kripke structure. So, there are two equivalent ways to interpret the basic epistemic language on the set $\mathcal{F}_{\kappa}(S)$ of κ -worlds. In the remainder of the paper, we will use whichever definition is most convenient.

We are interested in general maps between Kripke structures and knowledge structures. To this end, we fix a set of atomic propositions At and assume that the state space *S* is the set of propositional valuations of At, i.e. identifying a valuation with its characteristic function $S = \wp(At)$. To simplify our exposition, we identify $p \in At$ with $\{e \in S \mid p \in e\} \subseteq S$, i.e. the set of valuations containing *p*.

The key observation is that every Kripke structure can be naturally associated with a substructure of $\langle \mathcal{F}_{\omega}(S), \{\sim_i\}_{i \in I}, V \rangle$. The mapping is defined as follows.¹²

¹¹Quantifier depth is defined as usual by induction on the structure of $\varphi \in \mathcal{L}_{EL}$: Formally, qd(p) = 0, $qd(\neg \varphi) = qd(\varphi)$, $qd(\varphi \wedge \psi) = \max(qd(\varphi), qd(\psi))$, and $qd(K_i\varphi) = 1 + qd(\varphi)$.

¹²The mapping is a functional simulation but in general not a bisimulation onto its image. Nonetheless, it is a natural mapping in the sense that when applied to connected components \mathcal{K} of $\langle \mathcal{F}_{\omega}(S), \{\sim_i\}_{i \in I}, V \rangle$ it is simply the embedding of \mathcal{K} into $\langle \mathcal{F}_{\omega}(S), \{\sim_i\}_{i \in I}, V \rangle$.

Definition 2.7. Embedding from Kripke structures to knowledge structures Let $\mathcal{M} = \langle W, \{R_i\}_{i \in N}, V \rangle$ be a Kripke structure. We associate with each state $w \in W$ in \mathcal{M} an ω -world $\vec{f}_{\mathcal{M},w} = \langle f_0^w, f_1^w, f_2^w, \ldots \rangle$ where the f_{α}^w are defined by synchronous induction on all worlds $w \in W$:

- $f_0^w = \{p \mid w \in V(p)\}.$
- To define the function f_{k+1}^w assume inductively that f_0^x , f_1^x , f_2^x , ..., f_k^x have been defined for all worlds $x \in W$ (k a natural number). Then, $f_{k+1}^w(i) = \{\langle f_0^x, f_1^x, \dots, f_k^x \rangle | wR_i x\}$.

Define the map $r: W \to \mathcal{F}_{\omega}(\wp(\mathsf{At}))$ as $r(w) = \vec{f}_{\mathcal{M},w}$.

For every infinite ordinal λ we can continue the construction transfinitely to get a vector $\langle f_i^x | i < \lambda \rangle$. Thus this map naturally generalizes to maps $r_\lambda : W \to \mathcal{F}_\lambda(\wp(\mathsf{At}))$ for every ordinal λ . To simplify notation, assume for the rest of the paper that $S = \wp(\mathsf{At})$ and that *S* is finite. The map r_κ gives a precise way to connect the class of all Kripke structures to a single structure $\mathcal{M}^\kappa = \langle \mathcal{F}_\kappa(S), \{\sim_i\}_{i \in I}, V \rangle$ for any κ . The following observation is immediate from the relevant definitions.

Observation 1. Let $\mathcal{M} = \langle W, \{R_i\}_{i \in I}, V \rangle$ be a Kripke structure and \mathcal{M}^{κ} be the structure $\langle \mathcal{F}_{\kappa}(S), \{\sim_i\}_{i \in I}, V \rangle$.

- 1. The relation $wZ\vec{f}$ iff $r_{\kappa}(w) = \vec{f}$ is a functional simulation from \mathcal{M} into \mathcal{M}^{κ} , but, in general, is not a bisimulation.
- 2. There is an ordinal λ , depending on \mathcal{M} such that Z is a bisimulation if $\kappa \geq \lambda^{13}$
- In particular, if M is finite, then there is a bisimulation between M and r(M) = ⟨r[W], {~_i}, V⟩. Moreover, r(M) is the minimal bisimulation contraction of M, i.e. the Kripke model of minimal cardinality that allows for a total bisimulation to M.
- 4. Let $\varphi \in \mathcal{L}_{EL}$ with quantifier depth k and assume that $\kappa > k$. Then $\mathcal{M}, w \models \varphi$ iff $\mathcal{M}^{\kappa}, r_{\kappa}(w) \models \varphi$.

Proof. i) The functionality of Z is obvious, since r_{κ} is a function. Atomic harmony holds by definition of f_0^w . To see that zig holds let $v_0, v_1 \in M$ with $v_0R_iv_1$ and $w \in \mathcal{M}^{\kappa}$

¹³In fact, for $\mathcal{M} = \mathcal{F}_{\kappa}(S)$ we have $\lambda = \kappa$. Moreover the model \mathcal{M}^{κ} is a terminal object in the category of Kripke models over At with total simulations as morphisms. Though $\mathcal{F}_{\kappa}(S)$ and $\mathcal{F}_{\lambda}(S)$ are not bisimilar for $\kappa \neq \lambda$.

with $v_0 Z w$. Since Z is functional we have $w = \vec{f}_{\mathcal{M},v_0}$ An induction shows that $f_i^{v_0} = f_i^{v_1}$ for every $i \le \kappa$, thus $\vec{f}_{\mathcal{M},v_0}(i) = \vec{f}_{\mathcal{M},v_1}(i)$. Thus by definition of \sim_i we have $\vec{f}_{\mathcal{M},v_0} \sim_i \vec{f}_{\mathcal{M},v_1}$. By definition of Z we also have $v_1 Z \vec{f}_{\mathcal{M},v_1}$, thus zig holds. Example 3.10 of (Fagin et al. 1999) shows that that Z is in general not a bisimulation.

ii) Choose λ' such that for all $v, w \in \mathcal{M}$ holds: If there is some μ such that $r_{\mu}(v) \neq r_{\mu}(w)$, then $r_{\lambda'}(v) \neq r_{\lambda'}(w)$ and let $\lambda := \lambda' + \omega$. We have to show that zag holds: Let vZw with Z defined as above and let $w \sim_i w'$. We have to show that there is some $v' \in \mathcal{M}$ with $r_{\lambda}(v') = w'$. Indeed, since $w \sim_i w'$ we have for all $\mu < \lambda$ that $w' \upharpoonright \mu \in w_{\mu}(i)$. By the construction of r_{λ} this implies that for every $\mu < \lambda$ there is some $v' \in \mathcal{M}$ such that $w' \upharpoonright \mu = r_{\mu}(v')$. By the choice of λ' and the extendability conditon, we have that $\exists \mu \in [\lambda'; \lambda] : r_{\mu}(v') \in w_{\mu}(i)$ implies $\forall \mu \in [\lambda'; \lambda] : r_{\mu}(v') \in w_{\mu}(i)$. In particular we have by the limit condition that $r_{\lambda}(v') = w'$ as desired. See chapter 3 of (Fagin et al. 1999) for more details.

iii) Obvious from *ii*) and the definition of r_{ω} .

iv) straightforward induction over the quantifier depth of φ

3 Information dynamics on knowledge structures

Our aim is to examine natural transitions between types in a knowledge structure. These transitions are intended to represent some type of reasoning process or information update. For this initial study, we focus on the operation of product update (restricted to equivalence relations as in Definition 2.4).

3.1 Technical preliminaries: generalized product update

Our first contribution is to define a sequence of products \times_{N_n} between *Kripke struc*tures. The idea to apply product update between Kripke structures (rather than Kripke structures and event models) was initially proposed by Jan van Eijck and colleagues (van Eijck et al. 2010). We follow the same basic idea, although our approach differs in a technical, but crucial, way.

In order to generalize the product update operation so that it applies between two Kripke structures, we must replace the precondition function with something appropriate for merging two Kripke structures. Our approach is to explicitly mark which of the formulas we are interested in, and treat these formulas as atomic propositions.¹⁴ Fix a set I of players and At of atomic propositions (for simplicity assume both are finite).

Definition 3.1. Language extension

- Let ⊤ ⊆ *L*_{EL} with At ⊆ ⊤. For every φ ∈ ⊤ we introduce a new constant φ called the **name** of φ. Let ⊤ := {φ|φ ∈ ⊤}. The **language extension** with ⊤, denoted by *L*^T_{EL}, is the epistemic language with the same set of agents as *L* and ⊤ as atomic propositions. By a slight abuse of notation we write p instead of p for p ∈ At ⊆ ⊤. We denote the valuation function over the language *L*^T_{EL} by V_⊤. As usual, we omit the subscript when it is clear from the context.
- 2. Let $\mathcal{M} = \langle W, \{R_i\}_{i \in I}, V \rangle$ be a Kripke model with atomic propositions At and let $\top \subseteq \mathcal{L}_{EL}$ with At $\subseteq \top$. Then \mathcal{M} can naturally be interpreted as a Kripke model over \mathcal{L}_{EL}^{\top} by defining V_{\top} as: $w \in V_{\top}(\check{\varphi})$ iff $\mathcal{M}, w \models \varphi$. We denote \mathcal{M} viewed over \mathcal{L}_{EL}^{\top} by \mathcal{M}^{\top} .

In \oplus -updates every state *v* in the event model comes with a (generally complex) formula φ that is the precondition for *v* to occur. That is (w, v) is only defined if $\mathcal{M}, w \models pre(v)$. This is exactly the idea of the \times_{\top} update defined below: pairs of states are in the new model only if they agree on the formulas in \top .

Definition 3.2. i) Let $\top \subseteq \mathcal{L}_{EL}$ with At $\subseteq \top$. Let $\mathcal{M} = \langle W, \{R_i\}_{i \in I}, V \rangle$ and $\mathcal{M}' = \langle W', \{R'_i\}_{i \in I}, V' \rangle$ be two Kripke models over \mathcal{L}_{EL}^{\top} . The *product model* $\mathcal{M} \times \mathcal{M}' = \langle W'', \{R''_i\}_{i \in I}, V'' \rangle$ over \mathcal{L}_{EL}^{\top} is defined as follows:

- $W'' = \{(w, w') \mid w \in W, w' \in W' \text{ and for all } \check{\varphi} \in \check{T} : w \in V^{\mathcal{M}}_{\check{\tau}}(\check{\varphi}) \text{ iff } w' \in V^{\mathcal{M}'}_{\check{\tau}}(\check{\varphi});$
- $(w, w')R''_i(v, v')$ iff wR_iv and $w'R'_iv'$; and
- $(w, w') \in V''_{\check{\tau}}(\check{\varphi})$ iff $w \in V^{\mathcal{M}}_{\check{\tau}}(\check{\varphi})$ (and thus also $w' \in V'_{\check{\tau}}(\check{\varphi})$).

ii) The **generalized product update** of \mathcal{M} and \mathcal{M}' over \top , denoted by $\mathcal{M} \times_{\top} \mathcal{M}'$ is the model $\mathcal{M} \times \mathcal{M}'$ as defined above interpreted as a model over \mathcal{L}_{EL} . (That is: removing all atoms $\check{\varphi}$ with $\varphi \in \top \setminus At$ and identifying \check{p} with p for all $p \in At$.)

We write $\mathcal{M} \times_{\top} \mathcal{M}'$ where \mathcal{M} and \mathcal{M}' are Kripke models over \mathcal{L}_{EL} , meaning that we interpret \mathcal{M} and \mathcal{M}' as being models over $\check{\top}$ and do the \times_T -update as defined above. The procedure that we follow to compute this product runs as follows:

¹⁴In general, this type of language extension can be used to model agents with limited memory. For instance, this is needed for an analysis of situations such as the sum and product riddle involving the dialogues: *A: I don't know* φ . *B: I* knew you didn't know before you said that (cf. Sietsma and van Eijck, for an analysis of this puzzle in Public Announcement Logic).

- 1. Pick a set \top of statements to keep track of,
- 2. Build the Product in $\mathcal{L}_{EL}^{\mathsf{T}}$, and
- Remove the additional information, i.e., restrict the valuation function from ⁺ to At.

The following example demonstrates this procedure.

Example 1. Let $\top = \{p, K_1p, K_2p, K_1\neg p, K_2\neg p\}$. Then the product of the two models is calculated as follows.



Note that the reflexive and transitive arrows are not drawn in the above picture for simplicity. The set \mathcal{T} is rich enough to uniquely describe all knowledge assignments of level at most one. Thus, the product reflects a merging of models taking into account the agents' first-order information. The fragments of \top true at the individual worlds are:

$$\begin{array}{ll} \mathcal{M}, w_1 \models \{p, K_1 p\} & \mathcal{M}, w_2 \models \{K_1 \neg p\} & \mathcal{M}', v_1 \models \{p, K_1 p\} \\ \mathcal{M}', v_2 \models \emptyset & \mathcal{M}', v_3 \models p & \mathcal{M}', v_4 \models \{K_1 \neg p\} \end{array}$$

The only pairs satisfying the same fragment of \top are (w_1, v_1) and (w_2, v_4) . Observe that in the model $\mathcal{M} \times_{\top} \mathcal{M}'$ we have:

$$\mathcal{M} \times_{\top} \mathcal{M}', (w_1; v_1) \models \{p, K_1 p, K_2 p\}$$

which is different from $\{p, K_1p\}$, the fragment of \top satisfied by \mathcal{M}, w_1 .

In general, taking a generalized product update consists of two steps: The first is picking a set of statements $\top \supseteq$ At that one wants to keep track of and extending the language to \mathcal{L}_{EL}^{\top} . The second is to do generalized product update \times_{\top} , that is the normal

product × over \mathcal{L}_{EL}^{\top} followed by omitting all the information about the valuation of $\check{\top} \setminus \mathsf{At}$, i.e., making the newly created model an \mathcal{L}_{EL} model again. The above example shows that the \times_{T} product does not preserve higher order information.

Remark 1. There are epistemic models \mathcal{K}, w and \mathcal{L}, v over \mathcal{L}_{EL} a finite fragment $\top \subset \mathcal{L}_{EL}$ and some $\varphi \in \top \setminus$ At such that $(v, w) \in \mathcal{K} \times \mathcal{L}$ (the product over \mathcal{L}_{EL}^{\top}) and $\mathcal{K} \times \mathcal{L}, (v, w) \models \check{\varphi}$, but $\mathcal{K} \times \mathcal{L}, (v, w) \nvDash \varphi$. (Where in the last formula φ is evaluated as a formula of \mathcal{L}_{EL} .)

There is a close connection between generalized product update and the \oplus -update. In both cases, the result is not the complete cartesian product between the two state spaces, but a subset that is characterized by a certain set of formulae. The precise connection between the two concepts is clarified by the following lemma.

Lemma 1. For every event model \mathcal{E} there is some fragment $\top \subseteq \mathcal{L}_{EL}$ and a Kripke model \mathcal{M}' (for the language \mathcal{L}_{EL}^{\top}) such that $- \oplus \mathcal{E}$ is the same as $- \times_{\top} \mathcal{M}'$ (i.e., for all Kripke models $\mathcal{M}, \mathcal{M} \oplus \mathcal{E}$ is isomorphic to $\mathcal{M} \times_{\top} \mathcal{M}'$).

Proof. Let $\mathcal{E} = \langle E, \{Q_i\}_{i \in I}, pre \rangle$ be an event model. Let \top be the set $\{pre(e) \mid e \in E\} \cup At$. Construct the model $\mathcal{M}' = \langle W', \{R'_i\}, V' \rangle$ as follows: Let W' be the set of pairs (e, L_e) where $e \in E$ and $L_e \subseteq \top$ is a maximally consistent subset of \top containing pre(e). The relations R'_i are defined as $(e, L_e)R_i(e', L'_e)$ iff eQ_ie' , and the valuation V' is defined by L_e (i.e., $(e, L_e) \in V(\check{\varphi})$ provided $\check{\varphi} \in L_e$). It is easy to check that this \mathcal{M}' has the desired properties.

Corollary 1. If there is an upper bound for the quantifier depths of the preconditions in the event model \mathcal{E} (i.e., the set $\{qd(pre(e)) \mid e \in \mathcal{E}\}\$ has an upper bound) then the set \top in the above lemma can be chosen finite. This holds in particular if \mathcal{E} is finite.

Proof. Let n - 1 be an upper bound for the quantifier depths of {pre(e) | $e \in E$ }. Recall that $\mathcal{F}_n(\wp(At))$ is finite, and so, there are characteristic formulae φ_t of maximal quantifier depth n - 1 for every $t \in \mathcal{F}_n(\wp(At))$ (that is, $\mathcal{F}_n(\wp(At))$, $s \models \varphi_t \Leftrightarrow s = t$). Let $\top := \{\varphi_e \mid e \in \mathcal{F}_n(\wp(At))\} \cup At$ and construct the model \mathcal{M}' as follows:

$$W' := \{(e, t) | e \in E, t \in \mathcal{F}_n(\wp(\mathsf{At})) \text{ and } \mathcal{F}_n(\wp(\mathsf{At})), t \models \mathsf{pre}(e)\},\$$

let $(e, t)R'_i(e', s)$ if eQ_ie' , and define V' as:

$$(e, t) \in V'(\check{\varphi})$$
 iff $\mathcal{F}_n(\wp(\mathsf{At}), t \models \varphi)$.

The sets $S = \{\varphi_t \mid t \in \mathcal{F}_n(\varphi(At))\}$ chosen above are special in that these sets reflect all possible knowledge assignments up to depth n - 1. We denote the resulting set of formulas by N_n (i.e., $N_n = \{\varphi_t \mid t \in \mathcal{F}_n(\varphi(At))\} \cup At$).

- *Remark* 2. i) In the above proof, we can turn \mathcal{M}' into an event model \mathcal{E}' by letting $pre(e, t) = \varphi_t$. In this case we have $\mathcal{M} \times_{N_n} \mathcal{M}' = \mathcal{M} \oplus \mathcal{E}'$ for all \mathcal{M} . In particular the model \mathcal{E}' is a special event model that only has preconditions from N_n . This follows a general pattern: The initial strength of arbitrary event models is that they allow for a very intuitive description of events in a multi-agent setting. However, from a technical point of view arbitrary event models can be difficult to handle. Therefore it sometimes proves useful to translate arbitrary event models els into a certain subclass of event models which are easier to work with. For instance, van Ditmarsch et al. (2008) defined a class of canonical event models that are useful for studying when two event models are *equivalent*.
 - ii) The translation of an event model into a Kripke model blurs the distinction between static descriptions of situations and descriptions of events.

There is an interesting peculiarity of the \times_{T} -products. Obviously, \times_{T} is commutative, but the following example shows that it is not associative.¹⁵

Example 2. As in example 1 let $\top = \{p, K_1p, K_2p, K_1\neg p, K_2\neg p\}$. Consider the following \mathcal{L}_{EL} -models which we interpret as \mathcal{L}_{EL}^{\top} -models.



We now show that $(\mathcal{M}_1 \times_{\top} \mathcal{M}_2) \times_{\top} \mathcal{M}_3 \neq \mathcal{M}_1 \times_{\top} (\mathcal{M}_2 \times_{\top} \mathcal{M}_3)$. As we already noted in the previous example (Example 1), $\mathcal{M}_1 \times_{\top} \mathcal{M}_2 = \mathcal{M}_3$. In particular, $(\mathcal{M}_1 \times_{\top} \mathcal{M}_3)$

¹⁵In general, it is clear that the process of consecutive learning is not commutative. One's actions in some event *B* can depend on having learned *A* before. In our formalization, the non-associativity captures this intuition: $(A \times_S B) \times_S C$ is to be read as being in situation *A* and learning *B*, then *C*, whereas $A \times_S (B \times_S C) = A \times_S (C \times_S B)$ corresponds to learning *B* and *C* at a time. A similar phenomena has been noticed in the belief merging literature (Maynard-Reid II and Shoham 1998, Section 5.1).

 \mathcal{M}_2 $\times_{\top} \mathcal{M}_3 = \mathcal{M}_3 \times_{\top} \mathcal{M}_3 = \mathcal{M}_3$ where the last equivalence holds since u_1 and u_2 satisfy different formulas from \top .

On the other hand, note that the following formulas from \top are true at states in \mathcal{M}_3 :

$$\mathcal{M}_3, u_1 \models \{p, K_1 p, K_2 p\} \qquad \mathcal{M}_3, u_2 \models \{K_1 \neg p, K_2 \neg p\}$$

However, there are no states in \mathcal{M}_2 precisely these substes of \mathcal{T} , so $\mathcal{M}_2 \times_{\top} \mathcal{M}_3 = \emptyset$ and consequently $\mathcal{M}_1 \times_{\top} (\mathcal{M}_2 \times_{\top} \mathcal{M}_3) = \emptyset$. Thus, we have $(\mathcal{M}_1 \times_{\top} \mathcal{M}_2) \times_{\top} \mathcal{M}_3 \neq \mathcal{M}_1 \times_{\top} (\mathcal{M}_2 \times_{\top} \mathcal{M}_3)^{.16}$

The interpretation of this statement is that first learning \mathcal{E} and then learning \mathcal{E}' is different to learning \mathcal{E} and \mathcal{E}' at the same time. To be more precise, we have $(\mathcal{E} \times_{\top} \mathcal{F}) \times_{\top} \mathcal{G} \neq \mathcal{E} \times_{\top} (\mathcal{F} \times_{\top} \mathcal{G}) \neq \mathcal{E} \times_{\top} \mathcal{F} \times_{\top} \mathcal{G}^{17}$ This non-associativity shows that our framework is rich enough to distinguish between consecutive learning and receiving all information at once.

These observations should be contrasted with the theory developed in (van Eijck et al. 2010). There, the authors are concerned with updates where all preconditions are boolean combinations of the ground variables (describing non-epistemic facts about the state of the world). Learning facts about the world is associative (cf van Eijck et al. 2010, Theorem 1), whereas learning facts about the players' previous knowledge is not!

Van Eijck *et al.* study the monoid generated by \times_{At} products. Our primary goal in this paper is to understand how the \oplus -update works in type spaces. To that end, we first generalize a result from (van Ditmarsch and French 2009).

Theorem 1. Let \mathcal{M}_1 be a Kripke structure such that for any $v, w \in \mathcal{M}$ there is an epistemic formula φ distinguishing v and w (i.e. $\mathcal{M}, w \models \varphi$ and $\mathcal{M}, v \models \neg \varphi$). Let \mathcal{M}_2 be an arbitrary Kripke model. Then there is a set of formulas \top and \mathcal{L}_{EL}^{\top} -Kripke structure \mathcal{M}' such that $\mathcal{M}_1 \times_{\top} \mathcal{M}' \rightleftharpoons \mathcal{M}_2$ if and only if there is a total simulation from \mathcal{M}_2 to \mathcal{M}_1 . Furthermore, if the model \mathcal{M}_1 is finite the set \top can be chosen finite.

Proof. The direction from left to right is easy: Let \mathcal{M}' and \top be such that $\mathcal{M}_1 \times_{\top} \mathcal{M}' = \mathcal{M}_2$. It is easy to see that the map $\mathcal{M}_1 \times_{\top} \mathcal{M}' \to \mathcal{M}_1$ sending every pair (w, w') to w is a functional, hence total, simulation.

For the direction from right to left: Let Z be a total simulation from \mathcal{M}_2 to \mathcal{M}_1 . First we define a Kripke model $\mathcal{M}^\circ = \langle W^{\mathcal{M}^\circ}, \{R_i^{\mathcal{M}^\circ}\}_{i \in I}, V^{\mathcal{M}^\circ} \rangle$:

¹⁶There are examples where both $(\mathcal{M}_1 \times_{\mathsf{T}} \mathcal{M}_2) \times_{\mathsf{T}} \mathcal{M}_3$ and $\mathcal{M}_1 \times_{\mathsf{T}} (\mathcal{M}_2 \times_{\mathsf{T}} \mathcal{M}_3)$ are non-empty; however, they are more complicated while making the same point.

¹⁷Here $\mathcal{E} \times_T \mathcal{F} \times_T \mathcal{G}$ is the obvious generalization of \times_T where all tuples (e, f) in the definition are replaced by triples (e, f, g).

- $W^{\mathcal{M}^{\circ}} = \{(t_1, t_2) | t_i \in \mathcal{M}_i, i = 1, 2 \text{ and } t_1 Z t_2\}$
- $(t_1, t_2)R_i^{\mathcal{M}^\circ}(s_1, s_2)$ iff $t_1R_i^{\mathcal{M}_1}s_1$ and $t_2R_i^{\mathcal{M}_2}s_2$
- $(t_1, t_2) \in V^{\mathcal{M}^\circ}(p)$ iff $t_2 \in V^{\mathcal{M}_2}(p)$ (and thus also $t_1 \in V^{\mathcal{M}_1}(p)$)

First we show that the model \mathcal{M}° is bisimilar to \mathcal{M}_2 . We show that the projection map π_2 mapping every $(t_1, t_2) \in \mathcal{M}^{\circ}$ to $t_2 \in \mathcal{M}_2$ is a bitotal bisimulation (recall Definition 2.5). The atom condition is clear. For forth assume that $(t_1, t_2)\pi_2 t_2$ and that $(t_1, t_2)R_i^{\mathcal{M}^{\circ}}(s_1, s_2)$. By the definition of $R_i^{\mathcal{M}^{\circ}}$ we have $t_2R_i^{\mathcal{M}_2}s_2$ and by definition of π_2 we have $(s_1, s_2)\pi_2 s_2$, thus forth is fulfilled.

Similarly, for back assume that $(t_1, t_2)\pi_2 t_2$ and that $t_2 R_i^{\mathcal{M}_2} s_2$. Since Z is a total simulation and $t_1 Z t_2$ holds by the construction of \mathcal{M}° , there is some $s_1 \in \mathcal{M}_1$ with $s_1 Z s_2$ and $t_1 R_i^{\mathcal{M}_1} s_1$. But this means that $(s_1, s_2) \in \mathcal{M}^\circ$ and that $(t_1, t_2) R_i^{\mathcal{M}^\circ}(s_1, s_2)$, thus proving the back condition.

Since $\mathcal{M}_2 \simeq \mathcal{M}^\circ$, it suffices to show that there is some \mathcal{M}' with $\mathcal{M}_1 \times_{\mathsf{T}} \mathcal{M}' = \mathcal{M}^\circ$.

Note, that the projection $\pi_1 : \mathcal{M}^\circ \to \mathcal{M}_1$ sending each pair (t_1, t_2) to t_1 is a functional left simulation. The atom condition is clear, and the rest can be shown with arguments similar to ones given above.

Now, pick a set $\top^* \subseteq \mathcal{L}_{EL}$ that contains a distinguishing formula for any two worlds $v, w \in \mathcal{M}_1$ and let $\mathcal{T} := \mathcal{T}^* \cup At$. Turn \mathcal{M}° into an \mathcal{L}_{EL}^\top -model \mathcal{M}' by defining: $(t_1, t_2) \in V^\top(\check{\varphi})$ iff $\mathcal{M}_1, t_1 \models \varphi$. Since \top^* contains enough distinguishing formulae, $s_1 \in \mathcal{M}_1$ and $(t_1, t_2) \in \mathcal{M}^\circ$ satisfy the same $\check{\mathcal{T}}$ -formulae iff $s_1 = t_1$. Therefore $\mathcal{M}_1 \times_{\top} \mathcal{M}' = \mathcal{M}^\circ$ as desired. Furthermore, if \mathcal{M}_1 is finite, then the set \mathcal{T}^* can be chosen finite, thus proving the last statement.

Remark 3. Van Ditmarsch and French (2009) contains a proof for a similar statement about \oplus -updates in the *finite* case. However, the generalization to infinite Kripke models does not hold for the \oplus -update.

Remark 4. Note that the model \mathcal{M}' constructed in the right-to-left direction of the prove of Lemma 1 is in general *not* a \mathcal{L}_{EL} model that is simply interpreted as an \mathcal{L}_{EL}^{\top} model. That is: There is in general some $\varphi \in \top$ and some $w \in \mathcal{M}'$ such that $\mathcal{M}', w \models \check{\varphi}$ but $\mathcal{M}', w \nvDash \varphi$ (where $\check{\varphi}$ is an atom of \mathcal{L}_{EL}^{\top} and φ is a formula evaluated in \mathcal{M}' interpreted as a Kripke model over At (i.e. only containing atoms from $\{\check{p} \mid p \in At\}$). That is: to gain the expressive power of updating with an arbitrary event model, one needs the class of all \mathcal{L}_{EL}^{\top} -models. Interestingly enough, this is no longer true when we restrict ourselves to the class of *finite* Kripke structures. There, the full expressive power of the class of all \oplus -updates is already given by the class of all finite Kripke models over \mathcal{L}_{EL} together with the set of all \times_{N_n} products for $n \in \omega$. More formally, we have the following fact (whose straightforward, but tedious, proof we leave out). **Fact 3.1.** Let $\mathcal{K} = \langle W, (R_i)_i, V \rangle$ be a finite Kripke model without bisimilar points and let $\mathcal{L} = \langle W', R'_i, V' \rangle$ be a finite Kripke model such that \mathcal{L} is obtainable from \mathcal{K} by a product update. Then, there is some $\top \supseteq$ At and some Kripke model \mathcal{M} over \mathcal{L}_{EL} such that $\mathcal{K} \times_{\top} \mathcal{M} = \mathcal{L}$.

3.2 Characterization result

As discussed in the previous section, every \oplus -update can be written as a \times_{\top} -update over a language in which the formulas in \top are treated as atomic propositions. This will help us represent the product update in knowledge structures.

First, we need an equivalent to the extension of atomic propositions on types: For $n \in \mathbb{N}$ let S_n denote the set of all possible *n*-worlds, thus $S_n = \mathcal{F}_n(S)$ and $S_0 = S$). Technically, this is redundant, though it helps conceptually to distinguish $\mathcal{F}_n(S)$ as a type space generated by *S* and S_n which is the same type space reinterpreted as new set of atoms. By switching between those interpretations, every n + k world over *S* can be seen as a *k*-world over S_n and thus there is a canonical embedding $\mathcal{F}_{\omega}(S) \hookrightarrow \mathcal{F}_{\omega}(S_n)$.¹⁸

For any two Kripke models \mathcal{K} , v and \mathcal{L} , w we have defined the product update ($\mathcal{K} \times \mathcal{L}$, (v, w)) over the ground language \mathcal{L}_{EL} above. Furthermore, we have seen that there is some κ such that r_{κ} is a bisimulation of \mathcal{K} onto its image. Since $r_{\kappa}(v)$ is obviously in the image of $r_{\kappa}[\mathcal{K}]$ this implies that parts of \mathcal{K} are somehow coded in $r_{\kappa}(v)$. The idea of the following definition is that we can unravel enough information about \mathcal{K} and \mathcal{L} from $r_{\kappa}(v)$ and $r_{\kappa}(w)$ to determine $r_{\kappa}(v, w)$). We define a product \times_0 below and we will show later (lemma 2) that $r_{\kappa}((v, w)) = r_{\kappa}(v) \times_0 r_{\kappa}(w)$. As with the original definition of a κ -world (see 2.6), the definition is by inductive.

Definition 3.3. Suppose that $n \in \mathbb{N}$ and $\vec{f}, \vec{g} \in \mathcal{F}_{\omega}(S)$. Then the \times_0 -product $(\vec{f} \times_0 \vec{g}) \in \mathcal{F}_{\omega}(S) \cup \{\emptyset\}$ is defined as follows:

- $(\vec{f} \times_0 \vec{g})_0 = \langle f_0 \rangle$ iff $f_0 = g_0$ and \emptyset otherwise.
- $(\vec{f} \times_0 \vec{g})_m(i) = \{(\vec{f'} \times_0 \vec{g'})_{m-1} \mid \vec{f'} \in f_m(i), \vec{g'} \in g_m(i)\}$

This definition can be lifted to an analogue of the generalized product update: The operator \times_n will correspond to a product update with $\top = N_n$. First observe that the above definition of \times_0 works equally well if all *S* are replaced by S_n . As in the case of the generalized product update, the \times_n update implicitly consists of two steps: First a product update between two elements of $\mathcal{F}_{\kappa}(S_n)$ followed by a removal of information,

¹⁸Note that this map is not surjective for $n \ge 1$: For instance the introspection conditions of $\mathcal{F}_{k+1}(S)$ gives some limitations on which elements of $\mathcal{F}_2(S_k)$ can come from $\mathcal{F}_{k+1}(S)$.

i.e. a projection from S_n to S. As with general product updates, the definition contracts these two steps into one:

Definition 3.4. Let $\bar{\pi}$: $S_n \to S$ be the projection map sending the tuple $\langle f_0 \dots f_{n-1} \rangle$ to f_0 . Define $\times_n : \mathcal{F}_{\omega}(S_n) \times \mathcal{F}_{\omega}(S_n) \to \mathcal{F}_{\omega}(S)$ as follows:

• $(\vec{f} \times_n \vec{g})_0 = \langle s_0 \rangle$ iff $\bar{\pi}(f_0) = \bar{\pi}(g_0) = s_0$, and \emptyset otherwise.

•
$$(\vec{f} \times_n \vec{g})_m(i) = \{ (\vec{f'} \times_n \vec{g'})_{m-1} \mid \vec{f'} \in f_m(i), \vec{g'} \in g_m(i) \}.$$

The following lemma describes the relationship between the \times_{N_n} -product and the \times_n -product. Basically, the \times_{N_n} product of two Kripke models (\mathcal{K} , w) and (\mathcal{L} , v) carries the same information as the \times_n -product on the types r(v) and r(w).

For technical convenience we need a definition before we state the lemma: Recall that $N_n \setminus At$ was chosen to be a special set of characteristic formulae for $\mathcal{F}_n(S)$. Since the map r_n preserves the validity of all formulae in N_n (see observation 1 *iv*)), every state *w* in a Kripke structure \mathcal{K} over \mathcal{L}_{EL} satisfies exactly one formula of $N_n \setminus At$. By the the definition of \times_{\top} we thus get the following: In particular we get for any Kripke model \mathcal{K} over \mathcal{L}_{EL} and \mathcal{L} over $\mathcal{L}_{EL}^{N_n}$ that $(v, w) \in \mathcal{K} \times_{N_n} \mathcal{L}$ implies that there is exactly one $\check{\varphi} \in N_n \setminus At$ with $w \in V(\check{\varphi})$. Thus, we can restrict our study to Kripke models \mathcal{L} satisfying this property. We call Kripke models over $\mathcal{L}_{EL}^{N_n}$ satisfying this property admissible.

Furthermore, since every $e \in \mathcal{F}_n(S)$ satisfies exactly one formula from $N_n \setminus At$ we have that every world in an admissible Kripke model over \mathcal{L}_{EL}^{\top} corresponds to exactly one $e \in \mathcal{F}_n(S)$. We can define a map r' from admissible Kripke models to $\mathcal{F}_{\omega}(S_n)$ in the same way as we defined r.

Lemma 2. Let $n \in \omega$ and let \mathcal{K}, \mathcal{L} , be Kripke models over $\mathcal{L}_{EL}^{N_n}$. Let $v \in \mathcal{K}$, $w \in \mathcal{L}$ satisfaying the same N_n -formulae. Let $(v, w) \in \mathcal{K} \times_{N_n} \mathcal{L}$ denote the product of v and w in $\mathcal{K} \times_{N_n} \mathcal{L}$. Then we have $r((v, w)) = r'(v) \times_n r'(w)$, i.e., the following diagram commutes:



Proof. Let $n \in \mathbb{N}$ and $v \in \mathcal{K}$, $w \in \mathcal{L}$ satisfying the same N_n -formulae. We inductively show that $(r'(v) \times_n r'(w))_k = r(v, w)_k$. for k = 0 this is trival: If v and w satisfy the same atomic propositions over \check{N}_n we have $(r'(v) \times_n r'(w))_0 = r((v, w))_0 = \{p \in \mathsf{At} :$

 $v \in V^{\mathcal{K}}(p)$ }. If they satisfy different atomic propositions we have $(v, w) \notin \mathcal{K} \times_{N_n} \mathcal{L}$ and $r'(v) \times_n r'(w) = \emptyset$. Now assume the statement holds for k - 1 and let $i \in I$ (the set of agents).

First, we show $r(v, w)_k(i) \subseteq (r'(v) \times_n r'(w))_k(i)$. Let $x \in r((v, w))_k(i)$, thus x is a k - 1-world. By construction of the map r there is some \tilde{x} in $\mathcal{K} \times_{N_n} \mathcal{L}$ such that $\tilde{x}R_i(v, w)$ and $r(\tilde{x})_{k-1} = x$. Thus there are $x_1 \in \mathcal{K}$ and $x_2 \in \mathcal{L}$ such that the product of x_1 and x_2 in $\mathcal{K} \times_{N_n} \mathcal{L}$ is \tilde{x} - in particular x_1R_iv and x_2R_iw and x_1 and x_2 satisfy the same N_n -formulae. In particular, $r'(x_1) \times_n r'(x_2) \neq \emptyset$ and by induction we have that $(r(x_1, x_2))_{k-1} = (r'(x_1) \times_n r'(x_2))_{k-1}$. On the other hand, we have $r'(x_1)_{k-1} \in$ $r'(v)_k(i)$ and similarly for x_2 and w by the construction of r'. In particular, we have $x = (r'(x_1) \times_n r'(x_2))_{k-1} \in (r'(v) \times_n r'(w))_k(i)$ as desired, thus proving the first direction.

The argument for the reverse inclusion $r(v, w)_k(i) \supseteq (r'(v) \times_n r'(w))_k(i)$ is similar: Let $x \in (r('v) \times_n r'(w))_k(i)$. Then there are $\tilde{x}_1 \in r'(v)$ and $\tilde{x}_2 \in r'(w)$ such that $(r'(\tilde{x}_1) \times_n r'(\tilde{x}_2))_{k-1} = x$ and such that there are $x_1 \in \mathcal{K}, x_2 \in \mathcal{L}$ such that $r'(x_i) = \tilde{x}_i$ and x_1R_iv and x_2R_iw hold. Since $\tilde{x}_1 \times_n \tilde{x}_2$ exists, x_1 and x_2 satisfy the same N_n -formulae. In particular there is some (x_1, x_2) in $\mathcal{K} \times_{N_n} \mathcal{L}$ with $(x_1, x_2)R_i(v, w)$. By construction of r we have $r((x_1, x_2))_{k-1} \in r((v, w))_k$ and by induction we have $r((x_1, x_2))_{k-1} = x$, thus proving the reverse direction.

Note that the calculation of $\vec{f} \times_n \vec{g}$ from types \vec{f} and \vec{g} is computationally efficient: In order to calculate the *k*-th level of $\vec{f} \times_n \vec{g}$ only the first n + k levels of \vec{f} and \vec{g} are required.

The above definition of \times_n updates gives a way of modeling dynamics on a type space—thus, opening up the field of epistemic game theory to belief dynamics. Event models were designed as a very intuitive and natural tool for representing epistemic events in a multiagent setting. The translation of event models into the corresponding pair of Kripke models and a product relation \times_{N_n} , and further into a type and a relation \times_n allows us to calculate the change of epistemic status brought about by an event model \mathcal{E} .

On the other hand, every product update with a finite event model can be written as a \times_n -update, thus it suffices to understand the structure of \times_n to study product updates. Thus, $\mathcal{F}_{\omega}(\wp(S))$ is not only a universal Kripke model in the static sense, together with the products \times_n is also universal in that it incorporates all potential updates.

On Kripke structures, translating event models into types allows us to study updating events as separate entities without any reference to a ground type. Furthermore, the translation blurs the distinction between types as static descriptions of epistemic states and knowledge changing events.

One natural and important question is: Given two types \vec{f} and \vec{g} , is there a possible piece of incoming information that transforms \vec{f} into \vec{g} ?

The intuition behind the answer given by the following theorem is: In the entire model, the agents are assumed to be omniscient and non-forgetting. Thus, an event cannot add any uncertainty about the state of nature, it can only remove some states from the sets of possible states. In contrast, for the higher order information, essentially anything is possible as long as it is compatible with individuals gaining new information about the state of nature. In particular, an epistemic event may increase the uncertainty about other agents' types. This idea is captured by the following definition.

Definition 3.5. For a type $\vec{f} \in \mathcal{F}_{\alpha}(S)$ we say that a type \vec{g} is **admissible** for \vec{f} iff

- $f_0 = g_0;$
- for all agents $i: g_1(i) \subseteq f_1(i)$; and
- for $\alpha > 1$: If $\vec{h} \in g_{\alpha}(i)$ then there is a $\vec{h}' \in f_{\alpha}(i)$ such that *h* is admissible for \vec{h}' .

Our characterization theorem is similar to Theorem 1.

Theorem 2. Let $\vec{f}, \vec{g} \in \mathcal{F}_{\alpha}(S)$ be types such that \vec{g} is obtainable by an update from \vec{f} , *i.e.*, there is some *n* and some $\vec{h} \in \mathcal{F}_{\alpha}(S_n)$ such that $\vec{f} \times_n \vec{h} = \vec{g}$. Then \vec{g} is admissible for \vec{f} . If the submodel of $\mathcal{F}_{\omega}(S)$ generated by \vec{f} is finite also the converse holds true.

Before we can prove this theorem, we recall the following result from infinite combinatorics.

Theorem 3. (*König's Lemma*) Let T be an infinite, finitely branching tree. Then, T has an infinite branch.

Proof. Construct an infinite branch $\langle x_0, x_1, ... \rangle$ as follows: x_0 is the root. For i > 0: If $x_0, ..., x_i$ are already in the branch, pick a successor x_{i+1} of x_i such that $\{y|y > x\}$ is infinite. (since the tree is finitely branching such a successor always exists). Then $\langle x_0, x_1, ... \rangle$ is an infinite branch.

Prof of Theorem 2. The first statement is straightforward: Let \mathcal{F} and \mathcal{G} be the epistemic submodels of $\mathcal{F}_{\omega}(S)$ induced by \vec{f} and \vec{g} , respectively. Assume that there is some $\vec{h} \in \mathcal{F}_{\omega}(S_n)$ such that $\vec{f} \times_n \vec{h} = \vec{g}$. By Lemma 2, this is equivalent to saying that $\mathcal{F} \times_{N_n} \mathcal{H} = \mathcal{G}$, where $\mathcal{F}, \mathcal{G}, \mathcal{H}$ are the generated Kripke models (over $\mathcal{L}_{EL}^{N_n}$) from \vec{f}, \vec{g} , and \vec{h} . By Theorem 1 there is a total simulation *S* from *G* to *F*. We inductively show that every $\vec{g}' \in G$ is admissible for every $\vec{f}' \in F$ with $\vec{f}'S\vec{g}'$. The 0th-level is clear by definition of a simulation. Now it suffices to show that definition of admissibility is fulfilled at the 1st level: Since we do this for all $\vec{g}' \in G$ the rest follows from the inductive definition of admissibility and the map *r*. To see that admissibility is fulfilled

at the 1*st* level, let $\vec{h} \in G$ with $\vec{g}' \sim_i \vec{h}$. By definition, there is a $\vec{h}' \in \mathcal{F}$ with $\vec{f}' \sim_i \vec{h}'$. Thus, every state of nature that is conceivable for agent *i* in \mathcal{G} via \vec{h} is also conceivable in \mathcal{F} via \vec{h}' - this is exactly the definition of being admissible in the first level.

For the second statement let \vec{g} be admissible for \vec{f} and let the submodel of $\mathcal{F}_{\omega}(\varphi(S))$ generated by \vec{f} be finite. Again, let \mathcal{F} and \mathcal{G} be the Kripke submodels of $\mathcal{F}_{\omega}(S)$ induced by \vec{f} and \vec{g} . Define the Relation Z between \mathcal{F} and \mathcal{G} as $\vec{f'}Z\vec{g'}$ iff $\vec{g'} \in \mathcal{G}$ is admissible for $\vec{f'} \in \mathcal{F}$. We will show that Z is a total simulation from G to \mathcal{F} , thus showing that \mathcal{G} is obtainable by \mathcal{F} via update (again using Theorem 1 and Lemma 2). From the first clause in the definition of admissibility it follows that Z satisfied atomic harmony. Furthermore we have that \vec{g} is admissible for \vec{f} by assumption, thus Z is not empty. To see that Z satisfies the zig condition we have to show that whenever $\vec{g}' \in \mathcal{G}$ is admissible for $\vec{f'} \in \mathcal{F}$ and $\vec{g'} \sim_i \tilde{\vec{g}}$, then there is some $\vec{f'} \sim_i \tilde{\vec{f}}$ such that $\tilde{\vec{g}}$ is admissible for \vec{f} . Once we have that it is easy to see that Z is also total, since for every \vec{g}' in \mathcal{G} there is a chain $\vec{g} \sim_{i_1} \vec{g}_1 \sim_{i_2} \dots \sim_{i_n} \vec{g}'$ connecting \vec{g} with \vec{g}' . Thus to show (zig) let $\vec{g}' \in \mathcal{G}$ be admissible for $\vec{f}' \in \mathcal{F}$ and $\vec{g}' \sim_i \tilde{\vec{g}}$. We construct an ω -tree (T, \prec) as follows: Since $\vec{g}' \sim_i \tilde{\vec{g}}$ we have $\tilde{g}_{<k} \in g'_k(i)$ for every k. Since \vec{g}' is admissible for \vec{f}' we have that for there is some $h \in f'_k(i)$ such that $\tilde{g}_{<k}$ is admissible for h. Let the k^{th} level of the tree T consist of exactly these $h \in f'_k(i)$ that $\tilde{g}_{<k}$ is admissible for and let the three order \prec be the initial-segment preorder. Since \vec{g}' is admissible for \vec{f}' , every level of T is non-empty. On the other hand, since the state of nature is considered finite, every level of T finite, thus T is finitely branching. Thus, by König's lemma T has an infinite path *P*. By construction, $\vec{f} = \bigcup_{\vec{r} \in P} \vec{r}$ is a type, $\vec{f} \sim_i \vec{f'}$ and $\tilde{\vec{g}}$ is admissible for $\tilde{\vec{f}}$. Since we assumed that the submodel \mathcal{F} is the substructure of $\mathcal{F}_{\omega}(S)$ induced by \vec{f} (and thus by $\vec{f'}$) we have $\tilde{\vec{f}} \in \mathcal{F}$, thus the simulation Z relates $\tilde{\vec{g}}$ to $\tilde{\vec{f}}$.

Again, there is an obvious counterpart of Remark 4 allowing us to update with $\mathcal{F}(S)$ worlds rather than $\mathcal{F}(S_n)$ worlds, provided all the induced Kripke structures involved are finite. To be precise, we can show the following: Let $\vec{f}, \vec{g} \in \mathcal{F}_{\omega}(S)$ be such that the epistemic submodels of $\mathcal{F}_{\omega}(S)$ induced by \vec{f} and \vec{g} are finite. Then \vec{g} is admissible for \vec{f} if and only if there is some natural number *n* and some $\vec{h} \in \mathcal{F}_{\omega}(S)$ such that $\vec{f} \times_n \vec{h} = \vec{g}$.

4 Conclusion and future work

Many different formal models have been used to describe the players knowledge and beliefs in game-theoretic situations. The variety of models reflect different mathematical conventions used by the various sub-communities, as well as competing intuitions about how best to describe the players' beliefs and reasoning in a game situation. It is important to understand the precise relationship between the alternative modeling paradigms. In this paper, we focused on the two most prominent models found in the literature on the epistemic foundations of game theory: Kripke- or Aumann- structures and knowledge structures (non-probabilistic variants of Harsanyi type spaces).

There are two main contributions in this paper. The first is to initiate a study of "information dynamics" for knowledge structures in the style of recent work on *dynamic epistemic logic* (cf. van Benthem 2011). Such a theory would further illustrate the subtle relationship between type spaces and Kripke structures (updating the discussion initiated in Fagin et al. 1991; 1999). In particular, it allows us to combine the strengths of both approaches and use event models as a tool to describe epistemic events. The main technical contribution is the definition of a product operation \times_n on the type space $\mathcal{F}_{\omega}(S)$. We provide a procedure that allows us to translate arbitrary event models into types. Furthermore, we show that the \times_n product is powerful enough to simulate all updates by event models. We prove a characterization theorem (Theorem 2) showing when a type can be transformed into another type by updates with an event model.

This is only an initial study. We see our work here opening up many different avenues of future research. In particular, we plan on investigating the following issues in the future.

- What happens if we allow only updating types from a certain subclass of *F_α(S_n)* (for example, finite epistemic models ⟨*F_α(S_n)*, {~_i}_{i∈I}, V⟩)?
- What are the "behavioral" implications of our main characterization theorem (Theorem 2)? For example, if a strategy is rational for a type \vec{f} in a game *G*, does that strategy remain rational for all types that are admissible for \vec{f} ?
- How do we extend the ideas developed in this paper to Harsanyi type spaces, where the beliefs are represented by probability measures? The first step is to generalize the dynamic epistemic logic framework to settings where beliefs are represented by probabilities. Fortunately, this has largely been done (see van Benthem et al. 2009, Aceto et al. 2011, for details). A very interesting direction for future research is to explore how to use the probabilistic event models and product update operation of (van Benthem et al. 2009) to prove a result analogous to our main characterization theorem (Theorem 2) for Harsanyi type spaces.
- The relation "obtainable by an update" together with our extended theorem (see Remark 4) turns the set of finite induced submodels of $\mathcal{F}_{S}(w)$ into an algebra. Can we characterize this algebra?

Acknowledgements This research was funded by the NWO Vidi project 016.094.345. The authors would like to thank the anonymous referees for their extremely thorough comments which greatly improved this paper.

References

L. Aceto, W. van der Hoek, A. Ingólfsdóttir, and J. Sack. Sigma algebras in probabilistic epistemic dynamics. In *Proceedings of the Thirteenth conference on Theoretical Aspects of Rationality and Knowledge, ACM, 2011, pp. 191–199.*, pages 191–199, 2011.

R. Aumann. Interactive epistemology I: Knowledge. *International Journal of Game Theory*, 28:263–300, 1999.

J. van Benthem. *Logical Dynamics of Information and Interaction*. Cambridge University Press, 2011.

J. van Benthem, J. Gerbrandy, and B. Kooi. Dynamic update with probabilities. *Studia Logica: An International Journal for Symbolic Logic*, 93(1):67 – 96, 2009.

J. van Benthem, E. Pacuit, and O. Roy. Toward a theory of play: A logical perspective on games and interaction. *Games*, 2(1):52 – 86, 2011.

G. Bonanno and P. Battigalli. Recent results on belief, knowledge and the epistemic foundations of game theory. *Research in Economics*, 53(2):149–225, 1999.

A. Brandenburger. Epistemic game theory: Complete information. In S. N. Durlauf and L. E. Blume, editors, *The New Palgrave Dictionary of Economics*. Palgrave Macmillan, 2008.

A. Brandenburger. Origins of epsitemic game theory. In V. F. Hendricks and O. Roy, editors, *Epistemic Logic: Five Questions*, pages 59–69. Automatic Press, 2010.

A. Brandenburger and A. Friedenberg. Self-admissible sets. *Journal of Economic Theory*, 145:785 – 811, 2010.

H. van Ditmarsch and T. French. Simulation and information: Quantifying over epistemic events. In J.-J. Meyer and J. Broersen, editors, *Knowledge Representation for Agents and Multi-Agent Systems*, volume 5605 of *Lecture Notes in Computer Science*, pages 51–65. Springer Berlin / Heidelberg, 2009.

H. van Ditmarsch, W. van der Hoek, and B. Kooi. *Dynamic Epistemic Logic*. Synthese Library. Springer, 2007.

H. van Ditmarsch, J. Ruan, and R. Verbrugge. Sum and product in dynamic epistemic logic. *Journal of Logic and Computation*, 18:563–588, 2008.

J. van Eijck, Y. Wang, and F. Sietsma. Composing models. In *Proceedings of LOFT* 2010, 2010.

R. Fagin. A quantitative analysis of modal logic. *Journal of Symbolic Logic*, 59(1): 209 – 252, 1994.

R. Fagin and M. Vardi. An internal semantics for modal logic: Preliminary report. In *Proc. 17th ACM SIGACT Symposium on Theory of Computing*, pages 305 – 315, 1985.

R. Fagin, J. Halpern, and M. Vardi. A model-theoretic analysis of knowledge. *Journal of the ACM*, 91(2):382 – 428, 1991.

R. Fagin, J. Halpern, Y. Moses, and M. Vardi. *Reasoning about Knowledge*. The MIT Press, 1995.

R. Fagin, J. Geanakoplos, J. Halpern, and M. Vardi. The hierarchical approach to modeling knowledge and common knowledge. *International Journal of Game Theory*, 28(3):331 – 365, 1999.

J. Harsanyi. Games with incomplete informations played by 'bayesian' players. *Management Science*, 14:159–182, 320–334, 486–502, 1967.

P. Maynard-Reid II and Y. Shoham. From belief revision to belief fusion. In *Proceedings of LOFT-98*, 1998.

R. Myerson. Harsanyi's games with incomplete information. *Management Science*, 50(12):1818–1824, 2004.

A. Perea. *Epistemic Game Theory: Reasoning and Choice*. Cambridge University Press, 2012.

F. Sietsma and J. van Eijck. Action emulation between canonical models. preprint.

M. Siniscalchi. Epistemic game theory: Beliefs and types. In S. Durlauf and L. Blume, editors, *The New Palgrave Dictionary of Economics*. Palgrave Macmillan, Basingstoke, 2008.

W. van der Hoek and M. Pauly. Modal logic for games and information. In P. Blackburn, J. van Benthem, and F. Wolter, editors, *Handbook of Modal Logic*, volume 3 of *Studies in Logic*, pages 1077 – 1148. Elsevier, 2006.

Dependent Type Semantics: An Introduction

Daisuke Bekki

Ochanomizu University, Graduate School of Humanities and Sciences National Institute of Informatics CREST, Japan Science and Technology Agency bekki@is.ocha.ac.jp

Abstract

This paper introduces *dependent type semantics*, a framework for natural language semantics based on dependent type theory. The main features of dependent type semantics are as follows. 1) It is *dynamic*: it analyzes E-type/donkey anaphora with well-formed representations. 2) It is *proof-theoretic*: deductions between representations are available without recourse to models. 3) It is *compositional*: semantic representations of sentences are derived from lexicalized representations by a fixed number of combinatory rules. Lastly, 4) it explains *accessibility*: the accessibility or inaccessibility of anaphoric antecedents depends on the structural differences between proofs.

1 Introduction

1.1 Between dynamism and compositionality

Since 1980, the enterprise of dynamic semantics has pursued an alternative framework to Montagovian semantics, which compensates for the gap between syntactic structures of natural language sentences involving dynamic binding. The difficulty of this pursuit implies that there is tension between *dynamism* and *compositionality*, which have not yet been unified in a coherent semantic theory that accounts for both aspects.

This tension has been the driving force behind dynamic semantics, and in fact some theories have achieved partial success in unifying the two aspects. At this point, I

should clarify what I mean by dynamism and compositionality. Dynamic semantics explores various empirical data concerning dynamic binding, whose nature is exemplified by the two paradigms of *donkey sentences* in (1) by Geach (1962) and *E-type anaphora* in (2) by Evans (1980).

- (1) a. Every farmer who owns $[a \text{ donkey}]_{i}$ beats it_i.
 - b. If $[a \text{ farmer}]_{i}$ owns $[a \text{ donkey}]_{i}$, he_i beats it_j.
- (2) [A man]_{\rangle} entered. He_{*i*} whistled.

As discussed elsewhere, (1a), for example, is problematic in terms of compositionality. Compositional semantic theory is such that it provides a way to calculate any semantic representation of any target sentence from the semantic representations of its parts. The structural analogue of (1a) (and (1b)), which allows us to give a straightforward compositional analysis, is (3). However, it is not an appropriate semantic representation for (1a) since variable y occurs as a free variable outside of the scope of $\exists y$.

(3) $\forall x(\mathbf{Farmer}(x) \land \exists y(\mathbf{Donkey}(y) \land \mathbf{Own}(x, y)) \rightarrow \mathbf{Beat}(x, y))$

In the same way, the structural analogue of (2) is (4), which is not an appropriate representation for (2) since variable *x* in **Whistle**(*x*) is not bound by $\exists x$.

(4) $\exists x(\mathbf{Man}(x) \land \mathbf{Enter}(x)) \land \mathbf{Whistle}(x)$

Thus we have the following criteria for a successful theory of dynamic semantics.

Criterion #1: It gives sentences involving dynamic binding well-formed semantic representations.

On the other hand, semantic representations have to correctly represent their truth conditions, or the *entailment relation* in which they are involved. For example, (1a) may participate in the following syllogism.

Example 1 (Donkey Syllogism).

Every farmer who owns [a donkey], beats it_i. John is a farmer. Bill is a donkey. John owns Bill. John beats Bill. Furthermore, (2) may participate in the following syllogisms.

Example 2 (E-type Syllogisms).

| $[A man]_{i}$ entered. | $[A man]_{i}$ entered. |
|---------------------------|------------------------|
| He _i whistled. | He_i whistled. |
| A man entered. | A man whistled. |

Although these examples are far from exhaustive, they constitute central paradigms which can be used to check if a given analysis is not immediately falsified. Thus, a successful semantic theory should include a method for calculating such relations in one way or another.

Criterion #2: It calculates the entailment relation between sentences involving dynamic binding.

The first-order representations for sentences (1) and (2) necessary in order to correctly calculate their entailment relations are (5) and (6), respectively.

- (5) $\forall x(\mathbf{Farmer}(x) \rightarrow \forall y(\mathbf{Donkey}(y) \land \mathbf{Own}(x, y) \rightarrow \mathbf{Beat}(x, y)))$
- (6) $\exists x(\mathbf{Man}(x) \land \mathbf{Enter}(x) \land \mathbf{Whistle}(x))$

These represent proper information that the sentences (1) and (2) contain, in a sense that any proof system for first-order predicate logic will prove that the inferences in Example 1 and Example 2 are valid.

On the other hand, the structural similarity to the original sentences is lost in (5) and (6), so their direct decomposition does not lead to the respective lexicalized representations. In this way, there is always tension between dynamism and compositionality.

Criterion #3: It is compositional.

Moreover, one important empirical paradigm to be explained by dynamic semantics is that of *accessibility* constraints that restrict the access of anaphora to its antecedent, which is an issue reported by Karttunen (1976) in mid 1970's and given a first generalization in DRT (Kamp 1981, Kamp and Reyle 1993). Dynamic binding is licensed in some configurations but not in others, as exemplified in the following sentences.

- (7) Everybody bought $[a car]_{i}$. *It_{*i*} stinks.
- (8) If John bought $[a car]_{i}$, it must be a Porsche. *It stinks.
- (9) John didn't buy $[a car]_{i}$. *It_{*i*} stinks.

(10) John bought [a car], or didn't buy anything. *It_i stinks.

Dynamic semantics should thus correctly predict such accessibility constraints and should be able to *explain* why they appear as they do and where they come from.

Criterion #4: It predicts and explains accessibility constraints.

Therefore, a successful theory of dynamic semantics should satisfy all criteria #1, #2, #3 and #4.

1.2 "Type Theoretical" approaches

Sundholm (1986) noticed fairly early that Martin-Löf (or Constructive) Type Theory (Martin-Löf 1975; 1984) provides semantic representations for donkey sentences whose structures are parallel to their syntactic structures (Criterion #1 is satisfied) in a different way from DRT (Kamp 1981, Kamp and Reyle 1993), DPL (Groenendijk and Stokhof 1991), and their successors. Subsequently, the following three approaches have been proposed to obtain Sundholmian representations:

- 1. Ahn and Kolb (1990) provides a set of translation rules from Discourse Representation Structures (DRS) to Sundholmian representations.
- 2. Dávila-Pérez (1994) presented a reformulation of Montague Grammar (Montague 1973) in terms of Martin-Löf's Higher Type Theory, following the line of (Ranta 1991).
- 3. Ranta (1994) proposed a generative theory of grammar based on Martin-Löf's Lower Type Theory, known as Type-Theoretical Grammar (TTG).

In all these approaches, representations are incorporated into the proof system of Martin-Löf Type Theory (Criterion #2 is satisfied) and accessibility conditions are explained in a non-ad-hoc manner (Criterion #4 is satisfied: see (Fox 1994) among others), which I discuss in detail in Section 3.

Thus, type theoretical approaches are highly prospective in pursuing a unified theory which satisfies all the four criteria and serves as a proof-theoretic alternative to dynamic semantics. However, they face several problems with respect to compositionality (Criterion #3 is not satisfied) as matters stand. For example, TTG's derivations of the semantic representations for the sentences in (11) are given as (12).¹

¹In TTG, the sorts set and prop are synonyms of type. Man is a type by itself, while Enter(x) is a type depending on x.
(12)

(11) a. A man entered.

b. Every man entered.

 $\overline{x: \text{Man}}^{(1)} \qquad \overline{x: \text{Man}}^{(1)}$ $\vdots \qquad \vdots$ $(\Sigma I) \frac{\text{Man : set } \text{Enter}(x) : \text{set}}{(\Sigma x: \text{Man}) \text{Enter}(x) : \text{set}}^{(1)} \qquad (\Pi I) \frac{\text{Man : set } \text{Enter}(x) : \text{set}}{(\Pi x: \text{Man}) \text{Enter}(x) : \text{set}}^{(1)}$

The proof diagrams in (12) use the same set of axioms, and the only difference between them is the rules used in the last steps. It derives (11a) if the $(\pm I)$ rule is used, while it derives (11b) if the $(\diamond I)$ rule is used instead.

This point is puzzling from a perspective of formal semantics, making us wonder whether it counts as compositional analysis. At the very least, TTG is not compositional in the sense of Montagovian semantics since it provides no lexicalized representations for "A" and "Every", and as a result it provides no way to compose the representations in (12) from the representations of the words therein.

On the other hand, TTG analysis can be regarded as compositional from the perspective of analytic philosophy. That is, TTG analyzes the *concepts* that the sentences in (11) express, rather than the sentences in (11) themselves, and how such concepts can be composed from more primitive concepts such as **Man** : set and **Enter**(x) : set, with a fixed number of rules. In other words, TTG is a theory about compositionality of concepts, not of sentences.

Therefore, we have to be careful about the status of TTG as a grammar. TTG is not about the relation between sentences and meanings which Montagovian theories are about, but rather a theory of meaning which has its own way of verification.

1.3 Toward a dynamic, proof-theoretic and compositional theory of semantics

The aim of this paper is to provide a new framework of formal semantics that satisfies all four criteria in the subsection 1.1 in a single and unified setting. The new framework, *dependent type semantics*, is formulated based on dependent type theory that originates in Martin-Löf Type Theory and λ -cube (Barendregt 1992).

Although dependent type semantics is closely related to the three type theoretical approaches in the subsection 1.2, the analysis of presupposition and anaphora in dependent type semantics is vastly different in that it employs a mechanism of *context passing*, which is common in continuation semantics (cf. de Groote 2006).

This paper is organized as follows. In Section 2, I present the basic concept of dependent type semantics after informally introducing the notion of dependent types (the formal presentation of which is given in Bekki 2013). The basic concepts include the notion of truth based on the paradigm of Curry-Howard correspondence, the analysis of anaphora based on the relation between propositions and proofs (Criterion #1), the proof theory based on dependent type theory that calculates inferences involving dynamic binding (Criterion #2), the lexicalized semantic representations and a number of rules to combine these elements of the theory (Criterion #3). In Section 3, accessibility is uniformly explained in terms of dependent type semantics, without assuming any adhoc mechanism (Criterion #4). Similarly to the way that instances of inaccessibility are reduced to negation in DPL/HoDL/QDL, they are reduced to the Π-operator in dependent type semantics. In Section 4, in order to make the formulation of dependent type semantics more rigorous, I discuss the polymorphic nature of contexts and reformulate dynamic propositions in a polymorphic manner.

2 Dependent Type Semantics

2.1 What are Dependent Types?

The notion of dependent types originates from Martin-Löf Type Theory (Martin-Löf 1975; 1984), which was proposed as a foundation of constructive mathematics, and Calculus of Constructions (CoC, Coquand and Huet 1988), which was proposed as a foundation of functional programming and mathematical proofs. Lately, fragments of Martin-Löf Type Theory and CoC have been integrated into a general theory of the λ -cube (Barendregt 1992) and Pure Type Systems (PTS) (Berardi 1990, Barendregt 1991) with other type theories, such as Girard's *F* (Girard et al. 1989). Furthermore, CoC has been extended by the notion of inductive definition of types in Martin-Löf Type Theory, leading to Calculus of Inductive Constructions, which is known as an underlying language of proof assistants *Coq* (Bertot and Castéran 2004) and *Agda* (Nordström et al. 1990, Bove and Dybjer 2008).

For those familiar with typed lambda calculi and the Curry-Howard correspondence between logic and type theory, the difference between dependent type theory and simply-typed lambda calculi can be summarized as follows.²

- Types may depend on terms, i.e., there is a function from terms to types.
- Types that depend on term variables can be quantified (by Π as a universal quantifier and Σ as an existential quantifier)

²The formal presentation of the specific theory that we adopt in this paper (a fragment of CoC extended with Σ -types) is given in (Bekki 2013).

• $(\prod x:A)B$ is a generalized form of the type $A \to B$, and $(\sum x:A)B$ is a generalized form of the type $A \times B$ (thus, the quantifiers Π and Σ may appear within types).

2.2 Intra-sentential semantics: Truth as inhabitance

In dependent type semantics, dependent types are applied to natural language semantics in the following way. We start by considering the truth conditions of the simple sentence in (13).

(13) John is a man.

A proper name, such as "John", is considered to denote an entity. In dependent type semantics, it means that a semantic representation of "John" is a term *j* of type **Entity**.

Assume that John is in fact a man, then there must be John's state of being a man in the sense of Davidson (1967) and Parsons (1990), which serves as a *proof* for sentence (13). Such proof can be represented as a term of type $Man(j)^3$, which is dependent on *j*. Since it is obvious that someone's state of being a man does not serve as a proof that someone else is a man, the collection of (someone's) states of being a man is divided into subcollections of states of being a man in accordance with the bearers of the states. This is the reason why we should consider that the type for the states of being a man depends on the terms that represent the bearer of the states.

By Curry-Howard correspondence ("terms = proofs"), such a proof of (13) is identified with a term of type Man(j). Again, by Curry-Howard correspondence ("types = propositions"), the type Man(j) is also regarded as a proposition, which is the semantic representation of sentence (13) (tentatively, until the next section) in dependent type semantics.

According to the proof-theoretic paradigm, sentence (13) is true if and only if there exists a proof for the proposition that the sentence denotes, namely, the type Man(j). This condition can be rephrased as that the type Man(j) is *inhabited*, that is, there exists a term of type Man(j). Equivalently, sentence (13) is true if and only if there is a term that represents John's state of being a man.

Sentence (14) is another simple example.

(14) John ran.

Assume that John actually ran, then there must be a "running" event whose agent is John, again in the sense of Davidson (1967) and Parsons (1990). Such an event is represented as a term of type **Ran**(j), which is also dependent on j. Therefore, proofs

³The type of **Man**(*j*) itself is type, which plays the role of type *t* in Montagovian semantics, so the type of the one-place predicate **Ran** is **Entity** \rightarrow type.

for sentence (14) have a type $\mathbf{Ran}(j)$, which is also regarded as a proposition. The same argument as above applies here as well, and the type $\mathbf{Ran}(j)$ is (tentatively) a semantic representation of sentence (14), which is true if and only if the type $\mathbf{Ran}(j)$ is inhabited.

In dependent type semantics, a semantic representation of an declarative sentence is always of the sort type. Thus, these arguments for the simple sentences in (13) and (14) can be generalized as in Definition 2.1

Definition 2.1 (Truth Condition in Dependent Type Semantics). For any declarative sentence *S* whose semantic representation is *S'* of the sort type, the truth condition of *S* is stated as follows:⁴

S is true. \iff S' is inhabited.

Now, let us consider the truth condition of the existentially quantified sentence (15) in this setting.

(15) A man entered.

A qualified proof that "A man entered" (i.e. "there exists a man who entered") consists of at least 1) an entity, 2) a proof that the entity is a man (that is, a state of the entity's being a man), and 3) a proof that the entity entered (that is, an entering event whose agent is the entity). For example, suppose that j and b are John and Bill, m_1 and m_2 are John and Bill's respective states of being a man, and e_1 and e_2 are John and Bill's respective entering events, namely:

(16) j: Entity, m_1 : Man(j), e_1 : Enter(j)b: Entity, m_2 : Man(b), e_2 : Enter(b)

Then the tuples $((j, m_1), e_1)$ and $((b, m_2), e_2)$ are two possible proofs for "A man entered". Thus, the truth condition of (15) is that there exists a tuple consisting of x such that x: **Entity**, a proof of x's being a man and a proof of x's entering. The type of such proofs is (tentatively) (17), and so is the semantic representation for (15) in dependent type semantics.⁵

⁴This claim is rather controversial in a philosophcal viewpoint, which is related with anti-realism in (Dummett 1975; 1976; 1991) and (Prawitz 1980). See also (Dávila-Pérez 1996) which argued for a proof-theoretic semantics of type theoretical grammar.

⁵The tuples $(j, (m_1, e_1))$ and $(b, (m_2, e_2))$ are as much proofs for sentence (15) as $((j, m_1), e_1)$ and $((b, m_2), e_2)$, and thus their type $(\Sigma x: \mathbf{Entity})(\mathbf{Man}(x) \land \mathbf{Enter}(x))$ also qualifies as a semantic representation for the sentence (15). Although this looks simpler than (17), I do not adopt it due to the lack of compositionality.

(17) $(\Sigma u:(\Sigma x:Entity)Man(x))Enter(\pi_1(u))$

The fact that the representation (17) is in fact of the sort type is proved as follows.⁶

| (18) | $(\mathcal{A}\mathcal{X}) \overline{Entity: \text{type}}$ | $(\Pi E) \frac{(\mathcal{RX})}{Man : Entity \rightarrow type}}{Man(x) : type}$ | x : Entity (1) | | $(\mathcal{R}X) = \mathbf{Enter} : \mathbf{Entity} \to \mathbf{type}$ | $\underbrace{\frac{u:(\Sigma x:\mathbf{Entity})\mathbf{Man}(x)}{\pi_1(u):\mathbf{Entity}}}_{(\Sigma E)}$ | |
|--------------|---|--|----------------|--------|---|--|-----|
| (ΣF) (ΣF) | | $(\Sigma x: Entity)Man(x) : type$ | | -(1) (| Enter(| $\pi_1(u))$: type | (2) |

 $(\Sigma u: (\Sigma x: Entity)Man(x))Enter(\pi_1(u)): type$

Let us proceed to the truth condition of the universally quantified sentence (19).

(19) Every man entered.

A qualified proof for sentence (19) is a *function* that, for any pair of 1) an entity and 2) a proof that the entity is a man, returns a proof that the entity entered. The type of this function is not a simple form **Entity** \times **Man** \rightarrow **Enter** since the type **Man** depends on the element of **Entity**, and so does the type **Enter**. The type for the function, and thus the semantic representation of (19), is (tentatively) (20).

(20) $(\Pi u: (\Sigma x: \mathbf{Entity})\mathbf{Man}(x))\mathbf{Enter}(\pi_1(u))$

The representation in (20) is of the sort type, which is proved as in (21) (almost isomorphic to (18)).



The notion I introduced in this section, truth as inhabitance (of a proof), together with a bridging perspective of states and events as proofs, is a core notion of the truth condition within dependent type semantics. However, dependent type semantics gives a full play when in inter-sentential cases, which is discussed in the next section.

2.3 Inter-sentential semantics: Dynamics as dependence

Let us move on to the dynamic aspect of dependent type semantics. Consider an E-type pronoun in the sentences in (2).

(2) [A man]_{\rangle} entered. He_{*i*} whistled.

⁶For details about the rules used in (18), see the appendix of (Bekki 2013).

The semantic representation for the first sentence is (17) as we have discussed in the preceding section. Henceforth, I abbreviate **Man**(x) simply as **M**(x), **Enter**(x) as **E**(x), and so forth, for compactness. Moreover, I use the following abbreviated form for types of the form (Σx :**Entity**)M, which appears often.

Definition 2.2 (Abbreviated forms).

$$(\Sigma x)M \stackrel{\text{lif}}{=} (\Sigma x: \text{Entity})M \qquad M \stackrel{\text{lif}}{=} Man$$

$$E \stackrel{\text{lif}}{=} Enter \qquad W \stackrel{\text{lif}}{=} Whistle$$

The semantic representation in (17) can now be written as (22).

(22) $(\Sigma u:(\Sigma x)\mathbf{M}(x))\mathbf{E}(\pi_1(u)):$ type

In order to determine the semantic representation of (2), we need a representation for the anaphoric expression "He" in (2), which is related to its antecedent. The key in inter-sentential analysis of dependent type semantics is that we can always find the antecedent of an anaphoric expression in *proofs* of the preceding discourse.⁷ If $((j, s_1), e_1)$ is a proof for (22), then "He_i" in (2) must denote *j*. If $((b, s_2), e_2)$ is a proof for (22), then "He_i" in (2) denotes *b*. In general, if *v* is a proof for (22), namely $v : (\Sigma u: (\Sigma x) \mathbf{M}(x)) \mathbf{E}(\pi_1(u))$, then "He_i" in (2) denotes $\pi_1(\pi_1(v))$. In this way, the antecedent of an E-type pronoun can be accessed by means of (a composition of) projections from a proof of the representation of the sentence containing the antecedent.

Therefore, all we need is a mechanism that combines the representations of any two consecutive sentences by passing the *proof* for the representation of the first sentence to the representation of the second sentence. For this purpose, we employ a method which is similar to that in continuation semantics, and the semantic representation of the second sentence in (2) becomes as follows:

(23) $(\lambda c)\mathbf{W}(\mathbf{sel}_i(c))$: discourse \rightarrow type

For each index *i*, the *selection function* sel_i is a selected projection function (or a composition thereof)⁸. The choice of the selection function sel_i corresponds to the choice of indices in generative grammar. In this paper, subscripts in the sentences,

⁷This is in sharp contrast with the view of dynamic logic, which refers to the antecedent by means of *assignment functions*, while dependent type semantics refers to the antecedent by means of the proofs.

⁸The name **sel** is adopted from continuation semantics (de Groote 2006, Asher and Pogodalla 2011). More formally, **sel**_{*i*} is a member of the set of finite compositions of projections, namely $\{\pi_{i_1} \circ \cdots \circ \pi_{i_n} | i_1, \ldots, i_n \in \{1, 2\}, 1 \le n\}$. Similar ideas are also found in monadic semantics (Ogata 2008, Unger 2011).

such as *i* in "[A man]_i" and "He_{*i*}", are not a formal notation: they just indicate which expression is an antecedent of which anaphoric expression. The use of subscripts such as *i* in **sel**_{*i*} is also an informal notation: they just indicate that the selection function is properly chosen in order to establish the anaphoric link indicated by the same subscript.

The variable c in (23) takes the preceding *discourse*⁹, and in the discussion below a semantic representation for a sentence is a function from **discourse** to type, instead of just type. The actual type of **discourse** varies with respect to the content of previous discourses, but for a moment I simply specify it as **discourse**.¹⁰ This change should affect the semantic representation of (2) in a way that allows us to add the vacuous discourse variable c to (22), and we obtain (24).

(24)
$$(\lambda c)(\Sigma u:(\Sigma x)\mathbf{M}(x))\mathbf{E}(\pi_1(u)): \text{discourse} \rightarrow \text{type}$$

In (24), variable c does not appear in the scope of λc . This is a case when the sentence (2) contains no anaphoric expressions. Below, (24) serves as the semantic representation for the first sentence in (2).

The discourse variable *c* provides us with a nice way for combining semantic representations *M*, *N* for two consecutive sentences. They are composed by the *dynamic conjunction* ";", defined as follows:¹¹

Definition 2.3 (Dynamic Conjunction Operator).

.....

$$M; N \stackrel{\text{lift}}{\equiv} (\lambda c)(\Sigma u; M(c))(N((c, u)))$$

By means of the dynamic conjunction operator, we can define a CCG-style conjunction rule which can be applied to the two consecutive sentences.

Definition 2.4 (Dynamic Conjunction Rule).

$$:\frac{\mathsf{S}:\mathsf{M}}{\mathsf{S}:\mathsf{M}}:\mathsf{CON}\mathcal{J}:;\quad\mathsf{S}:\mathsf{N}}{\mathsf{S}:\mathsf{M};\mathsf{N}}$$

This operation can be generalized to be applied to any phrase that is "S-reducible"¹², following Gazdar's cross-categorical conjunction (Gazdar 1980) as in Definition 2.5.

⁹It is not a representation of the preceding discourse, but a proof of it. I revisit this issue in later sections. ¹⁰Rigorously, the type discourse should be formulated as a *polymorphic type*. I consider this issue again and elaborate on its analysis in Section 4.

¹¹In the notation N((c, u)), N is a one-place predicate which is fed a tuple (c, u). This is different from N(c, u), where N is a two-place predicate.

¹²The notion of "S-reducible category" (cf. Winter 1995) is recursively defined as follows:

^{1.} *S* is an *S*-reducible category.

Definition 2.5 (Generalized Dynamic Conjunction Rule).

$$:\frac{\mathsf{X}:\mathsf{M}}{\mathsf{X}:(\lambda \vec{x})(\lambda c)(\Sigma u:M\vec{x}(c))N\vec{x}((c,u))}$$

where *X* is a *S*-reducible category.

Now the representations (24) and (23) are combined by the dynamic conjunction rule, and the single representation for the two sentences in (2) is obtained by the following derivation.

(25)
$$\frac{S: (\lambda c)(\Sigma u: (\Sigma x)\mathbf{M}(x))\mathbf{E}(\pi_1(u)) \quad CON\mathcal{J}: S: (\lambda c)\mathbf{W}(\mathbf{sel}_i(c))}{S: (\lambda c)(\Sigma v: (\Sigma u: (\Sigma x)\mathbf{M}(x))\mathbf{E}(\pi_1(u)))\mathbf{W}(\mathbf{sel}_i((c, v)))}$$

The proper choice of **sel**_i here is such that **sel**_i(x) = $\pi_1\pi_1\pi_2(x)$. Since $\pi_1(\pi_1(\pi_2((c, v)))) = \pi_1(\pi_1(v))$, the pronoun "He" is analyzed as denoting the entity, for which a proof of being a man and a proof of entering exist.

2.4 Dependent Type Semantics is proof-theoretic

As mentioned in Section 1, the sentences in (2) participate in the entailment relations Example 2, repeated below.

(26) a. $[A man]_i$ entered. He_i whistled. \models A man entered.

b. $[A man]_{i}$ entered. He_i whistled. \models A man whistled.

In dependent type semantics, these are correctly calculated in a proof-theoretic manner, without recourse to their models. This shall be counted as one of the advantages of dependent type semantics.

The sentences in the premise of (26a), when they are subsequent to the discourse, whose proof is c (whatever c may be), have the following semantic representation with c applied to it.

(27) $(\Sigma v: (\Sigma u: (\Sigma x) \mathbf{M}(x)) \mathbf{E}(\pi_1(u))) \mathbf{W}(\pi_1 \pi_1(v)) : \text{type}$

The entailment in (26a) is proven in a straightforward manner since the consequence of (26a) is just a first projection of the premises of (26a). In proving (26a), we assume that (27) is inhabited. Assume that the term t is such an inhabitant.

^{2.} if X is an S-reducible category, X/Y and $X\setminus Y$ are also S-reducible categories, for any category Y.

^{3.} No other category is an S-reducible category.

(28)
$$\frac{t : (\Sigma v: (\Sigma u: (\Sigma x) \mathbf{M}(x)) \mathbf{E}(\pi_1(u))) \mathbf{W}(\pi_1(\pi_1(v)))}{\pi_1(t) : (\Sigma u: (\Sigma x) \mathbf{M}(x)) \mathbf{E}(\pi_1(u))}$$

The entailment in (26b) seems to be more complex, but still simple as in (29), where the use of the (CONV) rule (see the appendix of Bekki 2013) is somewhat abused.

$$(29) \underbrace{t : (\Sigma v: (\Sigma u: (\Sigma x) \mathbf{M}(x)) \mathbf{E}(\pi_{1}(u))) \mathbf{W}(\pi_{1}\pi_{1}(v))}_{(\Sigma E) \underbrace{\pi_{1}(t) : (\Sigma u: (\Sigma x) \mathbf{M}(x)) \mathbf{E}(\pi_{1}(u))}_{\pi_{1}\pi_{1}(t) : (\Sigma x) \mathbf{M}(x)} \underbrace{(\Sigma E) \underbrace{t : (\Sigma v: (\Sigma u: (\Sigma x) \mathbf{M}(x)) \mathbf{E}(\pi_{1}(u)))}_{\pi_{2}(t) : \mathbf{W}(\pi_{1}\pi_{1}(v))[\pi_{1}(t)/v]}_{= \mathbf{W}(\pi_{1}\pi_{1}\pi_{1}(t))}_{= \mathbf{W}(\pi_{1}\pi_{1}\pi_{1}(t))}_{= \mathbf{W}(\pi_{1}(u))[\pi_{1}\pi_{1}(t)/u]}$$

Thus, the following inferences hold, which is a proof-theoretic account of the data in (26).

(30) a. $(\Sigma v: (\Sigma u: (\Sigma x)\mathbf{M}(x))\mathbf{E}(\pi_1(u)))\mathbf{W}(\pi_1\pi_1(v)) \vdash (\Sigma u: (\Sigma x)\mathbf{M}(x))\mathbf{E}(\pi_1(u))$ b. $(\Sigma v: (\Sigma u: (\Sigma x)\mathbf{M}(x))\mathbf{E}(\pi_1(u)))\mathbf{W}(\pi_1\pi_1(v)) \vdash (\Sigma u: (\Sigma x)\mathbf{M}(x))\mathbf{W}(\pi_1(u))$

2.5 Dependent Type Semantics is compositional

In dependent type semantics, the semantic representation for the sentences in (2) can be naturally lexicalized and composed. In this paper, I adopt Combinatory Categorial Grammar (henceforth CCG) (Steedman 1996; 2000) as a syntactic theory in order to demonstrate semantic composition along with syntactic derivation. However, with minor modifications, it should work together with any other lexical grammar, such as type logical grammars, classical/abstract categorial grammars, LFG, LTAG and minimalist grammars, since dependent type semantics is independent of any specific feature of CCG.

I assume the following map [[–]] from the syntactic categories in CCG to the semantic types in dependent type semantics.

Definition 2.6 (Correspondence between Syntactic Categories in CCG and Semantic Types in Dependent Type Semantics).

| $\llbracket \mathcal{NP} \rrbracket$ | = | Entity | $\llbracket \mathcal{N} \rrbracket$ | = | Entity \rightarrow type |
|--------------------------------------|---|---|---------------------------------------|---|---|
| $\llbracket \mathcal{S} \rrbracket$ | = | discourse \rightarrow type | $\llbracket \overline{S} \rrbracket$ | = | discourse \rightarrow type |
| $\llbracket X/Y \rrbracket$ | = | $\llbracket Y \rrbracket \to \llbracket X \rrbracket$ | $\llbracket X \setminus Y \rrbracket$ | = | $\llbracket Y \rrbracket \to \llbracket X \rrbracket$ |

Theorem 1 states that combinatory rules in CCG *preserve* this correspondence, which allows us to conclude that every sentence has a semantic representation of type discourse \rightarrow type.

Theorem 1 (Soundness of Combinatory Rules with respect to the Category=Type Correspondence). For any combinatory rule in CCG of the following form (where X_1, \ldots, X_n , Y are syntactic categories and f_1, \ldots, f_n , g are semantic representations):

$$\frac{X_1:f_1 \quad \cdots \quad X_n:f_n}{Y:g}$$

if f_1 is of type $\llbracket X_1 \rrbracket$, ..., f_n is of type $\llbracket X_n \rrbracket$, then g is of type $\llbracket Y \rrbracket$.

The proof is routine. Then, a derivation of the sentence "A man entered" is as follows:

| (31) | А | man | |
|------|---|-------------------------------------|--------------------------------------|
| | $\mathcal{S}/(\mathcal{S}\backslash\mathcal{NP})/\mathcal{N}$ | N | |
| | $(\lambda n)(\lambda p)(\lambda c)(\Sigma u:(\Sigma x)nxc)(p(\pi_1(u))((c,u)))$ | $(\lambda x)(\lambda c)\mathbf{M}x$ | entered |
| | $>$ $S/(S \setminus NP)$ | | $\mathcal{S} \setminus \mathcal{NP}$ |
| | $(\lambda p)(\lambda c)(\Sigma u:((\Sigma x)\mathbf{M}x))p(\pi_1(u))((du))$ | $(\lambda x)(\lambda c)\mathbf{E}x$ | |
| > | S | | |
| | $(\lambda c)(\Sigma u:((\Sigma x)\mathbf{M}x))\mathbf{E}$ | $\mathcal{L}(\pi_1(u))$ | |

Intuitively, u denotes a proof of the subject noun phrase, and x denotes a proof of the representation of a head noun of the noun phrase.¹³

The variable c in nxc in the representation of "A" does not appear in the resulting representation above. This is because "A man" does not contain any modifiers that contain anaphora.

(32)
$$\frac{\text{He}_{i}}{S/(S\setminus\mathcal{NP})} \frac{\text{whistled}}{S\setminus\mathcal{NP}}$$

$$> \frac{(\lambda p)(\lambda c)p(\text{sel}_{i}(c))c}{S} \frac{(\lambda x)(\lambda c)Wx}{S}$$

$$(\lambda c)W(\text{sel}_{i}(c))$$

Thus, the derivation in (25) is resumed, this time entirely from the lexicalized representations.

$(\Sigma x:Man)Enter(x)$

I do not adopt this representation, however, for reasons of compositionality, and the analysis is more robust when we consider noun modifiers such as relative clauses.

¹³This representation may seem to be redundant because TTG would assign a representation as follows for sentence (2).

(33)

| _ | A man entered | He_i whistled |
|----|---|---|
| | S | S |
| | $(\lambda c)(\Sigma u:((\Sigma x)\mathbf{M}x))\mathbf{E}(\pi_1(u))$ | $(\lambda c) \mathbf{W}(\mathbf{sel}_i(c))$ |
| ;- | S | |
| | $(\lambda c)(\Sigma v:((\Sigma u:(\Sigma x)\mathbf{M}x)\mathbf{E}(\pi_1(u)$ | $))))\mathbf{W}(\mathbf{sel}_i((c, v)))$ |

3 Accessibility

DRT has boxes as a mechanism for controlling the accessibility of anaphora. In DPL, negation as a test plays the role of a box. The current implementation of dependent type semantics, where the Π -operator plays the role, inherits this idea from DPL, since in dependent type theory negation is defined via the Π -operator.

Let us check some of the classical paradigms of accessibility discussed in DRT and how dependent type semantics gives them a new life.

3.1 Universal quantification

The first case is universal quantification in the subject position, which blocks the link to indefinites within its scope from subsequent sentences, as in (7).

(7) Everybody bought $[a car]_{i}$. *It_{*i*} stinks.

The lexical item for "Everybody" is given by means of the Π -operator as follows.



$$(35) \qquad \underbrace{ \begin{array}{c} \text{Everybody bought [a car]}_{\mathcal{S}} & \varnothing \\ S \\ (\lambda c)(\Pi u:(\Sigma x)\mathbf{M}(x))(\Sigma v:(\Sigma y)\mathbf{C}(y))\mathbf{B}(\pi_1(u), \pi_1(v)) & ; \\ S \\ (\lambda c)(\Sigma w:(\Pi u:(\Sigma x)\mathbf{M}(x))(\Sigma v:(\Sigma x)\mathbf{C}(x))\mathbf{B}(\pi_1(v), \pi_1(u)))\mathbf{S}(\mathbf{sel}_i((c, w))) \\ \end{array}}_{\mathcal{S}}$$

In the derivation above, the projection $sel_i((c, w))$ has to pick an antecedent from a proof of the following type:

(36) $(\Pi u:(\Sigma x)\mathbf{M}(x))(\Sigma v:(\Sigma y)\mathbf{C}(y))\mathbf{B}(\pi_1(u),\pi_1(v))$

However, "a car" (denoted by a variable y) is not accessible from the subsequent discourse since w is a *function* from $(\Sigma x)\mathbf{M}(x)$ to $(\Sigma v:(\Sigma y)\mathbf{C}(y))\mathbf{B}(\pi_1(v), \pi_1(u))$, so that it is not possible to pick *a car* from w by (a composition of) projections alone. Thus, universal quantification behaves like a *box* in DRT.

3.2 Conditionals

The second case is conditional sentences. In a conditional sentence, an antecedent in the premise part is accessible from anaphora in the consequent part, but not from anaphora in subsequent sentences, as shown in (8).

(8) If John bought $[a car]_{i}$, it_i must be a Porsche. *It_i stinks.

The sentences in (8) are derived as follows. I analyze "If" as a material conditional for the sake of simplicity, which is represented by the Π -operator in dependent type semantics.

| (37) | If | John bought [a car] $_{\rangle}$ | | | |
|------|--|---|-----------------------------------|--|--|
| _ | S/S/S | <u>\$</u> | | | |
| | $(\lambda p)(\lambda q)(\lambda c)(\Pi w : pc)(q((c,w)))$ | $(\lambda c)(\Sigma u:(\Sigma y)\mathbf{C}(y))\mathbf{B}(j,\pi_1(u))$ | it _i must be a Porsche | | |
| > | S/S | S | | | |
| | $(\lambda q)(\lambda c)(\Pi w:(\Sigma u:(\Sigma y)\mathbb{C}(y)))$ | (λc) bePorsche (sel _i c) | | | |
| > | S | | | | |
| | w))) | | | | |

The antecedent "a car" is a proof *y* of type **Entity** that satisfies C(y), so "it_{*i*}" is accessible to the proof via *w*. On the other hand, "It_{*i*}" in the second sentence in (8) is not accessible to the proof *y*.

| (38) | If John bought $[a car]_{i}$, it _i must be a Porsche | Ø | *It _i stinks |
|-------|--|--|--|
| (37)- | S | CONJ | S |
| | $(\lambda c)(\Pi w:(\Sigma u:(\Sigma y)\mathbf{C}(y))\mathbf{B}(j,\pi_1(u)))(\mathbf{bePorsche}(\mathbf{sel}_i((c,w))))$ | ; | $(\lambda c)\mathbf{S}(\mathbf{sel}_i(c))$ |
| ;— | S | | |
| | $(\lambda c)(\Sigma v:(\Pi w:(\Sigma u:(\Sigma y)\mathbf{C}(y))\mathbf{B}(j,\pi_1(u)))(\mathbf{bePorsche}(\mathbf{sel}_i)$ | $((c, w))))\mathbf{S}(\mathbf{sel}_i)$ | (c, v))) |

As can be seen from the derivation in (38), the second \mathbf{sel}_i has to pick the antecedent from (c, u), however, the functional proof that u denotes encapsulates the participants within the first sentence, which includes a proof of $\mathbf{C}(y)$. Thus, the accessibility constraint in conditional sentences are predicted without any ad-hoc assumptions.

3.3 Negation

The third case is negation. Negations are known to block accessibility, as in (9).

(9) John didn't buy $[a car]_{i}$. *It_i stinks.

This constraint in dependent type semantics is due to the definition of negation via implication.

Definition 3.1 (Negation). $\neg A \stackrel{\text{[1]}}{=} (\Pi x:A) \bot$

The semantic representation of the first sentence in (9) is as follows:

(39)

| | didn't | buy a car | | | |
|---|---|--|--|--|--|
| | $\mathcal{S} \mathcal{NP} / (\mathcal{S} \mathcal{NP})$ | $\overline{\mathcal{S} \setminus \mathcal{NP}}$ | | | |
| John | $(\lambda p)(\lambda x)(\lambda c)\neg pxc$ | $(\lambda x)(\lambda c)(\Sigma u:(\Sigma y)\mathbf{C}(y))\mathbf{B}(x,\pi_1(u))$ | | | |
| $\overline{\mathcal{NP}}$ | > | $S \setminus NP$ | | | |
| j | $(\lambda x)(\lambda c)\neg$ | $D(\lambda c) \neg (\Sigma u: (\Sigma y) \mathbf{C}(y)) \mathbf{B}(x, \pi_1(u))$ | | | |
| < | S | | | | |
| $(\lambda c) \neg (\Sigma u: (\Sigma y) \mathbf{C}(y)) \mathbf{B}(j, \pi_1(u))$ | | | | | |

According to Definition 3.1, the semantic representation in the last line, if some context *c* is fed to it, is a functional type $(\Pi x: (\Sigma x) \mathbf{C}(j)) \mathbf{B}(j, \pi_1(u))) \perp$, which is the type of *functions* from $(\Sigma u: (\Sigma x) \mathbf{C}(j)) \mathbf{B}(j, \pi_1(u))$ to \perp . Therefore, it is not possible to pick up "a car" from its proof by means of a composition of projections. In other words, negation blocks accessibility due to its implicational nature.

Notice that the fact that universal quantifications, conditionals and negations block accessibility is not only correctly predicted but also explained in dependent type semantics in a uniform way. In other words, they are all represented by the Π -operator, whose proofs are functions, from which a mere composition of projections cannot pick up the elements involved. This is a deeper explanation on accessibility, which relies only on the structures of proofs between the Σ -operator and the Π -operator, which is widely adopted in type theory, without resource to any ad-hoc assumptions.

3.4 Nominal conjunction

Another support for dependent type semantics is found in cases of anaphoric links over a nominal conjunction, as in (40).

(40) [A monk], and his_i apprentice opened the gate.

Since $A \wedge B \stackrel{\text{[]}}{=} (\Sigma x:A)B$ (where $x \notin \{\sqsubseteq(B), \text{ see the appendix of Bekki 2013})$, the analysis of anaphora in conjunctions parallels with those in implications and universal

quantifications, in a sense that the proof of the first conjunct is passed to the second conjunct. The derivation of "A monk" follows that of "A man" in (31), and I assume the following lexical item for the possessive pronoun "his".

| (41) | | $\frac{\frac{\text{his}}{S/(S \setminus N\mathcal{P})/N}$ | | | | | |
|-----------------------|--|--|---|--|--|--|--|
| | | | | | | | |
| | $(\lambda n)(\lambda p)(\lambda c)(\Sigma v:(\Sigma y)(nyc))$ | $(\lambda n)(\lambda p)(\lambda c)(\Sigma v:(\Sigma y)(nyc \wedge \mathbf{of}(y, \mathbf{sel}_i(c))(p(\pi_1(v))((c, v))))$ | | | | | |
| | > | $S/(S \setminus NP)$ | | | | | |
| | $(\lambda p)(\lambda c)(\Sigma v:(\Sigma y)(\mathbf{A}(y) \land \mathbf{of}(y, \mathbf{sel}_i(c))(p(\pi_1(v))((c, v))))$ | | | | | | |
| | | | | | | | |
| (42) | A monk | and | his apprentice | | | | |
| \ | $\mathcal{S}/(\mathcal{S} \setminus \mathcal{NP})$ | CONJ | $\mathcal{S}/(\mathcal{S}ackslash\mathcal{NP})$ | | | | |
| $(\lambda p)(\lambda$ | c)(Σu :(Σx) $\mathbf{M}x$)($p(\pi_1(u))((c, u))$) | ; | $(\lambda p)(\lambda c)(\Sigma v:(\Sigma y)(\mathbf{A}(y)$ | $\wedge \mathbf{of}(y, \mathbf{sel}_i(c))(p(\pi_1(v))((c, v))))$ | | | |
| ;— | | S/(S | $\langle NP \rangle$ | | | | |
| | $(\lambda p)(\lambda c)(\Sigma w:(\Sigma u:(\Sigma x)\mathbf{M}x)(p(\pi_1)))$ | $u))((c, u))))(\Sigma v:($ | Σy (A (y) \wedge of (y, sel _i ((c, w)) | $))(p(\pi_1(v))(((c, w), v))))$ | | | |

The analysis of "and" as a category CONJ here is the same as that for the sentential conjunction in Definition 2.5. The essence is that the proof of the first conjunct is passed to the second conjunct, which may depend on it, i.e., whose member may be used as an antecedent by anaphoric expressions contained in the second conjunct.

If the selection function selects "A monk" (which is included in w), "his" gets bound in the intended way. Moreover, in the verb phrase, both "A monk" and "his apprentice" can serve as an antecedent as in (43), which are both derivable if we set **sel**_{*i*} appropriately.

(43) [A monk], and [his apprentice] met his_{i/i} father.

Anaphora has access to the proof of the preceding discourse, but the only method to pick the antecedent is a composition of projection functions. Thus, the pairwise proof introduced by a Σ -operator (existential quantification and conjunction) yields accessible referents, while the functional proof introduced by a Π -operator (universal quantification and implication) blocks access to the elements inside it. To sum up, accessibility amounts to reachability by projection.

4 Polymorphism of discourse

In the subsection 2.3, I defined the semantic type of declarative sentences as discourse \rightarrow type. However, this is not a precise formalization, as I have mentioned in Footnote 10, since the type discourse should vary according to the discourse that precedes the sentence. Recall the sentences in (2) and their semantic representations in (33), repeated here as (44) for convenience.

(2) $[A man]_{i}$ entered. He_i whistled.

(44) $\begin{array}{c|c} A \text{ man entered} & He_i \text{ whistled} \\ \hline S & \\ \hline (\lambda c)(\Sigma u:((\Sigma x)\mathbf{M} x))\mathbf{E}(\pi_1(u)) & \\ \hline S \\ \hline (\lambda c)(\Sigma v:((\Sigma u:(\Sigma x)\mathbf{M} x)\mathbf{E}(\pi_1(u))))\mathbf{W}(\mathbf{sel}_i((c,v))) \\ \hline \end{array}$

If the initial discourse i is of type **d**, the representation

$$(\lambda c)(\Sigma u:((\Sigma x)\mathbf{M}x))\mathbf{E}(\pi_1(u))$$

must be of type $\mathbf{d} \to \text{type}$. Then, the new discourse (which consists of "A man entered") (c, v) is passed to $(\lambda c)\mathbf{W}(\mathbf{sel}_i(c))$ where v is a proof of type $(\Sigma u:((\Sigma x)\mathbf{M}x))$, and thus the type of (c, v) is $(\Sigma c:\mathbf{d})(\Sigma u:((\Sigma x)\mathbf{M}x))\mathbf{E}(\pi_1(u))$.

One proper way to treat types of discourses is to introduce a *polymorphic type* available in λ -cube. A polymorphic type is a type depending on other types.

$$\begin{array}{c|c} (45) & A \text{ man} & \text{entered} \\ \hline S/(S \setminus \mathcal{NP}) & S \setminus \mathcal{NP} \\ & \xrightarrow{(\lambda p)(\lambda \delta : \text{type})(\lambda c : \delta)(\Sigma u: ((\Sigma x)\mathbf{M} x))p(\pi_1(u))(\delta)((c, u))} & (\lambda x : \text{Entity})(\lambda \delta : \text{type})(\lambda c : \delta)\mathbf{E} x \\ & \xrightarrow{S} & (\lambda \delta : \text{type})(\lambda c : \delta)(\Sigma u: ((\Sigma x)\mathbf{M} x))\mathbf{E}(\pi_1(u)) \end{array}$$

This representation is polymorphic, i.e., it works for any discourse type. Then, one can instantiate the type according to the initial discourse.

(46)
$$\frac{\mathbf{d}:\mathsf{type}}{(\Pi E)} \frac{(\lambda \delta:\mathsf{type})(\lambda c:\delta)(\Sigma u:((\Sigma x)\mathbf{M} x))\mathbf{E}(\pi_1(u)):(\Pi \delta:\mathsf{type})(\Pi c:\delta)\mathsf{type}}{(\lambda c:\mathbf{d})(\Sigma u:((\Sigma x)\mathbf{M} x))\mathbf{E}(\pi_1(u)):(\Pi c:\delta)\mathsf{type}}$$

One can also instantiate the type of discourse passed to the second sentence. The derivation (47) is a polymorphic version of (32).

$$(47) \qquad \underbrace{\frac{\text{He}_{i}}{S/(S \setminus \mathcal{NP})}}_{(\lambda p)(\lambda \delta: \text{type})(\lambda c: \delta)p(\text{sel}_{i}(c))\delta c} \qquad \underbrace{\frac{S \setminus \mathcal{NP}}{(\lambda x: \text{Entity})(\lambda \delta: \text{type})(\lambda c: \delta)Wx}}_{S}$$

$$(48) \qquad (\Pi E) \frac{(\Sigma u:((\Sigma x)\mathbf{M} x))\mathbf{E}(\pi_1(u)): \mathsf{type}}{(\lambda c:(\Sigma u)((\Sigma x)\mathbf{M} x)\mathbf{E}(\pi_1(u)):\mathbf{W}(\mathsf{sel}_i(c))):(\Pi c:(\Sigma u:((\Sigma x)\mathbf{M} x))\mathbf{E}(\pi_1(u)))\mathsf{type}}$$

On first look, we may get the impression that it is a bit complicated but it is not counterintuitive. This representation says that the discourse passed to the second sentence is a (generalized) pair composed of the discourse that precedes the first sentence and the proof of the first sentence.

This remedy is implemented in a simple manner by two modifications to our theory. First, we uniformly change the semantic representation for every proposition of the form (49) to (50).

- (49) $(\lambda c: \text{discourse})M: \text{discourse} \rightarrow \text{type}$
- (50) $(\lambda \delta : \text{type})(\lambda c : \delta)M : (\Pi \delta : \text{type})(\Pi c : \delta)M$

Second, we modify the dynamic conjunction rule in Definition 2.4 as follows:

Definition 4.1 (Dynamic Conjunction Rule (modified)).

| | S | CONJ | S |
|----|--|------|--|
| | $M: (\Pi \delta: type)(\Pi c: \delta)type$ | ; | $N: (\Pi \delta: type)(\Pi c: \delta)type$ |
| ;- | L J1 | S | |

$$M; N \stackrel{\text{\tiny thet}}{=} (\lambda \delta)(\lambda c')(\Sigma u: M \delta c')(N((\Sigma c: \delta) M \delta c')((c', u))) : (\Pi \delta: \mathsf{type})(\Pi c: \delta)\mathsf{type}$$

The proof of Theorem 1 must be updated accordingly.

5 Conclusion

This paper presented a new framework, referred to as dependent type semantics, which is based on dependent type theory and the context-passing mechanism in continuation semantics. I have demonstrated that dependent type semantics satisfies the four criteria required for a successful theory of dynamic semantics, which are discussed in the subsection 1.1, in the following way:

- **Criterion #1:** it proves well-formed representations for sentences involving dynamic binding.
- **Criterion #2:** it is proof-theoretic and thus predicts entailment relations involving dynamic binding.
- **Criterion #3:** it is compositional, and semantic representations for E-type and donkey sentences can be composed from lexicalized representations.
- **Criterion #4:** it explains accessibility in a uniform way based on the nature of implication/universal quantification.

The core concept of dependent type semantics, with respect to the notions of truth, dynamics and accessibility, is summarized by the following three concepts: i) truth as inhabitance, ii) dynamics as dependence on the contexts, and iii) accessibility as reachability by projections.

I hope that dependent type semantics not only serves as a dynamic, proof-theoretic and compositional framework which not only correctly predicts and explains both intrasentential and inter-sentential anaphora and their accessibility, but also provides a new conceptual perspective of the relation between representations, proofs and denotations in natural language semantics.

Acknowledgements My sincere thanks to the participants in my guest lectures at University of Amsterdam, Tilburg University and Utrecht University in October 2012, especially to Alexandru Baltag, Kohei Kishida, Reinhard Muskens, Rick Nouwen, Ben Rodenhaeuser and Sonja Smets for their valuable comments. I also thank Alastair Butler, Yuri Ishishita, Yuki Nakano and Ribeka Tanaka for many helpful discussions. This research is partially supported by a Grant-in-Aid for Young Scientists (A) (No. 22680013) from the Ministry of Education, Science, Sports and Culture.

References

R. Ahn and H.-P. Kolb. Discourse representation meets constructive mathematics. In L. Kalman and L. Polos, editors, *Papers from the Second Symposium on Logic and Language*. Akademiai Kiado, 1990.

N. Asher and S. Pogodalla. SDRT in continuation semantics. In T. Onoda, D. Bekki, and E. McCready, editors, *New Frontiers in Artificial Intelligence (JSAI-isAI 2010 Workshops, Tokyo, Japan, November 2010, Selected Papers from LENLS7)*, volume LNAI 6797, pages 3–15. Springer, Heidelberg, 2011.

H. P. Barendregt. Introduction to generalized type systems. *Journal of Functional Programming*, 1(2):125–154, 1991.

H. P. Barendregt. Lambda calculi with types. In S. Abramsky, D. M. Gabbay, and T. Maibaum, editors, *Handbook of Logic in Computer Science*, volume 2, pages 117–309. Oxford Science Publications, 1992.

D. Bekki. A type-theoretic approach to double negation elimination in anaphora. In *Logic and Engineering of Natural Language Semantics 10 (LENLS 10)*, Tokyo, 2013.

S. Berardi. *Type Dependence and Constructive Mathematics*. PhD thesis, Mathematical Institute, 1990.

Y. Bertot and P. Castéran. *Interactive Theorem Proving and Program Development*. Springer, 2004.

A. Bove and P. Dybjer. Dependent types at work, February - March 2008 2008.

T. Coquand and G. Huet. The calculus of constructions. *Information and Computation*, 76(2-3):95–120, 1988.

D. Davidson. The logical form of action sentences. In N. Rescher, editor, *The Logic of Decision and Action*. University of Pittsburgh Press, Pittsburgh, 1967.

R. Dávila-Pérez. Translating english into Martin-Löf's theory of types: A compositional approach. Technical report, University of Essex, May 1994 1994.

R. Dávila-Pérez. Comments on constructive semantics for natural language, 1996.

P. de Groote. Towards a montagovian account of dynamics. In M. Gibson and J. Howell, editors, *16th Semantics and Linguistic Theory Conference (SALT16)*, pages 148– 155, University of Tokyo, 2006. CLC Publications.

M. Dummett. What is a theory of meaning? In S. Guttenplan, editor, *Mind and Language*, pages 97–138. Oxford University Press, Oxford, 1975.

M. Dummett. What is a theory of meaning? (II). In Evans and McDowell, editors, *Truth and Meaning*, pages 67–137. Oxford University Press, Oxford, 1976.

M. Dummett. The Logical Basis of Metaphysics. Duckworth, London, 1991.

G. Evans. Pronouns. Linguistic Inquiry, 11:337–362, 1980.

C. Fox. Discourse representation, type theory and property theory. In H. Bunt, R. Muskens, and G. Rentier, editors, *the International Workshop on Computational Semantics*, pages 71–80, Institute for Language Technology and Artificial Intelligence (ITK), Tilburg, 1994.

G. Gazdar. A cross-categorial semantics for conjunction. *Linguistics and Philosophy*, 3:407–409, 1980.

P. Geach. *Reference and Generality: An Examination of Some Medieval and Modern Theories.* Cornell University Press, Ithaca, New York, 1962.

J.-Y. Girard, Y. Lafont, and P. Taylor. *Proofs and Types*. Cambridge Tracts in Theoretical Computer Science 7. Cambridge University Press, 1989.

J. Groenendijk and M. Stokhof. Dynamic predicate logic. *Linguistics and Philosophy*, 14:39–100, 1991.

H. Kamp. A theory of truth and semantic representation. In J. Groenendijk, T. M. Janssen, and M. Stokhof, editors, *Formal Methods in the Study of Language*. Mathematical Centre Tract 135, Amsterdam, 1981.

H. Kamp and U. Reyle. *From Discourse to Logic*. Kluwer Academic Publishers, 1993.

L. Karttunen. Discourse referents. In J. D. McCawley, editor, *Syntax and Semantics* 7: *Notes from the Linguistic Underground*, volume 7, pages 363–85. Academic Press, New York, 1976.

P. Martin-Löf. An intuitionistic theory of types. In H. E. Rose and J. Shepherdson, editors, *Logic Colloquium '73*, pages 73–118. North-Holland, Amsterdam, 1975.

P. Martin-Löf. *Intuitionistic Type Theory*, volume 17. Italy: Bibliopolis, Naples, 1984. Sambin, G. (ed.).

R. Montague. The proper treatment of quantification in ordinary english. In J. Hintikka, J. Moravcsic, and P. Suppes, editors, *Approaches to Natural Language*, pages 221–242. Reidel, Dordrecht, 1973.

B. Nordström, K. Petersson, and J. Smith. *Programming in Martin-Löf's Type Theory*. Oxford University Press, 1990.

N. Ogata. Towards computational non-associative lambek lambda-calculi for formal pragmatics. In the Fifth International Workshop on Logic and Engineering of Natural Language Semantics (LENLS2008) in Conjunction with the 22nd Annual Conference of the Japanese Society for Artificial Intelligence 2008, pages 79–102, Asahikawa, Japan, 2008.

T. Parsons. *Events in the Semantics of English: A Study in Subatomic Semantics*. The MIT Press, Cambridge MA, 1990.

D. Prawitz. Intuitionistic logic: A philosophical challenge. In G. von Wright, editor, *Logics and Philosophy*. Martinus Nijhoff, The Hague, 1980.

A. Ranta. Intuitionistic categorial grammar. *Linguistics and Philosophy*, 14:203–239, 1991.

A. Ranta. Type-Theoretical Grammar. Oxford University Press, 1994.

M. J. Steedman. *Surface Structure and Interpretation*. The MIT Press, Cambridge, 1996.

M. J. Steedman. *The Syntactic Process (Language, Speech, and Communication)*. The MIT Press, Cambridge, 2000.

G. Sundholm. Proof theory and meaning. In D. Gabbay and F. Guenthner, editors, *Handbook of Philosophical Logic*, volume III, pages 471–506. Kluwer, Reidel, 1986.

C. Unger. Dynamic semantics as monadic computation. In *the Eighth International Workshop on Logic and Engineering of Natural Language Semantics (LENLS8)*, pages 153–164, Takamatsu, Kagawa, Japan, 2011. JSAI International Symposia on AI 2011.

Y. Winter. Syncategorematic conjunction and structured meanings. In *SALT 1995*, 1995.

A Medieval Epistemic Puzzle

Sara L. Uckelman

Cluster of Excellence "Asia and Europe in a Global Context" Universität Heidelberg sara.uckelman@asia-europe.uni-heidelberg.de

Abstract

In this paper we present a dynamic epistemic language with names for propositions which we use to analyse an epistemic puzzle and associated disputation due to the medieval logician Paul of Venice.

1 Introduction

Epistemic logicians love puzzles, particularly puzzles involving reasoning about imperfect information in multi-agent settings. The literature on epistemic logic and dynamic epistemic logic is rife with examples of such puzzles¹, and many results in the field have arisen from attempts to formalize and solve these puzzles.

In this paper, we consider an interesting little epistemic puzzle with unusual origins. The puzzle is due to the medieval logician Paul of Venice, and occurs in the treatise *De scire et dubitare* 'On knowing and being uncertain'² of his monumental work *Logica Magna* 'Great Logic', which was published in the early 15th C.³ Paul introduces this puzzle as a precursor to presenting an *obligatio* concerning the contents of the puzzle. An *obligatio* is a type of formalized disputation which was developed

¹See (van Ditmarsch et al. 2007) for many examples.

²The translation of Latin *dubitare* and its cognates with 'to be uncertain' rather than 'to doubt' reflects the fact that *dubito A* 'I am uncertain about A' does not have the implication of belief in $\neg A$ that 'I doubt A' can have in English.

³For information about Paul's life and other works, see the intro of (Venice 1981).

in the Middle Ages and used as a method for testing students' abilities to recognize inferential relations and to reason under settings of changing information. As Paul explains in the treatise of the *Logica Magna* on *obligationes*, "the topic of *obligationes* is nothing other than the topic of inferences presented in a more subtle manner, in a way intended to test whether the respondent has a good head by setting a deceptive course before him" (Venice 1988, p. 33).⁴

Though topics in medieval logic are often considered far removed from the interests of modern logicians, it turns out that many of the puzzles medieval logicians were interested in remain interesting today for the same reasons. In this paper we use the tools of dynamic epistemic logic, suitably enriched to be sufficiently expressive to cover the relevant aspects of the puzzle and the ensuing disputation, to provide a formalization of and analysis of both. The resulting analysis displays a hitherto unnoticed inconsistency in Paul of Venice's epistemic theory, and also brings together different formal tools that are not ordinarily combined; as such, what we have to say here will be of interest to scholars of both medieval and modern logic.

We begin by giving the context of the disputation. The treatise *De scire et dubitare* is devoted to arguments for and against the claim that there is something known by someone which is uncertain to him. One of the arguments in favor of this claim rests on the following two assumptions:

(1) I assume (a) that you know that A is one of the two propositions 'God exists' and 'A human being is a donkey', and (b) that one A is every A, and (c) that it is hidden from you which of the propositions is A, but (d) you know perfectly well that the proposition 'God exists' is necessary and the other, 'A human being is a donkey', impossible (Venice 1981, p. 3).⁵

(2) Every proposition you consider which you do not know to be true and do not know to be false is uncertain to you (Venice 1981, p. 5).⁶

It must be noted, regarding the first assumption, that 'God exists' and 'A human being is a donkey' are being used as they typically are in medieval contexts; that is, they are being taken as a generic necessity and a generic impossibility, respectively. No part of

⁴Quod omnes regulae superius adsignatae in Tractatu Consequentiarum de consequentia bona vel non bona sunt hic fundamentaliter sustinendae. Et ratio quia materia obligationorum non est nisi materia consequentiarum stilo subtiliori procedens, et an respondens sit sani capitis gressu deceptorio temptativa (Venice 1988, p. 32).

⁵Et pono quod scias A esse alteram illarum 'Deus est' et 'Homo est asinus', et quod unum A est omne A, et lateat te quae illarum est A, sed bene scias quod illa est necessaria 'Deus est' et reliqua impossibilis 'Homo est asinus' (Venice 1981, p. 2).

⁶Omnis propositio de qua consideras quam non scis esse veram nec scis esse falsam sit tibi dubia (Venice 1981, p. 4).

the argument depends on the specific content of these propositions, it is rather the fact that (d) you know that one is necessity and the other is impossible that is relevant.

We will not discuss the details of the argument here, having done so elsewhere (Uckelman 2013b). Instead, we are interested in what follows after Paul has set up this background scenario.

2 Background information and other work

2.1 Obligationes

Though we do not assume that the reader is already familiar with the basics of medieval *disputationes de obligationibus* and their history, we cannot go into much detail here. Those wishing to know more are directed to (Uckelman 2012; 2013a) for further reading. We briefly recap Paul's rules and definitions for the basic variant of *obligationes*, called *positio*, found in his *De obligationibus*, Part II, Tract. 8 of the *Logica Magna* (Venice 1988).

There are two participants, Opponent and Respondent. Opponent begins by putting forward an initial statement, called the *positum*. When the Respondent admits a *positum*, he then binds himself to follow the rules of *obligationes* (Venice 1988, p. 3). It is this relation of rule following between the Respondent and the *positum* which is, on Paul's view, precisely what makes up the *obligatio*:

Definition 2.1. An *obligatio* is a relation limiting one to uphold some statement, or its equiform, in some way ... It is based on the obligater, by virtue of the *positio* or the *depositio*; and on the obligated, by reason of his *admissio*. (Venice 1988, pp. 7, 11).⁷

As is standard, Paul uses a notion of relevance. He first gives a general definition of relevance in terms of a relationship between propositions.

Definition 2.2 (Relevance). A proposition φ is relevant to a proposition ψ if φ follows from it or is inconsistent with ψ ; it is irrelevant otherwise (Venice 1988, pp. 24–25).

This general definition of relevance is broken down into a number of specific definitions, depending on whether the proposition φ is relevant (1) to the positing of the *positum*, (2) to the *positum* itself, (3) to both together, or, finally, (4) to the *positum* taken together with correctly granted propositions and negations of correctly denied propositions (Venice 1988, pp. 24–29). The 2nd definition is relevance in the Swyneshedian tradition; the 4th is the standard Burleyan definition of relevance, which Paul adopts.⁸

⁷Obligatio est relatio limitans ad aliquod enuntiabile vel sibi consimile aliqualiter sustinendum ... in obligante, ratione positionis vel depositionis; in obligato, vero ratione admissionis (Venice 1988, pp. 6, 10).
⁸See (Uckelman 2012, S4) for more on these two traditions.

The rest of the preliminary definitions and rules are standard: The *positum* should be admitted if there is no impediment arising in doing so (Venice 1988, pp. 48–49); any proposition once conceded must be conceded if it is ever put forward again (Venice 1988, pp. 34–35); relevant propositions must be conceded if they follow and denied if they do not (Venice 1988, pp. 54–63); irrelevant propositions must be conceded if neither (Venice 1988, pp. 64–65).

2.2 The language and its models

In other work we have developed a formal framework for the analysis of *obligationes* (Uckelman 2011a;b). For our present analysis, we use this framework along with the language introduced in (Uckelman 2013b), which is a combination of dynamic and epistemic logic, with additional alethic modalities, quantification over propositions, and names for propositions. The combination of dynamic logic with epistemic logic allows us to model both the actions of the players and their knowledge explicitly.

Definition 2.3 (Well-formed formulas). For sets Φ_0 of propositional letters, Ξ of variables, \mathcal{A} of agents (containing R and O, for Respondent and Opponent, respectively), and \mathcal{N} of names, the set Φ of well-formed formulas and Π of well-formed actions is defined by mutual induction:

$$\begin{array}{lll} \varphi & := & p \in \Phi_0 \mid x \in \Xi \mid x = y : x, y \in \Xi \mid x = p : x \in \Xi, p \in \Phi_0 \mid \\ & \neg \varphi \mid \varphi \lor \varphi \mid Tx : x \in \Xi \mid Tp : p \in \Phi \mid \exists x \varphi(x) \mid \\ & K_a \varphi : a \in \mathcal{A} \mid \Box \varphi \mid [\alpha] \varphi : \alpha \in \Pi \\ & \hat{a} : a \in \mathcal{N} \mid \hat{a} = \varphi \mid \hat{a} = x \mid \downarrow \varphi(\hat{a} \downarrow) \\ \alpha & := & \varphi! : \varphi \in \Phi \mid \varphi? : \varphi \in \Phi \end{array}$$

The other propositional operators, F, and \diamond are defined in the usual fashion. We define an uncertainty operator $U_a \varphi := \neg K_a \varphi \land \neg K_a \neg \varphi$; φ is uncertain for an agent a if he neither knows φ nor knows φ is false. Lastly, we also introduce a contingency operator, $\star \varphi := \neg \Box \varphi \land \neg \Box \neg \varphi$.

The action φ ! is a public announcement of φ (by any agent), and the action φ ? is an instruction to test for the truth of φ . As is usual, this language is interpreted on Kripke models.

Definition 2.4 (Model). A *model* is a structure $\mathfrak{M} = \langle W, w^*, \{\sim_a : a \in \mathcal{A}\}, V, v, n\rangle$ where

- *W* is a set, with $w^* \in W$ a designated point (representing the actual world).
- {~_a: a ∈ A} is a family of equivalence relations on W, one for each member of A. The relation w ~_a w' is interpreted as 'w and w' are epistemically equivalent for agent a'; that is, a cannot distinguish between w and w'.

- V : Φ₀ → 2^W is a valuation function associating atomic propositions with subsets of W.
- ν : Ξ → Φ₀ is an assignment of propositions to the variables in Ξ. The notion of an *x*-variant of ν is the standard one.
- $n : N \times W \to \Phi$ is a naming function associating names and propositions. If $n(a, w) = \varphi$, then we say that "a is a name of φ at w".

We designate the class of models by \mathcal{M} . The semantics for the dynamic modalities are defined in terms of two types of model modification, which require some preliminary definitions.

Definition 2.5. The *truth set* of $\varphi \in \Phi$ in a model \mathfrak{M} is $\llbracket \varphi \rrbracket^{\mathfrak{M}} = \{w : \mathfrak{M}, w \models \varphi\}$.

Definition 2.6 (Model reduction). The *reduction* of a model \mathfrak{M} by a formula $\varphi, \mathfrak{M} \upharpoonright \varphi = \langle W^{\mathfrak{M},\varphi}, w^*, \{\sim_a^{\mathfrak{M},\varphi} : a \in \mathcal{A}\}, V^{\mathfrak{M},\varphi}, v^{\mathfrak{M},\varphi}, n^{\mathfrak{M},\varphi} \rangle$, where $W^{\mathfrak{M},\varphi} := \{w \in W : \mathfrak{M}, w \models \varphi\}$, the actual world remains unchanged, and the relations and valuation functions are just restrictions of the originals.

Definition 2.7 (Uncertainty model). Given a model $\mathfrak{M} = \langle W, w^*, \{\sim_a : a \in A\}, V\rangle$, define the class of *a*-uncertainty models for $\varphi \mathcal{M}^{U_a \varphi}$ as follows:

If $\mathfrak{M}, w^* \models U_a \varphi$, then $\mathcal{M}^{U_a \varphi} = \{\mathfrak{M}\}.$

Otherwise, $\mathcal{M}^{U_a\varphi} = \{\mathfrak{M}^{U_a\varphi} = \langle W', w'^*, \{\sim_a' : a \in A\}, V' \rangle\}$ such that:

- $W' = W \cup \{v\}$, where $v \notin W$.
- $w'^* = w^*$.
- \sim'_a is the reflexive, transitive, and symmetric closure of $\sim_a \cup \langle w'^*, v \rangle$.
- For $a' \in A$, if $a' \neq a$, then $\sim'_{a'} = \sim_{a'}$.
- V' is any valuation function minimally extending V in such a way that
 - If $\mathfrak{M}, w^* \models K_a \varphi$, then $v \in \llbracket \neg \varphi \rrbracket$,
 - If $\mathfrak{M}, w^* \models K_a \neg \varphi$, then $v \in \llbracket \varphi \rrbracket$,

and for all other ψ , $v \in \llbracket \psi \rrbracket$ iff $w^* \in \llbracket \psi \rrbracket$.

Definition 2.8 (Model revision). Given a selection function $f : \mathcal{M} \times A \times \Phi \to \mathcal{M}$ such that $f(\mathfrak{M}, a, \varphi) \in \mathcal{M}^{U_a \varphi}$, the *f*-revision of a model \mathfrak{M} by a formula φ for an agent a, written $\mathfrak{M} \uparrow_f \varphi$, is defined as follows:

If φ is of the form $U_a\psi$: $\mathfrak{M}\uparrow_f \varphi = f(\mathfrak{M}, a, \psi)$ Otherwise: $\mathfrak{M}\uparrow_f \varphi = \mathfrak{M}\restriction \varphi$ **Definition 2.9** (Semantics). For a model \mathfrak{M} with valuation ν , the truth of a formula at a world $w \in W$ is defined recursively as follows:

| $\mathfrak{M}, w \models p$ | iff | $w \in V(p)$ |
|--|-----|---|
| $\mathfrak{M}, w \models x$ | iff | $\mathfrak{M}, w \models \nu(x)$ |
| $\mathfrak{M}, w \models x = y$ | iff | v(x) = v(y) |
| $\mathfrak{M}, w \models x = p$ | iff | v(x) = p |
| $\mathfrak{M}, w \models \exists x \varphi(x)$ | iff | there is an <i>x</i> -variant v' of v s.t. $\mathfrak{M}, v', w \models \varphi[v'(x)/x]$ |
| $\mathfrak{M}, w \models Tp$ | iff | $w \in V(p)$ |
| $\mathfrak{M}, w \models Tx$ | iff | $w \in V(v(x))$ |
| $\mathfrak{M}, w \models \Box \varphi$ | iff | $\forall w \in W, \mathfrak{M}, w \models \varphi$ |
| $\mathfrak{M}, w \models K_a \varphi$ | iff | $\forall w'(\langle w, w' \rangle \in \sim_a \text{ implies } \mathfrak{M}, w' \models \varphi)$ |
| $\mathfrak{M}, w \models \hat{a}$ | iff | $\mathfrak{M}, w \models \varphi \text{ and } n(a, w) = \varphi$ |
| $\mathfrak{M}, w \models x = \hat{a}$ | iff | v(x) = n(a, w) |
| $\mathfrak{M}, w \models \hat{a} = \varphi$ | iff | $n(a,w) = \varphi$ |
| $\mathfrak{M}, w \models [\varphi!]\psi$ | iff | $\mathfrak{M}, w \models \varphi \text{ implies } \mathfrak{M} \upharpoonright \varphi, w \models \psi$ |
| $\mathfrak{M}, w \models [\varphi?]\psi$ | iff | $\forall v \in \mathfrak{M} \uparrow \varphi, v \models \psi$ |
| $\mathfrak{M}, w \models \downarrow \varphi(\hat{a} \downarrow)$ | iff | $\mathfrak{M}, w \models \varphi[n(a, w)/\hat{a}\downarrow]$ |

where $\varphi[\nu(x)/x]$ is the result of replacing all occurrences of x in φ with $\nu(x)$, and where $\varphi[n(a, w)/\hat{a}\downarrow]$ is the result of replacing all occurrences of $\hat{a}\downarrow$ in φ with n(a, w).

As we discuss in (Uckelman 2013b), this language is richly expressive. Using it, we can express various types of scope distinctions (e.g., the difference between 'Of A you know that it is φ ' and 'You know that A is φ ' as well as talk about both propositions and names of propositions.

2.3 *Obligationes* in this framework

In this section we formally define *obligationes* disputations. We first specify the possible actions of R, which are defined using the dynamic test operator defined in the previous section:

Definition 2.10. (Actions of R) Let φ_n be a proposition put forward by O. The possible actions of R (designated Act) are:

| concede : | $[\varphi_n?]	op$ |
|-----------|-------------------------|
| deny : | $[\neg \varphi_n]?\top$ |
| doubt : | $[U_{R}\varphi_n?]$ T |

These are essentially tests for consistency; if R follows the rules in his responses, and does not concede contradictory statements, then he will always remain within a non-empty model. Note that as defined, doubting φ is the same as conceding that you are uncertain about φ . Thus, Paul's assertion that *dubitatio* is a variant of *positio* (Venice 1988, p. 39) is borne out.⁹

Definition 2.11. (*Obligatio*). An *obligatio* is a quadruple $\mathfrak{D} = \langle \Theta, R, \Gamma, \Gamma^R \rangle$ where

- Θ is a sequence of propositions, such that $\theta_0 \in \Theta$ is the *obligatum* and $\theta_n \in \Theta$ is the proposition put forward by O at round *n*.
- *R*: Θ × ℕ → 2^{Act} is an obligational rule. The intended interpretation of *R* is that for any statement proposed by O, the rule gives a set of correct responses for R. We often write *R*(φ_n) for *R*(φ, *n*) to simplify notation.
- Γ is a sequence of actions, formed by R's actual responses to each element of Θ .
- Γ^R is a sequence of actions, formed by the *correct* response of R to each element in Θ , as given by *R*.

The set of *obligationes* is denoted by O. We abuse notation and identify Γ and Γ^R with the formulas inside [], that is, if $\Gamma_1 = \langle [\theta_0?] \top, [\neg \theta_1?] \top \rangle$, we identify Γ_1 with $\langle \theta_0, \neg \theta_1 \rangle$. Further, for a set of ordered propositions Γ_n , let $\mathfrak{M} \uparrow \Gamma_n = \mathfrak{M} \uparrow \gamma_0 \uparrow \ldots \uparrow \gamma_n$, that is, $\mathfrak{M} \uparrow \Gamma_n$ is the result of the sequential reduction of \mathfrak{M} by the elements of Γ_n .

Different types of *obligationes* are modeled by changing *R*. Paul's rules for *positio* are defined as follows:

Definition 2.12. For a model \mathfrak{M} and *positum* $\theta_0 \in \Theta$:

| $R(\theta_{\alpha}) =$ | concede : | $\inf\mathfrak{M},w \models \langle \theta_0? \rangle \top$ |
|-----------------------------|-----------|---|
| $\mathbf{K}(\sigma_0) = \{$ | deny : | iff $\mathfrak{M}, w \models [\theta_0?] \perp$ |

For $\theta_n \in \Theta$, n > 0:

If $\mathfrak{M} \uparrow \Gamma_{n-1} \models \theta_n$: $R(\theta_n) =$ concede : If $\mathfrak{M} \uparrow \Gamma_{n-1} \models \neg \theta_n$: $R(\theta_n) =$ deny : Otherwise: If $\mathfrak{M}, w^* \models K_{\mathsf{R}}\theta_n$: $R(\theta_n) =$ concede : If $\mathfrak{M}, w^* \models K_{\mathsf{R}} \neg \theta_n$: $R(\theta_n) =$ deny : If $\mathfrak{M}, w^* \models \neg (K_{\mathsf{R}}\theta \lor K_{\mathsf{R}} \neg \theta_n)$: $R(\theta_n) =$ doubt :

⁹Note that this is not generally true of *dubitatio*; see (Uckelman 2011a, Uckelman et al. forthcoming).

Two things to note about Paul's rules. First, they are deterministic, which means that the sequence Γ^R is uniquely defined. Second, they are in a sense global or universal: for any formula φ_n which is conceded, either it was true in all the remaining worlds in $\mathfrak{M} \uparrow \Gamma_n$ or it was known in the original world, that is, it was true in all worlds accessible from the original world (and analogously for formulas which are denied).

Many *obligationes* take place against the backdrop of some explicit world-knowledge. This is done via the specification of a *casus* 'case'.

Definition 2.13 (*Casus*). Let $\text{Lit}(\Phi_0)$ be the set of literals of Φ_0 , $\text{Lit}(\mathcal{N})$ be the set of all formulas of the form $\hat{a} = \varphi$ or $\hat{a} \neq \varphi$, and $\text{Lit}^B(\Phi_0)$ be the closure of $\text{Lit}(\Phi_0)$ under the Boolean operators. A *casus* is any $C \subseteq \text{Lit}(\Phi_0) \cup \text{Lit}^B(\mathcal{N})$.

Definition 2.14. \mathfrak{M} models the casus if there is a partition $P = P_1 \cup P_2$ of W with $w^* \in P_1$, such that:

- if $w \sim_{\mathsf{R}} w^*$ or $w \sim_{\mathsf{O}} w^*$, then $w \in P_1$.
- for all $w, v \in P_1$, $w \sim_{\mathsf{R}} v$ and $w \sim_{\mathsf{O}} v$.
- for every positive literal $p \in C$, every negative literal $\neg q \in C$, and every $w \in P_1$, $w \in V(p)$ and $w \notin V(q)$.
- for every $\psi \in C \cap \operatorname{Lit}^{B}(\mathcal{N})$ and $w \in P_{1}, \mathfrak{M}, w \models \psi$.

Corollary 1. *Fix a model* \mathfrak{M} *and a* casus *C*. *Then, for every* $\varphi \in C$ *, if* \mathfrak{M} *models C, then* $\mathfrak{M}, w^* \models K_{\mathsf{R}}\varphi$ *and* $\mathfrak{M}, w^* \models K_{\mathsf{O}}\varphi$.

Proof. Follows straightforwardly from Def. 2.14.

3 Analysis of the disputation

With all of this in place, we can now turn to the analysis of the puzzle and ensuing disputation. Here, R stands for "you", that is, the reader, who plays the role of the Respondent in the disputation below. We let p := 'God exists' and q := 'A human being is a donkey', and then the assumptions are:

- (1a) $K_{\mathsf{R}}(\hat{a} = p \lor \hat{a} = q)$.
- (1b) $\forall x(\hat{a} = x \rightarrow (x = p \lor x = q)).$
- (1c) $\neg K_{\mathsf{R}}(\hat{a}=p) \land \neg K_{\mathsf{R}}(\hat{a}=q) \land \neg K_{\mathsf{R}}(\hat{a}\neq p) \land \neg K_{\mathsf{R}}(\hat{a}\neq q).$
- (1d) $K_{\mathsf{R}} \Box p \land K_{\mathsf{R}} \Box \neg q$.

(2) $\forall x((\neg K_{\mathsf{R}}x \land \neg K_{\mathsf{R}}\neg x) \rightarrow U_{\mathsf{R}}x).$

These assumptions set up the background against which the disputation is to be evaluated—that is, this is the presentation of a *casus*. (2) is simply the definition of uncertainty from S2.2. (1b) is a constraint on allowable n functions, requiring that each name is the name of a unique proposition. This constraint is already built in to our models (cf. Uckelman 2013b, fn. 12).

Thus, we must specify a *casus* which accounts for (1a), (1c), and (1d).

Proposition 1. Let $C = \{\hat{a} = p \lor \hat{a} = q, p, \neg q\}$, and let \mathfrak{M} model C. By Def. 2.14, it follows that $\mathfrak{M}, w^* \models \Box p$ and $\mathfrak{M}, w^* \models \Box \neg q$. Hence, by Cor. 1, $\mathfrak{M}, w^* \models K_{\mathsf{R}}(\hat{a} = p \lor \hat{a} = q)$ and $\mathfrak{M}, w^* \models K_{\mathsf{R}} \Box p \land K_{\mathsf{R}} \Box \neg q$.

This takes care of (1a) and (1d).

Proposition 2. Let $C = \{\hat{a} = p \lor \hat{a} = q, p, \neg q\}$, let \mathfrak{M} model C, and fix a suitable selection function f. Then, $\mathfrak{M}' = \mathfrak{M} \uparrow_f (\hat{a} = p \land \hat{a} = q)$ is such that $\mathfrak{M}', w^* \models \neg K_{\mathsf{R}}(\hat{a} = p) \land \neg K_{\mathsf{R}}(\hat{a} = q) \land \neg K_{\mathsf{R}}(\hat{a} \neq p) \land \neg K_{\mathsf{R}}(\hat{a} \neq q)$.

Proof. First, note that given the definition of U_{R} , (1c) is equivalent to $U_{\mathsf{R}}(\hat{a} = p) \land U_{\mathsf{R}}(\hat{a} = q)$, which in turn is equivalent to $U_{\mathsf{R}}(\hat{a} = p \land \hat{a} = q)$. There are two cases: (1) $\mathfrak{M}, w^* \models U_{\mathsf{R}}(\hat{a} = p \land \hat{a} = q)$. In this case, by Def. 2.8, $\mathfrak{M} = \mathfrak{M}'$, and we are done. (2) $\mathfrak{M}, w^* \nvDash U_{\mathsf{R}}(\hat{a} = p \land \hat{a} = q)$. There are a further four possibilities: (i) $\mathfrak{M}, w^* \models K_{\mathsf{R}}\hat{a} = p$; (ii) $\mathfrak{M}, w^* \models K_{\mathsf{R}}\hat{a} = p$; (iii) $\mathfrak{M}, w^* \models K_{\mathsf{R}}\hat{a} = q$; (iii) $\mathfrak{M}, w^* \models K_{\mathsf{R}}\hat{a} \neq p$; (iv) $\mathfrak{M}, w^* \models K_{\mathsf{R}}\hat{a} \neq q$, each of which are proven analogously, so we take case (i) only. Since \sim_{R} is reflexive, it follows that $n(a, w^*) = p$. By Def. 2.7, every $\mathfrak{M}' \in \mathcal{M}^{U_{\mathsf{R}}(\hat{a} = p \land \hat{a} = q)}$ is such that there is a $v \sim_{\mathsf{R}} w^*$ where $v \in [\![\hat{a} \neq p]\!]$; in fact, given that \mathfrak{M} models *C* and we have chosen a suitable *f*, we know that n(a, v) = q, and hence $v \in [\![\hat{a} = q]\!]$. Because $v \sim_{\mathsf{R}} w^*, \mathfrak{M}, w^* \models \neg K_{\mathsf{R}}\hat{a} = p$, as required.

The model given in Fig. 1 satisfies all the assumptions (1)–(2). (In fact—though we do not prove this here—it is a minimal model satisfying all the assumptions). Here, $\mathfrak{M} = \mathfrak{M} \uparrow_f (\hat{a} = p \land \hat{a} = q)$, because $\hat{a} = p$ and $\hat{a} = q$ are already uncertain for R, because he cannot distinguish between the actual world w^* where $\hat{a} = p$ and world w', where $\hat{a} = q$, and thus per Def. 2.7, $\mathcal{M}^{U_a \varphi} = \{\mathfrak{M}\}$.

Once Paul has set up the assumptions and the argument, he provides an *obligatio*like piece of reasoning. He says:

You could work out from this how to reply in the case under discussion. For example:



Figure 1

| 1 | A is true. | I am uncertain. | | | |
|----|--|-----------------|--|--|--|
| 2 | A is false. | I am uncertain. | | | |
| 3 | A is contingent. I deny it. | | | | |
| 4 | A is possible. | I am uncertain. | | | |
| 5 | A is necessary. | I am uncertain. | | | |
| 6 | A is impossible. | I am uncertain. | | | |
| 7 | You know that A is true. | I deny it. | | | |
| 8 | You know that A is false. | I deny it. | | | |
| 9 | A is uncertain to you. | I deny it. | | | |
| 10 | A is known by you. | I am uncertain. | | | |
| 11 | You know that A is known by you. | I deny it. | | | |
| 12 | You are uncertain that <i>A</i> is known by you. | I grant it. | | | |
| 13 | You know that A is uncertain to you. | I deny it. | | | |
| 14 | Of A you know that it is known by you. | I am uncertain. | | | |
| 15 | You are uncertain that A is true. | I grant it. | | | |
| 16 | Of A I am uncertain that it is true. | I deny it. | | | |
| 17 | About A I am uncertain. | I deny it. | | | |
| 18 | You are uncertain about A | I deny it. | | | |
| 19 | Of something true you know that it is A. | I deny it. | | | |
| 20 | Of A you know that it is something. | I grant it. | | | |
| 21 | You know that A is A. | I grant it. | | | |
| 22 | Of A you know that it is A. | I deny it. | | | |
| 23 | Of A you know that it is A | | | | |
| | or something other than A. | I grant it. | | | |
| 24 | Of A you know that it is something | | | | |
| | other than A. | I deny it. | | | |
| 25 | Of A you know that it is true or false. | I grant it. | | | |
| | | | | | |

| 26 | You know that A is necessary or impossible. | I grant it. |
|----|--|-----------------|
| 27 | Of A you know that it is possible | |
| | or contingent. | I am uncertain. |
| 28 | Of A you know that it is impossible | |
| | or contingent. | I am uncertain. |
| 29 | You know that A is possible or contingent. | I deny it. |
| 30 | You know that A is contingent or impossible. | I deny it. |
| | | |

And so on for innumerable other propositions, some of which should be granted, some of which should be denied, and about some of which one should say that one is uncertain, if one wants to examine the matter carefully¹⁰ (Venice 1981, pp. 19, 21).

(The numbers are my addition to make reference easier.) While Paul does not say explicitly that this is an obligational disputation, nor make explicit the rules involved, it is clear from the presentation—of propositions which the reader is responding to either by conceded, denying, or doubting—and from the initial assumptions that it is one, about the nature of \hat{a} . Thus, starting from an initial model where R concedes that he is uncertain about whether $\hat{a} = p$ or $\hat{a} = q$, if he correctly follows the rules, he will never be forced into admitting that he knows \hat{a} (cf. Uckelman 2011a, Theorem 24).

We now explore how this disputation plays out in the formalization we have defined. We begin with the statements of O, which make up the sequence Θ :

$$\Theta_0 = \langle \theta_0 \rangle = \langle \neg K_{\mathsf{R}} \hat{a} \rangle$$

| $\begin{array}{llllllllllllllllllllllllllllllllllll$ | Θ_{1-10} | = | $\langle T\hat{a}, F\hat{a}, \downarrow \star \hat{a} \downarrow, \downarrow \Diamond \hat{a} \downarrow, \downarrow \Box \hat{a} \downarrow, \downarrow \Box \neg \hat{a} \downarrow, K_{R}T\hat{a}, K_{R}F\hat{a}, \downarrow U_{R}\hat{a} \downarrow, \downarrow K_{R}\hat{a} \downarrow \rangle$ |
|---|------------------|---|--|
| $\begin{array}{llllllllllllllllllllllllllllllllllll$ | Θ_{11-17} | = | $\langle K_{R} \downarrow K_{R} \hat{a} \downarrow, U_{R} \downarrow K_{R} \hat{a} \downarrow, K_{R} \downarrow U_{R} \hat{a} \downarrow, \downarrow K_{R} \downarrow K_{R} \hat{a} \downarrow, U_{R} T \hat{a}, \downarrow U_{O} T \hat{a} \downarrow, \downarrow U_{O} \hat{a} \downarrow \rangle$ |
| $\begin{array}{lll} \Theta_{23-25} &= & \langle \downarrow K_{R}(\hat{a} \downarrow = \hat{a} \lor \hat{a} \downarrow \neq \hat{a}), \downarrow K_{R}(\hat{a} \downarrow \neq \hat{a}), \downarrow K_{R}(T\hat{a} \downarrow \lor F\hat{a} \downarrow) \rangle \\ \Theta_{26-27} &= & \langle K_{R}(\Box \hat{a} \lor \Box \neg \hat{a}), \downarrow K_{R}(\Diamond \hat{a} \downarrow \lor (\Diamond \hat{a} \downarrow \land \Diamond \neg \hat{a} \downarrow)) \rangle \\ \Theta_{28-29} &= & \langle \downarrow K_{R}(\Box \neg \hat{a} \downarrow \lor (\Diamond \hat{a} \downarrow \land \Diamond \neg \hat{a} \downarrow)), K_{R}(\Diamond \hat{a} \lor (\Diamond \hat{a} \land \Diamond \neg \hat{a})) \rangle \\ \Theta_{30} &= & \langle K_{R}((\Diamond \hat{a} \land \Diamond \neg \hat{a}) \lor \Box \neg \hat{a} \rangle \end{array}$ | Θ_{18-22} | = | $\langle U_{R}\hat{a}, \exists x(Tx \land K_{R}(x=\hat{a})), \downarrow K_{R}(\exists x=\hat{a}\downarrow), K_{R}(\hat{a}=\hat{a}), \downarrow K_{R}(\hat{a}\downarrow=\hat{a})\rangle$ |
| $\begin{array}{llllllllllllllllllllllllllllllllllll$ | Θ_{23-25} | = | $\langle \downarrow K_{R}(\hat{a} \downarrow = \hat{a} \lor \hat{a} \downarrow \neq \hat{a}), \downarrow K_{R}(\hat{a} \downarrow \neq \hat{a}), \downarrow K_{R}(T\hat{a} \downarrow \lor F\hat{a} \downarrow) \rangle$ |
| $\begin{array}{lll} \Theta_{28-29} &= & \langle \downarrow K_{R}(\Box \neg \hat{a} \downarrow \lor (\diamond \hat{a} \downarrow \land \diamond \neg \hat{a} \downarrow)), K_{R}(\diamond \hat{a} \lor (\diamond \hat{a} \land \diamond \neg \hat{a})) \rangle \\ \Theta_{30} &= & \langle K_{R}((\diamond \hat{a} \land \diamond \neg \hat{a}) \lor \Box \neg \hat{a} \rangle \end{array}$ | Θ_{26-27} | = | $\langle K_{R}(\Box \hat{a} \lor \Box \neg \hat{a}), \downarrow K_{R}(\Diamond \hat{a} \downarrow \lor (\Diamond \hat{a} \downarrow \land \Diamond \neg \hat{a} \downarrow)) \rangle$ |
| $\Theta_{30} = \langle K_{R}((\Diamond \hat{a} \land \Diamond \neg \hat{a}) \lor \Box \neg \hat{a} \rangle$ | Θ_{28-29} | = | $ \langle \downarrow K_{R}(\Box \neg \hat{a} \downarrow \lor (\Diamond \hat{a} \downarrow \land \Diamond \neg \hat{a} \downarrow)), K_{R}(\Diamond \hat{a} \lor (\Diamond \hat{a} \land \Diamond \neg \hat{a})) \rangle $ |
| | Θ_{30} | = | $\langle K_{R}((\Diamond \hat{a} \land \Diamond \neg \hat{a}) \lor \Box \neg \hat{a} \rangle$ |

¹⁰Due to reasons of space, I will not quote the original Latin here.

In addition to Θ , Paul also specifies Γ , the sequence of correct responses of R to the statements of O in Θ :

$$\begin{split} \Gamma_{1-30} &= \langle \textbf{doubt} : \theta_1, \textbf{doubt} : \theta_2, \textbf{deny} : \theta_3, \textbf{doubt} : \theta_4, \textbf{doubt} : \theta_5, \\ \textbf{doubt} : \theta_6, \textbf{deny} : \theta_7, \textbf{deny} : \theta_8, \textbf{deny} : \theta_9, \textbf{doubt} : \theta_{10}, \\ \textbf{deny} : \theta_{11}, \textbf{concede} : \theta_{12}, \textbf{deny} : \theta_{13}, \textbf{doubt} : \theta_{14}, \textbf{concede} : \theta_{15}, \\ \textbf{deny} : \theta_{16}, \textbf{deny} : \theta_{17}, \textbf{deny} : \theta_{18}, \textbf{deny} : \theta_{19}, \textbf{concede} : \theta_{20}, \\ \textbf{concede} : \theta_{21}, \textbf{deny} : \theta_{22}, \textbf{concede} : \theta_{23}, \textbf{deny} : \theta_{24}, \textbf{concede} : \theta_{25}, \\ \textbf{concede} : \theta_{26}, \textbf{doubt} : \theta_{27}, \textbf{doubt} : \theta_{28}, \textbf{deny} : \theta_{29}, \textbf{deny} : \theta_{30} \rangle \end{split}$$

We now show that Paul's Γ is correct; that is, R will not be led into error following his advice. We proceed through the first fifteen actions sequentially, omitting reference to \mathfrak{M} when context makes it clear.

The initial statement $\theta_0 = \neg K_R \hat{a}$ is satisfied at w^* in Fig. 1; for if it were the case that $w^* \models K_R \hat{a}$, then it would be the case that both $w^* \models \hat{a}$ and $w' \models \hat{a}$. But n(a, w') = q, and $w' \models \neg q$. Thus, $\mathfrak{M} \uparrow \Gamma_0 = \mathfrak{M}$. Thus, R can admit the initial thesis and the disputation begins.

- 1. Paul says that R should doubt $T\hat{a}$: If R is following the rules, this means that $\neg(K_{\mathsf{R}}T\hat{a} \lor K_{\mathsf{R}}\neg T\hat{a})$ should be true at w^* . We know from the preceding that $w^* \models \neg K_{\mathsf{R}}\hat{a}$, so it follows that $w^* \models \neg K_{\mathsf{R}}T\hat{a}$. A similar argument can be given for $w^* \models \neg K_{\mathsf{R}}\neg T\hat{a}$, and hence, by DeMorgan's, $w^* \models \neg(K_{\mathsf{R}}T\hat{a} \lor K_{\mathsf{R}}\neg T\hat{a})$. Therefore, R's doubting θ_1 is correct and $\mathfrak{M} \uparrow \Gamma_1 = \mathfrak{M}$.
- 2. Paul says that R should doubt $F\hat{a}$ as well: An analogous argument to the previous shows that this is correct, and $\mathfrak{M} \uparrow \Gamma_2 = \mathfrak{M}$.
- R should deny ↓ ★ â↓. There are two cases where this is correct: If M ↑ Γ₂ ⊨ ¬↓ ★ â↓ or if w* ⊨ K_R¬↓ ★ â↓. Both condition are satisfied; w* ⊨ ◊p, but w* ⊭ ◊¬p, while the reverse is true at w' with respect to q.
- 4. R should doubt $\downarrow \Diamond \hat{a} \downarrow$: This is correct if $w^* \models \neg (K_R \downarrow \Diamond \hat{a} \downarrow \lor K_R \neg \downarrow \Diamond \hat{a} \downarrow)$. It is easy to check the relevant witness worlds are w' for the first disjunct and w^* for the second, so R is correct in doubting θ_4 , and $\mathfrak{M} \uparrow \Gamma_4 = \mathfrak{M}$.
- 5. R should doubt $\downarrow \Box \hat{a} \downarrow$: This is correct if $\mathfrak{M}, w^* \models \neg (K_R \downarrow \Box \hat{a} \downarrow \lor \neg K_R \neg \downarrow \Box \hat{a} \downarrow)$; as in the previous case, the witness worlds are w' for the first disjunct and w* for the second, so R is correct in doubting θ_5 , and $\mathfrak{M} \uparrow \Gamma_5 = \mathfrak{M}$.
- R should doubt ↓□¬â↓: Completely analogous to the previous; R is correct in doubting θ₆, and M ↑ Γ₆ = M.

- 7. R should deny $K_{\mathsf{R}}T\hat{a}$: This is the case if either $\mathfrak{M} \uparrow \Gamma_6 \models \neg K_{\mathsf{R}}T\hat{a}$ or $w^* \models K_{\mathsf{R}}\neg K_{\mathsf{R}}T\hat{a}$. $\mathfrak{M} \uparrow \Gamma_6 \models \neg K_{\mathsf{R}}T\hat{a}$ iff $\mathfrak{M} \uparrow \Gamma_6, w^* \nvDash K_{\mathsf{R}}T\hat{a}$ and $\mathfrak{M} \uparrow \Gamma_6, w' \nvDash K_{\mathsf{R}}T\hat{a}$. Both conjuncts are true because $w^* \sim_{\mathsf{R}} w'$ and $w' \sim_{\mathsf{R}} w'$, and $w' \nvDash T\hat{a}$, since n(a, w') = q and $w' \notin V(q)$. The second conjunct is true because $w^* \sim_{\mathsf{R}} w^*$. R has responded correctly, and $\mathfrak{M} \uparrow \Gamma_7 = \mathfrak{M}$.
- R should deny K_RFâ: The argument is parallel to the previous. R has responded correctly, and M ↑ Γ₈ = M.
- 9. R should deny ↓U_Râ↓: This is the case if either M ↑ Γ₈ ⊨ ¬↓U_Râ↓ or w* ⊨ K_R¬↓U_Râ↓. The former holds if ¬↓U_Râ↓ holds at both worlds. w* ⊨ ¬↓U_Râ↓ iff w* ⊨ ¬(¬↓K_Râ↓ ∧ ¬↓K_R¬â↓), that is, iff w* ⊨ ↓K_Râ↓ ∨ ↓K_R¬â↓. w* ⊨ ↓K_Râ↓ iff w* ⊨ K_Rp iff w* ⊨ p and w' ⊨ p, which is the case. A straightforward argument shows that the opposite disjunct is true for w' because of q. R has responded correctly, and M ↑ Γ₉ = M.
- 10. R should doubt $\downarrow K_{\mathsf{R}}\hat{a}\downarrow$: This is true if $w^* \models \neg(K_{\mathsf{R}}\downarrow K_{\mathsf{R}}\hat{a}\downarrow \lor K_{\mathsf{R}}\neg\downarrow K_{\mathsf{R}}\hat{a}\downarrow)$, i.e., if $w^* \models \neg K_{\mathsf{R}}\downarrow K_{\mathsf{R}}\hat{a}\downarrow \land \neg K_{\mathsf{R}}\neg\downarrow K_{\mathsf{R}}\hat{a}\downarrow$.

 $w^* \models \neg K_{\mathsf{R}} \downarrow K_{\mathsf{R}} \hat{a} \downarrow$ iff either $w^* \models \neg \downarrow K_{\mathsf{R}} \hat{a} \downarrow$ or $w' \models \neg \downarrow K_{\mathsf{R}} \hat{a} \downarrow$, that is, either (1) $w^* \nvDash \downarrow K_{\mathsf{R}} \hat{a} \downarrow$ or (2) $w' \nvDash \downarrow K_{\mathsf{R}} \hat{a} \downarrow$. This means either (1') $w^* \nvDash K_{\mathsf{R}} p$ or (2') $w' \nvDash K_{\mathsf{R}} q$. But (2') is the case.

 $w^* \models \neg K_{\mathsf{R}} \neg \downarrow K_{\mathsf{R}} \hat{a} \downarrow$ iff either $w^* \models \downarrow K_{\mathsf{R}} \hat{a} \downarrow$ or $w' \models \downarrow K_{\mathsf{R}} \hat{a} \downarrow$. But $w^* \models K_{\mathsf{R}} p$, so the first disjunct is satisfied.

So R has responded correctly, and $\mathfrak{M} \uparrow \Gamma_{10} = \mathfrak{M}$.

- 11. R should deny $K_{\mathsf{R}} \downarrow K_{\mathsf{R}} \hat{a} \downarrow$: This is the case if either $\mathfrak{M} \uparrow \Gamma_{10} \models \neg K_{\mathsf{R}} \downarrow K_{\mathsf{R}} \hat{a} \downarrow$ or $w^* \models K_{\mathsf{R}} \neg K_{\mathsf{R}} \downarrow K_{\mathsf{R}} \hat{a} \downarrow$. The first disjunct holds since $w' \models \neg q$, and hence $w' \models \neg K_{\mathsf{R}} q$, that is, $w' \models \neg \downarrow K_{\mathsf{R}} \hat{a} \downarrow$, and thus this formula cannot be known by R at either w^* or w'. R has responded correctly, and $\mathfrak{M} \uparrow \Gamma_{11} = \mathfrak{M}$.
- 12. R should concede $U_R \downarrow K_R \hat{a} \downarrow$. This is the case if either $\mathfrak{M} \uparrow \Gamma_{11} \models U_R \downarrow K_R \hat{a} \downarrow$ or $w^* \models K_R U_R \downarrow K_R \hat{a} \downarrow$. In (10), R doubted $\downarrow K_R \hat{a} \downarrow$, the result of which is that $\mathfrak{M} \uparrow \Gamma_{10} \models U_R \downarrow K_R \hat{a} \downarrow$. But $\mathfrak{M} \uparrow \Gamma_{10} = \mathfrak{M} \uparrow \Gamma_{11}$, and hence the first disjunct holds.
- 13. R should deny $K_{\mathsf{R}} \downarrow U_{\mathsf{R}} \hat{a} \downarrow$: This is the case if either $\mathfrak{M} \uparrow \Gamma_{12} \models \neg K_{\mathsf{R}} \downarrow U_{\mathsf{R}} \hat{a} \downarrow$ or $w^* \models K_{\mathsf{R}} \neg K_{\mathsf{R}} \downarrow U_{\mathsf{R}} \hat{a} \downarrow$. The first disjunct holds; $w^* \nvDash K_{\mathsf{R}} \downarrow U_{\mathsf{R}} \hat{a} \downarrow$, since $w^* \models K_{\mathsf{R}} p$, and hence $w^* \models \downarrow K_{\mathsf{R}} \hat{a} \downarrow$, from which it follows that $w^* \models \neg \downarrow U_{\mathsf{R}} \hat{a} \downarrow$. So R has responded correctly, and $\mathfrak{M} \uparrow \Gamma_{13} = \mathfrak{M}$.

- 14. R should doubt $\downarrow K_R \downarrow K_R \hat{a} \downarrow$: This is the case if $w^* \models \neg (K_R \downarrow K_R \downarrow K_R \hat{a} \downarrow \lor K_R \neg \downarrow K_R \downarrow K_R \hat{a} \downarrow)$, that is, $w^* \models \neg K_R \downarrow K_R \hat{a} \downarrow \land \neg K_R \neg \downarrow K_R \downarrow K_R \hat{a} \downarrow$. It is easy to check that w' is a witness for the first conjunct, since n(a, w') = q and $w' \models \neg K_R q$, and w^* is a witness for the second conjunct, because $n(a, w^*) = p$. So R has responded correctly, and $\mathfrak{M} \uparrow \Gamma_{14} = \mathfrak{M}$.
- 15. R should concede $U_{\mathsf{R}}T\hat{a}$: This is the case if $\mathfrak{M} \uparrow \Gamma_{14} \models U_{\mathsf{R}}T\hat{a}$ or $w^* \models K_{\mathsf{R}}U_{\mathsf{R}}T\hat{a}$. In step (1), R doubted $T\hat{a}$, which means that $\mathfrak{M} \uparrow \Gamma_1 \models U_{\mathsf{R}}T\hat{a}$; and $\mathfrak{M} \uparrow \Gamma_1 = \mathfrak{M} \uparrow \Gamma_{14}$. So R has responded correctly, and $\mathfrak{M} \uparrow \Gamma_{15} = \mathfrak{M}$.

We leave the remaining actions as interesting exercises for the reader; they can all be worked out in similar fashion, though the level of complexity increases.

4 Observations

A number of interesting observations arise from the preceding, which we touch on by way of concluding. We begin with relatively specific, precise points, and end with conclusions of more general import.

First, consider the two sentences after where we left off, θ_{16} and θ_{17} :

Of A I am uncertain that it is true.
$$(\theta_{16})$$

About A I am uncertain.
$$(\theta_{17})$$

Historians writing on *obligationes* often point out the fact that O plays little substantive role as one of the puzzling properties of these disputations. There are no rules that govern O's behavior, and he seems to do little more than rotely spit out sentences. What is disputational about such exchanges? There is no substantive argument being made (cf., e.g., King 1991 and Uckelman 2013a, S3). These two sentences turn such an interpretation of O on its head. Here, just as "you" referred to the reader, i.e., R, "T' here can only refer to O (this role played by Paul). Thus, what Paul is asking R to do is reason about not only R's own knowledge, but also about *his*, that is, Paul or O's. Note that in no part of the discussion is it ever stated what "T' knows, only what "you" knows. The only way that R could have any knowledge about O's knowledge is by it arising from the disputational context: It is because this is set up as a disputation between two people, one of whom has just finished presenting the *casus* to the other (with the strong, albeit unstated, assumption that "T' know which sentence is on the other side of the card labeled A that I have just held up to you), that R is in any position to make the inference required to deny that O is uncertain about A or its truth value.

This strategic move—using the fact that something has been presented from the author to the reader, or O to R in a particular way to infer the presence of knowledge

Uckelman

in either participant—leads us to our second observation. Paul uses this strategic move in many different ways throughout *De scire et dubitare*. We have discussed elsewhere how the knowledge which is generated through this fits into traditional medieval approaches to epistemology, whereby knowledge is defined as "a mental grasp (notitia) of anything acquired by the most powerful demonstration" (Uckelman 2013b, pp. 3). Two aspects of this approach to epistemology are important in *De scire*: on the one hand, "mental grasp", akin to the contemporary concept of "awareness", is necessary for knowledge because one can only be properly said to know a proposition of which he is aware. One way to make someone aware of a proposition, as a precursor for knowledge, is to present a proof or demonstration of it; a simpler way is for one person simply to assert it to another, and that brings us to the second important aspect, namely that of demonstration. "Demonstration" of the most powerful type corresponds to a full-blooded proof, but medieval philosophers also allowed lesser kinds of knowledge corresponding to weaker types of demonstration, of which testimony by authority counted as one. Unfortunately, such a dynamic approach to logic is inconsistent with other principles that Paul ascribes to (specifically, a certain type of monotonicity), as we show in (Uckelman 2013b, S4.3).

Third, this example illustrates an interesting a general feature of *obligationes* which is not immediately obvious from the way the rules are presented, namely, that in a correctly-played *obligatio*, the only moves that change the model are the concessions and denials of irrelevant propositions, which can reduce the state-space. Since there were no such irrelevant sentences proposed, the model never changed as a result of R's responses that we analysed. Ordinarily, the difficulty in disputing in this fashion arises from keeping track of what is relevant and what is irrelevant, and how this changes given previous concessions and denials (on a static view of relevance such as Swyneshed's, the disputations are substantially easier). Here, the complexity comes from reasoning with multiply embedded modal operators; the models do not change, but the evaluation of formulas on these models is more complex the more complex the formula is.¹¹

Fourth, the entire excursion into the Middle Ages presented in this paper serves as yet another reminder that medieval logic, far from being dull syllogistic and irrelevant scholastic wrangling, handled interesting and complex puzzles with a sophistication that makes itself manifest when one attempts to provide a formal analysis of the puzzles. One need only look at the instructions for constructing well-formed formulas in Def. 2.3 to see that the language which we used is incredibly rich and varied; furthermore, every part of the language is required in order to be able to express the distinctions we need to express in order to formalize the argument. The very process of

¹¹Using an ordinary sense of 'complexity', not the computational notion of the same.

formalization thus brings to light not only interesting aspects of medieval puzzles, but also provides us with perhaps hitherto unexplored combinations of logical languages whose interaction the study of may prove of interest apart from anything medieval.

Acknowledgements Research for this paper was partially funded by the NWO project "Towards Logics that Model Natural Reasoning".

References

H. van Ditmarsch, W. van de Hoek, and B. Kooi. *Dynamic Epistemic Logic*, volume 337 of *Synthese Library Series*. Springer, 2007.

P. King. Medieval thought-experiments: The metamethodology of mediaeval science. In T. Horowitz and G. J. Massey, editors, *Thought Experiments in Science and Philosophy*, pages 43–64. Rowman & Littlefield, 1991.

S. L. Uckelman. Deceit and indefeasible knowledge: The case of *Dubitatio*. *Journal* of *Applied Non-Classical Logics*, 21, nos. 3/4:503–519, 2011a.

S. L. Uckelman. A dynamic epistemic logic approach to modeling *Obligationes*. In D. Grossi, S. Minica, B. Rodenhäuser, and S. Smets, editors, *LIRa Yearbook*, pages 147–172. Institute for Logic, Language & Computation, 2011b.

S. L. Uckelman. Interactive logic in the Middle Ages. *Logic and Logical Philosophy*, 21(3):439–471, 2012.

S. L. Uckelman. Medieval *Disputationes de obligationibus* as formal dialogue systems. *Argumentation*, 27(2):143–166, 2013a.

S. L. Uckelman. Paul of Venice on a puzzle about uncertainty. In submission, 2013b.

S. L. Uckelman, J. Maat, and K. Rybalko. The art of doubting in *Obligationes Parisienses*. In C. Kann, B. Löwe, C. Rode, and S. L. Uckelman, editors, *Modern Views of Medieval Logic*. Peeters, forthcoming.

P. Venice. *Pauli Veneti: Logica Magna, Prima Pars: Tractatus de Scire et Dubitare.* Oxford University Press, 1981. Ed. and trans. by Patricia Clarke.

P. Venice. *Pauli Veneti: Logica Magna, Secunda Pars: Tractatus de Obligationibus.* Oxford University Press, 1988. Ed. and trans. by E. Jennifer Ashworth.
