Logics for Dynamics of Information and Preferences: Seminar's yearbook 2008

Kurzen, L.; Velázquez-Quesada, F.R.

Publication date
2009

Document Version
Final published version

Logics for Dynamics of Information and Preferences

Seminar's yearbook 2008

The printing of this book was supported by the UP Fund of Johan van Benthem.

Cover design by Nina Gierasimczuk.

ISBN: 978–90–5776–199–7

Foreword

This collection contains selected papers that have been presented and discussed during the year 2008 at the seminar *"Logics for Dynamics of Information and Preferences"* (`http://www.illc.uva.nl/lgc/seminar/`), a regular event at the Institute for Logic, Language and Computation (ILLC) of the Universiteit van Amsterdam. It reflects the work not only of the regular participants of the seminar (PhD students at the ILLC), but also of visiting colleagues from around the world. Moreover, it contains selected papers from a closely related research community in China: the participants of the seminar *"Logical Dynamics of Information and Interaction"* (`http://staff.science.uva.nl/~johan/dynamicstest.html`) and the participants of the *"Dynamic Logic Seminar"* (`http://loriweb.org/?p=1050`), held at Tsinghua University, Beijing, in Autumn 2008 and April 2009, respectively.

We gratefully thank the authors for their contributions and their collaboration through the editing process. In particular, we thank Johan van Benthem for initiating this yearbook.

Lena Kurzen
Fernando R. Velázquez-Quesada
(eds.)

Contents

Preface

Rational agency is an important research area these days, where many disciplines meet. And these disciplines definitely include logic. But to play its role here, logic must address a number of topics beyond its traditional agenda. "Logical dynamics" as represented in this volume is about the following set of inter-connected issues:

(a) What *basic informational acts* do agents engage in? These include at least observation, but also inference, questions, and perhaps other sources. Taken together, these give our model of an agent, as an "intelligent computer".

(b) How do agents *update* their attitudes on the basis of these inputs? Agents are subjected to a constant stream of information, and this leads to continual adjustment. Describing this is a challenge to logical theory. There is the one-step dynamics of knowledge, but belief is also crucial.

(c) So far, logical theories have mostly described the information that we have about relevant situations, and how we can get more of it. But in addition to this basic phenomenon, there is another, equally fundamental one. Everything we do, and even much of what we think about, is determined by our goals, and these goals depend crucially on how we *evaluate* the relevant situations. Thus, information and evaluation go hand in hand, and are the two main driving forces of rational agency. This takes us to logics of preference, both static and dynamic.

(d) Next, how do agents form *groups* that share information and act together? What logical structures govern group decisions and social procedure?

(e) Finally, what happens as local changes in information and evaluation get repeated to form longer *temporal processes*? Procedural information and long-term goals have important properties of their own. Concrete examples where this complexity becomes concrete are strategic games, protocols for informational procedures, and learning methods "in the limit".

While these topics may look diverse, they all add to one logical picture of rational agents moving together intentionally through time. And this picture is still being painted, with many students adding strokes and retouches. The papers in this volume represent work at the ILLC, University of Amsterdam, in the long-running Dynamics Seminar `http://www.illc.uva.nl/lgc/seminar/`, that has already produced quite a few dissertations. But the volume also includes papers from congenial communities elsewhere, including our counterparts in China. A parallel seminar was run in April of this year at Tsinghua University, and you can read all about it at `http://loriweb.org/?p=1050`. The papers in this volume, many of them accepted conference submissions, testify to the liveliness of both these circles.

The editors are to be commended for their initiative in bringing these initial results to the attention of a larger audience. I hope that they will be successful, and start a broader tradition. Indeed, there are other places too where logics of rational agency are cooking, with Stanford University as one more example: `http://ai.stanford.edu/~epacuit/classes/logdyn-wkshp2009.html`

I, for one, am hoping for a Dynamics Yearbook in the future, where we can sample what this community is producing worldwide on such topics as logic and information flow, logic and games, intentions, and preference dynamics. But for now, there is just the immediate pleasure of reading a collection of remarkable papers.

Johan van Benthem
Amsterdam & Stanford

Suggestions for Strategies in Modeling the Role of Reasoning in Ensemble Coordination

Søren R. Frimodt-Møller
Institute for Philosophy, Education and the Study of Religions, University of Southern Denmark.
`srfm@ifpr.sdu.dk`

Abstract

Even though there is presumably a lot of direct communication going on in a performance situation (through facial expressions, bodily cues, the signals of a conductor, if any, etc.), there will inevitably be situations where the coordination of the musicians rests on an (often tacit) agreement on "what to do next". This agreement is in some cases only imagined (and hence the coordination happens by chance), in other cases the result of the musicians synchronously having the same expectations of what the other musicians will do, and in yet other cases the result of certain strategies being common knowledge (or a close approximation thereof) in the ensemble. This paper[1] will be an introduction to how one can model the role played by the reasoning of the individual musicians in achieving coordination of the entire group (more specifically, solving coordination problems). I move from a very simplified model in terms of a classic multi-agent system over a sketch of an analysis in terms of Michael Bacharach's use of "variable frames" in game theory to my present, very preliminary attempts of applying decision theory and theories of belief revision to the description. (The latter two may become important if we wish to describe how a composition or another set of rules for the performance can be moderated during the performance.)

1 Introduction

Consider the following situation in a music ensemble: We have three players, for the sake of desirable connotations let us denote them "the oboe", "the violin" and "the cello". They are playing a new piece of scored music that is

[1]This paper is a slightly expanded version of a paper presented at *CMMR 2008: The Genesis of Meaning in Sound and Music*, Copenhagen, May 19-23, 2008. The title of the original paper was "Modeling Coordination Problems in a Music Ensemble: Some Logical and Game Theoretical Considerations".

hence not part of their individual heritage as musicians.[2] Let us for simplicity consider 5 bars in this score, denoted bars 1-5 (although they may be thought to occur at a later occasion than the beginning of the piece). Still for simplicity, we decide that in these bars the three players each have two possible actions. An action is in this context a phrase to be played within a bar. To echo the theory of multi-agent systems as presented by Fagin et al. (2003)[3] we define the following.

Definition 1.1 (Local States in a Fictitious Ensemble)**.** Let us call the set of possible actions for a player i the set of possible local states for that player, L_i. We now define:

$L_{oboe} = \{phrase1, phrase2\}$
$L_{violin} = \{phrase3, phrase4\}$
$L_{cello} = \{phrase5, phrase6\}$

(To make the example more in accordance with reality, we could add a state Λ to each of the sets L_i, denoting that the player does not play anything. We will, however, not consider cases where such behavior is involved here, and therefore we omit these possible states. We could also have decided on a more general definition of a state to include any sort of event and subsequently added a set L_e of possible states for the environment, where we could have placed events external to the ensemble that may affect there actions, such as "a truck passes the concert hall". But due to the fairly short length of this paper we only consider the behavior of our three players in their interrelations.)

Now, according to the score, the three players are supposed to play their phrases in a rather staircase-like development: In bars 1-2, the oboe is supposed to play phrase 1, the violin phrase 3 and the cello phrase 5. In bar 3, the oboe is supposed to play phrase 2, the violin phrase 3 and the cello phrase 5. In bars 4-5 the oboe returns to playing phrase 1, but the violin plays phrase 4 in bar 4 and then returns to phrase 3 in bar 5, whereas the cello continues playing phrase 5 in bar 4 and then plays phrase 6 in bar 5. The situation is illustrated in Table 1.

	Bar 1	Bar 2	Bar 3	Bar 4	Bar 5
oboe	phrase 1	phrase 1	phrase 2	phrase 1	phrase 1
violin	phrase 3	phrase 3	phrase 3	phrase 4	phrase 3
cello	phrase 5	phrase 5	phrase 5	phrase 5	phrase 6

Table 1: An Example from a Fictious Score

Intuitively, the violin should wait for the oboe to play phrase 2 and then play phrase 4 at the following bar. The cello should wait for the violin to play phrase 4 and then play phrase 6 at the following bar. Now consider what happens,

[2] In previous conference presentations ("How Do Musicians Reach an Agreement? The Ensemble as a Multi-Agent System" at *Workshop on Deontic Logic*, Roskilde University, November 9, 2007 and participation in "Workshop on Academic Writing" at the annual graduate conference arranged by the Danish Research School in Philosophy, History of Ideas and History of Science, Sandbjerg Estate, December 7, 2007) I have described how the same sort of doubt may arise in a known piece of music, for instance bars 4 to 8 in Schubert's *Unfinished Symphony*. I have, however, found that my audience is less likely to accept that it can be problematic for skilled musicians to coordinate in such a (presumably) familiar context, therefore I have generalized the example.

[3] In which the theory is used to describe problems involving both communication and coordination such as the Problem of Coordinated Attack (see i.e. 109-122 and 190-199)

if the oboe plays phrase 1 three times in a row. The violin might either think "too bad for her, I'm proceeding to bar four anyway, or else the cello will not know what to do" or "I'd better wait for the oboe to play her phrase 2 and then interpret that bar as bar 3 and the following as my bar 4." But what will he think? This depends on how important he thinks the development in the phrasing of the oboe is in comparison with the development of his own phrasing not to mention that of the cello. Let us say that the violin chooses to pursue the second tactic, namely resume playing his phrase 3 until he hears the oboe playing phrase 2. What will the cello think? The cello might think, "The oboe and the violin have both got it wrong, but that is not my problem, I am going for the fifth bar in this development with my phrase 6 as planned, then they can adjust to what I am doing in the following bar." But she might also think "Oh, we should probably wait for the oboe to commence her phrase 2 and then continue the development as if that bar was bar 3." (She might actually also think "Never mind the oboe, I will wait for the violin to play his phrase 4 and then play phrase 6 at the next bar", but this will amount to the same line of action although the intention is different.) What she thinks depends on whether she thinks her own voice or that of the oboe (or, for completeness, that of the violin) is the most important in this section of the piece. Our troubles do not end here. The oboe might also be considering what to do next, e.g. wonder whether she should just think "Oh no, I blew it, but too bad, I just have to continue according to the score" or "the other musicians are waiting for my phrase 2, so I should play phrase 2 to get things going." As with the other two, what she chooses to do depends on how she conceives of the composition.

The score might quite probably give some clear normative guidelines as to what is the most important in the composition. (We think of the composition as something of which the score is an arrangement – in this way we are able to distinguish between different arrangements of a composition, although these arrangements will necessarily deviate in certain aspects from the original score (if any) of the composer.) But it might also be that the question of what the central parts of the composition are is a matter of interpretation, that is, relative to respective musicians.

In the following we will try to elucidate how the three musicians can navigate out of the situation through two different sorts of analysis. The first is in terms of a *multi-agent system* within epistemic logic and considers the case, where a set of guidelines being *common knowledge* in the ensemble will enable the musicians to solve their coordination problem. The second is in terms of Michael Bacharach's idea of the role of framing (Bacharach et al. 2006) in game theoretic problems and will address how musicians may be able to coordinate even though they do not have exactly the same opinion of the salient features of the composition.

2 The Coordination Problem Analyzed in Terms of a Multi-Agent System

Let us pick up our definition above of the local states of the players. Fagin et al. (2003, p.110-111) think of a state as an *information state*, that is, a state that contains information. Strictly speaking, if we want to follow this line of reasoning, we should add a number of possible local states for a player i containing information not only about what i is playing now, but also about what the other players are playing, and what everyone was playing at previous bars. In our example here, however, we assume that all players actually hear everything that happens, and that they have perfect memory. We therefore assume that everyone is always aware of what is happening at all local states, and for the sake of simplicity we choose to model the information state of player i as only containing information about the action of i at a given time.

Definition 2.1 (Global States in the Ensemble)**.** We now define a global state, $G = (s_{oboe}, s_{violin}, s_{cello})$, where s_i is the state for the player i (in our example, the phrase that the player is playing). Intuitively G expresses some situation where each of the players is playing a specific phrase from his or her respective set of possible states. We thus have a set of possible global states, $\mathcal{G}^{ensemble} = L_{oboe} \times L_{violin} \times L_{cello}$ (the Cartesian product of all the sets of possible local states).[4]

We would like to model G as a function of time. For the present purposes we think of time as being discreet and introduce a point in time m, $m \in \{0, 1 \ldots\}$. This is quite convenient because it allows us to think of steps in time as being synchronous with and equal to the length of developments from one bar to another, which is exactly what we will do.

Definition 2.2 (Runs and Systems)**.** We define *a run* to be a description of how the global state develops through time, more precisely, the global state as a function of m: $r(m) = (s_{oboe}, s_{violin}, s_{cello})$, such that r(0) is the initial global state, r(1) the next global state etc. We now define a multi-agent system $\mathcal{R}^{ensemble}$ over $\mathcal{G}^{ensemble}$ as a set of runs over $\mathcal{G}^{ensemble}$. A point (r, m) is the time-point m in the run r. We say that (r, m) is a point in the system $\mathcal{R}^{ensemble}$, if r$\in \mathcal{R}^{ensemble}$. $r_i(m) = s_i$, so that $r_i(m)$ is player i's local state at the point (r, m).

Before we can analyze our coordination problem above, we need to define what it means for a player to distinguish (or not be able to distinguish) between two global states:

Definition 2.3 (Distinguishability)**.**

- Let $s = (s_{oboe}, s_{violin}, s_{cello})$ and $s' = (s'_{oboe}, s'_{violin}, s_{cello})$ be two global states in $R^{ensemble}$. We say that player i cannot distinguish s from s′, notated $s \sim_i s'$, if player i has the same state in s and sï¿$\frac{1}{2}$, in other words if $s_i = s'_i$ (An important thing to notice here is that we think of a state not only as the action of a player but as a situation in which the player chooses the action that has given rise to the label of that state.)
- In accordance with this we say that player i cannot distinguish between two points (r, m) and (r', m), $(r, m) \sim_i (r', m)$ if $r(m) \sim_i r'(m)$, in other words if $r_i(m) = r'_i(m)$

In the epistemic logic of multi-agent systems described by Fagin et al. (2003), the notion of indistinguishability is used to define the operator K_i, which in our

[4]These formalizations are identical with the definitions given by Fagin et al. (2003, p.111-121). The following formalizations are my versions of definitions given in the same pages, only adapted to my own example.

case would intuitively mean "player i knows that ..." In this example we will not need to make statements about the players' knowledge of propositional facts, only their awareness of the global state and its relation to other global states, hence we omit the definition of the K_i-operator.[5]

With these formalities in place we can now describe the stepwise development of our coordination problem formulated as the system $\mathcal{R}^{ensemble}$. As hinted at before, we take the time variable m to be a stepwise development of one bar length. In a case where all three musicians follow the score perfectly (a specific run in $\mathcal{R}^{ensemble}$ which we choose to label r^{score}), m should therefore be perfectly synchronized with the bar numbers such that the global states develop in this way:

$$r^{score}(1) = (phrase1, phrase3, phrase5)$$
$$r^{score}(2) = (phrase1, phrase3, phrase5)$$
$$r^{score}(3) = (phrase2, phrase3, phrase5)$$
$$r^{score}(4) = (phrase1, phrase4, phrase5)$$
$$r^{score}(5) = (phrase1, phrase3, phrase6)$$

Now let us look at a case where the oboe forgets to play phrase 2 at bar 3. A number of different runs might then occur in which the first three steps would be

$$r^{late(u)}(1) = (phrase1, phrase3, phrase5)$$
$$r^{late(u)}(2) = (phrase1, phrase3, phrase5)$$
$$r^{late(u)}(3) = (phrase1, phrase3, phrase5)$$

(u should be read as a variable that can be substituted for a specific label.)

If phrase 2 and phrase 4 are strongly dissonant, the musicians would probably want to avoid a scenario where the two phrases occur at the same bar. In other words we would e.g. like to avoid the run $r^{lateoboe}$ where

$$r^{lateoboe}(4) = (phrase2, phrase4, phrase5)$$

Suppose that the violin chooses to wait for the oboe instead of proceeding according to the score. Then we would have a run $r^{lateviolin}$ where

$$r^{lateviolin}(4) = (phrase2, phrase3, phrase5)$$

But this run might continue in two different ways: One in which the cello adjusts to the other players and does not play phrase 6 until the bar after the violin has played phrase 4 (which would be a bar beyond our current example), and another in which the cello proceeds according to the score, that is, where we end up with

$$r^{lateviolin}(5) = (phrase1, phrase4, phrase6)$$

[5]We also omit describing the Kripke-structure associated with the interpreted system $\mathcal{I}^{ensemble}$ (the system $\mathcal{R}^{ensemble}$ along with an assignment of truth values to all propositions that occur in the system for each state in the system) as we will not have need for it here. For a discussion of this aspect of the semantics of multi-agent systems, see Fagin et al. (2003, 117-118).

Of course for all 1,307,674,368,000 possible deviations from the score, there is the possibility that everyone, including the player(s) with erroneous phrases, tries to keep following the score as closely as possible by playing the "right" phrase according to the score at the next m (thus interpreted as a bar number). For simplicity, we will not try to describe this general case formally here. For convenience, we may, however, add a run describing the situation where the oboe forgets to play phrase 2 at $m = 3$, but where everyone, including the oboe, continues according to the score:

$$r^{latescorevar1}(4) = (phrase1, phrase4, phrase5)$$
$$r^{latescorevar1}(5) = (phrase1, phrase3, phrase6)$$

And we can add a run describing the situation where the violin considers his own phrase 4 more important than the oboe's phrase 2, where he nevertheless forgets to play this at $m = 4$, but where everyone continues according to the score at $m = 5$:

$$r^{latescorevar2}(4) = (phrase1, phrase3, phrase5)$$
$$r^{latescorevar2}(5) = (phrase1, phrase3, phrase6)$$

To nearly complete the picture[6], let us describe the case where the oboe forgets to play phrase 2 at $m = 3$, but plays phrase 2 at a $m = t$, $t > 3$, where the violin chooses to wait for the oboe and reinterpret the bar where the oboe plays phrase 2 as bar 3 according to the score, and where the cello likewise interprets the bar where the violin plays phrase 4 as bar 4 according to the score:

$$r^{lateviolinwaits}(t) = (phrase2, phrase3, phrase5)$$
$$r^{lateviolinwaits}(t+1) = (phrase1, phrase4, phrase5)$$
$$r^{lateviolinwaits}(t+2) = (phrase1, phrase3, phrase6)$$

(We could also describe a situation where the violin does not wait for the oboe, but where the cello will wait for the violin. This is, however, not of relevance to our analysis of the example at this point.)

Now we can identify and formalize the situations of doubt the three players may experience when the oboe forgets to play phrase 2 at bar 3. At m=3, the violin does presumably realize that the other players are no longer proceeding according to rscore, but he does not know (in our current description of the full situation) whether the other players are proceeding according to $r^{lateoboe}$, $r^{lateviolin}$, $r^{latescorevar1}$, $r^{latescorevar2}$ or $r^{lateviolinwaits}$. Formally

$$(r^{lateoboe}, 3) \sim_{violin} (r^{lateviolin}, 3), \quad (r^{lateoboe}, 3) \sim_{violin} (r^{lateviolinwaits}, 3),$$
$$(r^{lateoboe}, 3) \sim_{violin} (r^{latescorevar1}, 3), \quad (r^{lateoboe}, 3) \sim_{violin} (r^{latescorevar2}, 3),$$
$$(r^{lateoboe}, 3) \sim_{violin} (r^{lateviolinwaits}, 3)$$

So how can he ever know what would be the appropriate way to proceed at m=4, except by picking a choice at random? In fact, this situation is the case for all of the players, hence

[6]Because the players can all hear each other, we have reason to believe that the oboe will understand that the other players have proceeded past bars 4 and 5 in the score, once she hears them play phrase 4 and phrase 6 respectively in succession.

$$(r^{lateoboe}, 3) \sim_i (r^{lateviolin}, 3), \quad (r^{lateoboe}, 3) \sim_i (r^{lateviolinwaits}, 3),$$
$$(r^{lateoboe}, 3) \sim_i (r^{latescorevar1}, 3), \quad (r^{lateoboe}, 3) \sim_i (r^{latescorevar2}, 3),$$
$$(r^{lateoboe}, 3) \sim_i (r^{lateviolinwaits}, 3)$$

because everyone has the same (local) state at m=3 no matter which of the runs is executed. (Strictly speaking, it is rather unlikely that any of the players should consciously choose to follow $r^{lateoboe}$ or $r^{lateviolin}$, but we will return to the discussion of what strategy a player is likely to choose later in this paper.)

In order for the players to be able to make a rational choice of what to play at m=4 and onwards, they must either have common knowledge of some rule that clearly states which of the runs is being executed, or they must have some way of getting about the problem of disagreement on the character of the run. The latter set of options is explored in the sections below. The aforementioned rule could be stated as a rather strict obligation to wait for the oboe's phrase 2 and then proceed according to $r^{lateviolinwaits}$, but a formalization of this will necessitate an introduction to deontic logic as well as temporal operators, for which we do not have the sufficient amount of space here. We will, however, dwell for a moment on the topic of what it means for such a rule to be common knowledge among the players.

2.1 Common Knowledge of Rules and Its Implications for the Ensemble

A statement p being common knowledge in a group G, notated $C_G p$, entails informally that everyone in the group knows p, and that the entire group is somehow aware of p's being known by everyone and being expected to be known by everyone. The formal representation of C_G in terms of the operator E_G, meaning "everyone in G knows that" is debated[7] , but all theories grant that $E_G p$ can be deduced from $C_G p$, and hence that $K_i p$ (i knows that p) can be deduced from $C_G p$, for all $i \in$ G.

Intuitively it should not be surprising that common knowledge in the group of a rule is required in a situation where coordination depends on the group members following the rule. In our example above, it is not enough that everyone in the ensemble knows that a rule p holds, if someone is in doubt whether the other ensemble members know that rule p holds. (We are of course still assuming that the players have no way of communicating that they follow p during a performance.) On the other hand, once p is common knowledge in the ensemble, that is, once it is part of the collective consciousness of the ensemble, it is safe to entail that everyone in the ensemble knows p. And since p is a rule that states what the ensemble should do when deviating from the score, knowing this rule combined with knowing that everyone else knows it and assumes that everyone else knows it, results in the individual ensemble member following the rule, thus ensuring coordination.

The idea of the ensemble being collectively conscious of a coordinating rule p, however, amounts to an idea of the ensemble having the same opinion of the salient features of the composition. Remember that in our description of the piece of music, we do not know whether the oboe, the violin or the cello

[7] Fagin et al. (2003) discuss at least two different interpretations of the notion, one in terms of a possible infinite iteration of the E_G-operator (23-25), another in terms of sets of information states in so-called Aumann structures (38-41).

has the most important role in the passage. It might be that the voice of the oboe is not only the initiator of a step-wise development in the voices but also an indispensable part of this development, for example if the sequence phrase 2 – phrase 4 – phrase 6 constitutes a melody that simply for the purpose of a fun effect has been distributed onto three different voices. But it might also be that the oboe's phrase 2 is just like a small prologue to a theme that actually begins with phrase 4 in the violin, and that phrase 4 for some reason is tightly knit to a rhythmic structure that develops over bars 1-3. A similar situation could be the case for the cello, if the oboe and the violin are merely adding small fills to the last of four bars that naturally precede phrase 6 in the cello. In any of the three cases, if we could point to a rule that, if common knowledge in the ensemble, would ensure safe conduct in the situation of doubt, this rule would indirectly be a statement of the compositional features to be regarded as salient by every musician. In other words, this account of coordination in the ensemble leaves no possibility of disagreement with respect to the interpretation of the composition. For a programmer simulating an en-semble as one virtual accompanist to one live soloist, this is not a big issue. We would generally like an accompanist that, at the worst, is only in disagreement with the soloist, not with itself also. For someone modeling the interactions of several independent players, modeling players with different initial perspectives on the music is, however, very important.[8]

In the following sections, we will examine what can be done for a formal description of ensemble coordination without imposing a structure where everyone has to have the same idea of the salient features of the composition.

3 Game Theory with Variable Frames

A great deal of effort has been put into explaining how people are able to coordinate in games where two or more players (here understood as players of the game, not musicians) only receive a payoff, if they are able to simultaneously choose the same of a number of options. For instance, in the introduction by Natalie Gold and Robert Sugden to Bacharach et al. (2006), we find the example of "Three Cubes and a Pyramid" (19). In this game two players have to choose the same out of four objects, a red cube, a blue cube, a yellow cube and a green pyramid. From an objective point of view, the probability that the two players coordinate on the same object is just 0.25, because there are 16 possible combinations of actions of the two and 4 possible ways they can choose the same object. But experimental studies show (according to Bacharach et al) that people actually tend to be much better at coordinating than that, and that the players tend to choose the green pyramid. The intuitive answer to this question (and the answer given by Schelling (1960, 64) in relation to similar experiments) is that the choice of the green pyramid is somehow more salient than the other options. But why?

First of all, the two players are not just picking at random without taking into consideration how they perceive the game and its four objects. They describe the game to themselves using predicates, and these predicates belong to what

[8]For one of many examples of the efforts being put into achieving alignment of a virtual accompanist's delimitation of what counts as instances of a given piece of music and the interpretation of the same composition by a soloist, see Fox and Quinn (2007)

Bacharach calls families Bacharach et al. (2006, 14-16). Formally, we define a set S of objects, a set P of predicates and a function E that assigns a (possibly empty) subset of S to each predicate in P, such that if φ is a predicate, then $E(\varphi)$ is the set of objects φ describes (or the extension of φ).[9] If we call the set of objects in the "Three Cubes and a Pyramid"-game $S_{objects} = \{x_1, x_2, x_3, x_4\}$, and decide that x_3 is the green pyramid, we have for instance $E_{objects}(cube) = \{x_1, x_2, x_4\}$ and $E_{objects}(pyramid) = \{x_3\}$. If the extension of a predicate has more than one member, such as "cube" in this case, we call the act of singling out one object to which that predicate applies, "picking". If the extension is a singleton, such as the extension of "pyramid", we call the act of singling out the object to which that predicate applies, "choosing". In other words, the players can "pick a cube" or "choose the pyramid" but not "choose a cube" or "pick a pyramid". The predicates can be arranged in families, understood as sets of predicates, where, if one comes to mind for the player, the other ones will come to mind as well. Hence we can define a shape family, $F_{shape} = \{cube, pyramid \ldots\}$ and a color family $F_{color} = \{blue, red, yellow, green \ldots\}$. We can also define a "generic family" $F_{thing} = \{thing\}$, where $E_{objects}(thing) = \{x_1, x_2, x_3, x_4\}$. We might be able to come up with other families and predicates, but let us stop here for the sake of clarity. Now, for each player, we can define a set of families that might come to mind for that player. We call such a set a *frame*. Such a set is a subset of the *universal frame F*, containing all families that can be taken into consideration in the example (thus the universal frame in "Three Cubes and a Pyramid" is $F_{objects} = \{F_{thing}, F_{shape}, F_{color}\}$). Each player assigns to his opponent (we are assuming a game of two players) a probability $v(F_i)$ that the opponent has a family F_i in his frame - this is also called the *availability* of F_i. For instance, a player may think that $v(F_{thing}) = 1$ for his opponent, $v(F_{color}) = 0.6$ and $v(F_{shape}) = 0.8$. So, if the player is right in how he considers the availability of the families for his opponent, the probability that the player will look upon the situation as choosing between shapes rather than "non-descript" objects Bacharach et al. (2006, 16) is 0.8. Because there are three cubes, the possibility of both players coordinating on the same cube if they both decide on the act-description "pick a cube" is 0.33 (1/3), and, if we take the availability of the shape family for granted, the possibility that they coordinate in general is 0.33*0.8 = 0.26. This is only marginally better than the chances of the players when just picking at random. If both of the players decide to "choose the pyramid", however, they have a 1*0.8 = 0.8 chance of perfect coordination, as there is only one pyramid. If we assume that the payoff for coordination is exactly the same no matter what the players agree to do, it seems that choosing the pyramid is a much better option than any other possible act, as the probability that the players coordinate is higher. (Actually, even if we assume that both players assign an availability of 1 to all families in their opponent's frame, "choose the pyramid" will still be the optimal choice. This is because the options "choose the blue", "choose the red", "choose the yellow" and "choose the green" are discarded due to what Bacharach calls the principle of *symmetry disqualification*[10]. This principle roughly entails that if there are two or more predicates from the same family that have exactly the same size

[9] This is my rendition of Bacharach et al. (2006, 10-11 and 14-20).

[10] The analysis of "Three Cubes and a Pyramid" is in essence the same as in (Bacharach et al. 2006, 19-22), although I have used a slightly different notation utilizing more transparent subscripts for the different variables.

of extension in the game, we have no reason for choosing one over the other, and hence we should disregard the family entirely. Another way of putting it in our case is that absence of a stand-out color choice converts the situation to an arbitrary "picking" between act-descriptions issued from the color family where the chances of coordinating are much smaller.)

We will now try to apply some of these ideas to our coordination problem in the music ensemble.

3.1 An Analysis of the Musical Coordination Problem in Terms of Variable Frame Theory

Our coordination problem as described in sections 1-2.1 can be interpreted as a coordination game such as the one we have just examined. The object of the "game" in our ensemble is to choose the same strategy as to which phrases should be played at what time and after which phrases. In our example we have roughly four different strategies: The first is where the musicians try to stick to the score as much as possible and disregard mistakes as unfortunate mishaps. The second is where all three musicians regard the oboe's phrase 2 as essential for the continuous development of the piece and thus wait for the oboe, if the oboe is late. The third is where the musicians regard the violin's phrase 4 as essential and therefore disregard the oboe's eventually being late as a source of confusion but wait for the violin to commence phrase 4 before proceeding according to bar 5 in the score. The fourth is where the musicians regard the cello's phrase 6 as essential, so that even if both the oboe and the violin is late, these players will continue playing their phrases 1 and 3 respectively until the cello commences phrase 6. Unless the cello is even later than both of the other players, the first and fourth strategies amount to the same: follow the score and just move on in case of errors. We can thus simplify our example a bit by eliminating the fourth strategy from our considerations. From the cello's point of view, however, the second and third strategies amount to the same line of action: wait for the violin to play phrase 4, then proceed to bar 5. On the other hand, since it is impossible for the oboe to wait for the violin, the oboe considers the first and third strategies similar with respect to her own line of action: in both cases, she should continue according to the score. So, to sum up, the only player for whom it really matters, if the violin's phrase 4 is most important of phrases 2, 4 and 6, is the violin. If we roughen our distinctions a bit, we could say that the violin really faces a problem of choosing between waiting for the oboe's phrase 2 and not waiting for the oboe's phrase 2. Not waiting for the oboe does not rule out the violin being late himself, if he follows the third strategy described above, but that does not change anything for the other two players. We can therefore describe the coordination problem as a game of coordinating on the same choice of strategy, where the two possible strategies are:

> "Wait" (meaning "wait for the oboe's phrase 2 (the oboe plays phrase 2 when ready)") and
>
> "Don't Wait" (meaning "do not wait for the oboe's phrase 2 (continue according to the score if the oboe does not play phrase 2 at bar 3)").

The "objective game" in Bacharach's terms Bacharach et al. (2006, 14), that is, the game without a representation of the players' frames looks like this: Each of the three players have a possibility of 0.25 of coordinating on the same strategy, whether "Wait" or "Don't Wait" (because there are 8 different combinations of strategies for the three players and 2 possible ways they can choose the same line of action). But the objective game only describes the situation as it would be, if the players picked their strategies at random. It is, however, more likely that they describe the two choices to themselves in terms of their qualities. For example a player could say that "Wait" is a more "melodic" solution with respect to phrasing, or s/he could say that "Don't Wait" "keeps the piece going rhythmically" understood such that this strategy is more in accordance with the overall rhythmical structure of the passage. Let us symbolize "Wait" by x_1 and "Don't Wait" by x_2. Then we can define a family of predicates F_{rhythm} = {keeps the piece going rhythmically, ...}, where E(keeps the piece going rhythmically) = $\{x_2\}$. We can also define a family $F_{melody} = \{melodic, \ldots\}$, where $E(melodic) = \{x_{10}\}$. If we once again include the generic family $F_{thing} = \{thing\}$ where $E(thing) = \{x_1, x_2\}$, we have the universal frame $F = \{F_{thing}, F_{rhythm}, F_{melody}\}$ for the coordination game. Now, because of the inclusion of F_{thing}, a player that has all three of the mentioned families in his frame can decide on one of these act-descriptions: "pick a thing (something)", "choose the option that keeps the piece going rhythmically" or "choose the melodic". I have deliberately simplified the amount of possible choices and predicates in this example, because our example has the complexity over "Four Cubes and a Pyramid" that there is an extra player. Each player assigns two availabilities for a family, that is, one for each of the other players. Let us say that the violin assigns the possibility $v_{oboe}(F_{melody}) = 0.7$ to the case where F_{melody} comes to mind for the oboe, $v_{oboe}(F_{rhythm}) = 0.3$ to the situation where F_{rhythm} comes to mind for the oboe, $v_{cello}(F_{melody}) = 0.6$ to the situation where F_{melody} comes to mind for the cello and $v_{cello}(F_{rhythm}) = 0.5$ to the case where F_{rhythm} comes to mind for the cello. If the violin is right about his estimates and decides to "choose the option that keeps the piece going rhythmically", he has a 0.3*0.5*1 = 0.15 chance of coordinating with the other musicians on this strategy. If on the other hand he decides to "choose the melodic", he has, provided his estimates are correct, a 0.7*0.6*1 = 0.42 chance of coordinating with them on this. This is still not an overwhelming safety, but if we grant that coordination on a strategy is good no matter the strategy, it seems reasonable for the violin to "choose the melodic" because he considers the probability of coordinating with the other two players higher than by picking at random. But does this ensure coordination in the ensemble? This is the subject of the next section.

3.2 What Does the Availability of a Frame Show Us?

There are at least two problems that some readers will notice immediately in the analysis above. The first is that it might be that the violin is wrong in his assignments of availabilities to families in the frames of his co-players. The second is that it might be that the other players have a different view of the availabilities of families in each other's frames, thus making the probability assessment even more complicated. It is important to note in connection with these two complications that what we have described above is how a player can rationally make a choice based on his or her expectations of how the other players may be likely to think. Even if the violin is for instance right in his assumption that $v_{oboe}(F_{rhythm}) = 0.3$, this does not mean that it can never occur that the oboe decides to "choose the option that kept the piece going rhythmically". But if his estimates of the availabilities are generally right, and if coordination, no matter the strategy, is still the objective, the violin will be foolish not to go for the strategy that gives him the highest probability of coordination. So the real trouble here is on what basis a player makes his estimates of the availabilities of families in the frames of his co-players. Intuitively, if an ensemble, such as the trio we are considering here, have been working together for a long time, it seems that it would be strange if the players deviated much from each other in their views of the availability of a family in a given player's frame. On the other hand, an ad hoc ensemble of musicians where no one knows each other, might have fairly the same expectations of the availabilities of different families in each other's frames, namely close to 0.5 for all families. The latter situation is, however, not likely to ensure very good coordination because the possibilities for coordination on a strategy will inevitably come out rather low. But both of the mentioned intuitions point to the relevance of musicians "knowing each other" prior to a performance.[11]

Of course, we can still improve the probabilities of coordination in the ensemble by strengthening the common knowledge or "consciousness" of certain rules inherent in the composition. In the above case, the violin would then probably assign the same availabilities to a family in all frames of his co-players. What we wanted to show in our analysis in terms of Bacharach's variable frame theory was, however, that the players might be able to make non-random decisions making coordination quite possible, even if they do not have common knowledge of the rules of the composition but only some expectations of each other's way of perceiving the situation. Such estimates as the one described in 3.1 does not ensure coordination, but makes coordination more possible than if everyone chooses at random.

4 Integrating Individual Interpretations of a Composition

If we view the composition as a set of rules guiding performance, not as a specific sonic outcome, we could define the musicians' interpretations of a composition in the following way: Say, that the musicians are working from a score by a composer, and that this score entails a list of rules (or instructions)."The composition" is obviously not identical with this list, since it would otherwise not make sense to speak of "arrangements" of a composition (these would by definition not be instances of the composition because they were not following the rules given by the score exactly). The composition is thus a selection of rules given by the score that are considered more important than other rules given by the score. Which rules are selected might be dependent on culture, context or even individual preferences. An *interpretation* is a pri-

[11]This in accordance with many of the musicians I know who either will not perform with other people without extensive rehearsal or will only perform with people they are familiar with in advance.

oritized ordering of the rules constituting the composition, such that in cases where not all of the instructions can be followed as originally intended (e.g. a coordination problem), the musician will try to satisfy some rules before others.

We could for instance describe the rankings of the interpretation in terms of a preference ordering for each player that determines which rules he/she will try to follow when having to choose between two or more different actions. This could also amount to modeling an extensive strategy for each player, that is, a description of what the player should do in any situation that might occur.[12] The possible benefit of such an approach would be that it allows a musician to reason about the strategies of other musicians without considering them capable of describing the different choices to themselves (as in variable frame theory).

4.1 Revision of Strategies

It seems to be a consequence of wanting to achieve coordination that some musicians – simply by being outnumbered – must occassionally give up their own strategy and adjust to those of other musicians (if we are looking at an improvisation context, revising ones own strategy is a constantly reoccuring phenomenon).

A model for a less drastic revision process could also be imagined in which a musician adds, removes or moves a rule to/from/in his or her interpretation[13] (still viewed as a ranking of rules) as the performance moves along, in order to either adjust to the other musicians or simply changes his or her mind about the relevant aspects of the music.

Of course, the boundaries between small changes in the strategy of the agent and a complete replacement with a new strategy are fluent. It is a point of further discussion how much an interpretation can be altered before it constitutes a new (or radically new view of the) composition. (This is a problem because of the possibility of adding new elements external to the initial composition.)

5 Conclusion

In order for perfect coordination to be certain to take place in the ensemble, some rule determining a prioritized ranking of instructions in the composition must be common knowledge. Yet, even if not everyone agrees on the ranking of instructions in the composition, coordination is still possible, because musicians reason according to their expectations of the actions of their co-players.

The above analysis is of course simplified but it points to a way of modeling ensemble relations that might be of relevance for researchers in computer music modeling. The idea is that when modeling two or more ensemble players, we should define their (possibly virtual) characteristics as musicians, that is, their musical background such as their tastes, their previous engagements

[12] A different approach I am now following is to try to define the interpretation in terms of the sonic outcome the musician wants. This intended outcome defines the possible strategies from which he can choose. (I am inspired here by the logic for intentions and their role in coordination described by Roy (2008))

[13] Rules added might include musical ideas external to the score

in other ensembles, their cultural heritage etc.[14] Some of these traits might be quasi-formalized, for instance a strong dependence to follow the score in a rhythmically strict way or a partiality to the execution of central melodic phrases. Depending on the outcome we want, we can make all of these characteristics known to all players, only some of them or none. We can then model how a rational player will navigate in situations of doubt (or, although this requires a different sort of analysis, situations where the musicians deviate from the score on purpose, such as in an improvisation) by computing his possibilities of coordinating with other players on a strategy given his estimate of what they are likely to choose. We can still include normative features of a composition (understood as something of which the score is merely one of many possible arrangements) in the model[15] if we want to and define which musicians know these features, but we do not need all of the players to agree on these features in advance for coordination to take place.

In this paper I have tacitly relied on my own experience as a violinist in several ensemble contexts. To achieve more accurate modeling of coordination processes such as the ones I have tried to describe here, it will of course be necessary to conduct experiments and interviews with several more ensemble musicians.

References

M. Bacharach, R. Sugden, and N. Gold. *Beyond Individual Choice: Teams and Frames in Game Theory.* Princeton University Press, Princeton, 2006.

R. Fagin, J. Y. Halpern, Y. Moses, and M. Y. Vardi. *Reasoning about knowledge*. The MIT Press, Cambridge, Mass., 2003.

C. Fox and J. Quinn. How to be lost: Principled priming and pruning with particles in score following. In K. Jensen, editor, *Proceedings of the 2007 International Computer Music Conference, vol. I*, pages 81–84, Copenhagen, Denmark, August 2007. The International Computer Music Association.

O. Roy. *Thinking Before Acting. Intentions, Logic, Rational Choice*. PhD thesis, Institute for logic, Language and Computation, Universiteit van Amsterdam, Amsterdam, The Netherlands, February 2008. ILLC Dissertation series DS-2008-03.

T. Schelling. *The Strategy of Conflict*. Harvard University Press, Harvard, 1960.

R. Sharpe. *Philosophy of Music: An Introduction*. Acumen, New York, 2004.

[14] I am thinking along the lines of how many chess engines are built. Here the player can choose his opponent among different profiles mimicking different human backgrounds.

[15] Indeed, following Sharpe (2004, 59-60), preserving characterizations of the composition in terms of normativity is essential for the clearest possible delimitation of the boundaries of a piece of music.

Reasoning about Cooperation, Actions and Preferences

Lena Kurzen
Institute for Logic, Language and Computation, Universiteit van Amsterdam.
L.M.Kurzen@uva.nl

Abstract

A logic for reasoning about coalitional power is developed which explicitly represents agents' preferences and the actions by which the agents can achieve certain results. A complete axiomatization is given and its satisfiability problem is shown to be decidable and EXPTIME-hard.

1 Introduction

Cooperation of agents plays a major role in many fields such as computer science, economics, politics, social sciences and philosophy. Agents can decide to cooperate and to form groups in order to share complementary resources or because as a group they can achieve something better than individually.

When analyzing interactive situations involving multiple agents, we are interested in what results agents can achieve – individually or together as groups. There can be many ways how agents can achieve some result. They can differ significantly, e.g. with respect to their feasibility, costs or side-effects. Hence, it is not only relevant what results groups of agents can achieve but also *how* exactly they can do so. In other words, plans and actions also play a central role if we want to reason about cooperation in an explicit way. However, cooperative ability of agents expressed only in terms of results and actions that lead to these results does not tell us *why* a group of agents would actually decide to achieve a certain result. We also need to take into account the preferences based on which the agents decide what to do. Summarizing, we can say that in interactive situations, the following three questions are of interest:

- ***What*** *results can groups of agents achieve?*
- ***How*** *can they achieve something?*
- ***Why*** *would they want to achieve a certain result?*

The above considerations show that coalitional power, actions/plans and preferences play a major role in interactive situations and are moreover tightly connected. Thus, we argue that a formal theory for reasoning about agents' cooperative abilities in an explicit way should also take into account actions/plans of agents and their preferences.

Modal logics have been used to develop formal models for reasoning about each of these aspects – mostly separately. Coalitional power has mainly been investigated within the frameworks of ATL (Alur et al. 1998), Coalition Logic (Pauly 2002a) and their extensions.

Recently, there have been some attempts to develop logics for reasoning about coalitional power that also take into account either agents' preferences or actions. One group of such logics looks at cooperation from the perspective of cooperative games (Ågotnes et al. 2007a). In a non-cooperative setting, preferences and strategic abilities have been considered in van Otterloo et al. (2004). Another path that has been taken in order to make coalitional power more explicit is to combine cooperation logics with (fragments of) action logics (Sauro et al. 2006, Borgo 2007, Walther et al. 2007).

In this paper, a logic for reasoning about cooperation, actions and preferences (CLA+P) is developed, which is obtained by combining the cooperation logic with actions CLA (Sauro et al. 2006) with a preference logic (van Benthem et al. 2006; 2007). Soundness and completeness are shown and the logic's expressivity and computational complexity are investigated.

The remainder of this paper is structured as follows. Section 2 gives an overview of CLA. In Section 3, CLA+P is developed, a complete axiomatization is given and its expressivity is discussed. Section 4 gives complexity results and Section 5 concludes this work.

2 Cooperation Logic with Actions (CLA)

In this section, we briefly present the cooperation logic with actions (CLA) developed by Sauro et al. (2006), which will be extended in the next section by combining it with a preference logic. The idea of CLA is to make coalitional power explicit by expressing it in terms of the ability to perform actions instead of expressing it directly in terms of the ability to achieve certain outcomes. CLA is a modular modal logic, consisting of an environment module for reasoning about actions and their effects, and an agents module for reasoning about agents' abilities to perform actions. By combining both modules, a framework is obtained in which cooperative ability can be made more explicit.

The environment is modeled as a transition system whose edges are labeled with sets of atomic actions.

Definition 2.1 (Environment Model (Sauro et al. 2006))**.** An environment model is a set-labelled transition system

$$E = \langle S, Ac, (\rightarrow)_{A \subseteq Ac}, V \rangle.$$

S is a set of states, Ac is a finite set of atomic actions, $\rightarrow_A \subseteq S \times S$ and V is a propositional valuation. Each $\rightarrow_A$ is required to be serial.

The intuition behing $s \rightarrow_A t$ is that if in s all actions and only the actions in A are performed concurrently, then the next state can be t.

Then a modal language is defined with modalities $[\alpha]$, for α being a propositional formula built from atomic actions. The intended meaning of $[\alpha]\varphi$ is that every transition $\rightarrow_A$ such that $A \models \alpha$ (using the satisfaction relation of propositional logic[1]) leads to a φ-state:

$E, s \models [\alpha]\varphi \quad$ iff $\quad \forall A \subseteq Ac, t \in S$: if $A \models \alpha$ and $s \rightarrow_A t$ then $E, t \models \varphi$.

Due to space restrictions, we cannot go into the underlying philosophy of actions but refer the reader to Broersen (2003) for a detailed discussion of action logics. The restriction to a finite set of actions is reasonable for modelling many concrete situations and also ensures that we have a finite axiomatization.

An environment logic Λ^E is developed, which is sound and complete (Sauro et al. 2006). It contains seriality axioms and the **K** axiom for each modality $[\alpha]$, for α being consistent. The environment logic can then be used for reasoning about the effects of concurrent actions.

Then an agents module is developed for reasoning about the ability of (groups of) agents to act. Each agent is assigned a set of atomic actions and a group is assigned the actions its members can perform.

Definition 2.2 (Agents Model (Sauro et al. 2006))**.** An agents model is a triple $\langle Ag, Ac, \mathsf{act}\rangle$, where Ag is a set of agents, Ac is a set of atomic actions and act is a function $\mathsf{act} : Ag \rightarrow \mathcal{P}(Ac)$ such that $\bigcup_{i \in Ag} \mathsf{act}(i) = Ac$. For $G \subseteq Ag$, define $\mathsf{act}(G) := \bigcup_{i \in G} \mathsf{act}(i)$.

We are also interested in agents' abilities to force more complex actions. A language is developed with expressions $\langle\!\langle G \rangle\!\rangle \alpha$, meaning that the group G can force that a concurrent action is performed that satisfies α. This means that G can perform some set of atomic actions such that no matter what the other agents do, the resulting set of actions satisfies α.

$\langle Ag, Ac, \mathsf{act}\rangle \models \langle\!\langle G \rangle\!\rangle \alpha \quad$ iff $\quad \exists A \subseteq \mathsf{act}(G) : \forall B \subseteq \mathsf{act}(Ag \setminus G) : A \cup B \models \alpha$.

Then a cooperation logic for actions is developed, which is very much in the style of Coalition Logic (Pauly 2002a) – the main difference being that it is concerned with the cooperative ability to force *actions*.

Definition 2.3 (Coalition Logic for Actions (Sauro et al. 2006))**.** The coalition logic for actions Λ^A is defined to be the logic derived from the following set of axioms, with rule of inference modus ponens.

1. $\langle\!\langle G \rangle\!\rangle \top$, for all $G \subseteq Ag$,
2. $\langle\!\langle G \rangle\!\rangle \alpha \rightarrow \neg \langle\!\langle Ag \setminus G \rangle\!\rangle \neg \alpha$,
3. $\langle\!\langle G \rangle\!\rangle \alpha \rightarrow \langle\!\langle G \rangle\!\rangle \beta$ if $\vdash \alpha \rightarrow \beta$ in propositional logic,
4. $\langle\!\langle G \rangle\!\rangle a \rightarrow \bigvee_{i \in G} \langle\!\langle \{i\} \rangle\!\rangle a$ for all $G \subseteq Ag$ and atomic $a \in Ac$,
5. $(\langle\!\langle G_1 \rangle\!\rangle \alpha \wedge \langle\!\langle G_2 \rangle\!\rangle \beta) \rightarrow \langle\!\langle G_1 \cup G_2 \rangle\!\rangle (\alpha \wedge \beta)$, for $G_1 \cap G_2 = \emptyset$,
6. $(\langle\!\langle G \rangle\!\rangle \alpha \wedge \langle\!\langle G \rangle\!\rangle \beta) \rightarrow \langle\!\langle G \rangle\!\rangle (\alpha \wedge \beta)$ if α and β have no common atomic actions,
7. $\langle\!\langle G \rangle\!\rangle \neg a \rightarrow \langle\!\langle G \rangle\!\rangle a$ for atomic $a \in Ac$,
8. $\langle\!\langle G \rangle\!\rangle \alpha \rightarrow \bigvee \{\langle\!\langle G \rangle\!\rangle \bigwedge \Psi | \Psi$ is a set of literals such that $\bigwedge \Psi \rightarrow \alpha\}$.

[1] That is, $A \models a$ iff $a \in A$, $A \models \neg\alpha$ iff $A \not\models \alpha$, and $A \models \alpha \wedge \beta$ iff $A \models \alpha$ and $A \models \beta$.

Axiom 5 says how groups can join forces. The coalition logic for actions is sound and complete with respect to the class of agents models (Sauro et al. 2006).

Next, agents are introduced as actors into the environment. This is done by combining the environment models with the agents models. Then the agents can perform actions which have the effect of changing the current state of the environment.

Definition 2.4 (Multi-agent System (Sauro et al. 2006))**.** A multi-agent system (MaS) is a tuple

$$M = \langle S, Ac, (\rightarrow)_{A \subseteq Ac}, V, Ag, \mathsf{act}\rangle,$$

where $\langle S, Ac, (\rightarrow)_{A\subseteq Ac}, V\rangle$ is an environment model and $\langle Ac, Ag, \mathsf{act}\rangle$ an agents model.

Now, we can reason about what states of affairs groups can achieve by performing certain actions. The corresponding language contains all expressions of the previously defined logics and additionally expressions for saying that a group has the power to achieve φ which means that the group can make the system move into a state where φ is true.

Definition 2.5 (Language for MaS (Sauro et al. 2006))**.** The language for multi-agent systems $\mathcal{L}_{cla}$ is generated by the following grammar:

$$\varphi ::= \quad p \mid \varphi \wedge \varphi \mid \neg\varphi \mid [\alpha]\varphi \mid \langle\!\langle G \rangle\!\rangle \alpha \mid \langle\!\langle G \rangle\!\rangle \varphi$$

for $G \subseteq Ag$ and α being an action expression.

$\langle\!\langle G \rangle\!\rangle \varphi$ means that G can force φ, i.e. G can perform a set of actions such that no matter what the other agents do, the system moves to a φ-state.

$$\begin{array}{lll} M, s \vDash \langle\!\langle G \rangle\!\rangle \varphi & \text{iff} & \exists A \subseteq \mathsf{act}(G) \text{ such that } \forall B \subseteq \mathsf{act}(Ag \setminus G), t \in S : \\ & & \text{if } s \rightarrow_{A \cup B} t, \text{ then } M, t \vDash \varphi. \end{array}$$

A complete axiomatization is obtained by combining the environment logic and the coalition logic for agents by adding two interaction axioms.

Definition 2.6 (Cooperation Logic with Actions (Sauro et al. 2006))**.** The cooperation logic with actions Λ^{CLA} combines the environment logic Λ^{E} and the coalition logic for actions Λ^{A} by adding

1. $(\langle\!\langle G \rangle\!\rangle \alpha \wedge [\alpha]\varphi) \rightarrow \langle\!\langle G \rangle\!\rangle \varphi$,
2. $\langle\!\langle G \rangle\!\rangle \varphi \rightarrow \bigvee\{\langle\!\langle G \rangle\!\rangle \alpha \wedge [\alpha]\varphi | \alpha$ is the conjunction of a set of atomic actions or their negations$\}$.

CLA provides us with a formal framework for reasoning about what states of affairs groups of agents can achieve and how they can do so. For a detailed discussion of CLA, the reader is referred to Sauro et al. (2006). Now, we proceed by adding preferences to CLA.

3 Cooperation, Actions and Preferences

In this section, a logic for reasoning about cooperation, actions and preferences is developed. This is done by adding a preference logic to CLA. For a more detailed discussion and proofs, see Kurzen (2007).

3.1 Preference Logic

There are various ways how preferences can be added to a logic for cooperation and actions. They could range over the actions that he can perform. Alternatively, we can think of each agent having preferences over the set of successor states of the current state.

In the current work, we consider preferences of individual agents ranging over the states of the environment. This is reasonable since by performing actions the agents can change the current state of the environment, and the preferences over those states can be seen as the base of how the agents decide how to act. Such a preference relation can also be lifted in several ways to one over formulas (van Benthem et al. 2007; 2006).

Definition 3.1 (Preference Model (van Benthem et al. 2006))**.** A preference model is a tuple

$$M^P = \langle S, Ag, \{\preceq_i\}_{i \in Ag}, V\rangle,$$

where S is a set of states, Ag is a set of agents, for each $i \in Ag, \preceq_i \subseteq S \times S$ is reflexive and transitive, and V is a propositional valuation.

We use a fragment of the preference language developed by van Benthem et al. (2007). It has strict and non-strict preference modalities.

Definition 3.2 (Preference Language)**.** Given a set of propositional variables and a finite set of agents Ag, define the preference language $\mathcal{L}_p$ to be the language generated by the following syntax:

$$\varphi := \quad p \mid \neg\varphi \mid \varphi \vee \varphi \mid \Diamond^{\preceq_i}\varphi \mid \Diamond^{\prec_i}\varphi.$$

$\Diamond^{\preceq_i}\varphi$ says that there is a state satisfying φ that agent i considers to be at least as good as the current one. The semantics is defined as follows.

$$M^P, s \vDash \Diamond^{\preceq_i}\varphi \text{ iff } \exists t : s \preceq_i t \text{ and } M^P, t \vDash \varphi.$$

Analogously for $\Diamond^{\prec_i}\varphi$. The preference relation $\preceq$ is a preorder and $\prec$ is its largest irreflexive subrelation. Hence, the following axiomatization.

Definition 3.3 (Preference Logic Λ^P)**.** For a given set of agents Ag, let Λ^P be the logic generated by the following axioms for each $i \in Ag$: For $\Diamond^{\preceq_i}$ and $\Diamond^{\prec_i}$, we have Duality and **K** and for $\Diamond^{\preceq_i}$ also reflexivity and transitivity axioms. Moreover, there are four interaction axioms:

1. $\Diamond^{\prec_i}\varphi \rightarrow \Diamond^{\preceq_i}\varphi$,
2. $\Diamond^{\preceq_i}\Diamond^{\prec}\varphi \rightarrow \Diamond^{\prec_i}\varphi$,
3. $\Diamond^{\prec_i}\Diamond^{\preceq_i}\varphi \rightarrow \Diamond^{\prec_i}\varphi$,
4. $\varphi \wedge \Diamond^{\preceq_i}\psi \rightarrow (\Diamond^{\prec_i}\psi \vee \Diamond^{\preceq_i}(\psi \wedge \Diamond^{\preceq_i}\varphi))$.

The inference rules are modus ponens, necessitation and substitution.

Transitivity for $\Diamond^{\prec_i}$ follows. We show soundness and completeness using the *bulldozing* technique (Blackburn et al. 2001) to deal with $\prec$. For details, we refer to van Benthem et al. (2007).

Theorem 1. *Λ^P is sound and complete with respect to the class of preference models.*

Proof. Follows from Theorem 1 in van Benthem et al. (2007). □

The preference logic is able to distinguish between weak and strict preference. This plays a major role in many concepts for reasoning about interaction in multi-agent systems. Due to the logic's simplicity it still has the modal character and talks about preferences from a local perspective. An additionally $\Diamond^{\succeq}$ modality would have increased the expressivity but would have resulted in a global existential modality with respect to all comparable states.

3.2 Environment Logic with Preferences

As an intermediate step towards a logic for cooperation, actions and preferences, we combine the preference logic and the environment logic. Their models are combined by identifying their states. The agents' preferences range over the states of the environment. The agents cannot act in the environment, but can rather be seen as observing the environment from the outside while having preferences over its states.

Definition 3.4 (Environment with Preferences)**.** An environment model with preferences is a tuple

$$E^{\preceq} = \langle S, Ac, (\rightarrow)_{A \subseteq Ac}, \{\preceq_i\}_{i \in Ag}, V \rangle,$$

where $\langle S, Ac, (\rightarrow)_{A \subseteq Ac}, \{\preceq_i\}_{i \in Ag}, V \rangle$ is an environment model and $\langle S, Ag, \{\preceq_i\}_{i \in Ag}, V \rangle$ is a preference model.

We combine the languages for environment and preferences and add expressions saying that "i (strictly) prefers every α-accessible state". This will later allow us to express statements saying by which actions groups can(not) achieve an outcome better for (some of) its members.

Convention 1. *We will write the symbol $\lhd$ in statements that hold for both $\preceq$ and $\prec$, each uniformly substituted for $\lhd$.*

Definition 3.5 (Environment Language with Preferences)**.** The language $\mathcal{L}_{ep}$ contains all expressions of the environment language and the preference language and additionally formulas of the forms $\alpha \subseteq\preceq_i$ and $\alpha \subseteq\prec_i$, for α being an action expression.
Boolean combinations and expressions of previously defined languages are interpreted as usual. For the newly introduced expressions, we have:
$E^{\preceq}, s \models \alpha \subseteq\lhd_i$ iff $\forall A \subseteq Ac, t \in S$: if $s \rightarrow_A t$ and $A \models \alpha$ then $s \lhd_i t$.

$\alpha \subseteq\lhd_i$ cannot be defined in the preference language and the environment language. because $\alpha \subseteq\preceq_i$ says that each state accessible by an α-transition is also accessible by $\preceq$. Thus, we would have to be able to refer to particular states. Therefore, we add two inference rules.

$$\text{(PREF-ACT)}\ \frac{\Box^{\preceq_i}\varphi \rightarrow [\alpha]\varphi}{\alpha \subseteq\preceq_i} \qquad \text{(STRICT PREF-ACT)}\ \frac{\Box^{\prec_i}\varphi \rightarrow [\alpha]\varphi}{\alpha \subseteq\prec_i}$$

In order to obtain a complete axiomatization, two axioms are added which correspond to the converse of the inference rules.

Theorem 2. *Let Λ^{EP} be the logic generated by all axioms of the environment logic Λ^{E}, all axioms of the preference logic Λ^{P}, and*

1. $\alpha \subseteq \preceq_i \rightarrow (\Box^{\preceq_i}\varphi \rightarrow [\alpha]\varphi)$,
2. $\alpha \subseteq \prec_i \rightarrow (\Box^{\prec_i}\varphi \rightarrow [\alpha]\varphi)$.

The inference rules are modus ponens, substitution, PREF-ACT and STRICT PREF-ACT. Then Λ^{EP} is sound and complete with respect to the class of environment models with preferences.

Proof. Soundness is straightforward, and completeness follows from completeness of the sublogics and the closure under the new rules. □

In Λ^{EP}, the performance of concurrent actions changes the current state of the system also with respect to the agents' "happiness": transitions can also be a transitions up or down in the agents' preference orderings.

3.3 Cooperation Logic with Actions and Preferences

Now, we introduce agents as actors by combining the environment models with preferences with agents models. The resulting model is then called a multi-agent system with preferences (henceforth MaSP).

Definition 3.6 (Multi-agent System with Preferences)**.** A multi-agent system with preferences (MaSP) $M^{\preceq}$ is a tuple

$$M^{\preceq} = \langle S, Ac, (\rightarrow)_{A \subseteq Ac}, Ag, \mathsf{act}, \{\preceq_i\}_{i \in Ag}, V \rangle,$$

where $\langle S, Ac, (\rightarrow)_{A \subseteq Ac}, V, Ag, \mathsf{act}\rangle$ is a MaS, $\langle S, Ag, \{\preceq_i\}_{i \in Ag}, V\rangle$ is a preference model and $\langle S, Ac, (\rightarrow)_{A \subseteq Ac}, \{\preceq_i\}_{i \in Ag}, V\rangle$ is an environment with preferences.

Remark 1. *Note that given a deterministic MaSP in which each preference relation $\preceq_i$ is total, we can consider each state s as having a strategic game (Osborne and Rubinstein 1994) $\mathcal{G}_s$ attached to it.*

$$\mathcal{G}_s = \langle Ag, (\mathcal{P}(\mathsf{act}(i)))_{i \in Ag}, (\lesssim_i)_{i \in Ag}\rangle,$$

$\times_{i=1}^{n} A_i \lesssim_i \times_{i=1}^{n} A'_i$ *iff* $t \preceq_i t'$ *for* $s \rightarrow_{\bigcup_{i \in Ag} A_i} t$ *and* $s \rightarrow_{\bigcup_{i \in Ag} A'_i} t'$.

Next, we introduce two expressions saying that a group can force the system to move into a φ-state that some agent (strictly) prefers.

Definition 3.7 (Language $\mathcal{L}_{cla+p}$)**.** The language $\mathcal{L}_{cla+p}$ extends $\mathcal{L}_{cla}$ by formulas of the form

$$\Diamond^{\preceq_i}\varphi \mid \Diamond^{\prec_i}\varphi \mid \alpha \subseteq \preceq_i \mid \alpha \subseteq \prec_i \mid \langle\!\lbrack G^{\preceq_i} \rbrack\!\rangle\varphi \mid \langle\!\lbrack G^{\prec_i} \rbrack\!\rangle\varphi.$$

The first four expressions are interpreted as in the environment logic with preferences and for the last two we have the following.

$M^{\preceq}, s \models \langle\!\lbrack G^{\triangleleft_i} \rbrack\!\rangle\varphi$ iff $\exists A \subseteq \mathsf{act}(G)$: $\forall B \subseteq \mathsf{act}(Ag \setminus G), t \in S$: if $s \rightarrow_{A \cup B} t$, then $M^{\preceq}, t \models \varphi$ and $s \triangleleft_i t$.

Let us now look at how coalitional power to achieve an improvement for an agent is made explicit in CLA+P. We can show that $\langle\!\![G^{\lhd_i}]\!\!\rangle\varphi$ is equivalent to the existence of an action expression α that G can force such that all transitions of type α lead to a φ-state preferred by i.

Observation 1. Given a MaSP $M^{\preceq}$ and a state s of its environment,

$M^{\preceq}, s \vDash \langle\!\![G^{\lhd_i}]\!\!\rangle\varphi$ iff there exists an action expression α such that $M^{\preceq}, s \vDash \langle\!\![G]\!\!\rangle\alpha \wedge [\alpha]\varphi \wedge (\alpha \subseteq \lhd_i)$.

Proof. Analogous to that of Observation 14 in Sauro et al. (2006). Use the action expression $\bigwedge \Phi(A, G) := \bigwedge(A \cup \{\neg a | a \in (\mathsf{act}(G) \setminus A), a \notin \mathsf{act}(Ag \setminus G)\})$ for A being the "witness" of G's ability to force φ. □

Now we need axioms establishing a relationship between the newly added formulas and the expressions of the sublogics.

Definition 3.8 (Cooperation Logic with Actions and Preferences)**.** Λ^{CLA+P} is defined to be the smallest logic generated by the axioms of the cooperation logic with actions, the environment logic with preferences and

1. $(\langle\!\![G]\!\!\rangle\alpha \wedge [\alpha]\varphi \wedge (\alpha \subseteq \lhd_i)) \to \langle\!\![G^{\lhd_i}]\!\!\rangle\varphi$,
2. $\langle\!\![G^{\preceq_i}]\!\!\rangle\varphi \to \bigvee\{\langle\!\![G]\!\!\rangle\alpha \wedge [\alpha]\varphi \wedge (\alpha \subseteq \preceq_i) | \alpha$ is a conjunction of action literals$\}$,
3. $\langle\!\![G^{\prec_i}]\!\!\rangle\varphi \to \bigvee\{\langle\!\![G]\!\!\rangle\alpha \wedge [\alpha]\varphi \wedge (\alpha \subseteq \prec_i) | \alpha$ is a conjunction of action literals$\}$.

The inference rules are modus ponens, necessitation for action modalities and preference modalities ($\Box^{\preceq_i}, \Box^{\prec_i}$), substitution of logical equivalents, PREF − ACT and STRICT PREF − ACT.

Theorem 3. *The logic Λ^{CLA+P} is sound and complete with respect to the class of MaSP's.*

Proof. Soundness of the axioms is straightforward and completeness follows from completeness of the sublogics. □

3.4 Expressivity of CLA+P

We now show that in CLA+P, we can express some concepts relevant for reasoning about game-like interaction in multi-agent systems.

Stability. Given a MaSP, $M^{\preceq} = \langle S, Ac, (\to_A)_{A \subseteq Ac}, Ag, \mathsf{act}, \{\preceq_i\}_{i \in Ag}, V\rangle$, the following formula characterizes the states that are *individually stable (group stable)*, i.e. no individual (group) has the power to achieve a strict improvement (for all its members).

$$\psi_{ind.\ stable} := \bigwedge_{i \in Ag} \neg\langle\!\![\{i\}^{\prec_i}]\!\!\rangle\top.$$

$$\psi_{gr.\ stable} := \bigwedge_{G \subseteq Ag} \bigwedge_{A \subseteq \mathsf{act}(G)} \left(\bigvee_{i \in G} \neg\left(\left(\bigwedge \Phi(A, G)\right) \subseteq \prec_i\right)\right).$$

Dictatorship. We can express that an agent d is a (strong) dictator in the sense that coalitions can only achieve what d (strictly) prefers.

$$\psi_{d=dict.} := \bigwedge_{G \subseteq Ag} \bigwedge_{A \subseteq \mathsf{act}(G)} \Big(\big(\bigwedge \Phi(G, A) \big) \subseteq \lhd_d \Big).$$

Then, we can also say that there is *no* (strong) dictator:

$$\psi_{no\ dict.} := \bigwedge_{i \in Ag} \neg \left(\bigwedge_{G \subseteq Ag} \bigwedge_{A \subseteq \mathsf{act}(G)} \Big(\big(\bigwedge \Phi(G, A) \big) \subseteq \lhd_i \Big) \right).$$

Enforcing Unanimity. In some situations we might want to impose the condition on a MaSP that groups should only be able to achieve something if they can do so by making all its members happy:

$$\langle\!\lbrack G \rbrack\!\rangle \varphi \to \left(\bigvee_{A \subseteq \mathsf{act}(G)} \left(\bigwedge_{i \in G} \Big(\big(\bigwedge \Phi(A, G) \big) \subseteq \prec_i \Big) \wedge \Big[\bigwedge \Phi(A, G) \Big] \varphi \right) \right).$$

Note that the length of the last four formulas is exponential in the number of agents (and atomic actions).

3.5 CLA+P and Coalition Logic

Let us now briefly discuss the relation between CLA+P and Coalition Logic (CL) (Pauly 2002a) in order to illustrate how CLA+P builds upon existing frameworks for reasoning about coalitional power and how exactly the underlying actions that are only implicitly represented in the semantics of CL are made explicit. Given a fixed set of agents Ag, a coalition model $M = \langle (S, E), V \rangle$ with S being a set of states, $E : S \to (\mathcal{P}(Ag) \to \mathcal{P}(\mathcal{P}(S)))$ being a playable effectivity function and V being a propositional valuation, we can we use Theorem 3.2 of Pauly (2002a) and obtain a corresponding *game frame*, i.e. each state has an associated strategic game form in which each outcome corresponds to some accessible state. Looking back at Remark 1, it is now easy to see how we can construct a corresponding MaS: We take the same set of states and add actions for each of the strategies in the attached games and define the accessibility relation in accordance with the outcome function. Finally, we can define a translation τ of formulas of CL to those of CLA in a straightforward way: formulas $[G]\varphi$ talking about coalitional power in CL are translated into $\langle\!\lbrack G \rbrack\!\rangle \tau(\varphi)$.

If we add preferences to the game forms in the game frame we obtained from the coalition model, then we can transform it into a MaSP in an analogous way.

This shows that the framework of CLA+P is a natural way to make coalitional power as modelled in CL and its extensions more explicit.

4 Complexity of CLA+P

In this section, we analyze the complexity of SAT of CLA+P.

4.1 Decidability of CLA+P

We show that SAT of CLA+P is decidable. The first step is to show that only a restricted class of models of CLA+P needs to be checked.

We start by looking at how we can restrict the class of models with respect to the set of agents. Let $Ag(\varphi)$ denote the set of agents occurring in φ. Now, we ask: Is every satisfiable φ also satisfiable in a MaSP with set of agents $Ag(\varphi)$? In Coalition Logic, the answer is negative: the formula $\varphi' = \neg\langle\!\langle G\rangle\!\rangle\{1\}p \wedge \neg\langle\!\langle\{1\}\rangle\!\rangle q \wedge \langle\!\langle\{1\}\rangle\!\rangle(p \vee q)$ is only satisfiable in models with at least two agents (Pauly 2002b). However, as in CLA+P the environment models can be nondeterministic, here φ' can indeed be satisfied in a model with only one agent, as the reader can check.

It can be shown that every satisfiable formula $\varphi \in \mathcal{L}_{cla+p}$ is satisfiable in a MaSP with set of agents $Ag(\varphi) \cup \{k\}$, for k being a new agent. k takes the role of all opponents of $Ag(\varphi)$ collapsed into one: k gets the ability to perform exactly the actions that agents not occurring in φ can perform as a group.

Theorem 4. *Every satisfiable formula $\varphi \in \mathcal{L}_{cla+p}$ is satisfiable in the class of MaSP's with at most $|Ag(\varphi)| + 1$ many agents.*

Proof. Assume that $M^{\preceq} = \langle S, Ac, (\rightarrow_A)_{A\subseteq Ac}, Ag, \mathsf{act}, \{\preceq_i\}_{i\in Ag}, V\rangle$ satisfies φ. If $Ag \supset Ag(\varphi)$, we construct $M'^{\preceq'} = \langle S, Ac, (\rightarrow)_{A\subseteq Ac}, Ag(\varphi) \cup \{k\}, \mathsf{act}', \{\preceq'_i\}_{i\in Ag(\varphi)\cup\{k\}}, V\rangle$, with $\mathsf{act}'(k) = \bigcup_{j\in Ag\setminus Ag(\varphi)} \mathsf{act}(j)$ and $\mathsf{act}'(i) = \mathsf{act}(i)$ for $i \neq k$. The preferences are defined as follows: $\preceq'_i = \preceq_i$ for $i \in Ag(\varphi)$ and $\preceq'_k = S \times S$. By induction, we can show that $M^{\preceq}, s \models \varphi$ iff $M'^{\preceq'}, s \models \varphi$. The case where φ is of the form $\langle\!\langle G\rangle\!\rangle\alpha$ follows from the definition of act'. Then the other cases involving coalition modalities follow. □

Next, we want to know how many actions a model needs for satisfying some formula. Consider e.g. $\varphi = \langle\!\langle G\rangle\!\rangle(p \wedge q) \wedge \langle\!\langle G\rangle\!\rangle(\neg p \wedge q) \wedge \langle\!\langle G\rangle\!\rangle(\neg p \wedge \neg q)$. It is only satisfiable in models with $|Ac| \geq 2$. The main task is to find "witnesses" for formulas of the form $\langle\!\langle G\rangle\!\rangle\psi$ in terms of concurrent actions. We can show that every satisfiable φ is satisfiable in a MaSP whose set of atomic actions consists of those in φ, one additional one (a dummy for ensuring that each agent can perform an action), and for every subformula $\langle\!\langle G\rangle\!\rangle\psi$ or $\langle\!\langle G^{\lhd_i}\rangle\!\rangle\psi$, one action for each of G's members.

The key step in transforming a model satisfying a formula φ into one whose set of actions satisfies the above condition is to appropriately define the action distribution and the accessibility relations. For every α occurring in φ, we have to ensure that two states are related by an α-transition in the new model iff they were in the original one. Additionally, for formulas $\langle\!\langle G\rangle\!\rangle\psi$ and $\langle\!\langle G^{\lhd_i}\rangle\!\rangle\psi$, the set of actions introduced for them serves for making explicit how G can force φ.

Theorem 5. *Every satisfiable formula $\varphi \in \mathcal{L}_{cla+p}$ is satisfiable in a MaSP with at most $|Ac(\varphi)| + (\sum_{\langle\!\langle G\rangle\!\rangle\psi\in Sub(\varphi)}|G|) + (\sum_{\langle\!\langle G^{\preceq_i}\rangle\!\rangle\psi\in Sub(\varphi)}|G|) + (\sum_{\langle\!\langle G^{\prec_i}\rangle\!\rangle\psi\in Sub(\varphi)}|G|) + 1$ many actions.*

Proof. Assume that $M^{\preceq} = \langle S, Ac, (\rightarrow_A)_{A\subseteq Ac}, Ag, \mathsf{act}, \{\preceq_i\}_{i\in Ag}, V\rangle$ satisfies φ. We construct a model $M'^{\preceq'} = \langle S, Ac', (\rightarrow')_{A'\subseteq Ac'}, Ag, \mathsf{act}', \{\preceq'_i\}_{i\in Ag}, V\rangle$ as follows.

$$Ac' := Ac(\varphi) \cup \bigcup_{\langle\!\langle G\rangle\!\rangle\psi\in Sub(\varphi)} A_{G\psi} \cup \bigcup_{\langle\!\langle G^{\preceq_i}\rangle\!\rangle\psi\in Sub(\varphi)} A_{G^{\preceq_i}\psi} \cup \bigcup_{\langle\!\langle G^{\prec_i}\rangle\!\rangle\psi\in Sub(\varphi)} A_{G^{\prec_i}\psi} \cup \{\hat{a}\}.$$

$A_{G\psi}$ and $A_{G^{\lhd_i}\psi}$ consist of newly introduced actions $a_{G\psi j}$, and $a_{G^{\lhd_i}\psi j}$ respectively, for each $j \in G$. Action abilities are distributed as follows:

$\mathsf{act}'(i) :=$ $(\mathsf{act}(i) \cap Ac(\varphi)) \cup \{\hat{a}\} \cup \{a_{Gi} | \langle\!\langle G \rangle\!\rangle \psi \in Sub(\varphi)$ or $\langle\!\langle G^{\lhd_i} \rangle\!\rangle \psi \in Sub(\varphi)$, for $i \in G\}$.

For defining the accessibility relation $\rightarrow_{A' \subseteq Ac'}$, we first define for any state s its set of successors.

$t \in T^s_{A'}$ iff the following conditions are satisfied:

1. $\forall [\alpha]\psi \in Sub(\varphi)$ such that $A' \models \alpha$: If $M^{\preceq}, s \models [\alpha]\psi$, then $M^{\preceq}, t \models \psi$,
2. $\forall \alpha \subseteq \lhd_i \in Sub(\varphi)$ such that $A' \models \alpha$: If $M^{\preceq}, s \models \alpha \subseteq \lhd_i$, then $s \lhd_i t$,
3. $\forall \langle\!\langle G \rangle\!\rangle \psi \in Sub(\varphi)$ such that $A' \models \bigwedge \Phi(A_{G\psi}, G)$, there is some $\bar{A} \subseteq \mathsf{act}(G)$ such that $s \rightarrow_A t$ for some $A \subseteq Ac$ such that $A \models \bigwedge \Phi(\bar{A}, G)$, and if $M^{\preceq}, s \models \langle\!\langle G \rangle\!\rangle \psi$ then $M^{\preceq}, s \models [\bigwedge \Phi(\bar{A}, G)]\psi$
4. $\forall \langle\!\langle G^{\lhd_i} \rangle\!\rangle \psi \in Sub(\varphi)$ such that $A' \models \bigwedge \Phi(A_{G^{\lhd_i}\psi}, G)$, there is some $\bar{A} \subseteq \mathsf{act}(G)$ such that $s \rightarrow_A t$ for some $A \subseteq Ac$ such that $A \models \bigwedge \Phi(\bar{A}, G)$, and if $M^{\preceq}, s \models \langle\!\langle G^{\lhd_i} \rangle\!\rangle \psi$ then $M^{\preceq}, s \models [\bigwedge \Phi(\bar{A}, G)]\psi$ and $M^{\preceq}, s \models (\bigwedge \Phi(\bar{A}, G) \subseteq \lhd_i)\}$.

For any $t \in T^s_{A'}$, we set $s \rightarrow'_{A'} t$. Then we can show by induction on $\psi \in Sub(\varphi)$ that $M^{\preceq}, s \models \psi$ iff $M'^{\preceq'}, s \models \psi$. □

The next step is to show that every satisfiable formula φ is satisfiable in a model with a certain number of states. Such results are usually obtained by transforming a model into a smaller one using a transformation that preserves the truth of subformulas of φ.

In the case of CLA+P, the irreflexivity of the strict preferences and the fact that also $\alpha \subseteq \preceq_i$ is not modally definable in a basic modal language call for a modification of the standard techniques.

We appropriately modify the method of *filtration* (Blackburn et al. 2001) and show that any satisfiable formula $\varphi \in \mathcal{L}_{cla+p}$ is satisfiable in a model with exponentially many states. The idea of a filtration is to transform a possibly infinite model into a finite one by identifying states that agree on the truth value of each subformula of the considered formula. So, given that we know that φ is satisfied in some MaSP $M^{\preceq}$ with states S, we construct an MaSP $\mathcal{M}^{\preceq^f}$ with set of states $S_{Sub(\varphi)} = \{|s|_{Sub(\varphi)} | s \in S\}$, where $|s|_{Sub(\varphi)}$ denotes the equivalence class of the states that in the model M agree with s on the truth values of all $\psi \in Sub(\varphi)$. The main task is to appropriately define the accessibility relations for actions and preferences in $\mathcal{M}^{\preceq^f}$ such that for $\psi \in Sub(\varphi)$, we then have that $M^{\preceq}, s \models \psi$ iff $\mathcal{M}^{\preceq^f}, |s| \models \psi$. Here, it is important to note that formulas of the form $\langle\!\langle G \rangle\!\rangle \psi$ and $\langle\!\langle G^{\lhd_i} \rangle\!\rangle \psi$ correspond to formulas of the form $\bigvee_{A \subseteq \mathsf{act}(G)} [\bigwedge \Phi(A, G)]\psi$ and $\bigvee_{A \subseteq \mathsf{act}(G)} ([\bigwedge \Phi(A, G)]\psi \wedge (\bigwedge \Phi(A, G) \subseteq \lhd_i))$, respectively – for $\bigwedge \Phi(A, G)$ as in the proof of Observation 1. Moreover, the transformation of the model does not change the underlying agents model. Thus, the truth of formulas of the form $\langle\!\langle G\rangle\!\rangle\alpha$ is preserved.

Theorem 6. *Every satisfiable $\varphi \in \mathcal{L}_{cla+p}$ is also satisfiable in a MaSP with $\leq 2^{|\varphi|}$ many states.*

Proof. Given that $M^{\preceq}, s \vDash \varphi$ for some $M^{\preceq} = \langle S, Ac, (\rightarrow_A)_{A\subseteq Ac}, Ag, \mathsf{act}, \{\preceq_i\}_{i\in Ag}, V\rangle$ and $s \in S$, we obtain $\mathcal{M}^{\preceq^f} = \langle S_{Sub(\varphi)}, Ac, (\rightarrow^f)_{A\subseteq Ac}, Ag, \mathsf{act}^f, \{\preceq_i^f\}_{i\in Ag}, V^f\rangle$ by filtrating $M^{\preceq}$ through $Sub(\varphi)$, where the accessibility relations for actions and preferences are defined as follows:

$|s| \rightarrow_A^f |t|$ iff the following conditions are satisfied:

1. $\forall[\alpha]\psi \in Sub(\varphi)$ such that $A \vDash \alpha$: if $M^{\preceq}, s \vDash [\alpha]\psi$, then $M^{\preceq}, t \vDash \psi$,
2. (a) $\forall\alpha \subseteq\preceq_i \in Sub(\varphi)$ such that $A \vDash \alpha$: if $M^{\preceq}, s \vDash \alpha \subseteq\preceq_i$, then $s \preceq_i t$,
   (b) $\forall\alpha \subseteq\prec_i \in Sub(\varphi)$ such that $A \vDash \alpha$: if $M^{\preceq}, s \vDash \alpha \subseteq\prec_i$, then $s \prec_i t$,
3. $\forall\langle\!\langle G\rangle\!\rangle\psi \in Sub(\varphi)$ such that $A \vDash \bigwedge\Phi(A', G)$ for some $A' \subseteq \mathsf{act}(G)$: if $M^{\preceq}, s \vDash [\bigwedge\Phi(A', G)]\psi$, then $M^{\preceq}, t \vDash \psi$,
4. (a) $\forall\langle\!\langle G^{\preceq_i}\rangle\!\rangle\psi \in Sub(\varphi)$ such that $A \vDash \bigwedge\Phi(A', G)$ for some $A' \subseteq \mathsf{act}(G)$: if $M^{\preceq}, s \vDash [\bigwedge\Phi(A', G)]\psi$ and $M^{\preceq}, s \vDash (\bigwedge\Phi(A', G) \subseteq \preceq_i)$, then $M^{\preceq}, t \vDash \psi$ and $s \preceq_i t$.
   (b) $\forall\langle\!\langle G^{\prec_i}\rangle\!\rangle\psi \in Sub(\varphi)$ such that $A \vDash \bigwedge\Phi(A', G)$ for some $A' \subseteq \mathsf{act}(G)$: if $M^{\preceq}, s \vDash [\bigwedge\Phi(A', G)]\psi$ and $M^{\preceq}, s \vDash (\bigwedge\Phi(A', G) \subseteq \prec_i)$, then $M^{\preceq}, t \vDash \psi$ and $s \prec_i t$.

$|s| \preceq_i^f |t|$ iff the following conditions hold:

1. (a) $\forall\Diamond^{\preceq_i}\psi \in Sub(\varphi)$: if $M^{\preceq}, t \vDash \psi \vee \Diamond^{\preceq_i}\psi$ then $M^{\preceq}, s \vDash \Diamond^{\preceq_i}\psi$,
   (b) If there is some $\Diamond^{\prec_i}\psi \in Sub(\varphi)$, then $s \preceq_i t$,
2. If there is some $\alpha \subseteq\preceq_i \in Sub(\varphi)$ or some $\alpha \subseteq\prec_i \in Sub(\varphi)$, then $s \preceq_i t$,
3. If there is some $\langle\!\langle G^{\preceq_i}\rangle\!\rangle\psi \in Sub(\varphi)$ or some $\langle\!\langle G^{\prec_i}\rangle\!\rangle\psi \in Sub(\varphi)$, then $s \preceq_i t$.

$V^f(p) := \{|s| | M, s \vDash p\}$, for all propositional letters $p \in Sub(\varphi)$. We can show by induction that for all $\psi \in Sub(\varphi)$ and $s \in S$ it holds that $M^{\preceq}, s \vDash \psi$ iff $M^{\preceq^f}, |s| \vDash \psi$. This follows from the definitions of $(\rightarrow^f)_{A\subseteq Ac}$ and $\preceq^f$, and the fact that we do not change the underlying agents model. The interesting cases are those involving strict preferences and those with formulas $\langle\!\langle G^{\preceq_i}\rangle\!\rangle\psi$ and $\alpha \subseteq\preceq_i$. Here, what makes the proof go through is that by conditions 1 b), 2 and 3 of $\preceq^f$, $|s| \preceq_i |t|$ implies $s \preceq_i t$. Similarly, due to conditions 2 and 4 of $\rightarrow_A^f$, the truth values of subformulas $\alpha \subseteq \lhd_i$ and $\langle\!\langle G^{\lhd_i} \rangle\!\rangle \psi$ is as in the original model. Moreover, $\mathcal{M}^{\preceq^f}$ is a proper MaSP since each $\preceq_i^f$ is reflexive and transitive, and each $\rightarrow_A^f$ is serial. By definition of $S_{Sub(\varphi)}$, $|S_{Sub(\varphi)}| \leq 2^{|\varphi|}$. $\square$

Now, we apply the constructions of the last three proofs successively.

Corollary 1. *Every satisfiable formula $\varphi \in \mathcal{L}_{cla+p}$ is satisfiable in a MaSP of size exponential in $|\varphi|$ satisfying the conditions $|Ag| \leq |Ag(\varphi)| + 1$ and $|Ac| \leq |Ac(\varphi)| + \sum_{\langle\!\langle G \rangle\!\rangle \psi \in Sub(\varphi)} |G| + (\sum_{\langle\!\langle G^{\preceq_i} \rangle\!\rangle \psi \in Sub(\varphi)} |G|) + (\sum_{\langle\!\langle G^{\prec_i} \rangle\!\rangle \psi \in Sub(\varphi)} |G|) + 1$.*

Having non-deterministically guessed a model of size exponential in $|\varphi|$, we can check in time exponential in $|\varphi|$ whether this model satisfies φ.

Theorem 7. *The satisfiability problem of CLA+P is in* NEXPTIME.

Proof. Given φ, we non-deterministically choose a model $M^{\preceq}$ of size exponential in $|\varphi|$ satisfying the conditions $|Ag| \leq |Ag(\varphi)| + 1$ and $|Ac| \leq |Ac(\varphi)| + \sum_{\langle\!\langle G \rangle\!\rangle \psi \in Sub(\varphi)} |G| + (\sum_{\langle\!\langle G^{\preceq_i} \rangle\!\rangle \psi \in Sub(\varphi)} |G|) + (\sum_{\langle\!\langle G^{\prec_i} \rangle\!\rangle \psi \in Sub(\varphi)} |G|) + 1$. Then, given this model, we can check in time $O(|\varphi| ||M^{\preceq}||)$, for $|M^{\preceq}|$ being the size of $M^{\preceq}$, whether $M^{\preceq}$ satisfies φ. Thus, given a model of size exponential in $|\varphi|$ that also satisfies the conditions on its sets of agents and actions explained earlier, it can be computed in time exponential in $|\varphi|$ whether it satisfies φ. Since it can be checked in time linear in the size of the model whether it is a proper MaSP, we conclude that SAT of CLA+P is in NEXPTIME. $\square$

This section has shown that SAT of CLA+P is in NEXPTIME. Now we show that the environment logic is already EXPTIME-hard.

4.2 Lower Bound

In order to show a lower bound for the complexity of SAT of CLA+P, we show that SAT of the environment logic is EXPTIME-hard. This is done by reduction from the Boolean modal logic $\mathbf{K}_m^{\neg\cup}$ (Lutz and Sattler 2001, Lutz et al. 2001).

Formulas of $\mathbf{K}_m^{\neg\cup}$ are interpreted in models $M = \langle W, R_1, \ldots R_m, V \rangle$, where W is a set of states, $R_i \subseteq W \times W$ and V is a valuation.

Definition 4.1. Let $\mathcal{R}_1, \ldots \mathcal{R}_m$ be atomic modal parameters. Then the set of modal parameters of $\mathbf{K}_m^{\neg\cup}$ is the smallest set containing $\mathcal{R}_1, \ldots \mathcal{R}_m$ that is closed under $\neg$ and $\cup$. The language $\mathcal{L}_m^{\neg\cup}$ is generated by the following grammar:

$$\varphi ::= \quad p \mid \varphi \wedge \varphi \mid \neg\varphi \mid \langle \mathcal{S} \rangle \varphi \qquad\qquad \mathcal{S} ::= \quad \mathcal{R}_i \mid \neg\mathcal{S} \mid \mathcal{S}_1 \cup \mathcal{S}_2.$$

The extension $\mathcal{E}(\mathcal{S}) \subseteq W \times W$ of a parameter $\mathcal{S}$ in a model is as follows.

$$\begin{array}{lcl} \mathcal{E}(\mathcal{R}_i) & = & R_i \\ \mathcal{E}(\neg\mathcal{S}) & = & (W \times W) \setminus \mathcal{E}(S) \\ \mathcal{E}(\mathcal{S}_1 \cup \mathcal{S}_2) & = & \mathcal{E}(\mathcal{S}_1) \cup \mathcal{E}(\mathcal{S}_2) \end{array}$$

Formulas of $\mathcal{L}_m^{\neg\cup}$ are interpreted in a model $M = \langle W, R_1, \ldots R_m, V \rangle$ as follows: Propositional letters and boolean combinations are interpreted in the standard way and for modal formulas we have

$$M, w \models \langle \mathcal{S} \rangle \varphi \quad \text{ iff } \quad \exists w' \in W : (w, w') \in \mathcal{E}(\mathcal{S}) \text{ and } M, w' \models \varphi.$$

We define a translation τ consisting of two components τ_1 for formulas and τ_2 for models. Let us extend the language $\mathcal{L}_e$ by a propositional letter $q \notin \mathcal{L}_m^{\neg\cup}$. Then τ_1 is defined as follows:

$$\begin{array}{rcl} \tau_1(p) & = & p \\ \tau_1(\varphi_1 \wedge \varphi_2) & = & \tau_1(\varphi_1) \wedge \tau_1(\varphi_2) \\ \tau_1(\neg\varphi) & = & \neg\tau_1(\varphi) \\ \tau_1(\langle S\rangle\varphi) & = & \neg[\tau^S(\mathcal{S})](q \vee \neg\tau_1(\varphi)) \end{array} \qquad \begin{array}{rcl} \tau^S(\mathcal{R}_i) & = & a_i \\ \tau^S(\mathcal{S}_1 \cup \mathcal{S}_2) & = & \tau^S(\mathcal{S}_1) \vee \tau^S(\mathcal{S}_2) \\ \tau^S(\neg\mathcal{S}) & = & \neg\tau^S(\mathcal{S}) \end{array}$$

τ_2 translates a model M of $\mathbf{K}_m^{\neg\cup}$ into an environment model $\tau_2(M) = \langle W \cup \{u\}, Ac, (\rightarrow)_{A\subseteq Ac}, V'\rangle$ with u being a newly introduced state and

$$w \rightarrow_A w' \text{ iff } A = \{a_i | (w, w') \in R_i\} \text{ or } w' = u.$$

Thus, each $\rightarrow_A$ is serial and $\tau_2(M)$ is an environment model. $V'(q) = \{u\}$, and for all $p \neq q$, $V'(p) = V(p)$. Before showing that for any $\varphi \in \mathcal{L}_m^{\neg\cup}$ and $M \in \mathbb{M}_m^{\neg\cup}$ for any state $w \in W : M, w \models \varphi$ iff $\tau_2(M), w \models \tau_1(\varphi)$, we proof a lemma saying that if in M w' is $\mathcal{S}$-accessible from w, then in $\tau_2(M)$, w' is accessible from w by a transition of type $\tau^S(\mathcal{S})$.

Convention 2. *For $M = \langle W, R_1, \ldots R_m, V\rangle \in \mathbb{M}_m^{\neg\cup}$ and $\tau_2(M) = \langle W \cup \{u\}, Ac, (\rightarrow)_{A\subseteq Ac}, V'\rangle$, define $A_{w,w'} := \{a_i \in Ac | (w, w') \in R_i\}$.*

Lemma 1. *Let $M = \langle W, R_1, \ldots R_m, V\rangle$ be a model of $\mathbf{K}_m^{\neg\cup}$. Then for any modal parameter $\mathcal{S}$ and for any states $w, w' \in W$ it holds that*

$$(w, w') \in \mathcal{E}(\mathcal{S}) \textit{ iff in } \tau_2(M) : \exists A \subseteq Ac : w \rightarrow_A w' \textit{ and } A \models \tau^S(\mathcal{S}).$$

Proof. Note that by definition of $(\rightarrow_A)_{A\subseteq Ac}$, the righthand side is equivalent to $A_{w,w'} \models \tau^S(\mathcal{S})$. Then the proof goes by induction on $\mathcal{S}$. □

Theorem 8. *For any formula $\varphi \in \mathcal{L}_m^{\neg\cup}$ and any model M of $\mathbf{K}_m^{\neg\cup}$, it holds that for any state w in M:*

$$M, w \models \varphi \textit{ iff } \tau_2(M), w \models \tau_1(\varphi).$$

Proof. By induction. Base case and boolean cases are straightforward. Let $\varphi = \langle S\rangle\psi$.

$(\Rightarrow)$ $M, w \models \langle S\rangle\psi \Leftrightarrow \exists w' : (w, w') \in \mathcal{E}(\mathcal{S})$ and $M, w' \models \psi$. By the previous lemma and induction hypothesis, $\tau_2(M), w \models \tau_1(\langle S\rangle\psi)$.

$(\Leftarrow)$ $\tau_2(M), w \models \tau_1(\langle S\rangle\psi) \Leftrightarrow \tau_2(M), w \models \neg[\tau^S(\mathcal{S})](q\neg\tau_1(\psi))$. Then $\exists w' \in W \cup \{u\}, \exists A \in Ac : A \models \tau^S(\mathcal{S}), w \rightarrow_A w'$ and $\tau_2(M), w' \models \neg q \wedge \tau_1(\psi)$. Thus, $w' \neq u$. By induction hypothesis and the previous lemma, $M, w' \models \psi$ and $(w, w') \in \mathcal{E}(\mathcal{S})$. Hence, $M, w \models \langle S\rangle\psi$.

□

Theorem 9. SAT *of Λ^E is* EXPTIME*-hard.*

Proof. We can polynomially transform any $\varphi \in \mathcal{L}_m^{\neg\cup}$ into $\tau_1(\varphi)$. Now, if $\tau_1(\varphi)$ is satisfiable in an environment model $E = \langle W, Ac, (\rightarrow)_{A\subseteq Ac}, V\rangle$ with $Ac = \{a_1, \ldots a_m\}$, then φ is satisfiable in a model $M = \langle W, R_1, \ldots R_m, V\rangle$, where $(w, w') \in R_i$ iff $\exists A \subseteq Ac$ such that $a_{i\in Ac}$ and $w \rightarrow_A w'$. This can be shown by induction on φ. If $\tau_1(\varphi)$ is not satisfiable in an environment model then it cannot be satisfiable in any model M of $\mathbf{K}_m^{\neg\cup}$ because otherwise by the previous theorem, $\tau_1(\varphi)$ would be satisfied in $\tau_2(M)$. □

Corollary 2. SAT *of CLA+P is* EXPTIME*-hard.*

This section has shown that the satisfiability problem of CLA+P is EXPTIME-hard but still decidable. This rather high complexity is due to the environment logic which itself is already EXPTIME-hard.

5 Conclusions and Future Work

We developed a modular modal logic that allows for reasoning about the coalitional power of agents, actions and their effects, and agents' preferences. The current approach is based on the logic CLA (Sauro et al. 2006) which is combined with a preference logic (van Benthem et al. 2007). The resulting logic CLA+P, which is shown to be sound and complete, allows us to make explicit how groups can achieve certain results. We can also express how a group can achieve that a transition takes place that is an improvement for some agent. In the framework of CLA+P, it can be expressed how the abilities to perform certain actions are distributed among the agents, what are the effects of the concurrent performance of these actions and what are the agents' preferences over those effects. Moreover, in CLA+P, we can distinguish between different ways how groups can achieve some result – not only with respect to the actions that lead to some result, but also with respect to the preferences. We can for instance express that a group can achieve some result in a way that is 'good' for all its members in the sense that after the achievement all of them are better off. This then also allows us to axiomatize properties that one might want to impose onto a multi-agent system, e.g. the restriction that groups can only achieve the truth of a certain formula if this can be done without making anybody worse off. Thus, CLA+P provides a framework for reasoning about interactive situations in an explicit way that gives us more insights into the cooperative abilities of agents. Comparing CLA+P to CL shows that CLA+P naturally builds on game frames underlying the semantics of CL and makes both the agents' actions and the preferences explicit that are only implicitly represented in the semantics of CL.

The satisfiability problem of CLA+P is shown to be decidable and EXPTIME-hard. Keeping in mind that using CLA+P we can talk about strict preferences, intersections of accessibility relations as well as the property of one relation being a subset of another, EXPTIME-hardness is not surprising. Even though the modular models of CLA+P are rather special, its complexity is in accordance with general results concerning the connection between expressive power and complexity of modal logics for reasoning about coalitional power and preferences (Dégremont and Kurzen 2009).

We showed that the satisfiability problem of the underlying environment logic is by itself already EXPTIME-hard. Thus, we identified one cause for the high complexity of CLA+P. It is mostly due to the fact that the accessibility relation of the models can be arbitrary: there does not need to be any relation between $\rightarrow_A$, $\rightarrow_B$ and $\rightarrow_{A\cap B}$. Whereas this generality allows us to model a lot of dynamic processes, from a computational viewpoint, it seems to be appealing to change the environment logic in order to decrease computational complexity. Also, when comparing our models to the game frames of CL, we can see that restricting ourselves to deterministic environment models can be reasonable. The same holds for considering only total preorders as preference relation. This would also increase the expressive power (Dégremont and Kurzen 2009).

There are several immediate ways to extend the logic developed in this paper. First of all, we can follow the ideas of Ågotnes et al. (2007b) and add a restricted form of quantification that allows statements of the form $\langle\!\langle P^{\preceq_i}\rangle\!\rangle\,\psi$ saying that there is some group G that has property P and $\langle\!\langle G^{\preceq_i}\rangle\!\rangle\,\psi$. In the current work, we chose a rather simple preference logic with unary modalities. In order to increase the expressive power with respect to the agents' preferences over outcomes that can be achieved, it seems very promising to consider more expressive preference logics that also contain binary preference modalities saying e.g. that an agent prefers every φ state over every ψ state (van Benthem et al. 2007).

Moreover, it will be interesting to develop a cooperation logic with actions and preferences based on a logic for reasoning about complex plans such as the the one developed by Gerbrandy and Sauro (2007).

Acknowledgements The author thanks Ulle Endriss, Eric Pacuit, Luigi Sauro, Johan van Benthem and Olivier Roy for their help and their useful comments. She also thanks the participants of the workshop *Logic and Intelligent Interaction* at *ESSLLI 2008* and the participants of the ILLC's seminar *Logics for Dynamics of Information and Preferences*. Their comments have helped to improve earlier versions.

References

T. Ågotnes, P. Dunne, W. van der Hoek, and M. Wooldridge. Logics for coalitional games. In J. van Benthem, S. Ju, and F. Veltman, editors, *A Meeting of the Minds*, number 8 in Texts in Computer Science, pages 3–20, London, UK, 2007a. College Publications.

T. Ågotnes, W. van der Hoek, and M. Wooldridge. Quantified coalition logic. In *IJCAI '07: Proceedings of the twentieth international joint conference on Artificial Intelligence*, pages 1181–1186, Hyderabad, India, 2007b.

R. Alur, T. A. Henzinger, and O. Kupferman. Alternating-time temporal logic. *Lecture Notes in Computer Science*, 1536:23–60, 1998.

P. Blackburn, M. de Rijke, and Y. Venema. *Modal Logic*. Number 53 in Cambridge Tracts in Theoretical Computer Science. Cambridge University Press, UK, 2001.

S. Borgo. Coalitions in action logic. In *IJCAI '07: Proceedings of the twentieth international joint conference on Artificial Intelligence*, pages 1822–1827, Hyderabad, India, 2007.

J. Broersen. *Modal Action Logics for Reasoning about Reactive Systems*. PhD thesis, Faculteit der Exacte Wetenschappen, Vrije Universiteit Amsterdam, 2003.

C. Dégremont and L. Kurzen. Modal logics for preferences and cooperation: Expressivity and complexity. In *KRAMAS 2008: Proceedings of the Workshop on Knowledge Representation for Agents and Multi-Agent Systems*, 2009. to appear.

J. Gerbrandy and L. Sauro. Plans in cooperation logic: a modular approach. In *Proceedings of the IJCAI Workshop on Nonmontonic Reasoning, Action and Change (NRAC 2007)*, Hyderabad (India), 2007.

L. Kurzen. Logics for Cooperation, Actions and Preferences. Master's thesis, Universiteit van Amsterdam, the Netherlands, 2007.

C. Lutz and U. Sattler. The complexity of reasoning with boolean modal logics. In F. Wolter, H. Wansing, M. de Rijke, and M. Zakharyaschev, editors, *Advances in Modal Logics Volume 3*. CSLI Publications, Stanford, 2001.

C. Lutz, U. Sattler, and F. Wolter. Modal logics and the two-variable fragment. In *Annual Conference of the European Association for Computer Science Logic CSL'01*, LNCS, Paris, France, 2001. Springer Verlag.

M. J. Osborne and A. Rubinstein. *A course in game theory*. MIT Press, Cambridge, MA, 1994.

M. Pauly. A modal logic for coalitional power in games. *Journal of Logic and Computation*, 12(1):149–166, 2002a.

M. Pauly. On the complexity of coalitional reasoning. *International Game Theory Review*, 4:237–254, 2002b.

L. Sauro, J. Gerbrandy, W. van der Hoek, and M. Wooldridge. Reasoning about action and cooperation. In *AAMAS '06: Proceedings of the fifth International Joint Conference on Autonomous Agents and Multi-agent Systems*, pages 185–192, Hakodate, Japan, 2006.

J. van Benthem, S. van Otterloo, and O. Roy. Preference logic, conditionals and solution concepts in games. In H. Lagerlund, S. Lindström, and R. Sliwinski, editors, *Modality Matters: Twenty-Five Essays in Honour of Krister Segerberg*, Uppsala Philosophical Studies 53. University of Uppsala, 2006.

J. van Benthem, O. Roy, and P. Girard. Everything else being equal: A modal logic approach to ceteris paribus preferences. *ILLC Prepublication Series, PP-2007-09*, 2007.

S. van Otterloo, W. van der Hoek, and M. Wooldridge. Preferences in game logics. In *AAMAS '04: Proceedings of the Third International Joint Conference on Autonomous Agents and Multiagent Systems*, pages 152–159. IEEE Computer Society, 2004.

D. Walther, W. van der Hoek, and M. Wooldridge. Alternating-time temporal logic with explicit strategies. In *TARK '07: Proceedings of the 11th conference on Theoretical aspects of rationality and knowledge*, pages 269–278. ACM, 2007.

Agreement Theorems in Dynamic-Epistemic Logic

Cédric Dégremont and Olivier Roy
Institute for Logic, Language and Computation, Universiteit van Amsterdam;
Faculty of Philosophy, University of Groningen.
cedric.uva@gmail.com, o.roy@rug.nl

Abstract

In this paper we bring Aumann's Agreement Theorem to dynamic-epistemic logic. We show that common *belief* of posteriors is sufficient for agreements in "epistemic-plausibility models", under common and well-founded priors, from which the usual form of agreement results, using common knowledge, follows. We do not restrict to the finite case, and show that in countable structures such results hold if *and only if* the underlying "plausibility ordering" is well-founded. We look at these results from a syntactic point of view, showing that neither well-foundedness nor common priors are expressible in a commonly used language, but that the static agreement result is finitely derivable in an extended modal logic. We finally consider "dynamic" agreement results, and show they have a counterpart in epistemic-plausibility models. We also show to which agreements one gets via "public announcements." A comparison of the two types of dynamic agreements reveals that they can indeed be different.

1 Introduction

In this paper we bring Aumann's Agreement Theorem Aumann (1976) and some of its subsequent extensions Geanakoplos and Polemarchakis (1982) and generalizations Cave (1983), Bacharach (1985) to dynamic-epistemic logic Baltag and Smets (2006), van Ditmarsch et al. (2007). We show that common *belief* of posteriors is sufficient for agreements in "epistemic-plausibility models", under common and well-founded priors, from which the usual form of agreement results, using common knowledge, follows. We do not restrict to the finite case, thus improving on known qualitative agreement theorems Bacharach (1985), and show that in countable structures such results hold if *and only if* the underlying "plausibility ordering" is well-founded. We then look at these results from a syntactic point of view, showing that neither well-foundedness nor common priors are expressible in the language proposed in Baltag and Smets (2008), even extended with a common belief operator, but we also show a finitary syntactic derivation of the static agreement result in an extended modal language. We finally consider "dynamic" agreement results. We show that "agreements via dialogues" Cave (1983), Bacharach (1985) have a counterpart in epistemic-plausibility models, and that one also gets agreements via "public announcements," a type of belief update that has so far not been considered in the agreement literature—see Bonanno and Nehring (1997) and Ménager (2006). A comparison of the two types of dynamic agreements reveals that in some situations they are indeed different.

These technical results answer an "internal" question for dynamic-epistemic logic, namely whether agreement results hold in this framework, but they also offer new insights on the contribution of agreement theorems to interactive epistemology. That common belief of posteriors is sufficient for agreements, under common and well-founded priors, strengthens one of the key lessons of agreement theorems, viz. that first-order information is highly dependent on higher-order information in situations of interaction Bonanno and Nehring (1997). Our inexpressibility results, on the other hand, support a qualm already voiced in the literature concerning the difficulty for agents to reason about static agreements Samet (In Press). The two dynamic results not only make a sharp distinction between two forms of belief changes, they also allow to more adequately capture the idea that agreements are reached via *public* dialogues. Bringing agreement theorems to dynamic-epistemic logic is thus of importance both technically and conceptually, and it helps to bridge the existing literature on agreements with the logical approaches to knowledge, beliefs and the dynamics of information.

2 Definitions

In this section we introduce the models in which we study the various agreement results, and the logical language used in Baltag and Smets (2008) to describe them.

2.1 Epistemic Plausibility Models

An epistemic plausibility model Baltag and Smets (2006) is a qualitative representation of the agents' beliefs as well as first- and higher-order information in a given interactive situation.

Definition 2.1 (Epistemic Plausibility Model)**.** Given a countable set of atomic propositions $\textsc{prop}$, an *epistemic plausibility model* $\mathcal{M} = \langle W, (\leq_i)_{i\in I}, (\sim_i)_{i\in I}, V\rangle$ has $W \neq \emptyset$ and is countable, $I = \{1, 2, \ldots, n\}$ is a finite set of *agents*, and for each $i \in I$, $\leq_i$ is a total (plausibility) pre-order on W and $\sim_i$ is binary equivalence relation on W, and $V : \textsc{prop} \to \wp(W)$. An epistemic plausibility *frame* $\mathcal{F}$ is an epistemic plausibility model with the valuation V omitted.

The total plausibility pre-order $\leq_i$ induces i's *priors*. If $w \leq_i w'$ we say that i considers w' at least as plausible as w. Given a set $X \subseteq W$, we say that $w \in X$ is $\leq_i$-minimal in X if $w \leq_i w'$ for all $w' \in X$. The relation $\sim_i$ induces i's *information partition* W. We write $\mathcal{K}_i[w]$ to denote the cell of this partition $\{v \in W \mid w \sim_i v\}$ to which w belongs. $\mathcal{K}_i[w]$ should be seen as i's (private) information at w. We write $|\mathcal{M}| = W$ for the domain of $\mathcal{M}$.

The next two assumptions are crucial in the following.

Definition 2.2 ((Local) well-foundedness)**.** A plausibility pre-order satisfies:

- **Local well-foundedness**. If for all $w \in W$ and $i \in I$, for all $X \subseteq \mathcal{K}_i[w]$, X has $\leq_i$-minimal elements.
- **Well-foundedness**. If for all $X \subseteq W$ and $i \in I$, X has a $\leq_i$-minimal elements.

$\mathcal{M}$ satisfies (*Local*) *Well-foundedness* if every plausibility pre-order has the corresponding property.

Definition 2.3 ((A priori/ a posteriori) Most plausible elements)**.**
- For all $X \subseteq W$, let $\beta_i(X) = min_{\leq_i}(X) = \{w : w \text{ is } \leq_i \text{-minimal in } X\}$.
- For all $w \in W$, let $\mathcal{B}_i[w] = \beta_i(\mathcal{K}_i[w])$.

We write $w \rhd_i^{\mathcal{B}} v$ iff $v \in \mathcal{B}_i[w]$, and $w \to_i^X v$ iff $v \in \beta_i(\mathcal{K}_i[w] \cap X)$.

Intuitively $\beta_i(X)$ are the *a priori* most plausible elements of a set, ignoring the information partitions. $\mathcal{B}_i[w]$ gives the states i considers the most plausible, conditional on the information he possesses at w, i.e. conditional on $\mathcal{K}_i[w]$. Observe that β_i is well-defined if the plausibility pre-order is well-founded, while local well-foundedness is sufficient for $\mathcal{B}_i$ to be well-defined.

Definition 2.4 (Common Prior)**.** There is *common prior beliefs* among group G in an epistemic plausibility model $\mathcal{M}$ when $\leq_i = \leq_j$ for all $i, j \in G$.

The reflexive-transitive closure of the union of the epistemic accessibility relations $\sim_i$ for all agents i in a group G is the model-theoretic counterpart of the notion of "common knowledge" in G Fagin et al. (1995), van Ditmarsch et al. (2007). We define "common belief" analogously.

Definition 2.5 (Common knowledge)**.** For each $G \subseteq I$, let $\sim_G^*$ be the reflexive-transitive closure of $\bigcup_{i \in G} \sim_i$. Let $[w]_G^* = \{w' \in W \mid w \sim_G^* w'\}$.

Definition 2.6 (Common belief)**.** For each $G \subseteq I$, let $\rhd_G^*$ be the reflexive-transitive closure of $\bigcup_{i \in G} \rhd_i^{\mathcal{B}}$.

2.2 Doxastic-Epistemic Logic

The logical language used in Baltag and Smets (2008) to describe epistemic-plausibility models is a propositional modal language with three families of modal operators, which we extend here with "common belief" operators.

Definition 2.7 (Epistemic Doxastic Language)**.** The language $\mathcal{L}_{EDL}$ is defined as follows:

$$\phi := p \mid \neg\phi \mid \phi \wedge \phi \mid K_i\phi \mid B_i^{\phi}\phi \mid C_G\phi \mid CB_G\phi$$

where i ranges over N, p over a countable set of proposition letters PROP and $\emptyset \neq G \subseteq I$.

The propositional fragment of this language is standard, and we write $\perp$ for $p \wedge \neg p$ and $\top$ for $\neg\perp$. A formula $K_i\phi$ should be read as "i knows that ϕ", $C_G\phi$ as "it is common knowledge among group G that ϕ", $CB_G\phi$ as "it is common belief among group G that ϕ." The formula $B_i^{\phi}\psi$, should be read " conditional on ϕ, i believes that ψ." These formulas are interpreted in epistemic plausibility models as follows:

Definition 2.8 (Truth definition)**.** We write $||\phi||^{\mathcal{M}}$ for $\{w \in |\mathcal{M}| \ : \ \mathcal{M}, w \Vdash \phi\}$. We omit $\mathcal{M}$ when it is clear from the context.

$$
\begin{array}{lll}
\mathcal{M}, w \Vdash p & \text{iff} & w \in V(p) \\
\mathcal{M}, w \Vdash \neg\phi & \text{iff} & \mathcal{M}, w \nVdash \phi \\
\mathcal{M}, w \Vdash \phi \wedge \psi & \text{iff} & \mathcal{M}, w \Vdash \phi \text{ and } \mathcal{M}, w \Vdash \psi \\
\mathcal{M}, w \Vdash K_i\phi & \text{iff} & \forall v \text{ (if } w \sim_i v \text{ then } \mathcal{M}, v \Vdash \phi) \\
\mathcal{M}, w \Vdash B_i^{\psi}\phi & \text{iff} & \forall v \text{ (if } w \rightarrow_i^{||\psi||^{\mathcal{M}}} v \text{ then } \mathcal{M}, v \Vdash \phi) \\
\mathcal{M}, w \Vdash C_G\phi & \text{iff} & \forall v \text{ (if } w \sim_G^{*} v \text{ then } \mathcal{M}, v \Vdash \phi) \\
\mathcal{M}, w \Vdash CB_G\phi & \text{iff} & \forall v \text{ (if } w \rhd_G^{*} v \text{ then } \mathcal{M}, v \Vdash \phi)
\end{array}
$$

Simple belief conditional only on i's information at a state w is definable using the conditional belief operator: $B_i\phi = B_i^{\top}\phi$, since: $\mathcal{M}, w \Vdash B_i^{\top}\phi$ iff $\forall v$ (if $w \rhd_i^{\mathcal{B}} v$ then $\mathcal{M}, v \Vdash \phi$).

3 Static Agreements and Well-foundedness

We first show that if an epistemic plausibility model is well-founded, then common belief that agent i believes that ϕ while j does not believe that ϕ implies that i and j have *different priors*, which is the contrapositive form of the agreement theorem.

Theorem 1 (Agreement theorem - Common Belief)**.** *If a* well-founded *epistemic plausibility model $\mathcal{M}$ satisfies $\mathcal{M}, w \Vdash CB_{\{i,j\}}(B_ip \wedge \neg B_jp)$ for some $w \in W$, then i and j have different priors in $\mathcal{M}$.*

This immediately implies the "common knowledge" agreement result below, because $C_G\phi \rightarrow CB_G\phi$ is a valid implication in epistemic plausibility models. Note, however, that this result can also have be shown independently, by application of Bacharach's Bacharach (1985) result on qualitative "decision functions", modulo generalization to the countable case.

Corollary 1 (Agreement theorem - Common Knowledge)**.** *If an epistemic plausibility model $\mathcal{M}$ satisfies* well-foundedness *and $\mathcal{M}, w \Vdash C_{\{i,j\}}(B_ip \wedge \neg B_jp)$ for one $w \in W$, then i and j have different priors in $\mathcal{M}$.*

Well-foundeness of the plausibility ordering is the crux of Theorem 1 and Corollary 1: if an epistemic-plausibility model is well-founded then common prior excludes the possibility of disagreements. Moreover the next result shows that the converse holds as well, and thus that *well-foundedness* cannot be weakened to *local well-foundedness*.

Proposition 1. *There exists a pointed epistemic plausibility model $\mathcal{M}, w$ which satisfies* local well-foundedness *and* common prior *such that $\mathcal{M}, w \Vdash C_{\{1,2\}}(B_1p \wedge \neg B_2p)$.*

Well-foundedness of the plausibility ordering is thus *the* safeguard against common knowledge of disagreement, once we drop the assumption that the state space is finite.

4 Expressive Power and Syntactic Proofs

$\mathcal{L}_{EDL}$ is a natural choice of language to talk about epistemic-plausibility models, and but we show here that it cannot express Theorem 1 nor Corollary 1, because it cannot express two of their key assumptions, common prior and well-foundedness.

Fact 2. *The class of epistemic plausibility frames that satisfies common prior is not definable in $\mathcal{L}_{EDL}$.*

This result confirms the idea that to reason about (common) priors the agents must make "inter-[information]-state comparisons" Samet (In Press), which they cannot do because their reasonings in $\mathcal{L}_{EDL}$ are local, i.e. bounded by the "hard information" van Benthem (2007) they have. This limitation also makes well-foundeness inexpressible, and with it the two static agreement results.

Fact 3. *There is no formula ϕ of $\mathcal{L}_{EDL}$ which is true in a pointed epistemic plausibility model $\mathcal{M}, w$ iff Theorem 1 or Corollary 1 holds in $\mathcal{M}, w$.*

The syntactical counterpart of the model-theoretic agreement results thus lives in more expressive languages. In the appendix we present a finite derivation in $\mathcal{H}(@, \downarrow, C_G)$, which extends the hybrid language $\mathcal{H}(@, \Downarrow)$ with a common knowledge modality C_G. The satisfiability problem for this language on the class of conversely well-founded frames is Σ^1_1-hard ten Cate (2005), ruling out any finite axiomatization of its validities. The derivation we show, however, is finite and uses only sound axioms. It is still unknown to us whether the agreement results of Section 3 could be derived in a less complex language. The fact that the syntactic derivation reported here pertains to such an expressive language nevertheless shows that reasoning explicitly about agreement results requires heavy expressive resources.

5 Agreements via dialogues

In this section we turn to"agreements-via-dialogues" Geanakoplos and Polemarchakis (1982), Bacharach (1985), which analyze how agents can reach agreement in the process of exchanging information about their beliefs by updating the latter accordingly.

5.1 Agreements via Conditioning

We first consider agreements by repeated belief conditioning. It is known that if agents repeatedly exchange information about each others' posterior beliefs about a certain event, and update these posteriors accordingly, these posteriors will eventually converge Geanakoplos and Polemarchakis (1982), Bacharach (1985). We show here that this result also holds for the "qualitative" form of beliefs conditionalization in epistemic plausibility models.

We call a *conditioning dialogue about ϕ* Geanakoplos and Polemarchakis (1982) at a state w of an epistemic plausibility model $\mathcal{M}$ a sequence of belief conditioning, for each agent, on all other agents' beliefs about ϕ. This sequence can be intuitively described as follows. It starts with the agents' simple belief about ϕ, i.e. for all i: $B_i\phi$ if $\mathcal{M}, w \Vdash B_i\phi$ and $\neg B_i\phi$ otherwise. Agent i's beliefs about ϕ at the next stage is defined by taking his beliefs about ϕ, conditional upon learning the others' belief about ϕ at that stage. Syntactically, this gives, $\mathbb{B}_{1,i} = B_i\phi$ if $\mathcal{M}, w \Vdash B_i\phi$ and $\mathbb{B}_{1,i} = \neg B_i\phi$ otherwise and, for two agents i, j, $\mathbb{B}_{n+1,i} = B_i^{\mathbb{B}_{n,j}\phi}\phi$ if $\mathcal{M}, w \Vdash B_i^{\mathbb{B}_{n,j}\phi}\phi$ and $\neg B_i^{\mathbb{B}_{n,j}\phi}\phi$ otherwise. This syntactic rendering is only intended to fix intuitions, though, since in countable models the limit of this sequence exceeds the finitary character of $\mathcal{L}_{EDL}$. We thus focus on models-theoretic conditioning.

Conditioning on a given event $A \subseteq W$ boils down to refining an agent's information partition by removing "epistemic links" connecting A and non-A states.

Definition 5.1 (Conditioning by a subset)**.** Given an epistemic plausibility model $\mathcal{M}$, the collection of epistemic equivalence relation of the agents is an element of $\wp(W \times W)^I$. Given a group $G \subseteq I$, the function $f_G : \wp(W) \to (\wp(W \times W)^I \to \wp(W \times W)^I)$ is a *conditioning function for G* whenever:

$$(w, v) \in f_G(A)(i)(\{\sim_i\}_{i\in I}) = \begin{cases} (w, v) \in \sim_i \text{ and } (w \in A \text{ iff } v \in A) & \text{if } i \in G \\ (w, v) \in \sim_i & \text{otherwise} \end{cases}$$

Given $\mathcal{M} = \langle W, (\leq_i)_{i\in I}, (\sim_i)_{i\in I}, V\rangle$ we write $f_G(A)(\mathcal{M})$ for the model $\langle W, (\leq_i)_{i\in I}, f_G(A)((\sim_i)_{i\in I}), V\rangle$.

It is easy to see that the relations $\sim_i$ in $f_G(A)(\mathcal{M})$ are equivalence relations. Here we are interested in cases where the agents condition their beliefs upon learning in which *belief state* the others are.

Definition 5.2 (Belief states)**.** Let $\mathcal{M}$ an epistemic plausibility model and $A \subseteq W$, we write

$$B_j^{\mathcal{M}}(A) \text{ for } \{w : \beta_j(\mathcal{K}_j^{\mathcal{M}}[w]) \subseteq A\} \text{ and}$$
$$\neg B_j^{\mathcal{M}}(A) \text{ for } W \setminus B_j^{\mathcal{M}}(A)$$

We define $\mathbb{B}_j^{\mathcal{M},w}(A)$ as follows:

$$\mathbb{B}_j^{\mathcal{M},w}(A) = \begin{cases} B_j^{\mathcal{M}}(A) & \text{if } w \in B_j^{\mathcal{M}}(A) \\ \neg B_j^{\mathcal{M}}(A) & \text{otherwise} \end{cases}$$

Observation 4. *For any plausibility epistemic model $\mathcal{M}$ indexed by a finite set of agents I, $\langle \wp(W \times W)^I, \subseteq\rangle$ is a chain complete poset. Moreover for all $A \subseteq W$, $w \in W$ and $G \subseteq I$, $f_G(A)$ is deflationary.*

Taking $f_I(\bigcap_{j\in I} \mathbb{B}_j^{\mathcal{M},w}(||\phi||^{\mathcal{M}}))$ as a mapping on models, it is easy to see from the preceding observation that conditioning by agents' beliefs about some event is deflationary with respect to the relation of epistemic-submodel. It follows then by the Bourbaki-Witt fixed-point Theorem Bourbaki (1949) (see appendix for an exact statement) that conditioning by agents' beliefs has a fixed point.

Given an initial pointed model $\mathcal{M}, w$ and some event $A \subseteq W$, we can construct its fixed point under conditioning by agents' beliefs as the limit of a sequences of models, that are the model-theoretic counterpart of the dialogues described above.

Definition 5.3. A *conditioning dialogue about* ϕ at the pointed plausibility epistemic model $\mathcal{M}, w$, with $\mathcal{M} = \langle W, (\leq_i)_{i \in I}, (\sim_i)_{i \in I}, V \rangle$ is the sequence of pointed epistemic plausibility models $(\mathcal{M}_n, w)$ with

$$(\mathcal{M}_0, w) = \mathcal{M}, w$$

$$(\mathcal{M}_{\beta+1}, w) = f_I(\bigcap_{j \in I} \mathbb{B}_j^{\mathcal{M}_\beta, w}(||\phi||^{\mathcal{M}}))(\mathcal{M}_\beta), w$$

$$(\mathcal{M}_\lambda, w) = \bigcap_{\beta < \lambda} (\mathcal{M}_\beta, w) \text{ for limit ordinals } \lambda$$

This extends to the countable case the standard representation of a dialogue about ϕ in the literature on dynamic agreements Geanakoplos and Polemarchakis (1982), Bacharach (1985). By Observation 4 we know that dialogues cannot last forever, i.e. that each such sequence has a limit.

Corollary 2. *For any pointed epistemic plausibility model $\mathcal{M}, w$ and $\phi \in \mathcal{L}_{EDL}$ there is a α^f such that, for all $i \in I$, $w \in W$ and $\alpha > \alpha^f$, $\mathcal{K}_{\alpha,i}[w] = \mathcal{K}_{\alpha^f,i}[w]$.*

Once the agents have reached this fixed-point α^f—possibly after transfinitely many steps—they have eliminated all higher-order uncertainties concerning the posteriors about ϕ of the others, viz. these posteriors are then common knowledge:

Theorem 5 (Common knowledge of beliefs about ϕ)**.** *At the fixed-point α^f of a conditioning dialogue about ϕ we have that for all $w \in W$ and $i \in I$, if $w \in B_i^{\mathcal{M}_{\alpha^f}, w}(||\phi||^{\mathcal{M}})$ then $w' \in B_i^{\mathcal{M}_{\alpha^f}, w}(||\phi||^{\mathcal{M}})$ for all $w' \in [w]^*_{\alpha^f, I}$, and similarly if $w \notin B_i^{\mathcal{M}_{\alpha^f}, w}(||\phi||^{\mathcal{M}})$.*

With this in hand we can directly apply the static agreement result for common knowledge (Corollary 1, Section 3) to obtain that the agents indeed reach agreements at the fixed-point of a dialogue about ϕ.

Corollary 3 (Agreement via conditioning dialogue)**.** *Take any dialogue about ϕ with common and well-founded, priors and α^f as in Corollary 2. Then for all w in W, either $[w]^*_{\alpha^f, I} \subseteq \bigcap_{i \in I} B_i^{\mathcal{M}_{\alpha^f}, w}(||\phi||^{\mathcal{M}})$ or $[w]^*_{\alpha^f, I} \subseteq \bigcap_{i \in I} \neg B_i^{\mathcal{M}_{\alpha^f}, w}(||\phi||^{\mathcal{M}})$.*

This result brings qualitative dynamic agreement results Cave (1983), Bacharach (1985) to epistemic plausibility models, and show that agents can indeed reach agreement via iterated conditioning, even when one drops the finite model assumption.

5.2 Agreements via Public Announcements

In this section we show that iterated "public announcements" lead to agreements, thus bringing a distinct form of information update to the agreement literature. Public announcements are "epistemic actions" van Ditmarsch et al. (2007) where truthful hard information is made public to the members of a group by a trusted source, in a way that no member is in doubt as whether the others received the same piece of information as him.

One extends a given logical language with public announcements by operators of the form $[\phi!]\psi$, meaning "after the announcement of ϕ, ψ holds" Plaza (1989), Gerbrandy (1999). A dialogue about ϕ via public announcements among the members of a group G thus starts, as before, with i simple beliefs about ϕ, for all $i \in I$. The agents' beliefs about ϕ at the next stage are then defined as those they would have after a public announcement of all agents' beliefs about ϕ at the first stage. Syntactically, this gives: $\mathbb{B}_{1,i}$ as in Section 5.1, and $\mathbb{B}_{n+1,i}$, as $[\bigcap_{j\in I} \mathbb{B}_{n,j}\phi!]B_i\phi$ if $\mathcal{M}, w \Vdash [\bigcap_{j\in I} \mathbb{B}_{n,j}\phi!]B_i\phi$ and as $[\bigcap_{j\in I} \mathbb{B}_{n,j}\phi!]\neg B_i\phi$ otherwise. For the same reason as in the previous section, we now move our analysis to the level of models.

The A-generated-submodel of a given epistemic plausibility model $\mathcal{M}$, the precise definition of which is in the appendix, is the model which results after of public announcement of A in $\mathcal{M}$. We write $Sub(\mathcal{M}) = \{\mathcal{M}'$ is the A-generated submodel of $\mathcal{M} \mid A \subseteq |\mathcal{M}|\}$ and $\mathcal{M}' \sqsubseteq \mathcal{M}$ whenever $\mathcal{M}' \in Sub(\mathcal{M})$.

Definition 5.4 (Relativization by agents beliefs)**.** Let $\mathbb{B}_i(\mathcal{M}, w, \phi)$ be defined as follows:

$$\mathbb{B}_i(\mathcal{M}, w, \phi) = \begin{cases} ||B_i\phi||^{\mathcal{M}} & \text{if } \mathcal{M}, w \Vdash B_i\phi \\ ||\neg B_i\phi||^{\mathcal{M}} & \text{otherwise} \end{cases}$$

Then given an epistemic-plausibility model $\mathcal{M} = \langle W, (\leq_i)_{i\in I}, (\sim_i)_{i\in I}, V\rangle$, the relativization $!B^{\phi}_w$ by agents' beliefs about ϕ at w (where $w \in |\mathcal{M}|$), takes $\mathcal{M}$ to $!B^{\phi}_w(\mathcal{M})$. Here $!B^{\phi}_w(\mathcal{M})$ is the $\bigcap_{i\in I} \mathbb{B}_i(\mathcal{M}, w, \phi)$-generated submodel $!B^{\phi}_w(\mathcal{M}) = \langle W^{!B^{\phi}_w}, \leq_i^{!B^{\phi}_w}, \sim_i^{!B^{\phi}_w}, V^{!B^{\phi}_w}\rangle$ of $\mathcal{M}$ such that:

- $W^{!B^{\phi}_w} = \bigcap_{i\in I} \mathbb{B}_i(\mathcal{M}, w, \phi)$

and for each $i \in I$

- $\leq_i^{!B^{\phi}_w} = \leq_i \cap (W^{!B^{\phi}_w} \times W^{!B^{\phi}_w})$
- $\sim_i^{!B^{\phi}_w} = \sim_i \cap (W^{!B^{\phi}_w} \times W^{!B^{\phi}_w})$
- For each $v \in W^{!B^{\phi}_w}$, $v \in V^{!B^{\phi}}(p)$ iff $v \in V(p)$

Note that by construction above the actual state w is never eliminated.

Observation 6. *For any plausibility epistemic model $\mathcal{M}$ indexed by a finite set of agents I, $\langle Sub(\mathcal{M}), \sqsubseteq\rangle$ is a chain complete poset. Moreover for all $\phi \in \mathcal{L}_{EDL}$, $w \in W$, $!B^{\phi}$ is deflationary.*

It follows then by Bourbaki-Witt Bourbaki (1949) Theorem (see appendix for an exact statement) that the process of public announcement of beliefs has a fixed point. Given an initial pointed model $\mathcal{M}, w$ and some formula $\phi \in \mathcal{L}_{EDL}$, we can construct this fixed point by taking the limit of a sequence of models, that we call a public dialogue.

Definition 5.5. A *public dialogue about ϕ* starting in $\mathcal{M}, w$ is a sequence of epistemic-doxastic pointed models $\{(\mathcal{M}_n, w)\}$ such that:

- $\mathcal{M}_0 = \mathcal{M}$ is a given epistemic-plausibility model.

- $\mathcal{M}_{\beta+1} = !B^{\phi}_{w}(\mathcal{M}_{\beta})$
- $(\mathcal{M}_{\lambda})$ is the submodel of $\mathcal{M}$ generated by $\bigcap_{\beta<\lambda} |\mathcal{M}_{\beta}|$ for limit ordinals λ

It is known that such a dialogue needs not to stop after the first round of announcements, in e.g. the "muddy children" case van Benthem (2006), but by Observation 6 we know that it will stop at some point.

Corollary 4 (Fixed-point). *Given an epistemic-plausibility model $\mathcal{M}_0$, w and a public dialogue about ϕ, there is a α^{ϕ} such that $(\mathcal{M}_{\alpha}, w) = (\mathcal{M}_{\alpha^{\phi}}, w)$ for all $\alpha \geq \alpha^{\phi}$.*

Moreover at $\mathcal{M}_{\alpha^{\phi}}, w$, which we call the *fixed point* of the public dialogue about ϕ, the posteriors of the agents about this formula are common knowledge, which means that they will reach an agreement on ϕ if they have common and well-founded priors.

Theorem 7 (Common knowledge at the fixed point). *At the fixed-point of a public dialogue $\mathcal{M}_{\alpha^{\phi}}, w$ about ϕ, for all $w \in W$ and $i \in I$, if $w \in ||B_i\phi||^{\mathcal{M}_{\alpha^{\phi}}}$ then $w' \in ||B_i\phi||^{\mathcal{M}_{\alpha^{\phi}}}$ for all $w' \in [w]^{*}_{\alpha^{\phi},I}$, and similarly if $w \notin ||B_i\phi||^{\mathcal{M}_{\alpha^{\phi}}}$.*

Corollary 5 (Agreements via Public Announcements). *For any public dialogue about ϕ, if there is common and well-founded priors then at the fixed-point $\mathcal{M}_{\alpha^{\phi}}, w$ either all agents believe that ϕ or they all do not believe that ϕ.*

This new form of dynamic agreements result is conceptually important because it fits better than iterated conditioning the intuitive idea of a *public* dialogue. Public announcements systematically change the higher-order information in a given interactive situation. This reflects the fact that the agents in a given group not only simultaneously receive the same piece of private information, as in conditioning dialogues, but that it is common knowledge that all agents receive it. The information sharing process induced by public announcements is thus genuinely public, and it leads to different agreement than conditioning, as we now show.

5.3 Comparing Agreements via Conditioning and Public Announcements

In this section we illustrate by way of an example that conditioning dialogues and dialogues via public announcements can lead to a different agreement.

Example 8. *Consider the model in Figure 1. The arrow represents* 1 *and* 2*'s common plausibility ordering, with $w_1 < w_2$, while the solid and dotted rectangles represent* 1 *and* 2*'s information partitions, respectively. Take a proposition letter p and assume that $V(p) = \{w_1\}$. Let $\phi := p \wedge \neg B_2 p$, i.e. "$p$ but* 2 *doesn't believes that p". Observe that ϕ holds at w_1, that* 1 *believes it but that* 2 *does not. The conditioning dialogue and the dialogue via public announcements, both about ϕ, reach their fixed point n^* after one round in this model, where $[w_1]_{n^*,1} = [w_1]_{n^*,2} = \{w_1\}$. The formula ϕ leads to an "unsuccessful update" by public announcement van Ditmarsch et al. (2007), and at the fixed point of the dialogue neither* 1 *nor* 2 *believe that ϕ. In conditioning dialogue, however, both agents* do *believe that ϕ at the fixed point.*

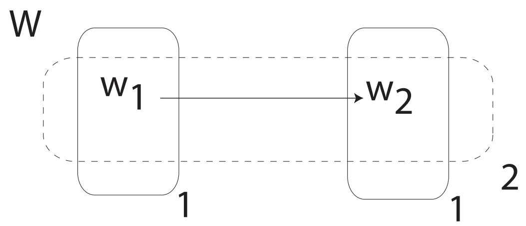


Figure 1: An epistemic plausibility model where conditioning leads to different agreements than public announcements

This example hinges on the fact that public announcement and belief conditioning threat higher-order information differently. In conditioning the truth value of the formula at hand remains fixed. If the formula contains epistemic (K_i or CB_G) or doxastic (B_i, C_G) operators, this means that the conditioning dialogue bears on the knowledge and beliefs of the agents anterior to the information exchange Baltag and Smets (2008). In dialogues via public announcements the truth value of the formula ϕ is dynamically adapted to the incoming new information, reflecting the fact that knowing that others receive the same piece of information might lead an agent to revise his higher-order information as well. Obviously, this only lead to different agreements when the dialogue is about an informational fact, but the point remains: the difference between public announcement and belief conditioning highlights the public character of the former in comparison to the latter, and shows the importance of both in the landscape of agreement results.

6 Conclusion

We studied agreement theorems from the point of view of dynamic-epistemic logic. We showed that both static and dynamic agreement results hold in epistemic plausibility models, answering an open question in the logic literature. We pointed the need for rather expressive logical languages to reason explicitly about static agreements results. We furthermore improved on existing qualitative agreement results, by proving that common belief in posteriors is sufficient to ensure agreement, under common and well-founded priors, and so for both finite and countable structures. Finally, we coped on the distinction between conditioning and public announcements to provide two dynamic agreement results, and argued that the later capture better the public character of dialogues. Bringing agreement theorems to dynamic-epistemic logic thus proves to be both technically and conceptually fruitful, and it bridges two important bodies of literature.

For future work one should put the full generality dynamic-epistemic logic Baltag et al. (1998), van Ditmarsch et al. (2007) to use, as well as recent developments on "softer" forms of belief updates van Benthem (2007), Baltag and Smets (200X), to analyze the possibility of agreements in a larger class of situations. It remains also open whether one can finitely axiomatize a logic which can derive the agreement results, in both their static and dynamic forms. Finally, the expressibility of alternative static agreement theorems, e.g. in Samet (In Press), should also be investigated from a logical point of view.

References

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

M. Bacharach. Some extensions of a claim of Aumann in an axiomatic model of knowledge. *Journal of Economic Theory*, 37(1):167–190, October 1985.

A. Baltag and S. Smets. A qualitative theory of dynamic interactive belief revision. In Giacomo Bonanno, Wiebe van der Hoek, and Michael Wooldridge, editors, *Logic and the Foundation of Game and Decision Theory (LOFT7)*, volume 3 of *Texts in Logic and Games*, pages 13–60. Amsterdam University Press, 2008.

A. Baltag and S. Smets. Learning by questions and answers: From belief-revision cycles to doxastic fixed points. Under Review, 200X.

A. Baltag and S. Smets. Conditional doxastic models: A qualitative approach to dynamic belief revision. In G. Mints and R. de Queiroz, editors, *Proceedings of WOLLIC 2006, Electronic Notes in Theoretical Computer Science*, volume 165, 2006.

A. Baltag, L.S. Moss, and S. Solecki. The logic of public announcements, common knowledge and private suspicions. In *TARK 98*, 1998.

J. van Benthem. One is a lonely number. In Z. Chatzidakis, P. Koepke, and W. Pohlers, editors, *Logic Colloquium '02*, Wellesley MA, 2006. ASL & A.K. Peters.

J. van Benthem. Dynamic logic for belief revision. *Journal of Applied Non-classical Logics*, 17(2), 2007.

P. Blackburn, M. de Rijke, and Y. Venema. *Modal Logic*, volume 53 of *Cambridge Tracts in Theoretical Computer Science*. Cambridge University Press, 2001.

Giacomo Bonanno and Klaus Nehring. How to make sense of the common prior assumption under incomplete information. *International Journal of Game Theory*, 28(03):409–434, August 1999.

Giacomo Bonanno and Klaus Nehring. Agreeing to disagree: a survey. Some of the material in this paper was published in Bonanno and Nehring (1999), 1997.

N. Bourbaki. Sur le théorème de Zorn. *Archiv der Mathematik*, 2, 1949.

B. ten Cate. *Model theory for extended modal languages*. PhD thesis, University of Amsterdam, 2005. ILLC Dissertation Series DS-2005-01.

J. A. K. Cave. Learning to agree. *Economics Letters*, 12(2):147–152, 1983.

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

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

J. Geanakoplos and H. M. Polemarchakis. We can't disagree forever. Cowles Foundation Discussion Papers 639, Cowles Foundation, Yale University, July 1982.

J. Gerbrandy. *Bisimulations on Planet Kripke*. PhD thesis, ILLC, Amsterdam, 1999.

L. Ménager. *Communication, common knowledge, and consensus*. PhD thesis, Universite Paris I Pantheon-Sorbonne, 2006.

J.A. Plaza. Logics of public communications. In M.L. Emrich, M.S. Pfeifer, M. Hadzikadic, and Z.W. Ras, editors, *Proceedings of the Fourth International Symposium on Methodologies for Intelligent Systems: Poster Session Program*, pages 201–216. Oak Ridge National Laboratory, 1989.

D. Samet. Agreeing to disagree: The non-probabilistic case. *Games and Economic Behavior*, In Press.

Appendix: Proofs

Proof of Theorem 1

Proof. We show that there is no pointed epistemic plausibility model $\mathcal{M}, w$ which satisfies *well-foundedness* and *common prior* such that $\mathcal{M}, w \Vdash CB_{\{i,j\}}(B_i p \wedge \neg B_j p)$. To do that we assume $\mathcal{M}, w \Vdash CB_{\{i,j\}}(B_i p \wedge \neg B_j p)$ and $\mathcal{M}$ satisfies *common prior* and show by induction that $\mathcal{M}$ must not be well-founded, by constructing an infinite descending chain $w_1 > w_2 > \ldots$, such that $w_1 \rhd^*_{\{1,2\}} w_n$ for every $n \in \omega$. Note that by common prior we have that $\leq_1 = \leq_2 = \leq$. Now assume that $\mathcal{M}, w \Vdash CB_{\{1,2\}}(B_1 p \wedge \neg B_2 p)$ (1) and suppose, towards contradiction, that $\leq$ is well-founded.

Base case. First of all we start by constructing a descending chain of length 2. By (1) we have in particular $\mathcal{M}, w \Vdash B_1(B_1 p \wedge \neg B_2 p)$. By assumption it follows from truth definition of B_1 (and $\leq_1 = \leq$-well-foundedness) that there is some state, call it w_0 such that $w_0 \in min_{\leq}\mathcal{K}_1[w]$, i.e. $w \rhd^{\mathcal{B}}_i w_0$ (2) and $\mathcal{M}, w_0 \Vdash B_1 p \wedge \neg B_2 p$ (3). In particular $\mathcal{M}, w_0 \Vdash \neg B_2 p$ (4). By the same argument as before that there must thus be a state, call it w_1, such that $w_1 \in min_{\leq}\mathcal{K}_2[w_0]$, i.e. $w_0 \rhd^{\mathcal{B}}_2 w_1$ (5) and $\mathcal{M}, w_1 \Vdash \neg p$ (6).

But by (1), (2) and (5) it follows that $\mathcal{M}, w_1 \Vdash B_1 p$ (7), i.e. $\{v \in W | v \in min_{\leq}\mathcal{K}_1[w_1]\} \subseteq V(p)$ (8). But then by the now usual argument it follows that there is a state, call it w_2 such that $w_2 \in min_{\leq}\mathcal{K}_1[w_1]$ (9) and $\mathcal{M}, w_2 \Vdash p$ (10). But then from (6), (8) and (9) it follows that $w_1 > w_2$.

Induction step. Assume that we have been able to construct a chain of length n, i.e. we have $w_1 > w_2 > \ldots > w_n$ such that $w_1 \rhd^*_{\{1,2\}} w_n$ (11) for every $n' \leq n$. Assume that there is no state v such that $w_n > v$ (12). Clearly w_n must be minimal within both $\mathcal{K}_1[w_n]$ (13) and $\mathcal{K}_2[w_n]$ (14). It is easy to see that by truth condition of common belief we have by (2), (5), (11) and (1) that $\mathcal{M}, w_n \Vdash B_1 p$ (15). But then by (13) we have $w_n \Vdash p$ (16). Similarly we have $\mathcal{M}, w_n \Vdash \neg B_2 p$ (17). It then follows WLOG that there must be state v_n such that $(v_n \leq w_n \ \& \ v_n \geq w_n)$ (18) such that $w_n \Vdash \neg p$ (19). It follows that $v_n \notin \mathcal{K}_1[w_n]$ (20). Moreover by common belief we have that $v_n \Vdash B_1 p$ (21). But it follows that $v_n \notin min_{\leq}\mathcal{K}_1[v_n]$ (22). Since this set is non-empty it follows that there is some state $w_{n+1} \in min_{\leq}\mathcal{K}_1[v_n]$ (23). But then we have by (22) and (23) that $v_n > w_{n+1}$ (24). But (24) and (18) implies that $w_n > w_{n+1}$ (25). Concluding the induction step and the proof. □

Proof of Theorem 1

Proof. Let the model $\mathcal{M}$ be defined as follows, with $I = \{1, 2\}$.

- $\mathcal{M} = \langle \mathbb{Z}, (\leq_i)_{i \in I}, (\sim_i)_{i \in I}, V \rangle$ such that:
  - $\mathbb{Z}$ is the set of integers.
  - For both agents $i \in I$, $x \leq_i y$ iff $x \geq y$.
  - For all $x, y \in \mathbb{Z}$: $x \sim_1 y$ is the smallest equivalence relation such that $x \sim_1 y$ whenever $y = x + 1$ and x is odd; $x \sim_2 y$ is the smallest equivalence relation such that $x \sim_2 y$ whenever $y = x + 1$ and x is even.
  - $V(p) = \{x : x \text{ is odd }\}$ and $V(q) = \emptyset$ for all $q \neq p$ in PROP.

It is easily checked that at every $x \in \mathbb{Z}$ we have $\mathcal{M}, x \Vdash (\neg B_1 p \wedge B_2 p)$, and so that $\mathcal{M}, x \Vdash C_{1,2}(\neg B_1 p \wedge B_2 p)$, and moreover that $\mathcal{M}$ satisfies *local well-foundedness* and *common prior*. □

Proof of Fact 2

Proof. Take $W = \{x, y\}$ and $\sim_1 = \sim_2 = \{(x, x), (y, y)\}$. We consider two epistemic plausibility frames $\mathcal{F}$ and $\mathcal{F}'$. $\mathcal{F} = \langle W, \sim_1, \sim_2, \leq_1, \leq_2 \rangle$ and $\mathcal{F}' = \langle W, \sim_1, \sim_2, \leq'_1, \leq'_2 \rangle$ where $\leq_1 = \leq_2 = \{(x, x), (x, y), (y, y)\}$ while $\leq'_1 = \{(x, x), (x, y), (y, y)\}$ and $\leq'_2 = \{(x, x), (y, x), (y, y)\}$. Clearly $\mathcal{F}$ satisfies common prior while $\mathcal{F}'$ does not. Now assume for contradiction that there is a formula $\psi \in \mathcal{L}_{EDL}$ that defines the class of epistemic plausibility frames with common prior. Then we have $\mathcal{F} \Vdash \psi$ (1) while $\mathcal{F}' \nVdash \psi$ (2). It follows from (2) that there is some valuation V and some state $s \in W = \{x, y\}$, such that $\mathcal{F}', V, s \nVdash \psi$ (3). But if follows from (1) that $\mathcal{F}, V, s \Vdash \psi$ (4).

We now prove by induction on the complexity of ϕ that for all $\phi \in \mathcal{L}_{EDL}$ and for all $s \in W$ we have $\mathcal{F}, V, s \Vdash \phi$ iff $\mathcal{F}', V, s \Vdash \phi$ which together with (3) and (4) gives us a contradiction. The base case for propositional letters is immediate. The cases for common knowledge and knowledge follows from the fact that the pointed models $\mathcal{F}, V, s$ and $\mathcal{F}, V, s$ are isomorphic with respect to $\sim_i$. The case for common belief is trivial due the fact that the information partitions are (isomorphic) singletons. Moreover the structures are fully isomorphic for agent 1. So it remains to consider the case of conditional belief for 2.

Take the state x (the proof is similar for y). Now assume that $\mathcal{F}, V, x \Vdash B_2^{\phi}\chi$ (5). We need to show that $\mathcal{F}', V, x \Vdash B_2^{\phi}\chi$. By IH we have $||\phi||^{\mathcal{M}} = ||\phi||^{\mathcal{M}'}$ (6), with $\mathcal{M} = \mathcal{F}, V$ and $\mathcal{M}' = \mathcal{F}', V$. Now (5) iff $\forall v$ (if $v \in \beta_2(\mathcal{K}_2[x] \cap ||\phi||^{\mathcal{M}})$ then $\mathcal{M}, v \Vdash$

χ) (7). Observe that by (6) we know that $\beta_2(\mathcal{K}_2[x] \cap ||\phi||^{\mathcal{M}}) = \beta_2'(\mathcal{K}_2[x] \cap ||\phi||^{\mathcal{M}'})$ (8), since $\mathcal{K}_2[x] = \{x\}$ in $\mathcal{M}$ and $\mathcal{M}'$. Now if $(\mathcal{K}_2[x] \cap ||\phi||^{\mathcal{M}}) = \emptyset$ we are done, and otherwise we use (7), (8), truth condition of $B_2^{\phi}\chi$ and IH for χ. Concluding the induction step and the proof. □

To show our next result we introduce a few more definitions.

Definition 6.1. Two models $\mathcal{M}$ and $\mathcal{M}'$ are *bisimilar* whenever there is a relation $\underline{\leftrightarrow} \subseteq W \times W'$ such that for all $w \in W$ and $v \in W'$, if $w \underline{\leftrightarrow} v$ then:

- For all $p \in$ PROP, $w \in V(p)$ iff $v \in V'(p)$.
- Back and forth for $\sim_i$.
  - If $w \sim_i w'$ then there is a $v' \in W'$ such that $v \sim_i' v'$ and $w' \underline{\leftrightarrow} v'$.
  - If $v \sim_i v'$ then there is a $w \in W$ such that $w \sim_i' w'$ and $v' \underline{\leftrightarrow} w'$.
- For all formulas ϕ, back and forth for $\rightarrow_i^{||\phi||}$. We write $\rightarrow_i^{\phi}$ in what follows.
  - If $w \rightarrow_i^{\phi} w'$ then there is a $v' \in W'$ such that $v \rightarrow_i'^{\phi} v'$ and $w' \underline{\leftrightarrow} v'$.
  - If $v \rightarrow_i'^{\phi} v'$ then there is a $w \in W$ such that $w \rightarrow_i^{\phi} w'$ and $v' \underline{\leftrightarrow} w'$.

Two pointed models $\mathcal{M}, w$ and $\mathcal{M}', v$ bisimilar, noted $\mathcal{M}, w \underline{\leftrightarrow} \mathcal{M}', v$, if $w \underline{\leftrightarrow} v$.

Fact 9. *For all models $\mathcal{M}$ and $\mathcal{M}'$, $w \in W$, $v \in W'$ and $\phi \in \mathcal{L}_{EDL}$, if $\mathcal{M}, w \underline{\leftrightarrow} \mathcal{M}', v$ then $\mathcal{M}, w \Vdash \phi$ iff $\mathcal{M}', v \Vdash \phi$.*

Proof of Fact 3 In the following proof we often write $[w]_1$ for $\mathcal{K}_1[w]$.

Proof. Let the model $\mathcal{M}$ be as in proof of Theorem 1 and $\mathcal{M}'$ be defined as follows, with $I = \{1, 2\}$ in both cases. $\mathcal{M}' = \langle W, I, (\leq')_{i \in I}, (\sim')_{i \in I}, V' \rangle$ such that:

- $W = \{w_o, w_e\}$.
- $w_e <_1' w_o$ and $w_o <_2' w_e$.
- For both $i \in I$ and $w \in W : [w]_i = W$.
- $V'(p) = \{w_o\}$ and $V(q) = \emptyset$ for all $q \neq p$ in PROP.

Observation 10. *In $\mathcal{M}$ there is common prior and $\mathcal{M}, x \Vdash CB_I(B_1(\neg p) \wedge B_2(p))$ for all $x \in \mathbb{Z}$. In $\mathcal{M}'$ the latter is common belief as well, at both $w \in W$, but 1 and 2 have different priors.*

Define the relation $\underline{\leftrightarrow} \subseteq \mathcal{M} \times \mathcal{M}'$ as follows: $x \underline{\leftrightarrow} w_o$ for all odd integers x, and $x \underline{\leftrightarrow} w_e$ for all even integers.

Claim 11. *The relation $\underline{\leftrightarrow}$ is a bisimulation.*

Proof. The propositional clause is trivial. It should also be clear that the clause for the relations $\sim_i$ and $\sim_i'$ is also satisfied. It remains to be shown that the clause for the families of relations $\rightarrow_i^{\phi}$ and $\rightarrow_i'^{\phi}$ are also satisfied. We show this by induction ϕ. In fact we show something stronger, namely that for all ϕ:

1. If $x \to_i^{\phi} y$ and $x \underline{\leftrightarrow} w$ then there is a w' such that $w \to_i^{'\phi} w'$ and $y \underline{\leftrightarrow} w'$.

2. If $w \to_i^{'\phi} w'$ then for all $x \underline{\leftrightarrow} w$ there is a y such that $x \to_i^{\phi} y$ and $w' \underline{\leftrightarrow} y$.

3. If x is odd then $x \in ||\phi||^{\mathcal{M}}$ iff $w_o \in ||\phi||^{\mathcal{M}'}$, and if x is even then $x \in ||\phi||^{\mathcal{M}}$ iff $w_e \in ||\phi||^{\mathcal{M}'}$.

Base Case $\phi \in$ PROP. We only have to consider p.

1. Assume that x is odd and that $x \to_i^{p} y$. Observe that by construction it can only be that $min_{\leq_i}([x]_i \cap ||p||) = \{x\}$, for both $i = 1, 2$. This means that $y = x$, and so we are done, since $w_o \to_i^{'p} w_o$ and $x \underline{\leftrightarrow} w_o$. Suppose that x is even. Then $x \to_1^{p} y$ iff $y = x - 1$, again by construction. But since $y \underline{\leftrightarrow} w_o$ and $w_e \to_1^{'p} w_o$, we are done. The case for $x \to_2^{p} y$ is similar, with taking here $y = x + 1$.

2. Consider first w_o, and suppose that $w_o \to_i^{'p} w'$. Observe again that this can only happen if $w' = w_o$. Now take any x such that $w_0 \underline{\leftrightarrow} x$. By definition any such x is odd, and thus $x \in V(p)$. But we know, furthermore, that $min_{\leq_i}([x]_i \cap ||p||) = \{x\}$ for both $i = 1, 2$, and so we are done. The case for w_e is entirely similar.

3. Follows directly from the definition of V and V'.

Inductive Step Our inductive hypothesis is that claims (1), (2) and (3) hold for all ϕ' of lower complexity than ϕ. We only show the cases for (1): the arguments for (2) are entirely symmetrical, and the ones for (3) are simple applications of the inductive hypothesis.

- $\phi := \neg\psi$.
  1. We only show the case where x is odd. The other one is similar, with 1 and 2 reversed. Suppose that $x \to_i^{\neg\psi} y$. This means that $\mathcal{M}, y \nVdash \psi$. Consider first the case where $x = y$. Then $\mathcal{M}, x \nVdash \psi$, and thus by our inductive hypothesis $\mathcal{M}', w_o \nVdash \psi$. This is enough to conclude that $w_o \to_2^{'\neg\psi} w_o$, simply because $w_e >'_2 w_o$. So consider 1, for which we have that $[x]_1 = \{x, x+1\}$. Since $x >_1 x+1$, it must be that $\mathcal{M}, x+1 \Vdash \psi$. This means, by the inductive hypothesis again, that $\mathcal{M}', w_e \Vdash \psi$. But since $w_o >'_1 w_e$, we have that $min_{\leq'_1}([w_o]_1 \cap ||\neg\psi||) = \{w_o\}$, and so that $w_o \to_1^{'\neg\psi} w_o$. The reasoning for $x \neq y$ is similar, again with 1 and 2 reversed.
- $\phi := \psi \wedge \xi$.
  1. We show again only the case where x is odd. Suppose that $x \to_i^{\psi\wedge\xi} y$. Since $x \underline{\leftrightarrow} w_o$, we have to show that there is a $w' \in W$ such that $w_o \to_i^{'\psi\wedge\xi} w'$ and $y \underline{\leftrightarrow} w'$. Observe that either $y = x$ or $y = x + 1$ if $i = 1$ and $y = x$ or $y = x - 1$ if $i = 2$, which means that in both cases $y \in min_{\leq_i}([x]_i \cap ||\psi|| \cap ||\xi||)$ iff
     - either (*) $y \in min_{\leq_i}([x]_i \cap ||\psi||)$, in which case, by the first clause of the inductive hypothesis, there is a w' such that $w_o \to_i^{'\psi} w'$ and $y \underline{\leftrightarrow} w'$;

- or (**) $y \in min_{\leq_i}([x]_i \cap ||\xi||)$ in which case, again by the first clause of the inductive hypothesis, there is a w'' such that $w_o \rightarrow_i^{\prime\psi} w''$ and $y \underline{\leftrightarrow} w'$.

One can check that for each agent there is a unique $w' \in [w_o]_i$ such that $y \underline{\leftrightarrow} w'$, whatever y is, and so if both (*) and (**) hold then it must be that $w' = w''$, which means that we are done. We show the case where only (*) holds. For agent 1, this can only happen when $x = y$. By (*) and the inductive hypothesis this means that $w_o \rightarrow_i^{\prime\psi} w_o$, because there is no other $w' \in [w]_i$ such that $x \underline{\leftrightarrow} w'$. By assumption, we know furthermore that $\mathcal{M}, x \Vdash \xi$, and so by the inductive hypothesis that $\mathcal{M}', w_o \Vdash \xi$. It remains to be shown to $\mathcal{M}', w_e \nVdash \psi$. Since (**) does not hold, it has to be that $\mathcal{M}, x+1 \Vdash \xi$: we know that $\mathcal{M}, x \Vdash \xi$ and $x >_1 x+1$. This means that $\mathcal{M}, x+1 \nVdash \psi$, for otherwise we would have $x+1 \in min_{\leq_1}([w]_i \cap ||\psi|| \cap ||\xi||)$, against the minimality of x. By the inductive hypothesis, then know that $\mathcal{M}', w_e \nVdash \psi$,as required. The case for agent 2 follows the same line, except that (**) can only fail if $x \neq y$.

- $\phi := K_i\psi$.

1. Suppose that x is odd and $x \rightarrow_i^{K_j\psi} y$. We only show the case for $i = 1$. Assume that $j = 1$ as well. Then $\mathcal{M}, y \Vdash K_1\psi$. By positive introspection of K_i, this means that $\mathcal{M}, y' \Vdash K_1\psi$ for all $y' \in [x]_1 = \{x, x+1\}$, and since $K_i\phi \rightarrow \phi$ is also valid for K_i, we get that $\mathcal{M}, y' \Vdash \psi$ for all such y'. By the inductive hypothesis this means that $\mathcal{M}', w_o \Vdash \psi$ and $\mathcal{M}', w_e \Vdash \psi$, and so that $\mathcal{M}', w_o \Vdash K_1\psi$ and $\mathcal{M}', w_e \Vdash K_1\psi$. Since y is either x or $x+1$, we get that for any $w' \underline{\leftrightarrow} y$, $w_o \rightarrow_1^{\prime K_1\psi} w'$. Suppose now that $j = 2$ and $y = x$. This means that $\mathcal{M}, x \Vdash K_2\psi$. Again by positive introspection and the truth axiom, we get that $\mathcal{M}, x \Vdash \psi$ and $\mathcal{M}, x-1 \Vdash \psi$. Using our inductive hypothesis twice, we conclude that $\mathcal{M}', w_o \Vdash \psi$ and $\mathcal{M}', w_e \Vdash \psi$. But this covers all $w' \in [w_o]_2 = [w_o]_1$, and in particular w_o, so we have that $w_o \rightarrow_1^{\prime K_2\psi} w_o$. The same reasoning applies *mutatis mutandis* when $y = x+1$, and if x is even.

- $\phi := B_j^{\xi}\psi$.

1. Suppose that x is odd and $x \rightarrow_i^{B_j^{\xi}\psi} y$. Assume that $i \neq j$, and suppose that $i = 1$, the argument for $i = 2$ being entirely symmetric. We first show that it cannot be the case that $y = x$, for it would imply that $\mathcal{M}', w_o \Vdash B_2^{\xi}\psi$ while $\mathcal{M}', w_e \Vdash \neg B_2^{\xi}\psi$, which is impossible since $[w_o]_2 = [w_e]_2$. If $x = y$ then by the minimality of x within $[x]_1 \cap ||B_2^{\xi}\psi||$ it must be that both:

- (*) $\mathcal{M}, x \Vdash B_2^{\xi}\psi$ and
- (**) $\mathcal{M}, x+1 \Vdash \neg B_2^{\xi}\psi$.

If (*) then for all $y' \in min_{\leq_2}([x]_2 \cap ||\xi||)$ we have that $\mathcal{M}, y' \Vdash \psi$. If $[x]_2 \cap ||\xi||$ is empty, then $\mathcal{M}, x \nVdash \xi$ and $\mathcal{M}, x \nVdash \xi$, which means by our inductive hypothesis that $\mathcal{M}', w_o \nVdash \xi$ and $\mathcal{M}', w_e \nVdash \xi$, and so that $\mathcal{M}', w_o \Vdash B^{\xi}_{2}\psi$ trivially. If $x \in min_{\leq_2}([x]_2 \cap ||\xi||) \subseteq ||\psi||$, then we know by the inductive hypothesis that $\mathcal{M}, w_o \Vdash \xi \wedge \psi$. But since w_o is $\geq'_2$-minimal in $[w_o]_2$, we can conclude that $\mathcal{M}', w_o \Vdash B^{\xi}_{2}\psi$ as well. Finally, if $x \notin min_{\leq_2}([x]_2 \cap ||\xi||) \subseteq ||\psi||$, it must be that $x \rightarrow^{\xi}_{2} x-1$. By the inductive hypothesis we know that $w_o \rightarrow^{\xi}_{2} w_e$, which can only happen if $\mathcal{M}', w_o \Vdash \neg\xi$, from which we can conclude that $\mathcal{M}', w_o \Vdash B^{\xi}_{2}\psi$. So from (*) we get be that $\mathcal{M}', w_o \Vdash B^{\xi}_{2}\psi$. Now, by (**) we know that there is a $y' \in min_{\leq_2}([x+1] \cap ||\xi||)$ such that $x \notin ||\psi||$. This y' is either $x+1 \underline{\leftrightarrow} w_e$ or $x+2 \underline{\leftrightarrow} w_o$. In the first case we get from our inductive hypothesis that $w_e \rightarrow'^{\xi}_{2} w_e$, which can only happen if $min_{\leq_2}([x+1] \cap ||\xi||) = \{w_e\}$, and thus if $\mathcal{M}', w_e \Vdash \neg B^{\xi}_{2}\psi$. In the second case we get by inductive hypothesis that $\mathcal{M}', w_o \Vdash \xi \wedge \neg\psi$, from which we also know that $\mathcal{M}', w_e \Vdash \neg B^{\xi}_{2}\psi$, since w_o is $\geq'_2$-minimal in $[w_o]_2$ and $B^{\phi}_{i}\phi' \rightarrow K_i B^{\phi}_{i}\phi'$ is valid in epistemic plausibility models. From (**) we thus know that both $\mathcal{M}', w_e \Vdash \neg B^{\xi}_{2}\psi$, which means in conjunction with (*) that it cannot be that $x = y$.

Assume thus that $x \neq y$. This means that $x \rightarrow^{B^{\xi}_{2}\psi}_{1} x+1$. We are done if we can show that $w_o \rightarrow'^{B^{\xi}_{j}\psi}_{i} w_e$, for which it is enough to show that $\mathcal{M}', w_e \Vdash B^{\xi}_{2}\psi$. That $x \rightarrow^{B^{\xi}_{2}\psi}_{1} x+1$ means that $\mathcal{M}, x_1 \Vdash B^{\xi}_{2}\psi$. From there we reach the intended conclusion by following the same steps as above for (*).

Suppose then that $i = j = 1$. Then $x \rightarrow^{B^{\xi}_{1}\psi}_{1} y$. This means that $\mathcal{M}, y \Vdash B^{\xi}_{1}\psi$ and $\mathcal{M}, x \Vdash B^{\xi}_{1}\psi$ because $B^{\phi}_{i}\phi' \rightarrow K_i B^{\phi}_{i}\phi'$ is valid in epistemic plausibility models. This means, first, $x \rightarrow^{\top}_{1} y$ and thus by the inductive hypothesis that there is a $w' \underline{\leftrightarrow} y$ such that $w_o \rightarrow'^{\top}_{1} y$, i.e. that w' is $\geq'_1$-maximal in $[w_o]_i$. If we can show that $\mathcal{M}', w_o \Vdash B^{\xi}_{1}\psi$ then we are thus done. If $\mathcal{M}, x \Vdash B^{\xi}_{1}\psi$ because $[x]_i \cap ||\xi|| = \emptyset$ then we are done. Otherwise, if $\mathcal{M}, x+1 \Vdash \xi$ then by the inductive hypothesis we know that $\mathcal{M}', w_e \Vdash \xi \wedge \psi$ and so we are done because w_e is $\geq'_1$-maximal in $[w_o]_i = [w_e]_i$. If finally $\mathcal{M}, x+1 \nVdash \xi$ but yet $[x]_i \cap ||\xi|| \neq \emptyset$ then it must be that $\mathcal{M}, x \Vdash \xi \wedge \psi$. But then by the inductive hypothesis we know that $\mathcal{M}', w_o \Vdash \xi \wedge \psi$ and $\mathcal{M}', w_e \nVdash \xi$, which is enough to show that $\mathcal{M}', w_o \Vdash B^{\xi}_{1}\psi$. The argument for $i = j = 2$ is symmetric.

- $\phi := C_G\psi$.

  1. The cases when G is a singleton boils down to knowledge. So we consider the case were $G = \{1, 2\}$. Assume $x \underline{\leftrightarrow} w$ (H) and $x \rightarrow^{C_{\{1,2\}}\psi}_{i} y$ (0). By definition of $\rightarrow^{C_{\{1,2\}}\psi}_{i}$ it follows that $\mathcal{M}, y \Vdash C_{\{1,2\}}\psi$. By definition of $\mathcal{M}$ it follows that for all $z \in |\mathcal{M}| = \mathbb{Z}$ we have $\mathcal{M}, z \Vdash \psi$. By IH it follows that for all $v \in |\mathcal{M}'|$ we have $\mathcal{M}', v \Vdash \psi$. Moreover in both models $C_{\{1,2\}}\psi$ is satisfied everywhere (1). Now assume that $i = 1$, and that x is even. It follows that $\mathcal{K}_1[x] = \{x-1, x\}$. Moreover x is the minimum of $\mathcal{K}_1[x]$ (2). So by (0), (1) and (2) we have $x = y$. Now we have

$x \underline{\leftrightarrow} w_e$ and in the second model $w_e \rightarrow_i^{C_{\{1,2\}}\psi} w_e$. Now assume that x is odd. It follows that $\mathcal{K}_1[x] = \{x, x+1\}$. Moreover $x+1$ is the minimum of $\mathcal{K}_1[x]$ (3). So by (0), (1) and (3) we have $x+1 = y$. Now we have $y = x+1 \underline{\leftrightarrow} w_e$ and in the second model $w_o \rightarrow_i^{C_{\{1,2\}}\psi} w_e$. Now assume $i = 2$ and that x is odd. It follows that $\mathcal{K}_2[x] = \{x-1, x\}$. Moreover x is the minimum of $\mathcal{K}_2[x]$ (2). So by (0), (1) and (2) we have $x = y$. Now we have $x \underline{\leftrightarrow} w_o$ and in the second model $w_o \rightarrow_i^{C_{\{1,2\}}\psi} w_o$. Suppose finally that x is even. It follows that $\mathcal{K}_2[x] = \{x, x+1\}$. Moreover $x+1$ is the minimum of $\mathcal{K}_2[x]$ (3). So by (0), (1) and (3) we have $x+1 = y$. Now we have $y = x+1 \underline{\leftrightarrow} w_o$ and in the second model $w_e \rightarrow_i^{C_{\{1,2\}}\psi} w_o$.

- $\phi := CB_G\psi$. We have that $x \rightarrow_i^{CB_G\phi} y$ iff $x \rightarrow_i^{C_G\phi} y$ in $\mathcal{M}$, and similarly in $\mathcal{M}'$.

□

This concludes the proof of the Claim and the whole argument. □

Existence of a finite syntactic derivation of the Agreement Theorem. In what follows we prove the existence of a finite syntactic derivation of Corollary 1, and additional facts about an interesting logic of agreement. The language we use authorizes the following basic programs.

$$\alpha := 1 \mid 2 \mid 1 \cup 2 \mid (1 \cup 2)^* \mid \geq_j \mid >_j$$

where j ranges over $\{1, 2\}$. We additionally authorize intersection of the basic program only.

$$\beta := \alpha \mid \alpha \cap \alpha$$

Finally we recursively define our language as follows:

$$\phi := p \mid i \mid x \mid \neg\phi \mid \phi \wedge \phi \mid \langle\beta\rangle\phi \mid @_i\phi \mid @_x\phi$$

where i ranges over a countable set of nominals NOM, x over a countable set of state variables SVAR and p over a countable set of proposition letters PROP. All these sets are assumed to be disjoint. Let us call this language $\mathcal{H}(\downarrow, @)[1, 2, (1 \cup 2), \geq_j, >_j, C_G, Res(\cap)]$. We immediately stress that our usage of intersection, union and of the strict modality does not increase the expressive power of $\mathcal{H}(\downarrow, @)[1, 2, \geq_j, C_G]$, i.e. of the fragment that does not allow intersection, union, or the strict modality. We give below reduction axioms that sustain this claim.

The programs $\{1, 2, \geq_j, >_j\}$ are interpreted in the obvious way. For example R_1 stands for $\sim_1$. The language $\mathcal{H}(\downarrow, @)[1, 2, (1 \cup 2), \geq_j, >_j, C_G, Res(\cap)]$ is interpreted on epistemic plausibility models together with an assignment function $g : \text{SVAR} \rightarrow W$ that maps states variables to states. The valuation function maps elements of NOM to singletons set of states. The following clauses cover the interpretation of the binder, of states variable and of the @ operator.

$$\begin{array}{lll}
\mathcal{M}, g, w \Vdash x & \text{iff} & g(x) = w \\
\mathcal{M}, g, w \Vdash i & \text{iff} & w \in V(i) \\
\mathcal{M}, g, w \Vdash @_x\phi & \text{iff} & \mathcal{M}, g, g(x) \Vdash \phi \\
\mathcal{M}, g, w \Vdash @_i\phi & \text{iff} & \mathcal{M}, g, v \Vdash \phi \text{ where } V(i) = \{v\} \\
\mathcal{M}, g, w \Vdash \downarrow x.\phi & \text{iff} & \mathcal{M}, g[g(x) := w], w \Vdash \phi
\end{array}$$

For the basic modalities $\{1, 2, \geq_j, >_j\}$ we have the classical scheme

$$\mathcal{M}, g, w \Vdash \langle\alpha\rangle\phi \quad \text{iff} \quad \exists v \text{ such that } (wR_\alpha v) \text{ and } \mathcal{M}, g, v \Vdash \phi$$

For the fragment of *PDL* we are using the clauses are:

$$\begin{array}{lll} \mathcal{M}, g, w \Vdash \langle 1 \cup 2\rangle\phi & \text{iff} & \exists v \text{ such that } (w \sim_1 v \text{ or } w \sim_2 v) \text{ and } \mathcal{M}, g, v \Vdash \phi \\ \mathcal{M}, g, w \Vdash \langle 1 \cup 2\rangle^*\phi & \text{iff} & \exists v \text{ such that } w \sim^*_{\{1\cup 2\}} v \text{ and } \mathcal{M}, g, v \Vdash \phi \end{array}$$

The second operator is nothing but a notational variation of common knowledge, in the sense that $C_{\{1,2\}}\phi \leftrightarrow \neg\langle 1 \cup 2\rangle^*\neg\phi$ which is useful to shorten our formulas when intersection is involved. The first one is the diamond version of the "everybody knows" modality. Finally we give the clause for intersection.

$$\mathcal{M}, g, w \Vdash \langle\alpha \cap \beta\rangle\phi \quad \text{iff} \quad \exists v \text{ such that } (wR_\alpha v \text{ and } wR_\alpha v) \text{ and } \mathcal{M}, g, v \Vdash \phi$$

Proposition 2. *The following reduction are axioms are sound on the class of epistemic plausibility models.*
1.$\langle<\rangle\phi \leftrightarrow \downarrow x.\langle\leq\rangle(\phi \wedge [\leq]\neg x)$ *where x does not occur in* ϕ.
2. $\langle\alpha \cap \beta\rangle\phi \leftrightarrow \downarrow x.\langle\alpha\rangle(\downarrow y.(\phi \wedge @_x\langle\beta\rangle y))$ *where* x, y *does not occur in* ϕ.
3. For $\alpha \in \{1, 2, (1 \cup 2), (1 \cup 2)^*, \geq_1\geq_2\}$ *and* $\beta \in \{1, 2, (1 \cup 2), (1 \cup 2)^*\}$*:*
$\langle\alpha \cap \beta\rangle\phi \leftrightarrow \downarrow x.\langle\alpha\rangle(\phi \wedge \langle\beta\rangle x))$ *where x does not occur in* ϕ.
4. For $\alpha \in \{1, 2, (1 \cup 2), (1 \cup 2)^*\}$*:*
$\langle>_j \cap \alpha\rangle\phi \leftrightarrow \downarrow x.\langle\alpha\rangle(\phi \wedge [\geq]\neg x \wedge \langle\alpha\rangle x))$ *where x does not occur in* ϕ.

Proof. Sahlqvist correspondence argument. (2) is valid for arbitrary programs on arbitrary relational structures. (3) and (4) uses the fact that the epistemic relation is symmetric. □

Let us note that the latter reduction axioms, which draw on the symmetry of the epistemic relations, are more efficient in terms of hybrid operator alternation and fresh variables we need. Therefore we will rather work with them in the syntactic proof.

Corollary 6. *On the class of epistemic plausibility models* $\mathcal{H}(\downarrow, @)[1, 2, \geq_j, C_G]$ *is at least as expressive* $\mathcal{H}(\downarrow, @)[1, 2, (1 \cup 2) \geq_j, >_j, C_G, Res(\cap)]$.

In addition to the reduction axioms given and the axiomatization of the pure hybrid logic $\mathcal{H}(\downarrow, @)$ (see ten Cate (2005)) we will make us of additional axioms in this proof. Their soundness is proved below.

Proposition 3. *The following axioms are valid on the class of well-founded epistemic-plausibility model.*
5. $[>]([>]p \rightarrow p) \rightarrow [>]p$
6. $\downarrow x.((\langle\geq\rangle(\neg\langle\geq\rangle x \wedge p)) \rightarrow (((\langle\geq\rangle((\neg\langle\geq\rangle x \wedge p) \wedge \neg \downarrow z.\langle\geq\rangle(\neg\langle\geq\rangle z \wedge p)))))$
7. For $\alpha \in \{1, 2, (1 \cup 2), (1 \cup 2)^*\}$*:* $\downarrow x.[\alpha]\langle\alpha\rangle x$
8. $\langle\alpha \cap \beta\rangle\phi \rightarrow (\langle\alpha\rangle\phi \wedge \langle\beta\rangle\phi)$
9. $\langle\alpha^*\rangle\phi \leftrightarrow (\phi \vee \langle\alpha\rangle\langle\alpha^*\rangle\phi)$

Proof. (5) is sound on the class of <-well-founded frames (see Blackburn et al. (2001)). For (6) note that by Ax.(1), (6) is equivalent to $\langle>\rangle p \rightarrow \langle>\rangle(p \wedge \neg\langle>\rangle p)$ which is equivalent on the level of frames to (5). (7) is sound on class of frames for which R_α is symmetric. (8) is obvious. See Blackburn et al. (2001). □

Theorem 12. *$\neg C_{\{1,2\}}(B_1 p \wedge \neg B_2 p)$ is a theorem of the logics of $\mathcal{H}(\downarrow, @)[\geq_j]$ extended by* Löb*'s axiom (item 5 and 6 in Proposition 3), the multi-agent S5-epistemic logic including C_G, S4 for $\geq_j$ and the Axiom of Common Prior: $\langle \geq_i \rangle \phi \leftrightarrow \langle \geq_j \rangle \phi$.*

Proof. For convenience we additionally use axioms 7 and 8 from Proposition 3 as useful shortcuts. (Löb(>)) is either Axiom 5 or 6 in Proposition 3. For the axiomatization $\mathcal{H}(\downarrow, @)[\geq_j]$ see ten Cate (2005). For the multi-agent S5-epistemic logic including C_G see Fagin et al. (1995).

In the following proof after the axiom of common prior has been applied, we drop the label, since the plausibility relation is thus the same for both agents. Our goal is to derive a contradiction from the assumption that disagreement is common knowledge.

(0)	$[(1 \cup 2)^*]\langle \geq \cap 2 \rangle(([> \cap 2]\bot) \wedge \neg p)$	Hypothesis.
(1)	$[(1 \cup 2)^*][\geq \cap 1](([> \cap 1]\bot) \rightarrow p)$	Hypothesis.
(2)	$[(1 \cup 2)^*][2][\geq \cap 1](([> \cap 1]\bot) \rightarrow p)$	From (0) by PDL.
(3)	$[(1 \cup 2)^*][\geq \cap 2][\geq \cap 1](([> \cap 1]\bot) \rightarrow p)$	From (2) by $\cap$.
(4)	$[(1 \cup 2)^*][\geq \cap 2][\geq \cap 1](\neg p \rightarrow \top)$	From (3) by PL.
(4′)	$[(1 \cup 2)^*][\geq \cap 2](\neg p \rightarrow \langle > \cap 1 \rangle \top)$	From (3) by Ref for $(\geq \cap 1)$.
(5)	$[(1 \cup 2)^*]\langle \geq \cap 2 \rangle(([> \cap 2]\bot) \wedge \neg p \wedge (\neg p \rightarrow \langle > \cap 1 \rangle \top))$	From (1) and (4′) by ML.
(6)	$[(1 \cup 2)^*]\langle \geq \cap 2 \rangle(([> \cap 2]\bot) \wedge \neg p \wedge \langle > \cap 1 \rangle \top)$	From (5) by PL and ML.
(7)	$\downarrow x.((\langle \geq \rangle(\neg \langle \geq \rangle x \wedge p)) \rightarrow (((\langle \geq \rangle((\neg \langle \geq \rangle x \wedge p) \wedge \neg \downarrow z.\langle \geq \rangle(\neg \langle \geq \rangle z \wedge p)))))$	Axiom. Löb for $>$

(8) $\downarrow x.((\langle\geq\rangle(\neg\langle\geq\rangle x \land \langle(1\cup 2)^*\rangle x)) \to$ $(((\langle\geq\rangle((\neg\langle\geq\rangle x \land \langle(1\cup 2)^*\rangle x) \land \neg\downarrow z.\langle\geq\rangle(\neg\langle\geq\rangle z$ $\land \langle(1\cup 2)^*\rangle x)))))$	From (7) by Uni Sub of p by $\langle(1\cup 2)^*\rangle x$.
(9) $[(1\cup 2)^*][\geq\cap 2]\downarrow x.(((\langle\geq\rangle(\neg\langle\geq\rangle x \land$ $(((\langle\geq\rangle((\neg\langle\geq\rangle x \land \langle(1\cup 2)^*\rangle x) \land$ $\neg\downarrow z.\langle\geq\rangle(\neg\langle\geq\rangle z \land \langle(1\cup 2)^*\rangle x)))))$	From (8) by Gen
(10) $\langle>\cap 1\rangle\phi \leftrightarrow\downarrow x.\langle\geq\rangle((\neg\langle\geq\rangle x \land \langle 1\rangle x) \land \phi)$	Reduction Axiom $<\cap 1$
(11) $\langle>\cap 1\rangle\top \leftrightarrow\downarrow x.\langle\geq\rangle(\neg\langle\geq\rangle x \land \langle 1\rangle x)$	From (10) by Uni Sub of ϕ by $\top$ and PL.
(12) $[(1\cup 2)^*][\geq\cap 2](\langle>\cap 1\rangle\top \leftrightarrow$ $\downarrow x.\langle\geq\rangle(\neg\langle\geq\rangle x \land \langle 1\rangle x))$	From (11) by Gen
(13) $[(1\cup 2)^*]\langle\geq\cap 2\rangle((([>\cap 2]\bot) \land \neg p \land \langle>\cap 1\rangle\top)$ $\land \downarrow x.\langle\geq\rangle(\neg\langle\geq\rangle x \land \langle 1\rangle x))$	From (6) and (12) by ML.
(14) $\langle 1\rangle x \to \langle(1\cup 2)^*\rangle x$	PDL
(15) $[(1\cup 2)^*]\langle\geq\cap 2\rangle((([>\cap 2]\bot) \land \neg p \land \langle>\cap 1\rangle\top) \land$ $\downarrow x.[(\langle\geq\rangle(\neg\langle\geq\rangle x \land \langle 1\rangle x)) \land (\langle\geq\rangle(\neg\langle\geq\rangle x \land$ $\langle(1\cup 2)^*\rangle x) \land \neg\downarrow z.\langle\geq\rangle(\neg\langle\geq\rangle z \land \langle(1\cup 2)^*\rangle x))])$	From (13), (14) and (9) by ML and PL.
(16) $[(1\cup 2)^*]\langle\geq\cap 2\rangle((([>\cap 2]\bot) \land \neg p \land \langle>\cap 1\rangle\top) \land$ $\downarrow x.[(\langle\geq\rangle(\neg\langle\geq\rangle x \land \langle 1\rangle x)) \land (\langle\geq\rangle(\neg\langle\geq\rangle x \land$ $\langle(1\cup 2)^*\rangle x) \land \downarrow z.[\geq]((\langle(1\cup 2)^*\rangle x \to \langle\geq\rangle z))])$	From (15) by ML and PL.
(16′) $[(1\cup 2)^*]\langle\geq\cap 2\rangle($ $\downarrow x.[(\langle\geq\rangle(\langle(1\cup 2)^*\rangle x) \land \downarrow z.[\geq]((\langle(1\cup 2)^*\rangle x \to \langle\geq\rangle z))])$	From (16) by ML and PL.
(17) $\downarrow x.(\langle\geq\rangle(\langle(1\cup 2)^*\rangle x \land \phi) \to \langle\geq\cap(1\cup 2)^*\rangle\phi)$	Axiom. By symmetry of $(1\cup 2)^*$ and $\cap$
(17′) $\downarrow x.(\langle\geq\rangle(\langle(1\cup 2)^*\rangle x \land \downarrow z.[\geq]((\langle(1\cup 2)^*\rangle x \to$ $\langle\geq\rangle z)) \to \langle\geq\cap(1\cup 2)^*\rangle\downarrow z.[\geq]((\langle(1\cup 2)^*\rangle x \to \langle\geq\rangle z))$	From (17) by Uni Sub of ϕ by $\downarrow z.[\geq]((\langle(1\cup 2)^*\rangle x \to \langle\geq\rangle z)$.
(18) $[(1\cup 2)^*]\langle\geq\cap 2\rangle($ $\downarrow x.\langle\geq\cap(1\cup 2)^*\rangle(\downarrow z.[\geq]((\langle(1\cup 2)^*\rangle x \to \langle\geq\rangle z))$	From (16′) and (17′) by ML and PL.
(19) $[(1\cup 2)^*][(1\cup 2)^*][\geq\cap 1](([>\cap 1]\bot) \to p)$	From (1) by PDL.
(20) $[(1\cup 2)^*][2][(1\cup 2)^*][2][\geq\cap 1](([>\cap 1]\bot) \to p)$	From (19) by PDL.
(21) $[(1\cup 2)^*][2\cap\geq][(1\cup 2)^*\cap\geq][2\cap\geq][\geq\cap 1]$ $(([>\cap 1]\bot) \to p)$	From (20) by $\cap$
(22) $[(1\cup 2)^*][2\cap\geq][(1\cup 2)^*\cap\geq][2\cap\geq][\geq\cap 1]$ $(\neg p \to (\langle>\cap 1\rangle\top))$	From (20) by PL and ML.
(21′) $[(1\cup 2)^*][2\cap\geq][(1\cup 2)^*\cap\geq]\langle\geq\cap 2\rangle$ $(([>\cap 2]\bot) \land \neg p)$	From (0) by a similar derivation.
(21.1) $[(1\cup 2)^*]\langle\geq\cap 2\rangle(\downarrow x.\langle\geq\cap(1\cup 2)^*\rangle[$ $(\downarrow z.[\geq]((\langle(1\cup 2)^*\rangle x \to \langle\geq\rangle z)\ ) \land$ $\langle\geq\cap 2\rangle(([>\cap 2]\bot) \land \neg p)]$	From (18) and (21′) by ML and PL.
(21.2) $\downarrow x.[(1\cup 2)^*]\langle(1\cup 2)^*\rangle x$	Axiom for symmetry.
(21.3) $\downarrow x.[(1\cup 2)^*][2]\langle(1\cup 2)^*\rangle x$	PDL.
(21.4) $\downarrow x.[(1\cup 2)^*][2\cap\geq]\langle(1\cup 2)^*\rangle x$	By $\cap$
(21.5) $[(1\cup 2)^*]\langle\geq\cap 2\rangle(\downarrow x.\langle\geq\cap(1\cup 2)^*\rangle[$ $(\downarrow z.[\geq\cap 2]((\langle(1\cup 2)^*\rangle x \to \langle\geq\rangle z))$	From (21.1) by ML and $\cap$
(21.6) $[(1\cup 2)^*]\langle\geq\cap 2\rangle(\downarrow x.\langle\geq\cap(1\cup 2)^*\rangle[$ $(\downarrow z.\langle\geq\cap 2\rangle([(([>\cap 2]\bot) \land \neg p) \land$ $((\langle(1\cup 2)^*\rangle x \land \langle\geq\rangle z)]$	From (21.1), (21.4) and (21.5) by ML and PL.
(21.7) $[(1\cup 2)^*]\langle\geq\cap 2\rangle(\downarrow x.\langle\geq\cap(1\cup 2)^*\rangle[$ $(\downarrow z.\langle\geq\cap 2\rangle([((\langle(1\cup 2)^*\rangle x \land \langle\geq\rangle z \land \langle>\cap 1\rangle\top) \land$ $([>\cap 2]\bot) \land \neg p)]$	From (21.1), (21.4) and (21.5) by ML and PL.
(21.8) $[(1\cup 2)^*]\langle\geq\cap 2\rangle(\downarrow x.\langle\geq\cap(1\cup 2)^*\rangle[$ $(\downarrow z.\langle\geq\cap 2\rangle[(\langle>\cap 1\rangle\top))$	From (21.8) by ML and PL.
(21.81) $[(1\cup 2)^*]\langle\geq\cap 2\rangle(\downarrow x.\langle\geq\cap(1\cup 2)^*\rangle[$ $(\downarrow z.\langle\geq\cap 2\rangle[(\langle\geq\cap 1\rangle(\neg\langle\geq\rangle z)))$	From (21.8) by ML and PL.
(21.82) $[(1\cup 2)^*]\langle\geq\cap 2\rangle(\downarrow x.\langle\geq\cap(1\cup 2)^*\rangle[$ $(\downarrow z.[\geq](\neg\langle\geq\rangle z \to \neg\langle(1\cup 2)^*\rangle x))]$	From (21.1) by ML and PL.
(21.83) $[(1\cup 2)^*]\langle\geq\cap 2\rangle(\downarrow x.\langle\geq\cap(1\cup 2)^*\rangle[$ $(\downarrow z.\langle\geq\cap 2\rangle[(\langle\geq\cap 1\rangle(\neg\langle(1\cup 2)^*\rangle x)))$	From (21.81) and (21.82) by ML and PL.
(23) $[(1\cup 2)^*]\langle 2\rangle\ (\downarrow x.\langle(1\cup 2)^*\rangle[(\downarrow z.\langle 2\rangle[$ $(\langle 1\rangle(\neg\langle(1\cup 2)^*\rangle x)))$	From (21.83) by $\cap$. (A contradiction)
(23.1) $[(1\cup 2)^*]\langle 2\rangle\ (\downarrow x.\langle(1\cup 2)^*\rangle(\neg\langle(1\cup 2)^*\rangle x)))$	From (23) by PDL. (A contradiction)

□

Bourbaki-Witt Theorem

Theorem 13 (Bourbaki-Witt Bourbaki (1949))**.** *Let X be a chain complete poset. If $f : X \to X$ is inflationary (deflationary), then f has a fixed point.*

Proof of Observation 4

Proof. Taking $\wp(W \times W)^I$ as a product, it is easy to see that $\langle \wp(W \times W)^I, \subseteq \rangle$ is a poset. The intersection of a decreasing sequence is the greatest lower bound of this sequence. Finally it is easy to see by inspection Definition 5.1 that $f_G(A)$ is deflationary. Indeed for every i and we have $f_G(A)(\sim_i) \subseteq \sim_i$ and thus by definition of a product $f_G(A)(\times_{i \in I}(\sim_i)) \subseteq \times_{i \in I}(\sim_i)$. □

Below we give the general definition of a generated submodel.

Definition 6.2 (Generated submodel)**.** Given an *epistemic plausibility model* $\mathcal{M}$ $= \langle W, (\leq_i)_{i \in I}, (\sim_i)_{i \in I}, V \rangle$ and a $A \subseteq W$. The submodel of $\mathcal{M}$ generated by A (or A-generated submodel), that we note $\mathcal{M}^A$ is defined as follows:
$\mathcal{M}^{\mathcal{A}} = \langle W^A, (\leq_i^A)_{i \in I}, (\sim_i^A)_{i \in I}, V^A \rangle$

- $W^A = W \cap A$

and for each $i \in I$

- $\leq_i^A = \leq_i \cap W^A \times W^A$
- $\sim_i^A = \sim_i \cap W^A \times W^A$
- For each $v \in W^A$, $v \in V^A(p)$ iff $v \in V(p)$

Proof of Observation 6

Proof. It is easy to see that $\langle Sub(\mathcal{M}), \sqsubseteq \rangle$ is a poset. Moreover taking the submodel of $\mathcal{M}$ generated by the intersection of the domain of each element any decreasing sequence is the greatest lower bound of this sequence. Finally it is easy to see by inspection Definition 5.4 that $!B^\phi$ is deflationary. □

Proof of Theorem 5 To save on notation we write $B_i^{\alpha^f}(A)$ for $B_i^{\mathcal{M}_{\alpha,f}}(A)$.

Proof. Let α^f be the fixed point existing by Corollary 5.5.

Given an arbitrary state w in the domain of $\mathcal{M}_{\alpha^f}$ we prove that for any $w' \in [w]^*_{\alpha^f,I}$ and for any $i \in I$ we have $w \in B_i^{\alpha^f}(||\phi||)$ iff $w' \in B_i^{\alpha^f}(||\phi||)$. The proof is by induction on length of the smallest chain $C = \langle w_1 \sim_{\alpha^f,x} \dots \sim_{\alpha^f,y} w_n \rangle$ where $x, y \in I$, $w_1 = w$ and $w_n = w'$.

Base case. For $|C| = 1$ is immediate by definition.

Induction step. We have two cases.

Case 1. $w \in B_{\alpha^f,i}(||\phi||)$. Assume that we have a chain $C = \langle w_1 \sim_{\alpha^f,x} \dots \sim_{\alpha^f,y} w_{n+1} \rangle$ where $x, y \in I$, $w_1 = w$ and $w_{n+1} = w'$ of length $n + 1$. By IH we have $w_n \in B_{\alpha^f,i}(||\phi||)$ (1). We have now two subcases. Subcase 1a: $w_n \sim_{\alpha^f,i} w'$ but then by epistemic introspection of beliefs and (1) we have $w' \in B_{\alpha^f,i}(||\phi||)$. Subcase 1b: $w_n \sim_{\alpha^f,j} w'$ in C, for some $j \neq i$ in I. Now assume for contradiction that $w' \notin B_{\alpha^f,i}(||\phi||)$. It follows then by definition of the conditioning function

$f_I(\bigcap_{j\in I}\mathbb{B}_j^{\mathcal{M},w}(\|\phi\|^{\mathcal{M}}))$ that $f_I(\bigcap_{j\in I}\mathbb{B}_j^{\mathcal{M},w}(\|\phi\|^{\mathcal{M}}))(\sim_{\alpha^f,j}) \subsetneq \sim_{\alpha^f,j}$. Contradicting the choice of α^f.

Case 2. The argument for the case of $w \notin B_{\alpha^f,i}(\|\phi\|)$ is entirely similar, except that we use negative introspection of beliefs (if $w \notin B_i(X)$ then $\mathcal{K}_i[w] \subseteq \neg B_i(X)$). □

Proof of Theorem 7 The proof follows the same line as for Theorem 5.

A Method for Solving Nash Equilibrium of Game based on Public Announcement Logic

Jianying Cui
Institute of Logic and Intelligence, Southwest University (Chongqing, China).
cjy85911@swu.edu.cn

Abstract

Describing the interactional behavior of rational agents and seeking equilibrium solutions are two main perspectives in game theory. In game theory, an agent is always assumed to be rational, i.e. to maximize her expected payoff, given her probabilistic beliefs about the strategies used by her opponents. However, Game theory itself can't precisely model the higher-order information change of mutual knowledge among agents, so these interpretations and expressions about rationality are vague in game theory. In this paper, we redefine rationality by incorporating epistemic ingredient, provide a method for helping to solve and refine Nash equilibriums based on public announcement logic, and indicate this method can reduce a game model with iterating some announcement of proper rationality assertion. Meanwhile, we show the iterating announcement of this rationality assertion characterizes iterated admissibility.

Key words: strategic-form game; iterated admissibility algorithm ; rationality ; public announce logic

1 Introduction

In general, game-theoretic solution is relevant to some specific algorithm. For instance, a perfect information extensive-form game can be solved by Backward Induction, Iterated Eliminations of Strictly Dominated strategy can be used for static-form game. Whether Backward Induction algorithm or Iterated Eliminations of Strictly Dominated strategy algorithm, are both based on a assumption which is all players are rational, and the rationality is postulated as common knowledge among players. However, lots of literatures show that this assumption is too ideal and crude.(Refer to Aumann (1999),Rubinstein (1998) and others). In this paper, based on the epistemic game model constructed by van Benthem (2006a), we give a deep analysis about players' rationality from the Dynamic Epistemic perspective, redefine rationality ,and indicate this rationality can be as a proper announcement assertion in public announcement logic(PAL). Consequently, bring forward a new method for Nash Equilibrium based on PAL, which characterizes just algorithm of iterated admissibility introduced by Gilli (2002) and Brandenburger et al. (2008), The algorithm is known to provide a valuable criterion for selecting among multiple equilibriums and to yield sharp predictions in finite games (Refer to Brandenburger et al. (2008)). Concretely, the classical concept of iterated admissibility, which is well established in the theory of finite game, is based on the notion of (weak) dominance. Considering two strategies s_i, r_i of a player i, s_i dominates r_i if, against any choice of stategies of the other players, s_i performs at least as well as r_i, but there are cases in which s_i performs strictly better than r_i. In terms of our rationality, atomic proposition "a player is rational" is just fails at the worlds which corresponding to the strategies are dominated. Accordingly, after public announcing this proposition for one time, we can remove simultaneously all these worlds. This leads to a new subgame model. In this subgame model, players possibly discover that some of their retaining strategies are again dominated owing to absence of some worlds. So, repeat the announcement, and remove continuously irrational worlds until this proposition holds at every world in some subgame model. For a game with finite strategy spaces, the procedure stabilities after a finite number of stages, the solution concept postulating that outcomes of a game should involve only strategies that survive the iterated elimination is called iterated admissibility (a precise definition follows). This procedure required extensive efforts. However, we will indicate the procedure which is constructed by removing worlds after repeated public announcements of some proper assertion is just corresponding to the procedure of iterated admissibility.

The paper is organized as follows. In the next section we review the definition of strategic-form game , admissibility and iterated deletion procedures mentioned by M.Gill. Reviewing for Public Announcement Logic is arranged in Section 3. In Section 4, we introduce an epistemic game model mentioned by J.van Benthem, and point out why we need incorporate epistemic ingredient in a notion of rationality in game. Thus, we redefine a rationality in this Section. In Section 5, we show it is justified to repeated announcing for this rationality can characterize the iterated admissibility. In Section 6,we discuss related, while Section 7 contains a summary and concluding remarks.

2 Game and dominance

In this paper we restrict attention to finite strategic-form games with pure strategies, which are defined as follows.

Definition 2.1. A finite strategic-form game is a quintuple

$$G = \langle N, \{S_i\}_{i\in N}, \{\succcurlyeq_i\}_{i\in N}, \{U_i\}_{i\in N}\rangle,$$

where

$N = \{1, 2, \cdots, n\}$ is a set of players;

S_i is a finite set of strategies of player $i \in N$;

$\succcurlyeq_i$ is player i's preference; [1]

U_i is a list of all players' payoffs function: $S \to R$ that give player i's vN-M utility $U_i(s)$ for each strategy profile $s \in S$. (where $S = S_1 \times \cdots \times S_n$)

Given a player i we denote by S_{-i} the set of strategy profiles of the players other than i, that is, $S_{-i} = S_1 \times \cdots \times S_{i-1} \times S_{i+1} \times \cdots \times S_n$.When we want to focus on player i we shall denote the strategy profile $s \in S$ by $(s_i, s_{-i})$where $s_i \in S_i$, and $s_{-i} \in S_{-i}$.

Definition 2.2. Given a game $G = \langle N, \{S_i\}_{i\in N}, \{\succcurlyeq_i\}_{i\in N}, \{U_i\}_{i\in N}\rangle$, let s_i' and s_i'' are available strategies for player i, and a set $S'_{-i} \subseteq S_{-i}$, we say s_i' is weakly dominated[2] by s_i'' on S'_{-i} if
$u_i(s_i'', s_{-i}) \geq u_i(s_i', s_{-i})$ for all $s_{-i} \in S_{-i}$ and $u_i(s_i'', s'_{-i}) > u_i(s_i', s'_{-i})$
for some $s'_{-i} \in S_{-i}$. For a set $S' \subseteq S$, we say that a strategy $s_i'' \in S_i$ is admissible with respect to S_i', if no strategy in S_i' dominates s_i'' on S'_{-i}. In additional, a dominated strategy r_i is also called inadmissible for a player i.

Definition 2.3. The iteratively weakly undominated strategies ($IWUS$)[3] is the following procedure. Given a game $G = \langle N, \{S_i\}_{i\in N}, \{\succcurlyeq_i\}_{i\in N}, \{U_i\}_{i\in N}\rangle$, let WUD be the set of iterated weakly undominated strategies (or iteratively admissible strategies) of G define recursively as follows.
$WUD = \prod_{i\in N} WUD_i$, where $WUD_i = \bigcap_{m\geqslant 0} WUD_i^m$, with $WUD_i^0 = S_i$ and for $m \geqslant 1, WUD_i^m = WUD_i^{m-1} \setminus WD_i^{m-1}$, where $WD_i^{m-1} = \{s_i \in WUD_i^{m-1} \mid s_i$ is inadmissible with respect to $WUD^{m-1}\}$.

Note that in definition 2.4, it is assumed that at each stage all dominated strategies are simultaneously deleted. In contrast to most equilibrium concepts, iterative admissibility yields a rectangular set of strategy profiles, i.e. a Cartesian product of sets. Accordingly, whether the choice of a particular player is rational in the sense does not depend on the choices of the other players. This $IWUS$ procedures is illustrated in the following figure 1.

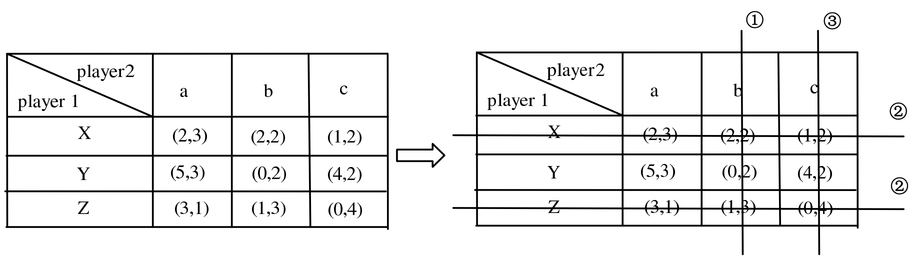


Figure 1: $IWUS$ procedure

[1]player i's preference is a partial relation, satisfied reflexivity,antisymmetry and transitivity. And the interpretation of $s_i \succcurlyeq_i s_i'$ is that, according to player i, the utilities obtained by selecting s_i is at least as good as that of selecting s_i'. Meanwhile, the strict ordering $\succ_i$ is defined as usual: $s_i \succ_i s_i'$ if and only if $s_i \succcurlyeq_i s_i'$ and not $s_i' \succcurlyeq_i s_i$. The interpretation of is that player i strictly prefers s_i to s_i'.

[2]Later, we will refer to dominated(or dominate) as weakly dominated(or weakly dominate).

[3]The procedure of IWUS is the algorithm of iterated admissibility introduced by Gilli (2002).

$WUD_1^0 = \{X, Y, Z\}, WD_1^0 = \emptyset, WUD_2^0 = \{a, b, c\}, WD_2^0 = \{b\}$

$WUD_1^1 = \{X, Y, Z\}, WD_1^1 = \{X, Z\}, WUD_2^1 = \{a, c\}, WD_2^1 = \emptyset$

$WUD_1^2 = \{Y\} = WUD_1, WD_1^2 = \emptyset, WUD_2^2 = \{a, c\}, WD_2^2 = \{c\}$

$WUD_2^3 = \{a\} = WUD_2$. Thus, $WUD = \{(Y, a)\}$

In Gilli (2002), the next proposition holds. In some sense, it is in accord with the announcement limits we mentioned later.

Theorem 1. *In every game G,there exists a finite number $K \in N$such that $\forall n \geq K, WD_i^n(G) = WD_i(G) \neq \emptyset$ for every $i \in N$.*[4]

3 Public Announcement Logic

The prominent application of logic is to model reasoning about agent's knowledge and belief in multi-agent system, although epistemic logic can basically finish this task. But the standard epistemic logic can not describe communication of knowledge and information. It's very important to model the change of agent's knowledge and information in a multi-agent system. So logicians develop many dynamic epistemic logics to model the change of high-order information of agents.(Refer to Baltag et al. (1999), Gerbrandy (1999), van der Meyden (2005) and van Ditmarsch et al. (2007) *et al*). Being a simple dynamic epistemic logic, public announcement logic can enhance expression power by adding a dynamic modality $[\varphi]$ to the standard epistemic logic, so as to describe and character the change of agent's information arose by agent's action. Apart from all the formation rules of standard epistemic logic, PAL include this dynamic modality $[\varphi]$,which means public announcing proposition $[\varphi]$. so, the meaning of formula $[\varphi]\psi$ is: after truthful public announcement of φ, formula ψ holds. The truth condition is that: $M, w \vDash [\varphi]\psi$ iff $M, w \vDash \varphi$ then $M|_\varphi, w \vDash \psi$.

With this language, we can say things like $[\varphi]K_i\psi$: after truthful public announcement of φ, agent i knows ψ, or $[\varphi]C_B\varphi$: after its announcement, φ has become common knowledge in the group B of agents and so on. Furthermore, the assertion of announcement can inclusive epistemic ingredient. The following formulas indicates some typical valid principles in PAL.

- $[\alpha]p \leftrightarrow (\alpha \rightarrow p)$ atomic permanence
- $[\alpha]\neg\beta \leftrightarrow (\alpha \rightarrow \neg[\alpha]\beta)$ announcement and negation
- $[\alpha](\beta \wedge \gamma) \leftrightarrow ([\alpha]\beta \wedge [\alpha]\gamma)$ announcement and conjunction
- $[\alpha]K_i\beta \leftrightarrow (\alpha \rightarrow K_i[\alpha]\beta)$ announcement and knowledge
- $[\alpha]\beta$ is valid if and only if $[\alpha]C_N\beta$ is valid.

Definition 3.1. For any model M and formula φ ,the announcement limit $\sharp(\varphi, M)$ is the first submodel in the repeated announcement sequence where announcing φ has no further effect.[5].

[4] Its proof refer to p15 in Gilli (2002).
[5] This definition is in van Benthem (2006a)

Consequently, J.van Benthem indicated for any model M we can keep announcing φ, retaining just those worlds where φ holds in van Benthem (2006a). This yields an sequence of nested decreasing sets, which must stop in finite models, i.e.$\sharp(\varphi, M)$.

4 Epistemic Game Model and Rationality

In order to give a dynamic epistemic analysis of game solution as model change, we provides a epistemic game model based the structure of some initial game, which is derived from van Benthem (2006a).

Definition 4.1. The full epistemic model over G is a multi-$S5_2$ epistemic structure M_G [6] whose worlds are all strategy profiles, and whose epistemic accessibility R_i for player i is defined as the equivalence relation of agreement of profiles in the i'th coordinate. That's to say, adding to Kripke frame $S5_n$ function σ_i: $\Omega \to S_i (i \in N)$, satisfying the following property: $R_i wv \Leftrightarrow \sigma_i(w) = \sigma_i(v)$.Meanwhile, incorporating the following valuation :

$M_G, w \models s_i$ if and only if $\sigma_i(w) = s_i$;

$M_G, w \models (s_i \succcurlyeq_i s_i')$ if and only if $u_i(s_i, \sigma_{-i}(w)) \geq u_i(s_i', \sigma_{-i}(w))$.

$M_G, w \models (s_i \succ_i s_i')$ if and only if $u_i(s_i, \sigma_{-i}(w)) > u_i(s_i', \sigma_{-i}(w))$.

Figure 2 represents a full epistemic game model from the game showed in figure 1. For convenience, we always take game model as epistemic game model, and make analysis on it directly. Here player 2's uncertainty relation R_2 runs along columns, because player 2 knows his own action, player 2, but not that of player 1. The uncertainty relation of player 1 runs among the rows.

$$\begin{tikzcd}
(X,a) \arrow[r, dash, "1"] \arrow[d, dash, "2"] & (X,b) \arrow[r, dash, "1"] \arrow[d, dash, "2"] & (X,c) \arrow[d, dash, "2"] \\
(Y,a) \arrow[r, dash, "1"] \arrow[d, dash, "2"] & (Y,b) \arrow[r, dash, "1"] \arrow[d, dash, "2"] & (Y,c) \arrow[d, dash, "2"] \\
(Z,a) \arrow[r, dash, "1"] & (Z,b) \arrow[r, dash, "1"] & (Z,c)
\end{tikzcd}$$

Figure 2: Epistemic game model

Definition 4.2. A general epistemic game model M_G' is any submodel of a full epistemic game model.

As van Benthem (2006a) observes, every $S5_2$-model is bisimular to a general game model .Thus, the logic corresponding a general epistemic game model is just $S5_2$. Hence, the following research and analysis are based on $S5_2$. But in order to express some conceptions of game theory with our epistemic logic languages, such as Nash Equilibrium (NE), the best response(Br), we need to

[6] In order to keep the general proposal as simple as possible, we focus on two-players strategic games with pure strategies, but our research results can extend to multi-players strategic games with pure strategies.

expand our languages with game languages by considering them as atomic propositions, and we can describe the concepts of game with these atomic proposition in epistemic language. The best response Br_i for player i says that i's utility cannot improve by changing her action in w, keeping the others'actions fixed. Formally:
$M_G, w \vDash Br_i \Longleftrightarrow M_G, w \vDash \wedge_{a \in S_i, a \neq \sigma_i(w)}(u_i(\sigma_i(w), \sigma_{-i}(w)) \geq u_i(a, \sigma_{-i}(w))$
Nash Equilibrium is expressed by the conjunction: $NE = \wedge_{i \in N} Br_i$, that's to say, $M_G, w \vDash NE \Longleftrightarrow M_G, w \vDash \wedge_{i \in N} Br_i$

The figure 3 describes the distributions of the players' best responses for figure 1. It's easy to prove that $M_G \vDash (\neg K_1 Br_1 \wedge \neg K_2 Br_2)$ from this figure, i.e., neither players knows that he plays a the best response. However, rationality means players aim at maximizing their utility, in other words, a rational player i (Ra_i) always choose his the best response (Br_i). Thus, we can deduce : $M_G \nvDash C_N R_{a_i}$. But this is in contradiction with the analysis principle of game theory, which is rationality is common knowledge among players. So it is necessary to modify the definition of rationality, and we need to add epistemic ingredient to definition of rationality.[7]

$$\begin{tikzcd}
B_{r_2} \arrow[r, dash, "1"] \arrow[d, dash, "2"] & B_{r_1} \arrow[r, dash, "1"] \arrow[d, dash, "2"] & o \arrow[d, dash, "2"] \\
B_{r_1}, B_{r_2} \arrow[r, dash, "1"] \arrow[d, dash, "2"] & o \arrow[r, dash, "1"] \arrow[d, dash, "2"] & B_{r_1} \arrow[d, dash, "2"] \\
o \arrow[r, dash, "1"] & o \arrow[r, dash, "1"] & B_{r_2}
\end{tikzcd}$$

Figure 3: Distributions of players' the best responses

Definition 4.3. Player i is rational at a world w if there is not strategy s_i , so that player i knows s_i to be at least as good as $\sigma_i(w)$, and she considers it possible thats_i is better than $\sigma_i(w)$. formally:
$M_G, w \vDash Ra_i \Longleftrightarrow \neg\exists s_i \in S_i, M_G, w \vDash K_i(s_i \succcurlyeq_i \sigma_i(w)) \wedge \langle K_i \rangle (s_i \succ_i \sigma_i(w)) (i \in N, \langle K_i \rangle$ is dual for K_i)

According to this definition, $M_G, w \vDash R_{a_i}$ means $\neg\exists s_i \in S_i$ so that, for $\forall v \in \Omega$, $R_i wv$ satisfied $M_G, v \vDash (s_i \succcurlyeq_i \sigma_i(w))$ and $\exists v' \in \Omega, R_i wv'$ so that $M_G, v' \vDash (s_i \succ_i \sigma_i(w))$. Furthermore, we have: $\neg\exists s_i \in S_i$ satisfied for $\forall v \in \Omega, R_i wv$ so that $u_i(s_i, \sigma_{-i}(v)) \geq u_i(\sigma_i(w), \sigma_{-i}(v))$ and $\exists v' \in \Omega, R_i wv'$ satisfied $u_i(s_i, \sigma_{-i}(v')) > u_i(\sigma_i(w), \sigma_{-i}(v'))$. Because $R_i wv$ means $\sigma_i(w) = \sigma_i(v)$, $M_G, w \vDash R_{a_i}$ is equal to $\neg\exists s_i \in S_i$ so that, for $\forall v \in \Omega$, $R_i wv$ satisfied $u_i(s_i, \sigma_{-i}(w)) \geq u_i(\sigma_i(w), \sigma_{-i}(w))$ and $\exists v' \in \Omega, R_i wv'$ satisfied $u_i(s_i, \sigma_{-i}(w)) > u_i(\sigma_i(w), \sigma_{-i}(w))$. Therefore, Ra_i fails exactly at the rows or columns with which weakly dominated strategies correspond for player i in a general game model. For instance, in figure 2, R_{a_2} fails at the states (X, b), (Y, b) and (Z, b).

[7]This part is motivated by van Benthem (2006a) and Blackburn et al. (2007.).

5 Solving for NE Based on PAL

It's known that the assertion which players announce publicly must be the statements which they know are true in PAL. The following theorems guarantee that the rationality which we define can be as a suitable assertion for public announcement. After each public announcement of the rationality, we can remove all the worlds which are corresponding to the weakly dominated strategies for players.

Theorem 2. *Every finite general game model has worlds with R_a true.* ($R_a = \wedge_{i \in N} R_{a_i}$).

Proof. According to the fact that atomic proposition R_{a_i} fails exactly at the rows or columns with which weakly dominated strategies correspond for player i in a general game model. Consider any of general game model M'_G, If there is not a weakly dominated action for all player in the M'_G, then R_a is true at all the worlds in it. Thus, iterated announcement of R_a can no more change the game model, and get stuck in cycles in this situation. If there is a weakly dominated action for some player in the game, but because of the relativity of the definition of weakly dominated strategy, i.e. if player i has a weakly dominated strategy a, then he must has a strategy which is weakly better than strategy a, let strategy b. Thus, R_{a_i} holds at all the worlds which are belong to the row or the column corresponding strategy b. On the other hand, for player j, if he has not a weakly dominated action, then also R_{a_j} holds at all the worlds, furthermore, R_{a_j} holds at the worlds which are belong to the row or the column corresponding strategy b. So, R_a holds in the general game model. But if player j has a weakly dominated action, accordingly he must has a weakly dominant action, let action Y, and R_{a_j} is true at at the worlds which are belong to the row or the column corresponding strategy Y. Therefore, R_a is satisfied at the world (Y, b).

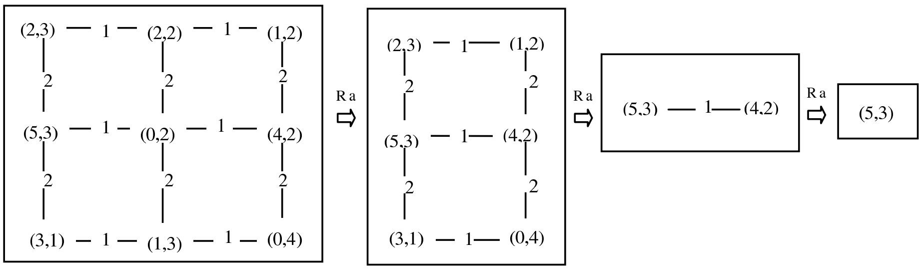
To sum up the above arguments, Every finite general game model has worlds with R_a true. ($R_a = \wedge_{i \in N} R_{a_i}$). □

Theorem 3. *Rationality is epistemically introspective. i.e. The formula $R_{a_i} \to K_i R_{a_i}$ is valid on a general game model.*

Proof. Given a general epistemic game model M'_G, an arbitrary w in M'_G, and $M'_G, w \vDash R_{a_i}$, but $M'_G, w \nvDash K_i R_{a_i}$ because $M'_G, w \nvDash K_i R_{a_i}$ which means $\exists v \in W, R_i wv$ so that $M'_G, v \nvDash R_{a_i}$ therefore $\exists s_i \in S_i$,satisfied $M'_G, v \vDash K_i(s_i \succcurlyeq_i \sigma_i(v)) \wedge \langle K_i \rangle (s_i \succ_i \sigma_i(v))$ i.e. for $\forall v', R_i vv'$ so that $M'_G, v' \vDash (s_i \succcurlyeq_i \sigma_i(v))$ and $\exists v'', R_i vv''$ satisfied $M'_G, v'' \vDash (s_i \succ_i \sigma_i(v))$. Then, $R_i wv'$ and $R_i wv''$, since R_i is equivalent relation. Thus, $M'_G, w \vDash K_i(s_i \succcurlyeq_i \sigma_i(w)) \wedge \langle K_i \rangle (s_i \succ_i \sigma_i(w))$. So, $M'_G, w \nvDash R_{a_i}$, by Definition 4.3, this is in contradiction with $M'_G, w \vDash R_{a_i}$. The formula $R_{a_i} \to K_i R_{a_i}$is valid on a general game model. □

Consequently, these theorems guarantee that we can remove the worlds at which $R_a (R_a = \wedge_{i \in N} R_{a_i})$ doesn't hold after each of players starts telling each other things they know about their behavior at some actual worlds at the same time.

In the figure 4, the left-most model is the model from figure 1. The other models are obtained by public announcements of R_a successively for three times. So, in the last submodel, we have:



$M_G,(Y,a) \models [R_a][R_a][R_a]C_B(NE)$. This formula indicates if the players iteratively announce that they are rational, the process of dominated strategy elimination leads them to a solution that is commonly known to be *NE*. And this procedure is closely to the *IWUS* procedures in some sense,we can say *IWUS* procedures can be characterized based on a Dynamic Epistemic logic.

Theorem 4. *Given a full epistemic game model M_G based a finite strategic-form G and an arbitrary world w, w is in a general epistemic game model M'_G which is stable by repeated announcement of R_a for all player if and only if $\sigma(w) \in WUD$. That's to say, $w \in \sharp(R_a, M_G) \Leftrightarrow \sigma(w) \in WUD$.*

Proof. (a) From left to right: if $w \in \sharp(R_a, M_G)$, i.e. $w \in M'_G$, then, $M'_G, w \models R_a$ i.e. $M'_G, w \models \wedge_{i\in N} R_{a_i}$. The proof is by induction.

First, we need to show that $\sigma_i(w) \notin WD_i^0$ (see Definition 2.5)for every player i. Suppose it doesn't hold, then there is a player j such that $\sigma_j(w) \in WD_j^0$. That is, strategy $\sigma_j(w)$ of player j is weakly dominated in G by some other strategy $s_j \in S_j$: for $\forall s_{-j} \in S_{-j}, u_j(s_j, s_{-j}) \geq u_j(\sigma_j(w), s_{-j})$, and $\exists s'_{-j} \in S_{-j}$, satisfied $u_j(s_j, s'_{-j}) > u_j(\sigma_j(w), s'_{-j})$. Then, for $\forall v$, satisfied $R_j wv$, we have $u_j(s_j, \sigma_{-j}(v)) \geq u_j(\sigma_j(w), \sigma_{-j}(v))$ and $\exists v', R_j wv'$, so that $u_j(s_j, \sigma_{-j}(v')) > u_j(\sigma_j(w), \sigma_{-j}(v'))$, so, for every $v, R_j wv$ satisfied $u_j(s_j, \sigma_{-j}(v)) \geq u_j(\sigma_j(v), \sigma_{-j}(v))$ and $\exists v', R_j wv'$ so that $u_j(s_j, \sigma_{-j}(v')) > u_j(\sigma_j(v'), \sigma_{-j}(v'))$, Because of $R_i wv \Leftrightarrow \sigma_i(w) = \sigma_i(v)$. It followed from Definition 4.3, $M'_G, w \models \neg R_{a_i}$. This contradicts the hypothesis that $M'_G, w \models R_a$. Thus, since, $\sigma(w)$[8]$\in S = WUD^0$, We have shown that, $\sigma(w) \in WUD^0 \setminus WD^0 = WUD^1$. Now, Fix an integer $m \geq 1$, and suppose $\sigma(w) \in WUD^m$. We want to show $\sigma(w) \notin WD^m$. Suppose, by contradiction, there exists a player i, $\sigma_i(w) \in WD_i^m$, this means $\sigma_i(w)$ is weakly dominated by some other $s'_i \in WUD_i^m$. By hypothesis, for every player $j, \sigma_j(w) \in WUD_j^m$, It follows, for every $s_{-j} \in WUD_{-j}^m, u_i(s'_i, s_{-j}) \geq u_i(\sigma_i(w), s_{-j})$ and $\exists s'_{-j} \in WUD_{-j}^m$, satisfied $u_i(s'_i, s'_{-j}) > u_i(\sigma_i(w), s'_{-j})$, Thus, by Definition 4.3, $M'_G, w \models \neg R_{a_i}$, which contradicts the initial hypothesis. Accordingly, for every player $i \in N, \sigma_i(w) \in \cap_{m\geqslant 0} WUD_i^m$. And therefore, $\sigma(w) \in WUD$.

(b) From right to left: Let $\sigma(w) \in WUD = \bigcap_{m\geqslant 0} WUD^m$, by Definition2.5, for $\forall i \in N, \neg\exists s_i \in WUD_i^{m-1}$, so that

$u_i(s_i, s_{-i}) \geq u_i(\sigma_i(w), s_{-i})$ *for every* $s_{-i} \in WUD_{-i}^{m-1}$;

$u_i(s_i, s_{-i}) > u_i(\sigma_i(w), s_{-i})$ *for some* $s_{-i} \in WUD_{-i}^{m-1}$

Then, for every v, $R_i wv, u_i(s_i, \sigma_{-i}(v)) \geq u_i(\sigma_i(w), \sigma_{-i}(v))$;

for some v', satisfied $R_i wv', u_i(s_i, \sigma_{-i}(v')) > u_i(\sigma_i(w), \sigma_{-i}(v'))$, by Definition 4.1, $R_i wv \Leftrightarrow \sigma_i(w) = \sigma_i(v)$, we have: $M'_G, w \models R_{a_i}$. Therefore, $w \in \sharp(R_a, M_G)$. □

[8] $\sigma(w) := (\sigma_1(w), \sigma_2(w), ..., \sigma_n(w))$is a strategies profile at the w.

6 Discussion

Many scholars has studied algorithms of iterated elimination, For example: Apt and Zvesper (2007), Christian (2002), Pearce (1984), van Benthem (2004) et al. In particular, Bonanno (2008) and van Benthem (2006a) are both to describe and characterize different algorithms in game theory by redefining rationality based on epistemic logic. The primary thought of this paper is derived from these papers. But there are many distinctions between them.

In van Benthem (2006a), J.van Benthem defined two types of rationality, weak rationality and strong rationality, which are denoted by *WR* and *SR*. Iterated public announcements of these rationality characterize respectively algorithms of iterated elimination strictly dominated strategy and of rationalizability, which is correspond to Bernheim's version of the rationalizability algorithm in Bernheim (1984). To compare the definition of our rationality (R_a) to these rationalities, we can conclude: as the worlds removed by announcing *WR* must be deleted by announcing R_a, so R_a is stronger than *WR* in some sense. But, there isn't the relation between R_a and *SR*. A *NE*, which can be solved by iterated announcement of *SR*, is not necessarily solved by iterated public announcement of R_a, and vice versa. For example, in the following games, G_1 can be only solved by repeating announcement of *SR*, but for G_2, we can only find the *NE* by announcement of R_a.

player / player 1	a	b	c
X	(1,2)	(1,0)	(1,1)
Y	(0,0)	(0,2)	(2,1)

G_1

player / player 1	a	b	c
X	(1,2)	(1,3)	(1,3)
Y	(0,2)	(0,1)	(2,0)

G_2

Figure 4: comparison between *SR* and R_a

In addition, J .van Benthem explained the two rationality based on fix-point logic, accordingly, more deeply analyzed the epistemic foundations on these algorithms. This is just our further work in the future.

In Bonanno (2008), the author also puts forward the two rationality, *WR′* and *SR′*, examined the implications of common belief and common knowledge of two, rather weak, notion of rationality, and showed that weaker axiom of rationality characterizes the iterated elimination strictly dominated strategy, while the stronger axiom characterizes the pure-strategy version of the algorithm introduced in Stalnaker (1997). As the same as the above reason, R_a is stronger than *WR′* in some sense. Comparing to *SR′*, although they are very closely in literal sense, because the epistemic game model which we defined is different from the game model defined by him, this leads to large difference in removing procedure and outcomes. The epistemic accessible relation in our epistemic game model is equivalent. It is built on the facts which are a definite game structure and this structure being common knowledge among players. Thereby, we provide epistemic game models to explain and analyze these iterated elimination algorithm from participants'(i.e. players themselves) views. However, the relation defined in Bonanno (2008) in the game logic is arbitrary, and what it models is major the structures of players' belief for a game. So, to a certain degree, he constructed epistemic game models from exterior modelers' view. Meanwhile, what we characterized for the algorithm of iterated admissibility is dynamic epistemic analysis, Bonanno focused on characterization for outcomes, which is a static description in some sense. But he provided characterization results in line with the notion of frame characterization in modal logic, this does not exist in our paper, and it is one of our future goals.

7 Conclusions

Iterated dominance is perhaps the most basic principle in game theory. The epistemic foundation of this principle is an assumption which is all players are rational. Many literatures have been developed to describe various of iterated dominance procedure in epistemic logic. Few people characterize the algorithm of iterated admissibility (weak dominance) from dynamic logic perspective, although iterated admissibility is a long-standing and attractive solution concept, make strong predictions in many games, e.g. the forward induction in signalling games and backward-induction in perfect-information. In this paper, we showed the first, characterize iterated admissibility procedure based on Public Announcement Logic, by redefining a rationality served as a proper assertion of announcement, thereby provided the dynamic epistemic foundations for iterated admissibility.

We have restricted attention to strategic-form games. In the future work, we intend to analyze the operator of rationality announcement from fix-point logic, and extend these analysis to extensive-form games with perfect information and the notion of backward induction.

Acknowledgements I am very grateful to Johan van Benthem.

References

K. R. Apt. The many faces of rationalizability. *The Journal of Theoretical Economics*, 7, 2007.

K. R. Apt and J. A. Zvesper. Common beliefs and public announcements in strategic games with arbitrary strategy sets, 2007. Available at `http://www.citebase.org/abstract?id=oai:arXiv.org:0710.3536`.

R. J. Aumann. Rationality and bounded rationality. *Games and Economic Behavior*, 53:293–291, 1999.

A. Baltag, L. S. Moss, and S. Solecki. The logic of public announcements, common knowledge and private suspicious. Technical Report SEN-R9922, CWI, Amsterdam, 1999.

B. D. Bernheim. Rationalizable strategic behavior. *Econometrica*, 52(4):1007–1028, 1984.

P. Blackburn, J. van Benthem, and F. Wolter. *Handbook of Modal Logic*. Elsevier Science Inc, 2007.

G. Bonanno. Epistemic foundations of game theory. Technical report, ILLC, University of Amsterdam, 2007.

G. Bonanno. A syntactic approach to rationality in games with ordinal payoffs. In G. Bonanno, W. van der Hoek, and M. Wooldridge, editors, *Logic and the Foundations of Game and Decision Theory (LOFT7)*, volume 3 of *Texts in Logic and Games*, pages 59–86. Amsterdam University Press, 2008.

A. Brandenburger, A. Friedenberg, and H. J. Kei. Admissibility in games. *Econometrica*, 76(2):307–352, 2008.

E. Christian. Iterated weak dominance in strictly competitive games of perfect information. *Games and Economic Behavior*, 2002.

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

J. Gerbrandy. Dynamic epistemic logic. In L. S. Moss, J. Ginzburg, and M. de Rijke, editors, *Logic, Language and Computation*, volume 2, pages 67–84. CSLI Publications, Stanford, CA, USA, 1999. ISBN 1-57586-181-X.

M. Gilli. Iterated admissibility as solution concept in game theory. Technical report, University of Milano-Bicocca, Department of Economics, 2002.

T.-H. Ho, C. Camerer, and K. Weigelt. Iterated dominance and iterated best response in experimentapl ”p-beauty contests”. *American Economic Review*, 88, 1998.

D. G. Pearce. Rationalizable strategic behavior and the problem of perfection. *Econometrica*, 52(4):1029–1050, 1984.

A. Rubinstein. *A Course in Game Theory*. The MIT Press, Cambridge, Mass., 1994.

A. Rubinstein. *Modeling Bounded Rationality*. The MIT Press, Cambridge, Mass., 1998.

R. Stalnaker. Extensive and strategic form: Games and models for games. *Research in Economics*, 21:2–14, 1997.

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

J. van Benthem. Rational dynamics and epistemic logic in games. *Games and Economic Behavior*, 2006a.

J. van Benthem. *Exploring Logical Dynamics*. ESLI Publications,Stanford, Cambridge, Mass., 1996.

J. van Benthem. Logic in games. Technical report, Amsterdam and Philosophy Stanford, 2004.

J. van Benthem. Epistemic logic and epistemology: The state of their affairs. *Philosophical Studies*, 128:49–76, March 2006b. doi: 10.1007/s11098-005-4052-0.

J. van Benthem, J. van Eijck, and B. P. Kooi. Common knowledge in update logics. In van der Meyden (2005), pages 253–261. ISBN 981-05-3412-4.

R. van der Meyden, editor. *Proceedings of the 10th Conference on Theoretical Aspects of Rationality and Knowledge (TARK-2005), Singapore, June 10-12, 2005*, 2005. National University of Singapore. ISBN 981-05-3412-4.

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

Keep “Hoping” for Rationality: A solution to the backward induction paradox

Alexandru Baltag, Sonja Smets and Jonathan A. Zvesper
Oxford University Computing Laboratoty, University of Oxford;
Department of Artificial Intelligence, University of Groningen;
Institute for Logic, Language and Computation, Universiteit van Amsterdam.
baltag@comlab.ox.ac.uk, S.J.L.Smets@rug.nl, jonathan@illc.uva.nl

Abstract

We formalise a notion of dynamic rationality in terms of a logic of conditional belief on (doxastic) plausibility models. Similarly to other epistemic statements (e.g. negations of Moore sentences and of Muddy Children announcements), dynamic rationality changes its meaning after every act of learning, and may become true after players learn it is false. Applying this to extensive games, we “simulate” the play of a game as a succession of dynamic updates of the original plausibility model: the epistemic situation when a given node is reached can be thought of as the result of a joint act of learning (via public announcements) that the node is reached. We then use the notion of “stable belief”, i.e. belief that is preserved during the play of the game, in order to give an epistemic condition for backward induction: *rationality and common knowledge of stable belief in rationality*. This condition is weaker than Aumann’s and compatible with the implicit assumptions (the ‘epistemic openness of the future’) underlying Stalnaker’s criticism of Aumann’s proof. The ‘dynamic’ nature of our concept of rationality explains why our condition avoids the apparent circular- ity of the ‘backward induction paradox’: it is consistent to (continue to) believe in a player’s rationality after updating with his irrationality.

> *Aumann has proved that common knowledge of substantive rationality implies the backward induction solution in games of perfect information. Stalnaker has proved that it does not.* Halpern (2001)

The jury is still out concerning the epistemic conditions for backward induction, the “oldest idea in game theory” (Aumann 1995, p. 635). Aumann (1995) and Stalnaker (1996) take conflicting positions in the debate: the former claims that common “knowledge” of “rationality” in a game of perfect information entails the backward-induction solution; the latter that it does not.[1] Of course there is

[1] Others agree with Stalnaker in disagreeing with Aumann: for example, Samet (1996) and Reny (1992) also put forwards arguments against Aumann’s epistemic characterisation of subgame-perfect equilibrium. Section 7 is devoted to a discussion of related literature.

nothing wrong with any of their relevant formal proofs, but rather, as pointed out by Halpern (2001), there are differences between their interpretations of the notions of knowledge, belief, strategy and rationality. Moreover, as pointed out by Binmore (1987, 1996), Bonanno (1991), Bicchieri (1989), Reny (1992), Brandenburger (2007) and others, the reasoning underlying the backward induction method seems to give rise to a fundamental paradox: in order even to start the reasoning, a player assumes that (common knowledge of, or some form of common belief in) "rationality" holds at all the last decision nodes (and so the obviously irrational leaves are eliminated); but then, in the next reasoning step (going backward along the tree), some of these (last) decision nodes are eliminated, as being incompatible with (common belief in) "rationality"! Hence, the assumption behind the previous reasoning step is now undermined: the reasoning player can now see, that *if* those decision nodes that are now declared "irrational" were ever to be reached, then the only way that this could happen is if (common belief in) "rationality" failed. Hence, she was *wrong* to assume (common belief in) "rationality" when she was reasoning about the choices made at those last decision nodes. This whole line of arguing seems to undermine itself!

Belief Dynamics In this paper we use as a foundation the relatively standard and well-understood setting of Conditional Doxastic Logic (CDL, Board (2002), Baltag and Smets (2006; 2008a;b)), and its "dynamic" version (obtained by adding to CDL operators for *truthful public announcements* $[!\varphi]\psi$): the logic PAL-CDL, introduced by van Benthem (2007a). In fact, we consider a slight extension of this last setting, namely the logic APAL-CDL, obtained by further adding dynamic operators for *arbitrary* (truthful) public announcements $[!]\psi$, as in Balbiani et al. (2008). We use this formalism to capture a novel notion of "dynamic rationality" and to investigate its role in decision problems and games. As usual in these discussions, we take a *deterministic stance*, assuming that the initial state of the world at the beginning of the game already fully determines the future play, and thus the unique outcome, irrespective of the players' (lack of) knowledge of future moves. We do *not*, however, require that the state of the world determines what *would* happen, if that state were *not* the actual state. That is, *we do not need to postulate the existence of any "objective counterfactuals"*. But instead, we only need "subjective counterfactuals": in the initial state, not only the future of the play is specified, but also the players' *beliefs* about each other, as well as their *conditional beliefs*, pre-encoding their *possible revisions of belief*. The players' conditional beliefs express what one may call their "propensities", or "dispositions", to revise their beliefs in particular ways, if given some particular pieces of new information.

Thus at the outset of a game, all is "done", including the future. But all is not necessarily "said". In a deterministic model, as time progresses the only thing that changes are the pictures of the world in the minds of the players: the information states of the players. This is *"on-line" learning*: *while* the game is being played, the players *learn* the played moves, and so they may change their minds about the situation. We can *simulate* this on-line learning (and its effect on the players' beliefs) via *off-line "public announcements"*: if, *before* the start of the game, the agents were *publicly told* that the game will reach some node u, then they would be in the *same epistemic state* as they would have been by (not having any such public announcement but instead) playing the game until node u was reached.

So in this paper we stress the importance of the *dynamics* of beliefs and rationality during a play of an extensive game, and we use dynamic operators in order to simulate the play of the game. Since we focus on games of *perfect information*, we only need *public* announcements to simulate the moves of the game. The idea of adding modalities for public announcements to epistemic logic was introduced and developed in Plaza (1989), Gerbrandy and Groeneveld (1997). Dynamic epistemic logic Baltag et al. (1999) provides for much richer dynamic modalities than just public announcements, capturing the effects of more complex and more "private" forms of learning. We think these could be applied to the case of games with *imperfect information*. However, for simplicity, we leave these developments for future work and consider for now only perfect information, and so only public announcements.

Games Using the terminology of Brandenburger (2007), ours is a *belief-based approach* to game theory (in the same category as the work of Battigalli and Siniscalchi (1999, 2002)), in contrast to the *knowledge-based approach* of Aumann (1995) and others. According to the belief-based approach, "only observables are knowable. Unobservables are subject to belief, not knowledge. In particular, other players' strategies are unobservables, and only moves are observables." (Brandenburger 2007, p. 489) This means that we take the players' *beliefs* (including conditional beliefs) as basic, instead of their *knowledge*. However, there is a notion of knowledge that naturally arises in this context: the "irrevocable knowledge", consisting of the beliefs that are *absolutely unrevisable*, i.e. believed under any conditions. This notion of knowledge is meant to apply only to the players' "hard information", obtained by observation or by undoubtable evidence. This is a much stronger condition than "certain belief" (subjective probability 1) or even "true belief", and as a result it may happen that very few things are "known" in this sense. One of the things we assume to be irrevocably known is the *structure of the game*: the possible outcomes, the players' preferences etc; also, in a game of perfect information, the *played moves* are observed, and thus known, after they are played; finally, another thing irrevocably known to a player is *her own beliefs*: by introspection, she knows what she believes and what not. Besides this, we do not assume much else to be known, although our general setting is in principle consistent with (common) knowledge of all the players' beliefs, their strategies, their rationality etc.

One thing we do *not* assume as known is the *future of the game*: no outcomes that are consistent with the structure of the game are to be excluded at the outset of the game. In fact, we make the opposite assumption: that it is common knowledge that nobody knows the future, i.e. nobody knows that some outcome will not be reached. This "open future" assumption seems to contradict common knowledge of rationality; but in fact, it is consistent with it, if by rationality we only mean "rational planning", leaving open the possibility that players may make mistakes or may change their minds. The players may certainly *believe* their rational plans will be faithfully carried out, but they have no way to "know" this in advance. We think of our "open future" assumption as being a realistic one, and moreover one that embodies the agents' "freedom of choice", as well as the "possibility of error", that underly a correct notion of rationality. An agent's rationality can be assessed only if she is given some options to freely choose from. There are certainly cases in which the future can be known, e.g. when it is determined by a known natural law. But it is an essential feature of rational agents that their own choices are not known to them to be thus determined; or else, they would have no real choices, and thus no rational choice. Any natural determinism is assumed to be absorbed in the definition of the game structure, which does pose absolute limits to choices. In a sense, this simply makes precise the meaning of our "knowledge" as "hard information", and makes a strict delimitation between the past and the future choices, delimitation necessary to avoid the various paradoxes and vicious circles that plague the notions of rational decision and freedom of choice: the agents may have "hard information" about the past and the present, but not about their own future free choices (although they may have "soft" information, i.e. "certain" beliefs, with probability 1, about their future choices).

Dynamic Rationality Maybe the most important original feature of our paper is our notion of *"dynamic" rationality*, which takes into account the dynamics of beliefs, as well as the dynamics of knowledge. On the one hand, following Stalnaker, Reny, Battigalli and Siniscalchi etc. (and in contrast with Aumann), we assess the rationality of a player's move at a node against the beliefs held *at the moment when the node is reached*. On the other hand, we incorporate the above-mentioned epistemic limitation to rationality: the rationality of an agent's move only makes sense when that move is *not already known* (in an irrevocable manner) to her. *Players cannot be held responsible for moves that they cannot choose or change any more (including their own past moves).* Since the players' knowledge increases during a game of perfect information, their set of available options decreases: passed options/nodes, or nodes that were by-passed, cannot be the objects of choice any more. As a result, our notion of rationality is *future-oriented*: at any stage of the game, *whether or not an agent is dynamically rational at that stage depends only on her current and future moves*. So a player can be rational *now* even if in the past she has made some "irrational" moves. In effect, performing such an irrational move in a game of perfect information is equivalent to a public announcement that "the player is (currently) *not* rational" (at the moment of moving)[2]. All the players jointly learn this "fact" (as a piece of 'hard' information), but the "fact" itself may vanish after being learnt: while previously "irrational" (since about to make a 'wrong' move), the player may "become rational" after the wrong move (simply because, for all the decisions that she can still make after that, she chooses the 'right' moves). So the *truth-value* of the sentence "player i is (dynamically) rational" *may change* after a move by player i. The way this is captured and explained in our formal setting is original and interesting in itself: *the meaning of our "rationality" changes in time, due to the change of beliefs and of the known set of options*. This is because the *"rationality of an agent" is an epistemic-doxastic concept*, so it is obviously affected by any changes in the information possessed by that agent (including the changes induced by the agent's own moves). In our Dynamic-Epistemic Logic setting, this is a *natural and perfectly standard feature*, an immediate consequence of the epistemic definition of rationality: in general,

[2]Technically, this claim is correct only for binary games, in which at any node there are only two possible moves; but a weak version of this claim holds in general.

epistemic sentences do not necessarily preserve their truth value after they are "learnt". Epistemic logicians are already familiar with this phenomenon, e.g. the examples of Moore sentences Moore (1942) and of the repeated public announcements of "ignorance" in the Muddy Children Scenario Fagin et al. (1995).

Our concept of dynamic rationality, developed on purely a priori grounds, solves in one move the "BI-paradox": the first reasoning step in the backward-induction argument (dealing with the last decision nodes of the game) is *not undermined* by the result of the second reasoning step, since the notion of "rationality" assumed in the first step is *not* the same as the "rationality" disproved in the second step! The second step only shows that some counterfactual nodes cannot be reached by rational play, and thus it implies that some agent must have been irrational (or must have had some doubts about the others' rationality, or must have made some "mistake") *before* such an "irrational" node was reached; but this doesn't contradict in any way the assumption that the agents *will* be rational *at* that node (and further in the future).

Stability Dynamics cannot really be understood without its correlative: *invariance* under change. Certain truths, or beliefs, *stay true* when everything else changes. We have already encountered an "absolute" form of invariance: "irrevocable knowledge", i.e. belief that is invariant under *any* possible information change. Now, we need a second, weaker form of invariance: "stability". A truth, or a belief, is *stable* if it remains true, or continues to be believed, after any *(joint) learning of "hard" information* (via some truthful public announcement). In fact, in the case of an "ontic" (non-doxastic) fact p, Stalnaker's favourite notion of "knowledge" of p Stalnaker (1996; 2006) (a modal formalisation of Lehrer and Klein's "defeasibility theory of knowledge"), corresponds precisely to the 'factive' (i.e. truth-entailing) version of *stable belief* in p, also called "safe belief" in Baltag and Smets (2008b). (But note that the two notions differ when applied to a doxastic-epistemic property, such as "rationality"!) Stability can be or not a property of a belief or a common belief: a proposition P is a "stable (common) belief" if the fact that P is (common) belief is a stable truth, i.e. P continues to be (common) belief after any (joint) learning of "hard" information.

What is required for achieving the backward induction outcome is *stable belief in dynamic rationality*, as a *default* condition (i.e. commonly known to hold for all agents). In some contexts, we can think of this condition as expressing an *"optimistic" belief-revision policy about the opponents' potential for rationality*: the players "keep hoping for rationality" with respect to everybody's current and future play, despite any past irrational moves. Of course, whether or not the words "hope" and "optimism" are appropriate depends on the players' payoffs: e.g. in common interest games (in which all players' payoffs are identical at all nodes), it indeed makes sense to talk about "hoping" for opponents' rationality; while in other games, it may be more appropriate to talk about "persistent cautiousness" and a "pessimistic" revision policy.

We can now give an informal statement of the main theorem of this paper:

> *Common knowledge of* (the game structure, of "open future" and of) *stable (common*[3]*) belief in dynamic rationality entails common belief in the backward induction outcome.*

[3] Adding the word "common" to this condition doesn't make a difference: common knowledge

Overview of the Paper To formalise stability and "stable common belief", we introduce in the next section Conditional Doxastic Logic CDL and its dynamic version APAL-CDL. Section 2 recalls the definition of extensive games and shows how to build models of those games in which the structure of the game is common knowledge, in our strong sense of "knowledge". In Section 3 we define "rationality" and "rational play", starting from more general decision-theoretic considerations, and arriving at a definition of *dynamic rationality* in extensive (aka "dynamic") games, which is in some sense a special case of the more general notion. Section 4 gives a formal statement of our main results, to whose proofs Section 5 is devoted. In Section 6 we consider a weaker condition that ensures the backward induction outcome, and is based on what we call *stable true belief*. Finally, Section 7 discusses connections between our work and some of the existing literature on the epistemic foundations of backward induction.

1 Conditional Doxastic Logic and its Dynamic Extensions

CDL models, also called "plausibility models" are essentially the "belief revision structures" in Board (2002), simplified by incorporating structurally the assumption of Full Introspection of Beliefs (which allows us to use binary plausibility relations on worlds for each agent, instead of ternary relations). But since we will also want to talk about the *actual change* under the effects of actions, like moves in a game, rather than just the *static* notion that is in effect captured by Board's models, we will enrich the language of CDL with model-changing *dynamic* operators for "public announcements", in the spirit of Dynamic Epistemic Logic (cf. Baltag et al. (1999), Benthem (2007a;b)).

The models are "possible worlds" models, where the worlds will usually be called *states*. Grove (1988) showed that the AGM postulates Alchourrón et al. (1985) for rational belief change are equivalent to the existence of a suitable pre-order over the state space.[4] The intended interpretation of the pre-order $\leq_i$ of some agent i is the following: $s \leq_i t$ means that, in the event $\{s, t\}$, i considers *s at least as plausible as t*.

In interactive situations, where there are several players, each player i has a doxastic pre-order $\leq_i$. In addition to having different *beliefs*, any two players might have different *knowledge*. We follow the mainstream in game theory since Aumann and model interactive knowledge using a partitional structure. However, as in Board (2002), we will derive i's partition from i's pre-order $\leq_i$. Let us be more precise: fix a set S and a relation $\leq_i \subseteq S \times S$; then we define the *comparability class* of $s \in S$ for $\leq_i$ to be the set $[s]_i = \{t \in S \mid s \leq_i t \text{ or } t \leq_i s\}$ of states $\leq_i$-comparable to s. Now we want the set of comparability classes to form a partition of S, so we will define a *plausibility frame* to be a sequence $(S, \leq_i)_{i \in N}$ in which S is a non-empty set of states, and each $\leq_i$ a pre-order on S such that for each $s \in S$, the restriction of $\leq_i$ to $[s]_i$ is a "complete" (i.e. "total" or "connected") pre-order.

that everybody has a stable belief in P is the same as common knowledge of common safe belief in P.

[4] A pre-order is any reflexive transitive relation. In Grove's representation theorem the pre-order must also be total and converse-well-founded.

Proposition 1. *In any plausibility frame, $\{[s]_i \mid s \in S\}$ forms a partition of S. We will interpret this as the* information partition *for player i (in the sense of "hard" information, to be explained below).*

So we can define player i's *knowledge* operator in the standard way, putting for any "proposition" $P \subseteq S$:

$$K_i P := \{s \in S \mid [s]_i \subseteq P\}$$

As explained below, this captures a notion of indefeasible, absolutely unrevisable knowledge. But we also want a notion of *belief* B, describing "soft" information, which might be subject to *revision*. So we want *conditional belief operators* B^P, in order to capture the *revised beliefs* given some new information P. If S is *finite*, let $min_{\leq_i}(P)$ denote the $\leq_i$-minimal P elements $\{s \in P \mid \forall t \in P, s \leq_i t\}$. So $min_{\leq_i}(P)$ denotes the set of states which i considers most plausible given P. Then $min_{\leq_i}(P \cap [s]_i)$ denotes the set of that states which i considers most plausible given both P and i's knowledge at state s. Thus we define player i's *conditional belief* operator as:

$$B_i^Q P := \{s \in S \mid min_{\leq_i}(Q \cap [s]_i) \subseteq P\}.$$

There is a standard way to extend this definition to total pre-orders on *infinite* sets of states, but we skip here the details, since we are mainly concerned with finite models. $B_i^Q P$ is the event that *agent i believes P conditional on Q*. Conditional belief should be read carefully: $B_i^Q P$ does *not* mean that after learning that Q, i will believe P; rather it means that after learning Q, i will believe that *P was the case before the learning*. This is a subtle but important point: the conditional belief operators do not directly capture the dynamics of belief, but rather as van Benthem (2007a) puts it, they 'pre-encode' it. We refer to Benthem (2007a), Baltag and Smets (2008b) for more discussion. The usual notion of *(non-conditional) belief* can be defined as a special case of this, by putting $B_i P := B_i^S P$. The notions of *common knowledge CkP* and *common belief CbP* are defined in the usual way: first, one introduces *general knowledge EkP* $:= \bigcap_i K_i P$ and *general belief EbP* $:= \bigcap_i B_i P$, then one can define $CkP := \bigcap_n (Ek)^n P$ and $CbP := \bigcap_n (Eb)^n P$.

It will be useful to associate with the states S some non-epistemic content; for this we use a *valuation function*. Assume given some finite set Φ of symbols, called *basic (or atomic) sentences*, and meant to describe ontic (non-epistemic, non-doxastic) "facts" about the (current state of the) world. A *valuation on Φ* is a function V that associates with each $p \in \Phi$ a set $V(p) \subseteq S$: V specifies at which states p is true. A *plausibility model* for (a given set of atomic sentences) Φ is a plausibility frame equipped with a valuation on Φ.

Interpretation: 'Hard' and 'Soft' Information Information can come in different flavours. An essential distinction, due to van Benthem (2007a), is between 'hard' and 'soft' information. *Hard information* is absolutely "indefeasible", i.e. *unrevisable*. Once acquired, a piece of 'hard' information forms the basis of the strongest possible kind of knowledge, one which might be called *irrevocable knowledge* and is denoted by K_i. For instance, the principle of Introspection of Beliefs states that (introspective) agents possess 'hard' information about their own beliefs: they know, in an absolute, irrevocable sense, what they believe and what not. *Soft information*, on the other hand, may in principle be defeated (even if it happens to be correct). An agent usually possesses only soft information about other agents' beliefs or states of mind: she may have beliefs about the others' states of mind, she may even be said to have a kind of 'knowledge' of them, but this 'knowledge' is defeasible: in principle, it could be revised, for instance if the agent were given more information, or if she receives misinformation.

For a more relevant, game-theoretic example, consider extensive games of perfect information: in this context, it is typically assumed (although usually only in an implicit manner) that, at any given moment, both the *structure of the game* and the players' *past moves* are 'hard' information; e.g. once a move is played, all players *know, in an absolute, irrevocable sense*, that it was played. Moreover, past moves (as well as the structure of the game) are common knowledge (in the same absolute sense of knowledge). In contrast, a player's 'knowledge' of other players' rationality, and even a player's 'knowledge' of her own future move at some node that is not yet reached, are not of the same degree of certainty: in principle, they might have to be revised; for instance, the player might make a mistake, and fail to play according to her plan; or the others might in fact play "irrationally", forcing her to revise her 'knowledge' of their rationality. So this kind of defeasible knowledge should better be called 'belief', and is based on players' "soft" information.[5]

In the 'static' setting of plausibility models given above, soft information is captured by the "belief" operator B_i. As already mentioned, this is defeasible, i.e. *revisable*, the revised beliefs after receiving some new information φ being pre-encoded in the conditional operator B_i^{φ}. Hard information is captured by the "knowledge" operator K_i; indeed, this is an absolutely unrevisable form of belief, one which can never be defeated, and whose negation can never be accepted as truthful information. This is witnessed by the following valid identities:

$$K_i P = \bigcap_{Q \subseteq S} B_i^Q P = B_i^{\neg P} \emptyset.$$

Special Case: Conditional Probabilistic Systems If, for each player i, we are given a *conditional probabilistic system* à la Renyi (1955) over a common set of states S (or if alternatively we are given a *lexicographic probability system* in the sense of Blume et al. (1991)), we can define subjective conditional probabilities $\mathrm{Prob}_i(P|Q)$ even for events of zero probability. When *S is finite and the system is discrete* (i.e., $\mathrm{Prob}(P|Q)$ is defined for all non-empty events Q), we can use this to define conditional belief operators for arbitrary events, by putting $B_i^Q P := \{s \in S : \mathrm{Prob}_i(P|Q) = 1\}$. It is easy to see that these are special cases of finite plausibility frames, by putting: $s \leq_i t$ iff $\mathrm{Prob}_i(\{s\}|\{s,t\}) \neq 0$. Moreover, the notion of conditional belief defined in terms of the plausibility relation is *the same* as the one defined probabilistically as above.

Dynamics and Information: 'Hard' Public Announcements Dynamic epistemic logic is concerned with the "origins" of hard and soft information: the

[5] By looking at the above probabilistic interpretation, one can see that the fact that an event or proposition has (subjective) probability 1 corresponds only to the agent having "soft" information (i.e. *believing* the event). "Hard" information corresponds to the proposition being true in *all* the states in the agent's information cell.

"epistemic actions" that can appropriately inform an agent. In this paper, we will focus on the simplest case of hard-information-producing actions: *public announcements*. These actions model the *simultaneous joint learning of some 'hard' piece of information* by a group of agents; this type of learning event is perfectly "transparent" to everybody: there is nothing hidden, private or doubtful about it. But dynamic epistemic logic Baltag et al. (1999) also deals with other, more complex, less transparent and more private, forms of learning and communication.

Given a plausibility model $\mathcal{M} = (S, \leq_i, V)_{i\in N}$ and a "proposition" $P \subseteq S$, the *updated model* $\mathcal{M} \upharpoonright P$ produced by a public announcement of P is given by *relativisation*: $(P, \leq_i \upharpoonright P, V \upharpoonright P)$, where $\leq \upharpoonright P$ is the restriction of $\leq$ to P and $(V \upharpoonright P)(p) = V(p) \cap P$. Notice that public announcements can change the knowledge and the beliefs of the players. So far we have, for readability, been writing events without explicitly writing the frame or model in question. However, since we are now talking about model-changing operations it is useful to be more precise; for this we will adopt a modal logical notation.

APAL-CDL: Language and Semantics Our language APAL-CDL is built recursively, in the usual manner, from atomic sentences in Φ, using the Boolean connectives $\neg\varphi$, $\varphi \wedge \psi$, $\varphi \vee \psi$ and $\varphi \Rightarrow \psi$, the epistemic operators $K_i\varphi$, $B_i^{\varphi}\psi$, $Ck\varphi$ and $Cb\varphi$ and the dynamic modalities $[!\varphi]\psi$ and $[!]\varphi$. (The language *CDL* of conditional doxastic logic consists only of the formulas of APAL-CDL that can be formed *without* using the dynamic modalities.)

For any formula φ of this language, we write $[\![\varphi]\!]_{\mathcal{M}}$ for the *interpretation* of φ, the event denoted by φ, in $\mathcal{M}$. We write $\mathcal{M}^{\varphi}$ for the updated model $\mathcal{M} \upharpoonright [\![\varphi]\!]_{\mathcal{M}}$ after the public announcement of φ. The interpretation map is defined recursively: $[\![p]\!]_{\mathcal{M}} = V(p)$; Boolean operators behave as expected; and the definitions given above of the epistemic operators in terms of events give the interpretation of epistemic formulae. Then the interpretation of the dynamic formulae, which include *public announcement modalities* $[!\varphi]\psi$, goes as follows:

$$[\![[!\varphi]\psi]\!]_{\mathcal{M}} = \{s \in S \mid s \in [\![\varphi]\!]_{\mathcal{M}} \Rightarrow s \in [\![\psi]\!]_{\mathcal{M}^{\varphi}}\}$$

Thus $[!\varphi]\psi$ means that after any true public announcement of φ, ψ holds. The *arbitrary (public) announcement modality* $[!]\varphi$ is to be read: after *every* (public) announcement, φ holds. Intuitively, this means φ is a *"stable" truth*: not only it is true, but it continues to stay true when any new (true) information is (jointly) learned (by all the players). There are some subtleties here: do we require that the new information/announcement be expressible in the language for example? This is the option taken in Balbiani et al. (2008), where the possible announcements are restricted to *epistemic* formulas, and a *complete axiomatisation* is given for this logic. In the context of *finite models* (as the ones considered here), this definition is actually equivalent to allowing *all formulas of our language* APAL-CDL as announcements. As a result, we can *safely use* the following *apparently circular definition*:

$$[\![[!]\varphi]\!]_{\mathcal{M}} = \{s \in S \mid \forall\psi\ s \in [\![[!\psi]\varphi]\!]_{\mathcal{M}}\}$$

Dynamic epistemic logic captures the "true" dynamics of (higher-level) beliefs after some learning event: in the case of public announcements, the beliefs of an agent i after a joint simultaneous learning of a sentence φ are fully expressed by the operator $[!\varphi]B_i$, obtained by composing the dynamic and doxastic operators. Note that this is *not* the same as the conditional operator B_i^{φ}, but the two are related via the following "Reduction Law", introduced in Benthem (2007a):

$$[!\varphi]B_i\psi \;\Leftrightarrow\; (\varphi \Rightarrow B_i^{\varphi}[!\varphi]\psi).$$

This is the precise sense in which the conditional belief operators are said to "pre-encode" the dynamics of belief.

Special Case: Bayesian Conditioning In the case of a conditional probability structure, the update $\mathcal{M} \upharpoonright P$ by a public announcement $!P$ corresponds to *Bayesian update (conditionalisation)*: the state space is reduced to the event P, and the updated probabilities are given by $\mathrm{Prob}_i'(Q) := \mathrm{Prob}_i(Q|P)$. So a dynamic modality $[!P]Q$ corresponds to the event that, after conditionalising with P, event Q holds. Similarly, the arbitrary announcement modality $[!]P$ is the event that P *stably holds*, i.e. it holds after conditionalising with any true event.

2 Models and Languages for Games

The notion of *extensive game with perfect information* is defined as usual (cf. Osborne and Rubinstein (1994)): Let N be a set of 'players', and G be a finite tree of 'decision nodes', with terminal nodes (leaves) $\mathcal{O}$ (denoting "possible outcomes"), such that at each non-terminal node $v \in G - \mathcal{O}$, some player $i \in N$ is the decision-maker at v. We write $G_i \subseteq G$ for the set of nodes at which i is the decision-maker. Add to this a payoff function h_i for each player i, mapping all the leaves $o \in \mathcal{O}$ into real numbers, and you have an extensive game. We write 'G' to refer both to the game and to the corresponding set of nodes. We also write $u \to v$ to mean that v is an immediate successor of u, and $u \rightsquigarrow v$ to mean that there is a path from u to v. A *subgame* of a game G is any game G', having a subset $G' \subseteq G$ as the set of nodes and having the immediate successor relation $\to'$, the set of decision nodes G_i' and the payoff function h_i' (for each player i) being given by restrictions to G' of the corresponding components of the game G (e.g. $G_i' = G_i \cap G'$ etc). For $v \in G$, we write G^v for the subgame of G in which v is the root. A *strategy* σ_i for player i in the game G is defined as usually as a function from G_i to G such that $v \to \sigma_i(v)$ holds for all $v \in G_i$. Similarly, the notions of *strategy profile*, of *the (unique) outcome determined by a strategy profile* and of *subgame-perfect equilibrium* are defined in the standard way (see e.g. Osborne and Rubinstein (1994)). Finally, we define as usually a *backward induction outcome* to be any outcome $o \in \mathcal{O}$ determined by some subgame-perfect equilibrium. We denote by BI_G the set of all backward-induction outcomes of the game G.

Consider as an example the "centipede" game G (cf. Rosenthal (1981)) given in Figure 1. This is a two-player game for a (Alice) and b (Bob).

Here, we represent the nodes of the game by dots and the possible moves by arrows. For each non-terminal node, the corresponding dot is labelled with the name of the node and the name of the player who decides the move at that node; while the dots corresponding to the terminal nodes (outcomes) are labelled with the name of the node (o_1, o_2, o_3, o_4) and with the players' payoffs,

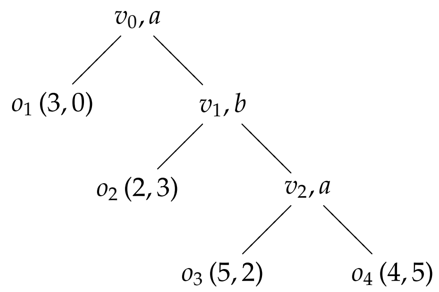


Figure 1: The "centipede" game G

written as pairs (p_a, p_b), where p_a is Alice's payoff and p_b is Bob's. Note that in this game there is one backward induction outcome, o_1, and furthermore that the unique backward induction strategy profile assigns to each v_m the successor o_{m+1}.

Language for Games For any given game G, we define a set of *basic (atomic) sentences* Φ_G from which to build a language. First, we require Φ_G to contain a sentence for each leaf: for every $o \in \mathcal{O}$, there is a basic sentence $\overline{o}$. For simplicity, we often just write o, instead of $\overline{o}$. In addition Φ_G contains sentences to express the *players' preferences over leaves*: for each $i \in N$ and $\{o, o'\} \subseteq \mathcal{O}$, Φ_G has a basic sentence $o \prec_i o'$. Our formal language for games G is simply the language APAL-CDL defined above, where the set of atomic sentences is the set Φ_G. To talk about the non-terminal nodes, we introduce the following abbreviation:

$$\overline{v} := \bigvee_{v \rightsquigarrow o} o\,,$$

for any $v \in G - \mathcal{O}$. As for terminal nodes, we will often denote this sentence by v for simplicity, instead of $\overline{v}$.

Plausibility Models for Games We now turn to defining models for games. A *plausibility model for game G* is just a plausibility model $(S, \leq_i, V)_{i\in N}$ for the set Φ_G. We interpret every state $s \in S$ as an *initial state in a possible play* of the game. Intuitively, the sentence $\overline{o}$ is true at a state s if *outcome o will be reached during the play* that starts at s; and the sentence $o \prec_i o'$ says that *player i's payoff at o is strictly smaller than her payoff at o'*.

Observe that nothing in our definition of models for G guarantees that states come with a unique outcome or that the players know the set of outcomes! To ensure this (and other desirable constraints), we later focus on a special class of plausibility models for a game, called "game models".

Examples Figures 2 and 3 represent two different plausibility models $\mathcal{M}_1$ and $\mathcal{M}_2$ for the centipede game G. Here, we use *labelled arrows for the converse plausibility relations* $\geq_a$ (going from less plausible to more plausible states), but for convenience *we skip all the loops and all the arrows that can be obtained by transitivity*.

Note that in the model $\mathcal{M}_2$, Alice (player a) *knows the state of the world*: in each state, she knows both the outcome and Bob's beliefs (and belief revision policy), i.e. the sentence $\bigwedge_{o\in\mathcal{O}}(\overline{o} \Rightarrow K_a\overline{o})$ holds at all states of $\mathcal{M}_2$. But this is

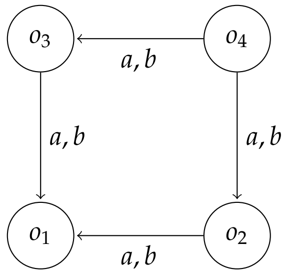


Figure 2: A game model $\mathcal{M}_1$ for the centipede game G

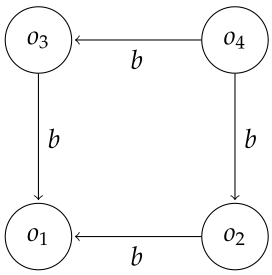


Figure 3: A plausibility model $\mathcal{M}_2$ for G which is *not* a "game model" (see below)

not true in model $\mathcal{M}_1$: on the contrary, in $\mathcal{M}_1$ (it is common knowledge that) nobody knows the outcome of the game, and moreover nobody can exclude any outcome. Intuitively, the future is "epistemically open" in $\mathcal{M}_1$, but not in $\mathcal{M}_2$. However, we can also intuitively see that, in both models, (it is common knowledge that) all the players *know the (structure of the) game*: the available outcomes, the structure of the tree, the payoffs etc.

We now want to formalise our intuitions about open future and about having common knowledge of the structure of the game. To do this, we will focus on a special class of models, that we call "game models". Intuitively, each state of a game model comes with a complete play of the game, and hence it should have a uniquely determined outcome, and the set of possible outcomes as well as the players' preferences over them should be common knowledge. However, the players in this (initial) state should not have non-trivial knowledge about the outcome of the play. Indeed, they should have "freedom of choice" during the play, which means they can in principle play any move, so that at the outset of the play they cannot exclude a priori any outcomes.

Game Models The class of *game models for G*, denoted by $\mathfrak{M}_G$, is the class of all plausibility model for G satisfying the following conditions (for all players $i \in N$):

1. $\forall s \in S\, \exists! o \in \mathcal{O} : s \in V(o)$

2. $V(o <_i o') = \begin{cases} S & \text{if } h_i(o) < h_i(o') \\ \emptyset & \text{otherwise} \end{cases}$

3. $\forall s \in S\, \forall o \in O : V(o) \cap [s]_i \neq \emptyset$

The first condition entails that there is common knowledge of the set of possible outcomes, as well as of the fact that to each state is associated *a unique actual outcome*. This reflects the fact that the future, for each particular play (state), is determined. The second condition entails that the preferences over outcomes are commonly known. Finally, the third condition says that (it is common knowledge that) the *future is epistemically open*: in the initial state of any play, no player has "knowledge" (in the strong sense of "irrevocable", absolutely unrevisable knowledge) that any outcome is impossible. This is meant to apply even to the states that are incompatible with that player's plan of action.

Open Future We take condition (3) to embody the players' *freedom of choice*, as well as the *possibility of error*: in principle, players might always change their minds or make mistakes, hence any belief excluding some of the outcomes may have to be revised later. Even if we would assume (as is usually assumed) that players *(irrevocably) know their own strategy*, i.e. even if they are not allowed to change their minds, and even if we assume (as postulated by Aumann) that *they have common knowledge of "rationality"* (and so that they can exclude some obviously irrational choices), it still would not follow that they can completely exclude any outcome: mistakes can always happen, or players may always lose their rationality and become temporarily insane; so a rational plan does not necessarily imply a rational play, and hence the future still remains open.

Condition (3) is natural given our interpretation of the "knowledge" operator K as representing *hard information*, that is absolutely certain and irrevocable. If any node is "known" (in this sense) to be unreachable, then that node should simply be deleted from the game tree: this just corresponds to playing a *different game*. So if a player i were to irrevocably *know* that a given node is unreachable, then the structure of the game would not really be common knowledge: i would in fact know that she is playing another game than G. Thus, one can consider the "open future" postulate as a natural strengthening of the "common knowledge of the game" assumption.

A different way to proceed would be to impose the above conditions only locally, at the "real" (initial) state of the play. Let Struct_G be the following sentence, describing the "structure of the game" G:

$$\bigvee_{o \in O} o \wedge \bigwedge_{o \neq o' \in O} \neg(o \wedge o') \wedge \bigwedge_{\substack{i \in N, o, o' \in O \\ \text{s.t. } h_i(o) < h_i(o')}} o <_i o' \wedge \bigwedge_{\substack{i \in N, o, o' \in O \\ \text{s.t. } h_i(o) \geq h_i(o')}} \neg o <_i o'$$

Similarly, let $\text{F}_G := \bigwedge_{o \in O, i \in N} \neg K_i \neg o$ be the sentence saying that at the outset of game G the future is epistemically open. Then our proposed "local" requirement is that in the initial state s we have "common knowledge of the structure of the game and of open future", i.e. s satisfies the sentence $Ck(\text{Struct}_G \wedge \text{F}_G)$. Then it is easy to see that this "local" requirement is equivalent to the above global requirement of having a "game model": for every state s in any plausibility model $\mathcal{M}$ for G, s satisfies $Ck(\text{Struct}_G \wedge \text{F}_G)$ iff it is bisimilar[6] to a state in some game model $\mathcal{M}' \in \mathfrak{M}_{\mathcal{G}}$.

Examples Note that the model $\mathcal{M}_1$ from Figure 2 is a *game model*, while $\mathcal{M}_2$ from Figure 3 is *not*: indeed, in $\mathcal{M}_2$ it is common knowledge that Alice always *knows the outcome*, which contradicts the "Open future" assumption.

Encoding Strategies as Conditional Beliefs If a player adopts a particular (pure) *strategy*, our language can encode this in terms of *the player's conditional beliefs about what she would do at each of her decision nodes*. For instance, we say that Alice "adopts the backward induction strategy" in a given state s of a model for the Centipede Game in Figure 1 iff the sentences $B_a o_1$ and $B_a^{v_2} o_3$ hold at state s. Similarly, we can express the fact that Bob adopts a particular strategy, and by putting these together we can capture *strategy profiles*. A given profile is realized in a model if the correspondent sentence is true at a state of that model.

Note that, in our setting, *nothing forces the players to adopt (pure) strategies*. Strategies are "complete" plans of action prescribing a unique choice (a belief that a particular move will be played) for each decision node of the player. But the players might simply consider all their options as equi-plausible, which essentially means that they do not have a strategy.

Examples In (any state of) model $\mathcal{M}_1$ from Figure 2, it is common knowledge that *both players adopt their backward induction strategies*. In contrast, in the model $\mathcal{M}_3$ from Figure 4, it is common knowledge that *no player has a strategy* (at any node):

$$
\begin{tikzcd}
o_3 \arrow[rr, leftrightarrow, "{a,b}"'] \arrow[d, leftrightarrow, "{a,b}"] & & o_4 \arrow[d, leftrightarrow, "{a,b}"] \\
o_1 \arrow[rr, leftrightarrow, "{a,b}"'] & & o_2
\end{tikzcd}
$$

Figure 4: A game model $\mathcal{M}_3$ in which players don't have strategies

So the assumption that *players have (pure) "strategies"* is an extremely strong assumption, which *we will not need*. There is no a priori reason to assume (and there are good empirical reasons to reject) that players play according to fully-determined strategies. Our models are general enough to dispense with this assumption; indeed, our work shows that *this assumption is not needed for proving (common belief) that the backward induction strategy is played*.

Intentions as Beliefs In the above discussion, we identified an agent's *intentions* with *her beliefs about what she is going to do*, and so we represented the decision

[6]Here, "bisimilarity" is the standard notion used in modal logic, applied to plausibility models viewed as Kripke models with atomic sentences in Φ and with relations $\leq_i$. The important point is that our language APAL-CDL cannot distinguish between bisimilar models and states.

maker's plan of action as a belief about her (future) action. This identification is philosophically debatable, since agents may be aware of the possibility of mistakes, and so they may doubt that their intentions will be realized. But one can also argue that, in the context of Game Theory, such distinctions will be of very limited significance: indeed, an intention that is not believed to be enforceable is irrelevant for strategic planning (though see Roy (2008) for a discussion of intentions in game theory). The players only need to know each others' beliefs about their future actions and about each others' beliefs etc., in order to make their own rational plans; whether or not they are being informed about each others' (completely unenforceable and not believed to be enforceable) "intentions" will not make any difference. So, for the purposes of this paper, we can safely adopt the simplifying assumption that the *agents believe that they will be able to carry out their plans*. Given this assumption, an agent's "intentions" can be captured by her beliefs about her (future) actions.

Representing Players' Evolving Beliefs Recall that we think of every state of a game model $\mathcal{M}_G \in \mathfrak{M}_G$ as an *initial state (of a possible play)* of the game G. As the play goes on, the players' hard and soft information, their knowledge and beliefs, *evolve*. To represent this evolution, we will need to successively change our model, so that e.g. when a node v is reached, we want to obtain a corresponding model of the subgame G^v. That is precisely, in this perfect information setting, what is achieved by *updating the model with public announcements*: indeed, in a game of perfect information, every move, say from a node u to one of its immediate successors u', can be "simulated" by a public announcement $!u'$. In this way, for each subgame G^v of the original model $\mathcal{M}$, we obtain a model $\mathcal{M}^v$, that *correctly describes the players' knowledge and beliefs at the moment when node v is reached during a play*. This is indeed a model of the corresponding subgame G^v:

Proposition 2. *If $\mathcal{M} \in \mathfrak{M}_G$ then $\mathcal{M}^v \in \mathfrak{M}_{G^v}$.*

Example Consider a play of the Centipede game G that starts in the initial situation described by the model $\mathcal{M}_1$ in Figure 2, and in which the real state of the world is the one having outcome o_2: so Alice first plays "right", reaching node v_1, and the Bob plays "down", reaching the outcome o_2. The model $\mathcal{M}_1$ from Figure 2 gives us the initial situation, the model $\mathcal{M}_1^{v_1}$ in Figure 5 describes the epistemic situation after the first move, and then the model $\mathcal{M}_1^{o_2}$ in Figure 6 gives the epistemic situation at the end of the play:

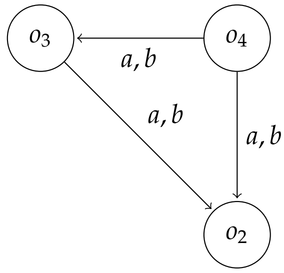
In this way, for each given initial state s (of a given play $v_0, v_1, \ldots, o$ of the game, where o is the unique outcome such that $s \in V(o)$), we obtain a *sequence of evolving game models*

$$\mathcal{M} = \mathcal{M}^{v_0}, \mathcal{M}^{v_1}, \ldots, \mathcal{M}^o ,$$

describing *the evolving knowledge and beliefs of the players* during any play. Each model $\mathcal{M}^v$ accurately captures the players' beliefs at the moment when node v is reached. Note also that every such sequence ends with a model $\mathcal{M}^o$ consisting of only one node (a leaf o); this reflects the fact that *at the end of the game, there is no uncertainty left*: the outcome, as well as the whole history of the game, are now common knowledge.

Simulating Moves by Public Announcements Using the dynamic "public announcement" modalities in constructs such as $[!v]B_i$, we can talk, *at the initial*

Figure 5: The model $\mathcal{M}_1^{v_1}$



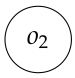


Figure 6: The model $\mathcal{M}_1^{o_2}$

state $s \in \mathcal{M}$ and *without leaving the original model* $\mathcal{M} \in \mathfrak{M}_G$, about all these future, evolving beliefs of the players at nodes v other than the initial node v_0. Indeed, in a game of perfect information, all the moves are *public*. So the epistemic effect of a move to node v is the same as that of a truthful public announcement $!v$ (saying that the node v is reached during the play). In other words, *we can "simulate" moves in games of perfect information by truthful public announcements*.[7]

3 Dynamic Rationality in Decisions and Games

We now define our fundamental notions of *dynamic rationality* and *rational play*. First we will look at *single-agent (one-step) decision situations*, and then at interactive decision situations, i.e. *games*.

3.1 Single Agent Decision Problems

Given a *one-step decision problem* $\mathcal{P}$ with a set of *outcomes* O, the *decision-maker* i selects one of the outcomes $o \in O$. The decision-maker may have various hard and soft information about which outcomes can actually be realized and which not. This will determine her knowledge and her beliefs. We assume that *her "hard" knowledge restricts her possible choices*: she *can* only select outcomes that she doesn't *know* to be *impossible*.

What this amounts to is the following: for the decision maker i, the "true" set of possible outcomes is $\{o \in O \mid \neg K_i \neg o\}$, i.e. the set of all the *"epistemically possible" outcomes*. So her selected option must satisfy: $o \in \{o \in O \mid \neg K_i \neg o\}$. This allows us to capture the "selection" problem using epistemic operators.

[7]We believe that the more general case, of games of imperfect information, can also be handled by using other kinds of epistemic actions proposed in Dynamic Epistemic Logic Baltag et al. (1999). But we leave this development for future work.

To assess whether the decision is "rational" or not, one considers the decision-maker's subjective preferences, modelled as a total pre-order $\preccurlyeq_i$ on O. We assume that *agents know their preferences*; indeed, these are interpreted as "doxastic" preferences: beliefs about what's best. Given this interpretation, the CDL postulation of Full Introspection (of beliefs) implies that agents know their preferences.

Rational Choice Rationality, in this case, corresponds to requiring that the selected option is *not worse than* any other (epistemically) possible alternative. In other words, i's solution of the decision problem $\mathcal{P}$ is *rational* if she does not choose any option that is strictly less preferable than an option she doesn't know to be impossible:

$$\mathrm{R}_i^{\mathcal{P}} := \bigwedge_{o,o' \in O} (o \prec_i o' \wedge \neg K_i \neg o' \Rightarrow \neg o).$$

The main difference between our definition and the standard definition of rational decision-making is the *epistemic limitation of the choice set*. The epistemic operators are used here to delimit *what is currently known about the availability of options*: i's choice should only be compared against options that are not known to be unavailable. This is an *important difference*, and its importance becomes clear when we generalise our definition to extensive games, cf. the difference between 'dynamic' rationality and traditional 'substantive' rationality, described below.

3.2 Extensive Games

We now aim to extend the above definitions to the case of *multi-agent many-stage decisions*, i.e. extensive games (of perfect information). Recall that in an extensive game we are given the players' subjective preferences $\preccurlyeq_i$ only over the leaves. However, at all the intermediate stages of the game, players have to make local choices, not between "final" outcomes, but between "intermediary" outcomes, that is: between other nodes of the game tree.

So, in order to assess players' rationality, we need to *extend* the subjective preference relations to *all the nodes* of the game tree. Fortunately, given the above doxastic interpretation of preferences, there is an obvious (and natural) way to define these extensions. Namely, a player considers a node u to be *strictly less preferable* to a node u' if she *believes* the first to be *strictly dominated* by the second. More precisely, if every outcome that she believes to be achievable given that u is reached is worse than every outcome that she believes to be achievable given that u' is reached:

$$u \prec_i u' := \bigwedge_{o,o' \in O} (\neg B_i^u \neg o \wedge \neg B_i^{u'} \neg o' \Rightarrow o \prec_i o').$$

By the Full Introspection of beliefs (a postulate of the logic *CDL*), it follows that we still have that *players know their extended preferences* over all the nodes of the game.

Rationality at a Node Each node $v \in G_i$ can be considered as a (distinct) decision problem, in which the decision-maker is i, the set of outcomes is the set $\{u \in G : v \to u\}$ of all immediate successors of v, and the subjective preference relation is given by the (restriction of the) extended relation $<_i$ defined above (to the set $\{u \in G : v \to u\}$). So we can define the rationality of a player i at a node $v \in G_i$ as rationality for the corresponding decision problem, i.e. the player's selection at each decision node consists only of "best answers". Note that, as before, *the player's choice is epistemically limited*: if she has "hard knowledge" excluding some successors (for instance, because those nodes have already been bypassed), then those successors are excluded from the set of possible options. The only difference is that the "knowledge" involved is the one the agent *would have at that decision node*, i.e. it is *conditional on that node being reached*. Formally, we obtain:

$$\mathrm{R}_i^v := \bigwedge_{u,u' \leftarrow v} (u <_i u' \wedge \neg K_i^v \neg u' \Rightarrow \neg u)$$

where $K_i^{\varphi}\psi := K_i(\varphi \Rightarrow \psi)$.

Dynamic Rationality Let R_i be the sentence

$$\mathrm{R}_i = \bigwedge_{v \in G_i} \mathrm{R}_i^v.$$

If R_i is true, we say that player i satisfies *dynamic rationality*. By unfolding the definition, we see it is equivalent to:

$$\mathrm{R}_i = \bigwedge_{v \in G_i} \bigwedge_{u,u' \leftarrow_i v} (u <_i u' \wedge \neg K_i^v \neg u' \Rightarrow \neg u).$$

As we'll see, asserting this sentence at a given moment is a way of saying that *the player will play rationally from that moment onwards*, i.e. she will make the best move at any *current or future* decision node.

In the following, "Dynamic Rationality" denotes the sentence

$$\mathrm{R} := \bigwedge_i \mathrm{R}_i$$

saying that all players are dynamically rational.

Comparison with Substantive Rationality To compare our notion with Aumann's concept of "substantive rationality", we have to first adapt Aumann's definition to a belief-revision context. This has already been done by a number of authors e.g. Battigalli and Siniscalchi (1999, 2002), resulting in a definition of "rationality at a node" that differs from ours only by *the absence of epistemic qualifications to the set of available options* (i.e. the absence of the term $\neg K_i^v \neg u'$). The notion of *substantive rationality* is then obtained from this in the same way as dynamic rationality, by quantifying over all nodes, and it is thus equivalent to the following definition:

$$\mathrm{SR}_i = \bigwedge_{v \in G_i} \bigwedge_{u,u' \leftarrow_i v} (u <_i u' \Rightarrow \neg u).$$

It is obvious that *substantive rationality implies dynamic rationality*

$$\mathrm{SR}_i \Rightarrow \mathrm{R}_i,$$

but *the converse is in general false*. To better see the difference between SR_i and R_i, recall that a formula being true in a model $\mathcal{M} \in \mathfrak{M}_G$ means that it is true at the first node (the root) of the game tree G. However, we will later have to evaluate the formulas R_i and SR_i at *other* nodes w, i.e. in *other models* of the form M^w (models for *subgames* G^w). Since *the players' knowledge and beliefs evolve during the game*, what is (not) known/believed conditional on v *in model* M^w differs from what was (not) known/believed conditional on v *in the original model* (i.e. at the outset of the game). In other words, the meaning of both dynamic rationality R_i and substantive rationality SR_i *will change during a play*. But they *change in different ways*. At the initial node v_0, the two notions are equivalent. But, once a node v has been bypassed, or once the move at v has already been played by a player i, that player is counted as *rational at node* v according to our definition, while according to the usual (non-epistemically qualified) definition the player may have been irrational at v.

In other words, the epistemic limitations we imposed on our concept of dynamic rationality make it into a *future-oriented* concept. At any given moment, the rationality of a player depends only on her current beliefs and knowledge, and so only on the options that she *currently considers possible*: past, or by-passed, options are irrelevant. Dynamic Rationality simply expresses the fact that the player's decision in any future contingencies is rational (given her future options and beliefs). Unlike substantive rationality, our concept has nothing to do with the past or with contingencies that are known to be impossible: a player i may still be "rational" in our sense at a given moment/node v even when v could only have been reached if i has already made some "irrational" move. The (knowledge of some) past mistake(s) may of course affect the others' *beliefs* about this player's rationality; but it doesn't directly affect her rationality, and in particular it doesn't automatically render her irrational.

Solving the BI Paradox As explained above, our concept is very different from (and, arguably, more realistic than) Aumann's and Stalnaker's substantive rationality, but also from other similar concepts in the literature (for example Rabinowicz's (1998) "habitual" or "resilient" rationality, etc). The difference becomes more apparent if we consider the assumption that *"rationality" is common belief, in the strongest possible sense,* including *common "strong" belief* (in the sense of Battigalli and Siniscalchi (2002)), common persistent belief, or even common "knowledge" in the sense of Aumann. As correctly argued by Stalnaker and Reny, these assumptions, if applied to the usual notions of rationality in the literature, bear no relevance for what the players would do (or believe) at the nodes that are incompatible with these assumptions! The reason is that, if these counterfactual nodes were to be reached, then by that time the belief in "rationality" would have already been *publicly disproved*: we cannot even entertain the possibilities reachable by irrational moves except by suspending our belief (or "knowledge") in rationality. Hence, the above assumptions cannot tell us anything about the players' behaviour or rationality at such counterfactual nodes, and thus they cannot be used to argue for the plausibility of the backward induction solution (even if they logically imply it)! In contrast, our notion of dynamic rationality is *not* automatically disproved when we reach a node excluded by common belief in it: a player *may* still be rational with respect to her current and future options and decisions *even after* making an "irrational" move. Indeed, the player may have been playing irrationally in the past, or may have had a moment of temporary irrationality, or may have made some mistakes in carrying out her rational plan; but she may have recovered now and may play rationally thereafter. Since our notion of rationality is future-oriented, no information about past moves will necessarily and automatically shatter belief in rationality (although of course it *may* still shatter it, or at least weaken it). So it is perfectly consistent (although maybe not always realistic) to assume that players maintain their common belief in dynamic rationality despite all past failures of rationality. In fact, *this is our proposed solution to the BI paradox*: we will show that such a "stable" common belief in dynamic rationality (or more precisely, common *knowledge* of the stability of the players' common belief in rationality) is exactly what is needed to ensure common belief in the backward induction outcome!

Rational Planning A weaker condition requires only that, for each decision node v, the option that the decision-maker *is planning at v to select (at v)* is the best, given the other (epistemically) possible alternatives. By identifying as above the players' plans of actions with their beliefs about their actions, we can thus say that a decision maker is a *rational planner* in the game G if at each decision node she *believes* that she will take "the best decision", even if in the end she may accidentally make a wrong choice:

$$\mathrm{RP}_i := \bigwedge_{v \in G_i} B_i^v \mathrm{R}_i^v.$$

By unfolding the definition, we see it is equivalent to:

$$\mathrm{RP}_i = \bigwedge_{v \in G_i} \bigwedge_{u,u' \leftarrow_i v} (u \prec_i u' \wedge \neg K_i^v \neg u' \Rightarrow B_i^v \neg u).$$

No Mistakes RP_i only states that the decision maker i has a *rational plan* for current and future contingencies. But mistakes can happen, so if we want to ensure that the decision that is actually taken is rational we need to require the player makes *no mistakes* in carrying out her plan:

$$\text{No-Mistakes}_i := \bigwedge_{v \in G_i} \bigwedge_{u \leftarrow v} (B_i^v \neg u \Rightarrow \neg u)$$

The sentence No-Mistakes_i says that player i's decision are always consistent with her "plan": she never plays a move that, at the moment of playing, she believed won't be played.

As expected, the conjunction of "rational planning" and "no mistakes" entails "rational playing" (i.e., "dynamic rationality"):

$$\mathrm{RP}_i \wedge \text{No-Mistakes}_i \Rightarrow \mathrm{R}_i.$$

4 Backward Induction in Games of Perfect Information

It is easy to see that Aumann's theorem stating that common knowledge of substantive rationality implies the backward induction outcome Aumann (1995) can be strengthened to the following

Proposition 3. *In any state of any plausibility model for a game of perfect information, common knowledge of dynamic rationality implies the backward induction outcome.*

Unfortunately, common knowledge of (either dynamic or substantive) rationality *can never hold in a game model*: it is simply incompatible with the "Epistemically-Open Future" condition. By requiring that players have "hard" information about the outcome of the game, Aumann's assumption *does not allow them to reason hypothetically or counterfactually about other possible outcomes*, at least not in a consistent manner.[8] This undermines the intuitive rationale behind the backward induction solution, and it is thus open to Stalnaker's criticism.

So in this section, we are looking for natural conditions that can be satisfied on game models, but that still imply the backward induction outcome (or at least common belief in it). One such condition is *common knowledge of (general) stable belief* in (dynamic) rationality: $Ck[!]EbR$. This is in fact a "strong" form of common belief, being equivalent to $Ck[!]CbR$, i.e. to *common knowledge of stable common belief* in rationality.

Theorem 1. *The following holds in any state s of any game model* $\mathcal{M} \in \mathfrak{M}_G$*:*

$$Ck[!]Eb\mathrm{R} \;\Rightarrow\; Cb(BI_G)\,,$$

where $BI_G := \bigvee\{o \mid o \in BI_G\}$ *is the sentence saying that the current state determines a backward-induction outcome in the game G. Equivalently, the following formula is valid over plausibility frames for the game G:*

$$Ck(Struct_G \wedge F_G \wedge [!]Cb\mathrm{R}) \;\Rightarrow\; Cb(BI_G).$$

In English: assuming common knowledge of the game structure and of open future, if it is common knowledge that, no matter what new (truthful) information the players may (jointly) learn during the game (i.e. *no matter what is played*), general belief in rationality will be maintained, then it is common belief that the backward induction outcome will be reached. If we define "stable common belief" in a proposition P as $[!]CbP$, then we can give a more concise English formulation of the above theorem: *common knowledge of the game structure, of open future and of stable common belief in dynamic rationality implies common belief in the backward-induction outcome.*

Although rationality cannot be common knowledge in a game model, *rational planning can be*. When this is the case, we obtain the following

Corollary 1. *In a game model, common knowledge of "rational planning" and of stable belief in "no mistakes" implies the backward-induction outcome; i.e. the formula*

$$Ck(\mathrm{RP} \wedge [!]Eb\text{No-Mistakes}) \;\Rightarrow\; Cb(BI_G)$$

is valid on game models.

The above results only give us *common belief* in the backward-induction outcome, but nothing ensures that this belief is correct. If we want to ensure that the backward-induction outcome is actually played, we need to add the requirement that the (stable common) belief in rational play assumed in the premise is correct, i.e. that *players actually play rationally*:

Theorem 2. *The following holds in any state s of any game model* $\mathcal{M} \in \mathfrak{M}_G$*:*

$$\mathrm{R} \wedge Ck[!]Eb\mathrm{R} \;\Rightarrow\; BI_G$$

[8] Indeed, if o is the backward induction outcome, then the above Proposition entails $K_i o$ for all players i, and thus for every other outcome $o' \neq o$ and every proposition P, we have $B_i^{o'}P$: the players *believe everything* (including inconsistencies) conditional on o'.

No strategies! Observe that we did not assume that the players have complete (pure) "strategies". That is, we do not insist that they have fully determined plans of action, uniquely specifying one move for at each decision node, but only that they have *partial* plans, i.e. incomplete beliefs about what moves they should play: at each decision node they choose a *set* of moves rather than one unique move. So an important side-result of our work is that the assumption that players have (complete, pure) strategies is *not necessary* for proving backward-induction results.

Ensuring Backward-Induction Strategy Profile If, however, we want to postulate that *every player does have a (complete, pure) strategy*, we need to say that, for each node v of her choice, there exists a (unique) immediate successor u that she *believes* will be played if v is reached (i.e. she *plans to play u* at v):

$$\text{Strategies} := \bigwedge_i \bigwedge_{v \in G_i} \bigvee_{u \leftarrow_i v} B_i^v u.$$

In cases where Strategies is common knowledge as well, we can strengthen the Theorem 1 to:

Corollary 2. *The following holds in any state s of any game model $\mathcal{M} \in \mathfrak{M}_G$:*

$$Ck(\mathit{Strategies} \wedge [!]Eb\mathrm{R}) \;\Rightarrow\; Cb(\mathit{BI\text{-}Profile}_G)$$

where $\mathit{BI\text{-}Profile}_G$ is the sentence saying that the strategies given by each player's conditional beliefs in the initial state s form a backward-induction profile.

Finally, the following theorem ensures that above results are not vacuous:

Theorem 3. *For every extensive game G, there is a game model $\mathcal{M} \in \mathfrak{M}_G$ and a state $s \in \mathcal{M}$ satisfying the sentence*

$$\text{No-Mistakes} \wedge Ck(\mathrm{RP} \wedge \mathit{Strategies} \wedge [!]Eb\text{No-Mistakes}).$$

As a consequence, the sentence $\mathrm{R} \wedge Ck[!]Eb\mathrm{R} \wedge Ck\mathit{Strategies}$ *is also satisfied.*

The *proofs* of these theorems are in the next section. Some alternative (weaker) conditions ensuring the backward induction outcome are given in Section 6.

5 Proofs

Definition 5.1. For a finite set O of "outcomes" and a finite set P of "players", we denote by $\mathfrak{Games}(O, P)$ the class of all perfect information games having any subset of O as their set of outcomes and having any subset of P as their set of players.

Definition 5.2. For any sentence φ of our language,
φ is *valid on a game G* if φ is true at every state s of every game model $\mathcal{M} \in \mathfrak{M}_G$.
φ is *valid over $\mathfrak{Games}(O, P)$* if φ is valid on every game $G \in \mathfrak{Games}(O, P)$.

When the game G is implicit from the context, we will often abbreviate BI_{G^u}, i.e. the name for the formula that defines all the backward-induction outcomes in the subgame of G that starts at the node u, to BI^u.

Lemma 1. *For every perfect information game G, if we denote the root of G by v_0, the first player of G (playing at v_0) by i and the first move of i (the successor node played at v_0) by v_1, then the sentence*

$$R_i^{v_0} \wedge \bigwedge_{u \leftarrow v_0} B_i^u [!u] BI^u \wedge [!v_1] BI^{v_1} \Rightarrow BI$$

is valid on G.

Proof. This follows directly from the definition of rationality at a node and the definition of BI. The assumption that $B_i^u[!u]BI^u$ is true at s means that all the states (deemed as "most plausible by i conditional on u ") in the set $s_i^u := min_{\leq_i}(\overline{u} \cap [s]_i)$ have only outcomes that are backward induction outcomes in the corresponding subgame: i.e. we have $o(t) \in BI_{G^u}$ for all $t \in s_i^u$. Given that all these outcomes $\{u : u \leftarrow v_0\}$ are consistent with i's knowledge (since we are in a game model), the fact that i is rational at v_0 implies that the successor node v_1 chosen by i must be one that maximises her payoff $h_i(o(s_i^u))$ among all the outcomes in $\bigcup_{u \leftarrow v_0} BI_{G^u}$. But, by the definition, such a node v_1 is exactly the choice prescribed at v_0 by the backward induction strategy! Given this backward-induction choice (v_1) of i at node v_0, and given the fact (ensured by the condition $[!v_1]BI^{v_1}$) that starting from node v_1 everybody *will* play the backward induction choices, we can conclude that the outcome $o(s)$ belongs to the backward induction set of outcomes $BI_{G^{v_0}} = BI_G$ for the game G. Hence s satisfies BI_G. □

The *Main Lemma* underlying our results is the following:

Lemma 2. *(**"Main Lemma"**) Fix a finite set O of outcomes and a finite set P of players. Let φ be any sentence in our language APAL-CDL having the following property: for every game $G \in \mathfrak{Games}(O, P)$, if we denote the root of G by v_0, the first player of G (playing at v_0) by i and the first move of i (the successor node played at v_0) by v_1, then the sentence*

$$\varphi \Rightarrow R_i^{v_0} \wedge \bigwedge_{u \leftarrow v_0} B_i^u [!u]\varphi \wedge [!v_1]\varphi$$

is valid on G.

Given this condition, we have that the sentence

$$\varphi \Rightarrow BI_G$$

is valid over $\mathfrak{Games}(O, P)$.

Proof. We need to prove that, for every game $G \in \mathfrak{Games}(O, P)$, the sentence $\varphi \Rightarrow BI_G$ is valid on G. The proof is *by induction on the length of the game G*.

For games of length 0 (only one outcome, no available moves), the claim is trivial (since the only possible outcome is by definition the backward induction outcome).

Let G be now a game of length $n > 0$, and assume the claim is true for all games of smaller length. Let v_0 be the root of G, i be the first player of G, $\mathcal{M} \in \mathfrak{M}_G$ be a game model for G and s be a state in $\mathcal{M}$ such that $s \models_{\mathcal{M}} \varphi$.

Let u be any *arbitrary immediate successor* of v_0 (i.e. any node such that $u \leftarrow v_0$). By the property assumed in the statement of this Lemma, we have that $s \models_{\mathcal{M}} B_i^u[!u]\varphi$, and so (if s_i^u is the set defined in the proof of the previous Lemma, then) we have $t \models_{\mathcal{M}} [!u]\varphi$ for all $t \in s_i^u$. Hence, we have $t \models_{\mathcal{M}^u} \varphi$ for all $t \in s_i^u \cap \overline{u}$. By the induction hypothesis, we must have $t \models_{\mathcal{M}^u} BI^u$ (since $\mathcal{M}^u$ is a game model for G^u, which has length smaller than G, and so the implication $\varphi \Rightarrow BI^u$ is valid on $\mathcal{M}^u$), for all $t \in s_i^u \cap \overline{u}$. From this we get that $t \models_{\mathcal{M}} [!u]BI^u$ for all $t \in s_i^u$, and hence that $s \models_{\mathcal{M}} B_i^u[!u]BI^u$.

Let v_1 be now the first move of the game in state s (i.e. the unique immediate successor $v_1 \leftarrow v_0$ such that $s \models_{\mathcal{M}} v_1$). By the property assumed in this Lemma, we have that $s \models_{\mathcal{M}} [!v_1]\varphi$. By the same argument as in the last paragraph, the induction hypothesis gives us that $s \models_{\mathcal{M}} [!v_1]BI^{v_1}$. Putting together with the conclusion of the last paragraph and with the fact (following from the Lemma's hypothesis) that $\varphi \Rightarrow R_i^{v_0}$ is valid on $\mathcal{M}$, we infer that $s \models_{\mathcal{M}} R_i^{v_0} \wedge \bigwedge_{u \leftarrow v_0} B_i^u[!u]BI^u \wedge [!v_1]BI^{v_1}$. The desired conclusion follows now from Lemma 1. □

Lemma 3. *The sentence*

$$\varphi \;:=\; \mathrm{R} \wedge Ck[!]Eb\mathrm{R}$$

has the property assumed in the statement of Lemma 2.

Proof. The claim obviously follows from the following three sub-claims:

1. dynamic rationality is a "stable" property, i.e. the implication $\mathrm{R} \Rightarrow \bigwedge_u [!u]\mathrm{R}$ is valid;
2. the implication $Ck[!]Eb\psi \Rightarrow B_i^u[!u]Ck[!]Eb\psi$ is valid, for all formulas ψ and all nodes $u \in G$;
3. the implication $Ck[!]Eb\psi \Rightarrow [!u]Ck[!]Eb\psi$ is valid, for all formulas ψ and all nodes u.

All these claims are easy exercises in dynamic-epistemic logic. The first follows directly from the definition of dynamic rationality.

The second sub-claim goes as follows: assume that we have $Ck[!]Eb\psi$ at some state of a given model; then we also have $Ck[!u][!]Eb\psi$ for any node u (since $[!]\theta$ implies $[!u][!]\theta$), and so also $K_iCk[!u][!]Eb\psi$ (since common knowledge implies knowledge of common knowledge), from which we get $B_i^uCk[!u][!]Eb\psi$ (because knowledge implies conditional belief under any conditions). This is the same as $B_i^u(u \rightarrow Ck[!u][!]Eb\psi)$, which implies $B_i^u(u \rightarrow Ck^u[!u][!]Eb\psi)$ (since common knowledge implies conditional common knowledge). But this last clause is equivalent to $B_i^u[!u]Ck[!]Eb\psi$ (by the Reduction Law for common knowledge after public announcements).

The third sub-claim goes as follows: assume that we have $Ck[!]Eb\psi$ in some state of a given model; then as before we also have $Ck[!u][!]Eb\psi$, and thus $Ck^u[!u][!]Eb\psi$ (since common knowledge implies conditional common knowledge). From this we get $u \rightarrow Ck^u[!u][!]Eb\psi$ (by weakening), which is equivalent to $[!u]Ck[!]Eb\psi$ (by the Reduction Law for common knowledge after public announcements). □

Theorems 2 and 1

Proof. Theorem 2 follows now from Lemma 2 and Lemma 3. Theorem 1 follows from Theorem 2, by applying the operator $Ck[!]Eb$ to both its premiss and its conclusion, and noting that the implication

$$Ck[!]Eb\psi \Rightarrow Ck[!]EbCk[!]Eb\psi$$

is valid. □

6 An Alternative Condition: Common Stable True Belief in Dynamic Rationality

The epistemic condition $\mathrm{R} \wedge Ck[!]Eb\mathrm{R}$ given above is *not* the weakest possible condition that ensures the backward induction outcome. Any property φ satisfying the condition of our Main Lemma (Lemma 2) would do it. In particular, there exists a *weakest* such condition (the smallest event $E \subseteq S$ such that $E \subseteq R_i^{v_0} \cap \bigcap_{u \leftarrow v_0} B_i^u[!u]E \cap [!v_1]E$), but it is a very complicated and unnatural condition. The one given above seems to be the simplest such condition expressible in our language APAL-CDL.

However, one can give weaker simple conditions if one is willing to go a bit beyond the language APAL-CDL, by adding fixed points for other (definable) epistemic operators.

Let *stable true belief* be a belief that is known to be a stable belief and it is also a stably true belief. Formally, we define:

$$Stb_i\varphi \ := \ K_i[!]B_i\varphi \wedge [!]\varphi.$$

Thus stable true belief is *stable* belief in a *stably* true proposition. Stable true belief is a form of "knowledge", since *it implies truth and belief:*

$$Stb_i\varphi \Rightarrow \varphi \wedge B_i\varphi.$$

We can also think of stable true belief as an epistemic attitude towards the *stability* of a proposition: clearly it implies stably truth ($Stb_i\varphi \Rightarrow [!]\varphi$), but furthermore, *knowledge that something is stably true implies stable true belief in it.*

$$K_i[!]\varphi \Rightarrow Stb_i\varphi.$$

Stable true belief is inherently a "positively introspective" attitude, i.e.

$$Stb_i\varphi \Rightarrow Stb_iStb_i\varphi,$$

but it is not positively introspective with respect to ("hard") knowledge:

$$Stb_i\varphi \not\Rightarrow K_iStb_i\varphi.$$

Stable true belief is *not* negatively introspective, neither inherently nor with respect to knowledge.

If we restrict our attention to only *ontic* (i.e. non-doxastic) facts p, then we cannot detect the subtleties of stable true belief, and the difference between this concept, "stable belief" (simpliciter) and "safe belief". Notice in particular that, *when when applied to ontic facts p*, stable true belief of p is just the same as stable belief of p and the same as the "safe belief" in p from Baltag and Smets (2008b). (which is the same as what Stalnaker calls "knowledge" Stalnaker (2006)). However, it is typical of *interactive epistemology* that one is not in general interested in epistemic/doxastic attitudes towards ontic facts, but in attitudes towards propositions that in turn depend on other attitudes. Examples of such higher-level attitudes are the important game-theoretic notions of "common knowledge of (or common belief in) rationality", "common knowledge of stable belief in rationality" and "common stable true belief in rationality": exactly the notions that interest us in this paper!

We can define *common stable true belief* in the same way as common knowledge: first define *general stable true belief*

$$Estb\varphi = \bigwedge_{i \in P} Stb_i\varphi$$

("everybody has stable true belief"), then put

$$Cstb\varphi = \bigwedge_{n} (Estb)^n\varphi.$$

Note that this definition, although semantically meaningful, is not a definition in our language APAL-CDL, since it uses infinite conjunctions. Indeed, we conjecture that common stable true belief is undefinable in the language APAL-CDL, since it doesn't seem to be expressible as a combination of common knowledge, common belief and dynamic operators.

Lemma 4. *The sentence $Cstb\mathrm{R}$ satisfies the condition of our Main Lemma (Lemma 2).*

As an immediate consequence, we have:

Theorem 4. The sentence

$$Cstb\mathrm{R} \Rightarrow BI_G$$

is valid over game models. In English: (if we assume common knowledge of the structure of the game and of open future, then) *common stable true belief in (dynamic) rationality implies the backward induction outcome .*

7 Comparison with Other Work

The game-theoretic issues that we deal with in this paper originate in the work of Aumann (1995), Stalnaker (1994, 1996, 1998) and Reny (1992), and have been investigated by a number of authors Binmore (1987; 1996), Bicchieri (1989), Battigalli (1997), Battigalli and Siniscalchi (1999; 2002), Bonanno (1991), Brandenburger (2007), Halpern (2001), Samet (1996), Clausing (2003) etc. Our work obviously owes a great deal to these authors for their illuminating discussions of the topic.

The logic CDL of conditional belief was first introduced and axiomatised by Board (2002), in a slightly more complicated form. The version presented here is due to Baltag and Smets (2006, 2008b). The dynamic extension of CDL obtained by adding the public announcements modalities (coming from the public announcement logic PAL, originally developed by Plaza (1989)) has been developed by van Benthem (2007a) and, independently, by Baltag and Smets

Baltag and Smets (2006). The extension of PAL with arbitrary announcement modalities $[!]\varphi$ is due to Balbiani et al (2008). The belief-revision-friendly version of APAL presented here (obtained by combining APAL with CDL) is an original contribution of our paper.

The work of Battigalli and Siniscalchi (2002) is the closest to ours, both through their choice of the basic setting for the "static logic" (also given by conditional belief operators) and through the introduction of a strengthened form of common belief ("common strong belief") as an epistemic basis for a backward-induction theorem. Strong belief, though different from our "stable" belief, is another version of persistent belief: belief that continues to be maintained unless and until it is contradicted by new information. However, their notion of rationality is only "partially dynamic": although taking into account the dynamics of beliefs (using conditional beliefs given node v to assess the rationality of players' choices at v), it does not fully take into account the limitations posed to the set of possible options by the dynamics of "hard knowledge". In common with most other previous notions of rationality, it requires agents to make rational choices at all nodes, including the past ones and the ones that have already been bypassed. As a result, it is enough for a player to make only one "irrational" move to completely shatter the (common) belief (however strong) in rationality; and as a consequence, common strong belief in rationality does not by itself imply backward induction. To obtain their theorem, Battigalli and Siniscalchi have to add another assumption: that the game model is a *complete type structure*, i.e. it contains, in a certain sense, every possible epistemic-doxastic "type" for each player. This means that the players are assumed to have *absolutely no "hard" information, not only about the outcomes or about the other players' strategies, but also about the other players' beliefs*, so that they have to consider as epistemically possible *all* consistent (probabilistic) belief assignments for the other players! This is an extremely strong (and, in our opinion, unrealistic) "completeness" assumption, one that can only be fulfilled in an infinite model. In contrast, the analogue completeness assumption in our approach is the much weaker "Open Future" assumption, postulating that (at the beginning of the game) players have no non-trivial "hard" information about the outcomes (except the information given by the structure of the game): they cannot foretell the future, cannot irrevocably know the players' freely chosen future moves (though they *do* irrevocably know the past, and they *may* irrevocably know the present, including all the beliefs and the plans of action of all the players). Our more realistic postulate is weak enough to be realized on finite models. In particular, it can be realized on models as small as the set of terminal nodes of the game tree (having one state for each terminal node), and in which all the plans of action are common knowledge, so that the only uncertainty concerns possible mistakes in playing (and hence the final outcome).

Samet (1996) introduces a notion of *hypothetical knowledge*, in order to develop an epistemic characterisation of backward induction. Hypothetical knowledge looks prima facie similar to conditional belief, except that the interpretation of the hypothetical knowledge formula $K_i^{\varphi}\psi$ is different: "Had φ been the case, i would have known ψ" (Samet (1996), p. 237). This mixture of counterfactual conditionals and knowledge is specifically introduced in Samet (1996) only to discuss backward induction, and it has not occurred before or subsequently in the literature. In contrast, our approach is grounded in the relatively standard and well-understood foundations of Conditional Doxastic Logic, independently studied by logicians and philosophers. While Samet does make what we agree is the important point that some form of counterfactual reasoning is of vital importance to the epistemic situation in extensive games, his model and conditions seem to us more complex, less transparent and less intuitive than ours.

We are aware of only one prior work that uses dynamic epistemic logic (more precisely, the logic of public announcements, but in the context of "classical DEL", i.e. dealing only with knowledge update and not with belief revision) for the analysis of solution concepts in extensive games: Benthem (2007b). That work takes Aumann's "static" notion of rationality as given, and accepts Aumann's classical result as valid, and so it does not attempt to deal with the cases in which Aumann's assumptions do not apply, nor to address the criticism and the issues raised by Stalnaker, Reny and others. Instead, van Benthem's contribution focuses on the *sources* of knowledge, on explaining *how* complex epistemic conditions of relevance to Game Theory (such as Aumann's common knowledge of rationality) can *be brought about*, via repeated public announcements of rationality. So van Benthem does *not* use public announcements in order to simulate a play of the game. Public announcements in van Benthem's approach represent *off-line learning*, i.e. *pre-play* or *inter-play* learning, whereas the public announcements in our present approach simulate *on-line learning*, i.e. learning that takes place *during the play* of the game. A very interesting open question is to address the same issue answered by van Benthem, but for the case of the dynamic-epistemic condition proposed here, instead of Aumann's condition: find some off-line communication or learning protocol that can achieve common knowledge of stable common belief in rational play.

We should say how our result resolves the apparently conflicting positions of Aumann and Stalnaker. Under sympathetic interpretations of those authors, we would say that we agree with both of them:

We have already commented on the differences between our approach and that of Aumann (1995). However, in order to find a similarity, notice that if we say that strategies *are* beliefs, then the condition we give begins to look a little like common knowledge of rational strategies. (This identification of strategies with beliefs was not possible in Aumann's framework, so even from this perspective our work would be a considerable advance.)

Stalnaker writes that "[t]he rationality of choices in a game depend[s] not only on what players believe, but also on their policies for revising their beliefs" (1998, p. 31). He then gives a condition on belief revision policies in terms of "epistemic independence" of the players. We agree entirely with the sentiment in the quoted sentence. Indeed game models provide a specification of exactly how players will revise their beliefs, including their beliefs about other players' beliefs, so that these beliefs remain consistent no matter how the play of the game goes. Theorem 2 goes further and *specifies* conditions necessary on such models, purely in terms of epistemic and doxastic attitudes towards rationality, that ensure the backward induction outcome. Stable belief in dynamic rationality is in effect a partial description of an "optimistic" belief revision policy, that says: "when you revise your beliefs, maintain at all costs a belief in the opponents' rational potential, despite their past deviations from rationality".

In fact, as mentioned in the Introduction, whether this policy can appropriately be called "optimistic" or "pessimistic" depends on the game and the players' payoffs. In many contexts, such an "incurably optimistic" (or "persistently pessimistic") revision policy may seem naïve, but our point is that *only such a policy can offer a rational doxastic justification to backward induction*. The well-known examples of "catastrophic" BI outcomes can thus been seen to illustrate the dangers of "rational" pessimism, while the examples of "desirable" BI outcomes illustrate the saving power of "incurable" optimism.

Acknowledgements The authors would like to give special thanks to Johan van Benthem for his insights and the stimulating discussions that lead to this paper. The last two authors would also like to thank Larry Moss for his help, including for organising the Bloomington workshop at which a draft of this paper was first presented. We are grateful to the two anonymous referees for their comments. The research of the first author was partially supported by the Netherlands Organisation for Scientific Research (NWO), grant number B 62-635, which is herewith gratefully acknowledged. The second author acknowledges the support by the Flemish Fund for Scientific Research in the early stage of this paper and support by the University of Groningen via a Rosalind Franklin research position in the last stage of this paper. The last author was supported by a Marie Curie Early Stage Research fellowship in the project GLoRiClass (MEST-CT-2005-020841).

References

C. E. Alchourrón, P. Gärdenfors, and D. Makinson. On the logic of theory change: Partial meet contraction and revision functions. *Journal of Symbolic Logic*, 50(2):510–530, 1985.

R. Aumann. Backward induction and common knowledge of rationality. *Games and Economic Behavior*, 8:6–19, 1995.

P. Balbiani, A. Baltag, H. van Ditmarsch, A. Herzig, T. Hoshi, and T. de Lima. 'Knowable' as 'known after an announcement'. *The Review of Symbolic Logic*, 1(3):305–334, Oct. 2008. Published online by Cambridge University Press 15 Dec 2008 doi:10.1017/S1755020308080210.

A. Baltag and S. Smets. The logic of conditional doxastic actions. In R. van Rooij and K. R. Apt, editors, *New Perspectives on Games and Interaction*, volume 4 of *Texts in Logic and Games*, pages 9–31. Amsterdam University Press, 2008a.

A. Baltag and S. Smets. A qualitative theory of dynamic interactive belief revision. In G. Bonanno, W. v. d. Hoek, and M. Wooldridge, editors, *Logic and the Foundations of Game and Decision Theory (LOFT 7)*, volume 3 of *Texts in Logic and Games*, pages 9–58. Amsterdam University Press, 2008b.

A. Baltag and S. Smets. Conditional doxastic models: A qualitative approach to dynamic belief revision. *Electronic Notes in Theorerical Computer Science*, 165: 5–21, 2006. ISSN 1571-0661.

A. Baltag, L. S. Moss, and S. Solecki. The logic of public announcements, common knowledge and private suspicions. Technical Report SEN-R9922, Centrum voor Wiskunde en Informatica, 1999.

P. Battigalli. On rationalizability in extensive games. *Journal of Economic Theory*, 74(1):40–61, May 1997.

P. Battigalli and M. Siniscalchi. Hierarchies of conditional beliefs and interactive epistemology in dynamic games. *Journal of Economic Theory*, 88(1): 188–230, September 1999. URL `http://ideas.repec.org/a/eee/jetheo/v88y1999i1p188-230.html`.

P. Battigalli and M. Siniscalchi. Strong belief and forward induction reasoning. *Journal of Economic Theory*, 106(2):356–391, October 2002. URL `http://ideas.repec.org/a/eee/jetheo/v106y2002i2p356-391.html`.

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

J. v. Benthem. Rational dynamics and epistemic logic in games. *International Game Theory Review*, 9(1):13–45, 2007b. (Erratum reprint, 9(2), 377–409).

C. Bicchieri. Self-refuting theories of strategic interaction: a paradox of common knowledge. *Erkenntnis*, 30:69–85, 1989.

K. Binmore. Modeling rational players, part I. *Economics and Philosophy*, 3: 179–214, 1987.

K. Binmore. A note on backward induction. *Games and Economic Behavior*, 17 (1):135–137, November 1996.

L. Blume, A. Brandenburger, and E. Dekel. Lexicographic probabilities and equilibrium refinements. *Econometrica*, 59(1):81–98, January 1991.

O. Board. Dynamic interactive epistemology. *Games and Economic Behavior*, 49: 49–80, 2002.

G. Bonanno. The logic of rational play in games of perfect information. *Economics and Philosophy*, 7:37–65, 1991.

A. Brandenburger. The power of paradox: some recent developments in interactive epistemology. *International Journal of Game Theory*, 35 (4):465–492, April 2007. URL `http://ideas.repec.org/a/spr/jogath/v35y2007i4p465-492.html`.

T. Clausing. Doxastic conditions for backward induction. *Theory and Decision*, 54:315–336, 2003.

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

J. D. Gerbrandy and W. Groeneveld. Reasoning about information change. *Journal of Logic, Language, and Information*, 6:147–169, 1997.

A. Grove. Two modellings for theory change. *Journal of Philosophical Logic*, 17 (2):157–170, 1988.

J. Y. Halpern. Substantive rationality and backward induction. *Games and Economic Behavior*, 37:425–435, 2001.

G. Moore. A reply to my critics. In P. Schilpp, editor, *The Philosophy of G.E. Moore*, volume 4 of *The Library of Living Philosophers*, pages 535–677. Northwestern University, Evanston IL, 1942.

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

J. A. Plaza. Logics of public communications. In M. L. Emrich, M. S. Pfeifer, M. Hadzikadic, and Z. W. Ras, editors, *Proceedings of the 4th International Symposium on Methodologies for Intelligent Systems*, pages 201–216, 1989.

W. Rabinowicz. Grappling with the centipede: defense of backward induction for bi-terminating games. *Philosophy and Economics*, 14:95–126, 1998.

P. Reny. Rationality in extensive form games. *Journal of Economic Perspectives*, 6:92–100, 1992.

A. Rényi. On a new axiomatic theory of probability. *Acta Mathematica Academiae Scientiarum Hungaricae*, 6:285–335, 1955.

R. Rosenthal. Games of perfect information, predatory pricing, and the chain store. *Journal of Economic Theory*, 25:92–100, 1981.

O. Roy. *Thinking before acting: intentions, logic, rational choice*. PhD thesis, ILLC, Amsterdam, 2008.

D. Samet. Hypothetical knowledge and games with perfect information. *Games and Economic Behavior*, 17:230–251, 1996.

R. C. Stalnaker. Knowledge, beliefs and counterfactual reasoning in games. *Economics and Philosophy*, 12:133–163, 1996.

R. C. Stalnaker. Belief revision in games: Forward and backward induction. *Mathematical Social Sciences*, 36:31–56, 1998.

R. C. Stalnaker. On the evaluation of solution concepts. *Theory and Decision*, 37: 49–73, 1994.

R. C. Stalnaker. On logics of knowledge and belief. *Philosophical Studies*, 128: 169–199, 2006.

Learning by Erasing in Dynamic Epistemic Logic

Nina Gierasimczuk
Institute for Logic, Language and Computation, Universiteit van Amsterdam.
n.gierasimczuk@uva.nl

Abstract

This work provides a comparison of learning by erasing (Lange et al. 1996) and iterated epistemic update (Baltag and Moss 2004) as analyzed in dynamic epistemic logic (see e.g. van Ditmarsch et al. 2007). We show that finite identification can be modelled in dynamic epistemic logic and that the elimination process of learning by erasing can be seen as iterated belief-revision modelled in dynamic doxastic logic.

1 Introduction

There have been many formal attempts to grasp the phenomenon of epistemic change. In this paper we will discuss two of them. On the one hand we have the formal learning theory (LT) framework (see e.g. Jain et al. 1999), with its direct implications for analysis of scientific discovery, on the other — belief-revision theory in its interrelation with dynamic epistemic logic (DEL). In learning theory, the classical framework of identification in the limit (Gold 1967) was motivated mostly by the problem of language acquisition. It turned out to be very useful for modelling the process of grammar inference, and found numerous applications in the area of syntax. Initially the idea of identification was unappreciated in semantic considerations, but eventually also this direction has started to be developed resulting in applications to the acquisition of semantics of natural language (Tiede 1999, Costa Florêntio 2002, Gierasimczuk 2007) as well as in modelling the process of scientific inquiry (Kelly 1996). The serious step towards involving more semantics was coupled with the design of model-theoretic learning (Osherson et al. 1997) and its application to belief-revision theory (Martin and Osherson 1998).

Other, very prominent directions that explicitly involve notions of knowledge and belief have been developed in the area of epistemology. First, a precise language to discuss epistemic states of agents has been established in (Hintikka 1962). After that the need of formalizing dynamics of knowledge emerged. The belief-revision AGM framework (Alchourrón et al. 1985) constitutes an attempt to talk about the dynamics of epistemic states. Belief-revision policies thus explained have been successfully modelled in dynamic epistemic logic (see van Ditmarsch et al. 2007) and in the above-mentioned model-theoretic learning (Martin and Osherson 1998).

In the present paper we show how those two important traditions, LT and DEL, can be merged. We explain this connection by joining iterated epistemic update as modelled in DEL with a special case of learning in the limit — learning by erasing (Freivalds and Zeugmann 1996, Lange et al. 1996).

We will proceed according to the following plan. First we explain the ideas of dynamic epistemic logic (DEL) from a strictly semantic point of view. We will also mention an important modification of DEL, namely dynamic doxastic logic (DDL). As we will see the part 'dynamic' in those names refers to the fact that those logics include operators which modify models. With respect to those modifications we discuss the notions of epistemic and doxastic update. In particular, we focus on public announcement as a special case. Then we leave this logical subject and move to briefly recall the basics of formal learning theory in its set-theoretical version. After that the definition of learning by erasing is provided. The last two parts present a way to model finite identification in DEL and learning by erasing in DDL.

2 Dynamic Epistemic Logic. Semantic Perspective

In general, dynamic epistemic logic has been introduced to formalize knowledge change. In this section basic notions of DEL will be provided. The definitions are based on (van Benthem et al. 2007). Let us take *Atom* to be a set of atomic propositions and A — a set of agents.

Definition 2.1 (Epistemic Model)**.** Epistemic model M is a triple

$$\langle W, \{\sim_i\}_{i\in A}, V\rangle\,,$$

where W is a set of possible worlds, for each $i \in A$, $\sim_i \subseteq W \times W$ is an indistinguishability relation and $V : Atom \rightarrow \wp(W)$ is a valuation.

Intuitively, M formalizes the epistemic situation of all agents from A. The indistinguishability relation models their uncertainty about which of the possible worlds is the actual one.

Definition 2.2 (Event Model)**.** An event model E is a triple $\langle S, \{\rightarrow_i\}_{i\in A}, pre\rangle$, where S is a set of worlds, for each $i \in A$, $\rightarrow_i \subseteq S \times S$, and $pre : S \rightarrow Atom$ is a pre-condition function which indicates what *pre-condition* a world has to satisfy to enable the event to take place.

Event model describes the epistemic content of the event. Relation $\rightarrow_i$ directly corresponds to the indistinguishability relation $\sim_i$ of epistemic model.

Definition 2.3 (Product Update)**.** Let M, E be such that $M = \langle W, \{\sim_i\}_{i\in A}, V\rangle$ and $E = \langle S, \{\rightarrow_i\}_{i\in A}, pre\rangle$. The product update $M \otimes E$ is the epistemic model $M' = \langle W', \{\sim'_i\}_{i\in A}, V'\rangle$ such that:

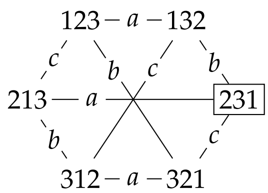
- $W' = \{(w, s) | w \in W, s \in S \text{ and } M, w \models pre(s)\}$,
- $(w, s) \sim_i (w', s')$ iff $w \sim_i w'$ and $s \rightarrow_i s'$,

- $V'((w,s)) = V(w)$.

Definition 2.4 (Public Announcement (Batlag et al. 1998))**.** The public announcement of a formula φ is the event model $E_\varphi = \langle S, \{\rightarrow_i\}_{i\in A}, pre\rangle$, such that $S = \{e\}$ and for each $i \in A$, $e \rightarrow_i e$ and $pre(e) = \varphi$.

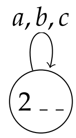
The major result of updating an epistemic model M with public announcement of φ is a submodel of M containing only the states that satisfy φ.

Example 1. Let us take the set of agents $A = \{a \text{ (Anne)}, b \text{ (Bob)}, c \text{ (Carl)}\}$ and the deck of cards consisting of: 1, 2, 3. Each person gets one card. We can represent the situation after dealing as a triple xyz, where x, y, z are cards and the first position in the triple assigns the value to a (Anne), second to b (Bob), etc. For instance, 231 means that Anne has 2, Bob has 3 and Carl has 1. All possible situations after a deal are: 123, 132, 213, 231, 312, 321. We assume that all the players are witnessing the fact of dealing but they do not know the distribution of the cards. The epistemic model M of this situation is illustrated in the figure.



Let us then assume that as a result the actual world is 231. Obviously each player's knowledge does not allow certainty about which is the actual world. In the model the uncertainty of the agent x about the worlds w and w' is symbolized by the following: $w \sim_x w'$ (in the Figure this relation is depicted by two states being joined by a line labeled with x).

Let us now assume that Anne shows her card to all the players publicly, i.e., all the players see her card and all of them know that all of them see it. This event is modelled by $E = (S, \{\rightarrow_i\}_{i\in A}, pre)$, where $S = \{s\}$, for each $x \in A$, $s \rightarrow_x s$ and $pre(s) = 2\_\,\_$ ('Anne has 2').



The public announcement of 'Anne has 2' results in the epistemic situation, which can be presented as $M' = M \otimes E$ (depicted below).

$$213 \overset{a}{\text{———}} \boxed{231}$$

Event E is an example of a public announcement, in this case: 'Anne has 2'. In dynamic epistemic logic the public announcement of φ is represented by '$!\varphi$' and corresponds to the elimination of all those possible worlds that do not satisfy φ. In other words, public announcement works as relativization of the model to those worlds that satisfy the content of the announcement.

3 Dynamic Doxastic Logic

The objective of dynamic doxastic logic (DDL) is to formalize the notion of belief change. This is usually done by introducing preference relations over the possible worlds. Each agent has his own preference relation. Belief of agent a is determined by the set of his most preferred states.

Definition 3.1 (Epistemic Plausibility Model)**.** Let *Atom* be a set of atomic propositions and A — a set of agents. Epistemic plausibility model E is a quadruple: $\langle W, \{\sim_i\}_{i\in A}, \{\leq_i\}_{i\in A}, V\rangle$, where W is a set of possible worlds, for each $i \in A$, $\sim_i \subseteq W \times W$ is an indistinguishability relation, $\leq_i \subseteq W \times W$ is a preference relation and $V : Atom \to \wp(W)$ is a valuation.

Definition 3.2 (Plausibility Event Model)**.** An event model E is a quadruple: $\langle S, \{\to_i\}_{i\in A}, \{\leq_i\}_{i\in A}, pre\rangle$, where S is a set of worlds, for each $i \in A$, $\to_i \subseteq S \times S$, $\leq_i \subseteq S \times S$ and $pre : S \to Atom$ is a pre-condition function.

For completeness' sake we add the definition of priority update.

Definition 3.3 (Priority Update)**.** The priority update works analogously to the epistemic update. The additional condition is for the $\leq_i$ relation:

- for $w \in W$ and $s \in S$, $(w, s) \leq'_i (w', s')$ iff $s <_i s'$, or $s \simeq_i s'$ and $w \leq_i w'$, where $s \simeq_i s'$ iff $s \leq_i s'$ and $s' \leq_i s$.

4 Learning Theory

4.1 Identification in the Limit

Learning theory is concerned with the process of inductive inference (Gold 1967). We can think of it as of a game between Scientist and Nature. In the beginning we have a class of possible worlds together with a class of hypotheses (possible descriptions of worlds). Different hypotheses may describe the same world. We assume that both Scientist and Nature know what all the possibilities are, i.e., they both have access to the initial classes. Nature chooses one of those possible worlds to be the actual one. Scientist has to guess which it is. Scientist receives information about the world in an inductive manner. The stream of data is infinite and contains only and all the elements from the chosen reality. Each time Scientist receives a piece of information he answers with one of the hypotheses from the initial class. We say that Scientist identifies Nature's choice in the limit if after some finite number of guesses his answers stabilize on a correct hypothesis. Moreover, to discuss more general identifiability, we require that the same is true for all the possible worlds from the initial class, i.e., regardless of which element from the class is chosen by Nature to be true, Scientist can identify it in the limit on the basis of data about the actual world.

To formalize this simple setting we need to make the notion of *stream of data* clear. In learning theory such streams are often called 'environments'[1].

Let us consider E — the set of all computably enumerable sets. Let $C \subseteq E$ be some class of c.e. sets. For each S in C we consider Turing machines h_n which generate S and in such a case we say that n is an index of S. The Turing

[1]We are concerned here only with sequences of positive information (texts).

machines will function as the conjectures that Scientist makes. It is well-known that each S has infinitely many indices. Let us take I_S to be the set of all indices of the set S, i.e, $I_S = \{n | h_n \textit{ generates } S\}$.

Definition 4.1 (Environment)**.** By environment of S, ε, we mean any infinite sequence of elements from S such that:

1. ε enumerates all the elements from S;
2. ε enumerates only the elements from S;
3. ε allows repetitions.

Definition 4.2 (Notation)**.** We will use the following notation:

- ε_n is the n-th element of ε;
- $\varepsilon|n$ is a sequence $(\varepsilon_0, \varepsilon_1, \ldots, \varepsilon_{n-1})$;
- SEQ denotes the set of all finite initial segments of all environments;
- $set(\varepsilon)$ is a set of elements that occur in ε;
- h_n will refer to a hypothesis, i.e., a finite description of a set, a Turing machine generating S;
- L is a learning function — a map from finite data sequences to indices of hypotheses, $L : SEQ \to I_{H_C}$.

The structure of the identifiability in the limit can be formulated by the following chain of definitions:

Definition 4.3 (Identification in the limit, LIM)**.** We say that a learning function L:

1. identifies $S \in C$ in the limit on ε iff there is a number k, such that for co-finitely many m, $L(\varepsilon|m) = k$ and $k \in I_S$;
2. identifies $S \in C$ in the limit iff it identifies S in the limit on every ε for S;
3. identifies C in the limit iff it identifies in the limit every $S \in C$.

The notion of identifiability can be strengthened in various ways. One radical case is to introduce a finiteness condition for identification.

Definition 4.4 (Finite identification, FIN)**.** We say that a learning function L:

1. finitely identifies $S \in C$ on ε iff, when successively fed ε, at some point L outputs a single k, such that $k \in I_S$, and stops;
2. finitely identifies $S \in C$ iff it finitely identifies S on every ε for S;
3. finitely identifies C iff it finitely identifies every $S \in C$.

4.2 Learning by Erasing

Learning by erasing (Lange et al. 1996, Freivalds et al. 2002) is an epistemologically intuitive modification of identification in the limit. Very often the cognitive process of converging to a correct conclusion consists of eliminating those possibilities that are falsified during the inductive inquiry. Accordingly, in the formal model the outputs of the learning function are negative, i.e., the function each time eliminates a hypothesis, instead of explicitly guessing one that is supposed to be correct. A special case of learning by erasing is co-learning (Freivalds and Zeugmann 1996). The set $S \in C$ is co-learnable iff there is a function which stabilizes by eliminating all indices from I_{H_C} except just one from I_S. The difference between this approach and the usual identification is in the interpretation of the positive guess of the learning function. In learning by erasing there is always some ordering of the initial hypothesis space. This allows to interpret the actual positive guess of the learning-by-erasing function to be the least hypothesis (in a given ordering) not yet eliminated.

Let us give now the two definitions that explain the notion of learning by erasing.

Definition 4.5 (Function Stabilization)**.** In learning by erasing we say that a function stabilizes to number k on environment ε if and only if for co-finitely many $n \in \mathbb{N}$:

$$k = min\{\mathbb{N} - \{L(\varepsilon|0), \ldots, L(\varepsilon|n)\}\}.$$

Definition 4.6 (Learning by Erasing, E-learning)**.** We say that a learning function, L:

1. learns $S \in C$ by erasing on ε iff L stabilizes to k on ε and $k \in I_S$;
2. learns $S \in C$ by erasing iff it learns by erasing S from every ε for S;
3. learns C by erasing iff it learns by erasing every $S \in C$.

A variety of additional conditions for learning can be defined. Let us mention the following conditions on e-learning function L (Lange et al. 1996).

1. L erases all but one, correct hypothesis (co-learning, e-ALL);
2. L erases only hypotheses that are incorrect (e-SUB);
3. L erases exactly all hypotheses that are incorrect (e-EQ);
4. L erases all hypotheses that are incorrect but may also erase some that are correct (e-SUPER);

Let us cite two theorems (Lange et al. 1996) that establish the relationships between various types of learning: e-learning, finite identifiability and identifiability in the limit.

Theorem 1. *FIN* $\subset$ *e-EQ* $\subset$ *e-SUB* $\subset$ *LIM*

Theorem 2. e-ALL, e-SUPER = *LIM*

5 Finite Identification in DEL

The word 'learning' is used in epistemology to cover a variety of epistemic processes. One of them is the epistemic update in the form of one-step *learning that* φ, followed by a direct modification of the set of beliefs, as we have seen in sections 2 and 3. In the learning-theoretic setting the incoming information is of a different nature than the actual thing being learned. This feature has an important consequence for modelling learning in DEL. We are forced to provide two-sorted models, with one sort for pieces of incoming information and another for the hypotheses. To establish a bridge between those two different ontologies we treat a hypothesis as the set of events that it predicts, e.g., if we take a hypothesis h to be 'There are all natural numbers except 3' it predicts that the environment will enumerate all the natural numbers except 3.

The possible worlds in our epistemic model are identified with hypotheses. Unlike in the classical DEL approach, the event models are announcements of data corresponding to elements of the sets being learned, and not hypotheses themselves.

A further difference is in the number of agents. In sections 2 and 3 we provided definitions for multi-agent epistemic cases. Although science as well as learning seem to be at least a two-player game, in the present paper we are concerned only with the role of Scientist (Learner). By implication, we assume Nature (Teacher) to be an objective machine that makes an arbitrary choice and gives out random data, she does not have any particular strategy, is neither helping the learner, nor obstructing his attempts to identify a correct hypothesis. We recognize the possibility and potential of analyzing two or more agents in the contexts of inductive inference. However, for the sake of simplicity our DEL and DDL models are going to account only for one agent.

Let us again fix C to be a class of sets, and for each $S_n \in C$ we consider h_n to be a hypothesis that describes S_n. In learning by erasing we can take the initial epistemic model to represent the background knowledge of Scientist together with his uncertainty about which world is the actual one. Let us take the initial epistemic frame to be

$$M = \langle H_C, \sim \rangle ,$$

where H_C is a possibly infinite[2] set of worlds (hypotheses that are considered possible) and $\sim \subseteq H_C \times H_C$ is an uncertainty relation for Scientist. Since we assume that the initial hypothesis space is arbitrary, we also do not require any particular preference of the scientist over H_C. Hence, we take the relation $\sim$ to be a universal, equivalence binary relation over H_C. The initial epistemic state of the Scientist is depicted in Figure 1. This model corresponds to the starting point of the scientific discovery process. Each world represents a hypothesis from the initial set determined by the background knowledge. In the beginning Scientist considers all of them possible. The model also reflects the fact that Scientist is given the class of hypotheses H_C. In other words he knows what the alternatives are.

Next, Nature decides on some state of the world by choosing one possibility from C. Let us assume that as a result h_3 correctly describes the chosen world.

[2]We can effectively deal with the epistemic update and identification in infinite domains by using special enumeration strategies (for explanation and examples see Gierasimczuk 2009).

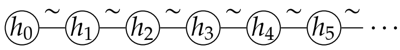


Figure 1: Initial epistemic model

Then, she decides on some particular environment ε, of the elements from the world. We picture this enumeration in Figure 2 below.

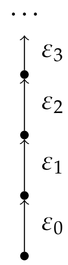


Figure 2: Environment ε consistent with h_3

The sequence ε is successively given to Scientist. Let us focus now on the first step of the procedure. We have the uncertainty range of Scientist, it runs through the whole set of hypotheses H_C. A piece of data ε_0 is given to Scientist. This fact can be represented by the event model $E_0 = \langle \{e\}, \rightarrow, pre \rangle$, where $e \rightarrow e$ and $pre(e) = \varepsilon_0$ (see Figure 3).

Figure 3: Event model E_0 of the announcement of ε_0

Scientist, when confronted with the announcement of ε_0 updates his epistemic state accordingly. We will represent the process formally by the product update $M \otimes E_0$. The result of the product update is again an epistemic model $M' = \langle H_C{}', \sim' \rangle$, where:

1. $H_C{}' = \{(h_n, e) | h_n \in H_C \; \& \; pre(e) \in S_n)\}$;

2. $\sim' = \sim |H_C{}'$.

We use here event models similar in spirit to those of public announcement (Batlag et al. 1998). They consist in only one state with a pre-condition determined by the piece of data that is given. In Figure 4 Scientist's confrontation with ε_0 is depicted.

Scientist tests each hypothesis with ε_0. If a hypothesis is consistent with it, it remains as a possibility, if it is not consistent, it is eliminated (see figure 5). Let us assume that ε_0 is not consistent with h_2.

This epistemic update can be iterated infinitely many times along ε resulting in an infinite sequence of models which according to the lines of DEL can be called ε-generated epistemic model (see e.g. van Benthem et al. 2007).

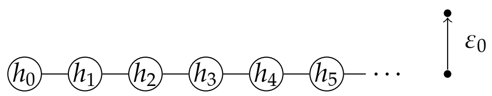


Figure 4: Confrontation with data

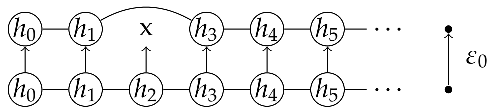


Figure 5: Epistemic update

Definition 5.1 (Generated Epistemic Model)**.** The generated epistemic model $(M)^{\varepsilon}$, with $\varepsilon = \varepsilon_0, \varepsilon_1, \varepsilon_2, \ldots$, is the result of epistemic update $M \otimes E_0 \otimes E_1 \otimes E_2 \otimes \ldots$, where for each n, the event E_n corresponds to the announcement of ε_n.

Let us now see a simple example of finite identification of a single hypothesis.

Example 2. Let us take $H_C = \{h_0, h_1, h_2\}$, such that $h_n = \{0, ..., n\}$. Nature makes her choice regarding what the world is like. We assume that as a result h_2 holds. Then, Nature chooses an enumeration $\varepsilon = 0, 1, 0, 2, 1, \ldots$. After the first piece of data, 0, the uncertainty range of Scientist includes the whole H_C. After the second, 1, Scientist eliminates h_0 since it does not contain the event 1 and now he hesitates between h_1 and h_2. The third piece, 0, does not change anything, however the next one, 2, eliminates h_1. Uncertainty is eliminated as well. He knows that the only hypothesis that can be true is h_2. Therefore, we can say that he learned it conclusively, with certainty.

The above discussion suggests the following thesis.

Thesis 1. *Finite identifiability can be modelled within the DEL framework, using:*

- *epistemic states for hypotheses;*
- *infinite sequences of announcements for environments;*
- *epistemic update for the progress in eliminating uncertainty over hypothesis space.*

Scientist succeeds in finite identification of S from ε if and only if there is a finite initial segment of ε, $\varepsilon|n$, such that the domain of the $\varepsilon|n$-generated model contains only one hypothesis h_k and $k \in I_S$. In other words, there is a finite step of the iterated epistemic update along ε, that eliminates Scientist's uncertainty.

6 Learning by Erasing in DDL

From Scientist's point of view the process of learning has a few components that are very important in logical modelling. The first is of course the current conjecture — a hypothesis that is considered appropriate in a given step of the procedure. The second is the set of those hypotheses that were used in the past and have already been discarded. The third part is the set of hypotheses that are still considered to be possible, but for some reasons less probable than the chosen one.

Let us consider the following example of a learning scenario, in which the uncertainty is never eliminated.

Example 3. As you probably observed, in the Example 2 Scientist was very lucky. Let us assume for a moment that nature had chosen h_1, and had fixed the enumeration $\varepsilon = 0, 1, 0, 1, 1, 1, 1, \ldots$ In this case Scientist's uncertainty can never be eliminated.[3]

This example indicates that the central element of the identification in the limit model is the unavoidable presence of uncertainty. The limiting framework allows however the introduction of some kind of *operational* knowledge, which is uncertainty-proof.

To model the algorithmic nature of the learning process that includes actual guess and other not-yet-eliminated possibilities, we enrich the epistemic model with some preference relation $\leq: H_C \times H_C$. The relation $\leq$ represents some preference over the set of hypotheses, e.g., if Scientist is an occamist, the preference would be defined according to the simplicity of hypotheses. In the initial epistemic state the uncertainty of the scientist again ranges over all of H_C. This time however the class is ordered and Scientist current belief is the most preferred hypothesis. Therefore, we consider the initial epistemic state of Scientist to be:

$$M = \langle H_C, \sim, \leq \rangle .$$

The procedure of erasing hypotheses that are inconsistent with successively incoming data is the same as in the previous section. This time however let us introduce the current-guess state which is interpreted as the actual guess of the Scientist. It is always the one that is most preferred — the smallest one according to $\leq$. In doxastic logic a set of most preferred hypotheses is almost invariably interpreted as the one that the agent *believes* in. Let us go back to Example 2, where Nature chose a world consistent with h_1. After seeing 1 and eliminating h_0, Scientist's attention focuses on h_1, then h_1 is his current belief. It is the most preferred hypothesis, and as such it can be reiterated as long as it is consistent with ε. In this particular case, since Nature chose a world consistent with h_1, it will never be contradicted, so Scientist will always be uncertain between h_1 and h_2. However, his preference directs him to believe in the correct hypothesis, without him being aware of the correctness. Therefore, we claim the following.

Thesis 2. *Learning by erasing can be modelled within the DDL framework, using:*

- *epistemic states for hypotheses;*
- *infinite sequences of announcements for environments;*
- *epistemic update for the progress in eliminating uncertainty over the hypothesis space;*

[3]As we are interested here in learning by erasing, we assume a suitable underlying ordering of hypothesis space. In this case it is: h_0, h_1, h_2. However, note that this type of identification is not order-independent. If the initial ordering was: h_0, h_2, h_1, then Scientist would not stabilize on the correct hypothesis.

- *preference relation for the underlying hypothesis space;*
- *in each step of the procedure, the most preferred hypothesis for the actual positive guess of the learning function.*

Scientist learns S by erasing from ε if and only if there is n such that for every $m > n$, the most preferred state of the domain of the $\varepsilon|m$-generated epistemic model is h_k, and $k \in I_S$.

7 Conclusions and Further Work

In this paper we argued that the process of inductive inference can be modelled in dynamic epistemic logic and dynamic doxastic logic. To support our claim we provided a translation of the components of learning into a two-sorted semantics for DEL and DDL. In particular, we see DEL as an appropriate framework to analyze the notion of finite identifiability. Learning by erasing, a special case of identifiability in the limit, is based on the existence of an underlying ordering of hypothesis space. Therefore, in logical modelling it requires adding to the epistemic model a preference relation over possible worlds. This indicates that it should be formalized in DDL, where the preference relation is a standard element of any model.

The above-presented conceptual work has many implications and possible continuations. After establishing a correspondence on the semantic level, it is possible to formulate axioms of epistemic logic for inductive inference. We find this project promising and potentially fruitful for both DEL and LT. Moreover, modal analysis of the process of learning can be continued in the following directions:

- formulating LT theorems as validities in epistemic and temporal logic;
- analyzing the inductive inference process in game-theoretical terms, and discussing strategies for learning and teaching;
- studying the notion of non-introspective operational knowledge and uncertainty that are involved in the process of inductive inference;
- comparing formal learning theory and belief-revision theory in a systematic way.

References

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(2):510–530, 1985.

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

A. Batlag, L. S. Moss, and S. Solecki. The logic of public announcements and common knowledge and private suspicions. In *Proc. TARK'98*, pages 43–56, San Francisco, CA, USA, 1998. Morgan Kaufmann Publishers Inc.

J. van Benthem, J. Gerbrandy, and E. Pacuit. Merging frameworks for interaction: DEL and ETL. In *Proc. TARK'07*, pages 72–81, New York, NY, USA, 2007. ACM Press.

C. Costa Florêntio. Learning generalized quantifiers. In *Proc. 7th ESSLLI Student Session*, 2002.

R. Freivalds and T. Zeugmann. Co–learning of recursive languages from positive data. In *Perspectives of System Informatics, Proc. 2nd Andrei Ershov Memorial Conference*, volume 1181 of *LNCS*, pages 122–133. Springer, 1996.

R. Freivalds, M. Karpinski, C. H. Smith, and R. Wiehagen. Learning by the process of elimination. *Information and Computation*, 176(1)(1):37–50, 2002.

N. Gierasimczuk. The problem of learning the semantics of quantifiers. In *Logic, Language, and Computation, 6th TbiLLC 2005*, volume 4363 of *LNAI*, pages 117–126. Springer, 2007.

N. Gierasimczuk. Identification through inductive verification. In *7th TbiLLC 2007*, volume 5422 of *LNAI*, pages 193–205. Springer, 2009.

E. M. Gold. Language identification in the limit. *Information and Control*, 10: 447–474, 1967.

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

S. Jain, D. Osherson, J. S. Royer, and A. Sharma. *Systems that Learn*. MIT Press, Chicago, 1999.

K. Kelly. *The Logic of Reliable Inquiry*. Oxford University Press, Oxford, 1996.

S. Lange, R. Wiehagen, and T. Zeugmann. Learning by erasing. In *Proc. 7th Int. Workshop on Algorithmic Learning Theory*, volume 1160 of *LNAI*, pages 228–241. Springer-Verlag, 1996.

E. Martin and D. Osherson. *Elements of Scientific Inquiry*. MIT Press, Cambridge, 1998.

D. Osherson, D. de Jongh, E. Martin, and S. Weinstein. Formal learning theory. In J. van Benthem and A. Ter Meulen, editors, *Handbook of Logic and Language*, pages 737–775. MIT Press, Cambridge, 1997.

H. J. Tiede. Identifiability in the limit of context-free generalized quantifiers. *Journal of Language and Computation*, 1:93–102, 1999.

H. van Ditmarsch, W. van der Hoek, and B. Kooi. *Dynamic Epistemic Logic*. (Synthese Library). Springer, 2007.

Multi-agent belief dynamics: bridges between dynamic doxastic and doxastic temporal logics

Johan van Benthem and Cédric Dégremont

Institute for Logic, Language and Computation, Universiteit van Amsterdam and Department of Philosophy, Stanford University;
Institute for Logic, Language and Computation, Universiteit van Amsterdam.
johan.vanbenthem@uva.nl, cedric.uva@gmail.com

Analyzing the behavior of agents in a dynamic environment requires describing the evolution of their knowledge as they receive new information. Moreover agents entertain beliefs that need to be revised after learning new facts. I might be confident that I will find the shop open, but once I found it closed, I should not crash but rather make a decision on the basis of new consistent beliefs. Such beliefs and information may concern ground-level facts, but also beliefs about other agents. I might be a priori confident that the price of my shares will rise, but if I learn that the market is rather pessimistic (say because the shares fell by 10%), this information should change my higher-order beliefs about what other agents believe.

Tools from modal logic have been successfully applied to analyze knowledge dynamics in multi-agent contexts. Among these, Temporal Epistemic Logic Parikh and Ramanujam (2003), Fagin et al. (1995)'s Interpreted Systems, and Dynamic Epistemic Logic Baltag et al. (1998) have been particularly fruitful. A recent line of research Benthem and Pacuit (2006), Benthem et al. (2007; 2008) compares these alternative frameworks, and Benthem et al. (2007) presents a representation theorem that shows under which conditions a temporal model can be represented as a dynamic one. Thanks to this link, the two languages also become comparable, and one can merge ideas: for example, a new line of research explores the introduction of protocols into the logic of public announcements PAL, as a way of modeling informational processes (see Benthem et al. (2008)).

To the best of our knowledge, there are no similar results yet for multi-agent belief revision. One reason is that dynamic logics of belief revision have only been well-understood recently. But right now, there is work on both dynamic doxastic logics Benthem (2007), Baltag and Smets (2006) and on temporal frameworks for belief revision, with Bonanno (2006) as a recent example. The exact connection between these two frameworks is not quite like the case of epistemic update. In this paper we make things clear, by viewing belief revision as priority update over plausibility *pre-orders*. This correspondence allows for similar language links as in the knowledge case, with similar precise benefits.

We start in the next section with background about earlier results and basic terminology. In section 2 we give the main new definitions needed in the paper. Section 3 presents the key temporal doxastic properties that we will work with. In section 4 we state and prove our main result linking the temporal and the dynamic frameworks, first for the special case of *total* pre-orders and then in general. We also discuss some variations and extensions. In Section 5 we introduce formal languages, providing an axiomatization for our crucial properties, and discussing some related definability issues. In Section we introduce the idea of protocols-based belief revision and give a Henking-style completeness proof for an axiomatization of the class of doxastic temporal models generated by protocols-based lexicographic upgrade. We state our conclusions and mention some further applications and open problems in the last section.

1 Introduction: background results

Epistemic temporal trees and dynamic logics with product update are complementary ways of looking at multi-agent information flow. Representation theorems linking both approaches were proposed for the first time in Benthem (2001). A nice presentation of these early results can be found in (Liu 2008, ch5). We start with one recent version from Benthem et al. (2008), referring the reader to Benthem et al. (2008) for a proof, as well as generalizations and variations.

Definition 1.1 (Epistemic and Event Models, Product Update)**.**

- An *epistemic model* $\mathcal{M}$ is of the form $\langle W, (\sim_i)_{i\in N}, V\rangle$ where $W \neq \emptyset$, for each $i \in N$, $\sim_i$ is a relation on W, and $V : Prop \to \wp(W)$.
- An *event model* $\epsilon = \langle E, (\sim_i)_{i\in N}, \mathtt{pre}\rangle$ has $E \neq \emptyset$, and for each $i \in N$, $\sim_i$ is a relation on W. Finally, there is a precondition map $\mathtt{pre} : E \to \mathcal{L}_{EL}$, where $\mathcal{L}_{EL}$ is the usual language of epistemic logic.
- The *product update* $\mathcal{M} \otimes \epsilon$ of an epistemic model $\mathcal{M} = \langle W, (\sim_i')_{i\in N}, V\rangle$ with an event model $\epsilon = \langle E, (\sim_i)_{i\in N}, \mathtt{pre}\rangle$, is the model whose worlds are pairs (w, e) with the world w satisfying the precondition of the event e, and accessibilities defined as:

$$(w, e) \sim_i' (w', e') \text{ iff } e \sim_i e', w \sim_i w'$$

Intuitively epistemic models describe what agents currently know while the product update describe the new multi-agent epistemic situation after some epistemic event has taken place. Nice intuitive examples are in Baltag and Moss (2004).

Next we turn to the epistemic temporal models introduced by Parikh and Ramanujam (2003). In what follows, Σ^* is the set of finite sequences on any set Σ, which naturally forms a branching 'tree'.

Definition 1.2 (Epistemic Temporal Models)**.** An *epistemic temporal model* (*ETL* model for short) $\mathcal{H}$ is of the form $\langle \Sigma, H, (\sim_i)_{i\in N}, V\rangle$ where Σ is a finite set of events, $H \subseteq \Sigma^*$ and H is closed under non-empty prefixes. For each $i \in N$, $\sim_i$ is a relation on H, and there is a valuation $V : Prop \rightarrow \wp(H)$.

The following epistemic temporal properties drive Benthem et al. (2008)'s main theorem.

Definition 1.3. Let $\mathcal{H} = \langle \Sigma, H, (\sim_i)_{i\in N}, V\rangle$ be an *ETL* model. $\mathcal{H}$ satisfies:

- **Propositional stability** whenever h is a finite prefix of h', then h and h' satisfies the same proposition letters.
- **Synchronicity** iwhenever $h \sim h'$, we have $len(h) = len(h')$.

Let $\sim^*$ be the reflexive transitive closure of the relation $\bigcup_{i\in N} \sim_i$:

- **Local Bisimulation Invariance** whenever $h \sim^* h'$ and h and h' are epistemically bisimilar[1], we have $h'e \in H$ iff $he \in H$.
- **Perfect Recall** whenever $ha \sim_i h'b$, we also have $h \sim_i h'$.
- **Local No Miracles** whenever $ga \sim g'b$ and $g \sim^* h \sim h'$, then for every $h'a, hb \in H$, we also have $h'a \sim hb$.

These properties describe the idealized epistemic agents needed in:

Theorem 1.1 (van Benthem et al. Benthem et al. (2008))**.** *Let $\mathcal{H}$ be an ETL model, $\mathcal{M}$ an epistemic model, and the 'protocol' P a set of finite sequences of pointed events models closed under prefixes. We write $\otimes$ for* product *update. Let $Forest(M, P) = \bigcup_{\vec{\epsilon}\in P} M \otimes \vec{\epsilon}$ be the 'epistemic forest generated by' $\mathcal{M}$ and sequential application of the events in P.* [2] *The following are equivalent:*

- *$\mathcal{H}$ is isomorphic to $Forest(M, P)$.*
- *$\mathcal{H}$ satisfies propositional stability, synchronicity, local bisimulation invariance, Perfect Recall, and Local No Miracles.*

Thus, epistemic temporal conditions describing idealized epistemic agents characterize just those trees that arise from performing iterated product update governed by some protocol. Benthem et al. (2008) and (Liu 2008, ch5) have details.

Our paper extends this analysis to the richer case of belief revision, where plausibility orders of agents evolve as they observe possibly surprising events. First we prove two main results (in Section 4 and Section 5 respectively), with variations and extensions:

Theorem. *Let $\mathcal{H}$ be a doxastic temporal model, $\mathcal{M}$ a plausibility model, $\vec{\epsilon}$ a sequence of event models, and $\otimes$* priority *update. The following are equivalent, where the notions will of course be defined later:*

1. *$\mathcal{H}$ is isomorphic to the forest generated by $\mathcal{M} \otimes \vec{\epsilon}$*

[1] The reader is referred to subsection 3.1 for a definition of bisimulation invariance.
[2] For a more precise definition of this notion, see Section 2 below.

2. *$\mathcal{H}$ satisfies propositional stability, synchronicity, invariance for bisimulation, as well as principles of Preference Propagation, Preference Revelation and Accommodation.*

Theorem. *Preference Propagation, Preference Revelation and Accommodation are definable in an extended doxastic modal language.*

We then consider protocol-based belief revision and prove completeness of an axiomatization of the class of doxastic temporal trees generated by protocols-based lexicographic upgrade. Usually in belief revision any information can received. Adding a protocol corresponds to the intuition is that the incoming signal might be constrained by some external rules and by physical laws. To be precise in Section 6 we prove the following theorem:

Theorem (Completeness). `Stbr` *is sound and strongly complete with respect to the class of DoTL models generated by a sequence of protocol-based lexicographic upgrade.*

2 Definitions

We now turn to the definitions needed for the simplest version of our main representation theorem, postponing matching formal languages to Section 5. Let $N = \{1, \ldots, n\}$ be a finite set of agents.

2.1 Plausibility models, event models and priority update

As for the epistemic case, we first introduce static models that encode the current prior (conditional) beliefs of agents. These carry a pre-order $\leq$ between worlds encoding a plausibility relation. Often this relation is taken to be total, but when we think of elicited beliefs as *multi-criteria decisions*, a pre-order allowing for incomparable situations may be all we get Eliaz and Ok (2006). We will therefore assume reflexivity and transitivity, but not totality.

As for notation: we write $a \simeq b$ ('indifference') if $a \leq b$ and $b \leq a$, and $a < b$ if $a \leq b$ and $b \nleq a$.

The following definition strongly relates to the models introduced in Board (2004), Ditmarsch (2005), Baltag and Smets (2006).

Definition 2.1 (Doxastic Plausibility Models). A *doxastic plausibility model* $\mathcal{M} = \langle W, (\leq_i)_{i \in N}, V \rangle$ has $W \neq \emptyset$, for each $i \in N$, $\leq_i$ is a pre-order on W, and $V : Prop \to \wp(W)$.

There are at least two ways to think and work with the preceding models. Either they can be taken to encode the beliefs of the agents (including their conditional belief), or they can be taken to encode the *prior* beliefs of the agent and should be combined with an epistemic relation, taking the minimal elements of each information partition to encode the posterior belief of the agents. We don't decide between the two approaches and discuss both throughout the paper.

We now consider how such models evolve as agents observe events.

Definition 2.2 (Plausibility Event Model)**.** A *plausibility event model* (event model, for short) ϵ is a tuple $\langle E, (\preceq_i)_{i \in N}, \texttt{pre} \rangle$ with $E \neq \emptyset$, each $\preceq_i$ a pre-order on E, and $\texttt{pre} : E \to \mathcal{L}$, where $\mathcal{L}$ is a doxastic language. [3]

The reader should note immediately that a plausibility event model content way more structure than a simple formula. The contrast with standard belief revision is that the signal is really what triggers the revision. Strong signals will alter our current belief in a strong manner while weaker signals or signals coming from less trusted sources might keep them much less changed. Classical belief revision theories consider different ways agents can update their belief with an incoming formula. By doing so they take the type of update to be determined by the kind of agents. Different agents, different belief update style. The current approach consider this approach philosophically implausible: the same agent can update in different ways depending on the signals she received. In this sense there are no choice to be made between different update rule. The last event come with all the instructions as how prior beliefs should be overridden. The following is therefore the natural choice of update rule:

Definition 2.1 (Priority Update; Baltag and Smets (2006))**.** Priority update *of a plausibility model* $\mathcal{M} = \langle W, (\preceq_i)_{i \in N}, V \rangle$ *and an event model* $\epsilon = \langle E, (\preceq_i)_{i \in N}, \texttt{pre} \rangle$ *is the plausibility model* $\mathcal{M} \otimes \epsilon = \langle W', (\preceq'_i)_{i \in N}, V' \rangle$ *defined as follows:*

- $W' = \{(w, e) \in W \times E \mid \mathcal{M}, w \Vdash \texttt{pre}(e)\}$
- $(w, e) \preceq'_i (w', e')$ *iff either* $e \prec_i e'$, *or* $e \simeq_i e'$ *and* $w \preceq_i w'$
- $V'((s, e)) = V(s)$

The idea behind the following definition is clear. Priority is given to the current event plausibility order. If the agent is indifferent, the old plausibility order applies. More motivation can be found in Baltag and Smets (2006), Benthem (2008). But we insist that the reader don't think that the preceding rule allows only very strong signals. In fact many of the classical update rule can be simulated by a suitable event model. Example of such simulations can be found in Baltag and Smets (2008).

2.2 Doxastic Temporal Models

We now turn to the global temporal perspective on multi-agent belief revision. Our models can be seen as a natural counterpart doxastic counterpart to Parikh and Ramanujam (2003)'s *ETL* models (and equivalent models such as (Fagin et al. 1995, ch. 4)' Interpreted Systems, cf. Pacuit (2007)). And they are also related to Bonanno (2006; 2007)'s temporal doxastic models.

Definition 2.3 (Doxastic Temporal Models)**.** A *doxastic temporal model* (*DoTL* model for short) $\mathcal{H}$ is of the form $\langle \Sigma, H, (\preceq_i)_{i \in N}, V \rangle$, where Σ is a finite set of events, $H \subseteq \Sigma^*$ is closed under non-empty prefixes, for each $i \in N$, $\preceq_i$ is a pre-order on H, and $V : Prop \to \wp H$.

Our task is to identify just when a doxastic temporal model is isomorphic to the 'forest' generated by a sequence of priority updates:

[3] This definition is incomplete without specifying the relevant language, but all that follows can be understood by considering the formal language as a 'parameter'.

2.3 Dynamic Models Generate Doxastic Temporal Models

First of all let us see how a sequence of priority updates of plausibility models generate doxastic temporal forests.

Definition 2.2 (*DoTL* model generated by a sequence of updates)**.** *Each initial plausibility model* $\mathcal{M} = \langle W, (\leq_i)_{i\in N}, V\rangle$ *and sequence of event models* $\epsilon_j = \langle E_j, (\leq_i^j)_{i\in N}, \mathtt{pre}_j\rangle$ *yields a* generated *DoTL* plausibility model $\langle \Sigma, H, (\leq_i)_{i\in N}, \mathbf{V}\rangle$ *as follows:*

- *Let* $\Sigma := \bigcup_{i=1}^{m} e_i$.
- *Let* $H_1 := W$ *and for any* $1 < n \leq m$ *let* $H_{n+1} := \{(we_1 \dots e_n) | (we_1 \dots e_{n-1}) \in H_n$ and $\mathcal{M} \otimes \epsilon_1 \otimes \dots \otimes \epsilon_{n-1} \Vdash \mathtt{pre}_n(e_n)\}$. *Finally let* $H = \bigcup_{1\leq k\leq m} H_k$.
- *If* $h, h' \in H_1$, *then* $h \leq_i h'$ *iff* $h \leq_i^{\mathcal{M}} h'$.
- *For* $1 < k \leq m$, $he \leq_i h'e'$ *iff 1.* $he, h'e' \in H_k$, *and 2. either* $e <_i^k e'$, *or* $e \simeq_i^k e'$ *and* $h \leq_i h'$.
- *Let* $wh \in \mathbf{V}(p)$ *iff* $w \in V(p)$.

Now come the key doxastic temporal properties of our idealized agents.

3 Crucial Frame Properties for Priority Updaters

We first introduce the notion of bisimulation, modulo a choice of language.

3.1 Bisimulation Invariance

Definition 3.1 ($\leq$-Bisimulation)**.** Let $\mathcal{H}$ and $\mathcal{H}'$ be two *DoTL* plausibility models $\langle H, (\leq_1, \dots, \leq_n)$
$, V\rangle$ and $\langle H', (\leq_1', \dots, \leq_n'), V'\rangle$ (for simplicity, assume they are based on the same alphabet Σ). A relation $Z \subseteq H \times H'$ is a $\leq$-Bisimulation if, for all $h \in H$, $h' \in H'$, and all $\leq_i$ in $(\leq_1, \dots, \leq_n)$,

(prop) h and h' satisfy the same proposition letters,

(zig) If hZh' and $h \leq_i j$, then there exists $j' \in H'$ such that jZj' and $h' \leq_i' j'$,

(zag) If hZh' and $h' \leq_i' j'$, then there exists $j \in H$ such that jZj' and $h \leq_i j$.

If Z is a $\leq_n$-bisimulation and hZh', we call h and h' are $\leq$-bisimilar.

Definition 3.2 ($\leq$-Bisimulation Invariance)**.** A *DoTL* model $\mathcal{H}$ satisfies $\leq$-*bisimulation invariance* if, for all $\leq$-bisimilar histories $h, h' \in H$, and all events e, $h'e \in H$ iff $he \in H$.

3.2 Agent-Oriented Frame Properties

In the following we drop agent labels and the "for each $i \in N$" for the sake of clarity. Also, when we write ha we will always assume that $ha \in H$. We will make heavy use of the following notion:

Definition 3.3 (Accommodating Events)**.** Two events $a, b \in \Sigma$ are *accommodating* if, for all $ga, g'b$, ($g \leq g' \leftrightarrow ga \leq g'b$) and similarly for $\geq$, i.e., a, b preserve and anti-preserve plausibility.

Definition 3.4. Let $\mathcal{H} = \langle \Sigma, H, (\leq_i)_{i \in N}, V \rangle$ be a *DoTL* model. $\mathcal{H}$ satisfies:

- **Propositional stability** whenever h is a finite prefix of h', then h and h' satisfy the same proposition letters.
- **Synchronicity** whenever $h \leq h'$, we have $len(h) = len(h')$.

The following three properties trace the belief revising behavior of agents in doxastic trees.

- **Preference Propagation** whenever $ja \leq j'b$, then $h \leq h'$ implies $ha \leq h'b$.
- **Preference Revelation** whenever $jb \leq j'a$, then $ha \leq h'b$ implies $h \leq h'$.
- **Accommodation** a and b are *accommodating* whenever both $ja \leq j'b$ and $ha \nleq h'b$.

These properties - and in particular the last one - are somewhat trickier than in the epistemic case, reflecting the peculiarities of priority update in settings where incomparability is allowed. But we do have:

Fact 3.1. *If $\leq$ is a* total *pre-order and $\mathcal{H}$ satisfies Preference Propagation and Preference Revelation, then $\mathcal{H}$ satisfies Accommodation.*

Proof. From left to right. Assume that $g \leq g'$ and $ja \leq j'b$. By Preference Propagation, $ga \leq g'b$. Now assume that $ha \nleq h'b$. Then by totality, $h'b \leq ha$. Since $g \leq g'$, it follows by Preference Propagation that $gb \leq g'a$.

From right to left, assume that $gb \leq g'a$ and that $ja \leq j'b$. It follows by Preference Revelation that $g \leq g'$. Now assume that $ga \leq g'b$ (1) and $ha \nleq h'b$ (2). From (2), it follows by totality that $h'b \leq ha$ (3). But if (3) and (1), then by Preference Revelation we have $g \leq g'$. □

We can also prove a partial converse without assuming totality:

Fact 3.2. *If $\mathcal{H}$ satisfies Accommodation, it satisfies Preference Propagation.*

Proof. Let $ja \leq j'b$ (1) and $h \leq h'$ (2). Assume that $ha \nleq h'b$. Then by Accommodation, for every $ga, g'b$, $g \leq g' \leftrightarrow ga \leq g'b$. So, in particular, $h \leq h' \leftrightarrow ha \leq h'b$. But since $h \leq h'$, we get $ha \leq h'b$: a contradiction. □

No similar result holds for Preference Revelation. An easy counter-example shows that, even when $\leq$ is total:

Fact 3.3. *Accommodation does not imply Preference Revelation.*

4 The Main Representation Theorem

We start with a warm-up case, with plausibility a *total* pre-order.

4.1 Total pre-orders

Theorem 4.1. *Let $\mathcal{H}$ be a total doxastic-temporal model, $\mathcal{M}$ a total plausibility model, $\vec{\epsilon}$ a sequence of total event models, and let $\otimes$ stand for* priority *update. The following are equivalent:*

- *$\mathcal{H}$ is isomorphic to the forest generated by $\mathcal{M} \otimes \vec{\epsilon}$.*
- *$\mathcal{H}$ satisfies propositional stability, synchronicity, bisimulation invariance, Preference Propagation, and Preference Revelation.*

Proof.

Necessity We first show that the given conditions are indeed satisfied by any *DoTL* model generated through successive priority updates along some given protocol sequence. Here, *Propositional stability* and *Synchronicity* are straightforward from the definition of generated forests.

Preference Propagation Assume that $ja \leq j'b$ (1). It follows from (1) plus the definition of priority update that $a \leq b$ (2). Now assume that $h \leq h'$ (3). It follows from (2), (3) and priority update that $ha \leq h'b$.

Preference Revelation Assume that $jb \leq j'a$ (1). It follows from (1) and the definition of priority update that $b \leq a$ (2). Now assume $ha \leq h'b$ (3). By the definition of priority update, (3) can happen in two ways. Case 1: $a < b$ (4). It follows from (4) by the definition of $<$ that $b \nleq a$ (5). But (5) contradicts (2). We are therefore in Case 2: $a \simeq b$ (6) and $h \leq h'$ (7). But (7) is precisely what we wanted to show.

Note that we did not make use of totality here.

Sufficiency Given a *DoTL* model $\mathcal{M}$, we first show how to construct a *DDL* model, i.e., a plausibility model and a sequence of event models.

Construction Here is the initial plausibility model $\mathcal{M} = \langle W, (\leq_i)_{i \in N}, \hat{V} \rangle$:

- $W := \{h \in H \mid len(h) = 1\}$.
- Set $h \leq_i h'$ iff $h \leq_i h'$.
- For every $p \in Prop$, $\hat{V}(p) = V(p) \cap W$.

Now we construct the j-th event model $\epsilon_j = \langle E_j, (\leq_i^j)_{i \in N}, \mathtt{pre}_j \rangle$:

- $E_j := \{e \in \Sigma \mid$ there is a history$he \in H$ with $len(h) = j\}$
- For each $i \in N$, set $a \leq_i^j b$ iff there are $ha, h'b \in H$ such that $len(h) = len(h) = j$ and $ha \leq_i h'b$.

- For each $e \in E_j$, let $\mathtt{pre}_j(e)$ be the formula that characterizes the set $\{h \mid he \in H$ and $len(h) = j\}$. By general modal logic, *bisimulation invariance* guarantees that there is such a formula, though it may be an infinitary one in general.

Now we show that the construction is correct in the following sense:

Claim 4.2 (Correctness)**.** *Let $\leq$ be the plausibility relation in the given doxastic temporal model. Let $\preccurlyeq^F_{DDL}$ be the plausibility relation in the forest induced by priority update over the just constructed plausibility model and matching sequence of event models. We have:*

$$h \leq h' \text{ iff } h \preccurlyeq^F_{DDL} h'.$$

Proof of the claim The proof is by induction on the length of histories. The base case is obvious from the construction of our initial model $\mathcal{M}$. Now for the induction step. As for notation we will write $a \leq b$ for $a \leq^n_i b$ with n the length for which the claim has been proved, and i an agent.

From *DoTL* **to** *Forest*(*DDL*) Assume that $h_1 a \leq h_2 b$ (1). It follows that in the constructed event model $a \leq b$ (2). Case 1: $a < b$. By priority update we have $h_1 a \preccurlyeq^F_{DDL} h_2 b$. Case 2: $b \leq a$ (3). This means that there are $h_3 b, h_4 a$ such that $h_3 b \leq h_4 a$. But then by *Preference Revelation* and (1) we have $h_1 \leq h_2$ (in the doxastic temporal model). It follows by the inductive hypothesis that $h_1 \preccurlyeq^F_{DDL} h_2$. But then by priority update, since by (2) and (3) a and b are indifferent, we have $h_1 a \preccurlyeq^F_{DDL} h_2 b$.

From *Forest*(*DDL*) **to** *DoTL* Next let $h_1 a \preccurlyeq^F_{DDL} h_2 b$. The definition of priority update has two clauses. Case 1: $a < b$. By definition, this implies that $b \nleq a$. But then by the above construction, for all histories $h_3, h_4 \in H$ we have $h_3 b \nleq h_4 a$. In particular we have $h_2 b \nleq h_1 a$. But then by *totality*[4], $h_1 a \leq h_2 b$. Case 2: $a \simeq b$ (4) and $h_1 \preccurlyeq^F_{DDL} h_2$. For a start, by the inductive hypothesis, $h_1 \leq h_2$ (5). By (4) and our construction, there are $h_3 a, h_4 b$ with $h_3 a \leq h_4 b$ (6). But then by *Preference Propagation*, (5) and (6) imply that we have $h_1 a \leq h_2 b$. □

Next, we turn to the general case of pre-orders, allowing incomparability.

4.2 The general case

While the argument went smoothly for *total* pre-orders, it gets somewhat more interesting when incomparability enters the stage. In the case of pre-orders we need the additional axiom of Accommodation as stated below:

Theorem 4.3. *Let $\mathcal{H}$ be a doxastic-temporal model, $\mathcal{M}$ a plausibility model, $\vec{\epsilon}$ be a sequence of event models while $\otimes$ is* priority *update. The following assertions are equivalent:*

- *$\mathcal{H}$ is isomorphic to the forest generated by $\mathcal{M} \otimes \vec{\epsilon}$,*
- *$\mathcal{H}$ satisfies bisimulation invariance, propositional stability, synchronicity, Preference Revelation and Accommodation.*

[4]Note that this is the only place in which we make use of totality.

By Fact 3.2, Accommodation also gives us Preference Propagation.

Proof. **Necessity of the conditions**
The verification of the conditions in the preceding subsection did not use totality. So we concentrate on the new condition:

Accommodation Assume that $ja \leq j'b$ (1). It follows by the definition of priority update that $a \leq b$ (2). Now let $ha \not\leq h'b$ (3). This implies by priority update that $a \not< b$ (4). By definition, (2) and (4) means that $a \simeq b$ (5). Now assume that $g \leq g'$ (6). It follows from (5), (6) and priority update that $ga \leq g'b$. For the other direction of the consequent assume instead that $g \not\leq g'$ (7). It follows from (5), (7) and priority update that $ga \not\leq g'b$.

Sufficiency of the conditions Given a *DoTL* model, we again construct a *DDL* plausibility model plus sequence of event models:

Construction The plausibility model $\mathcal{M} = \langle W, (\preceq_i)_{i \in N}, \hat{V} \rangle$ is as follows:

- $W := \{h \in H \mid len(h) = 1\}$,
- Set $h \preceq_i h'$ whenever $h \leq_i h'$,
- For every $p \in Prop$, $\hat{V}(p) = V(p) \cap W$.

We construct the j-th event model $\epsilon_j = \langle E_j, (\preceq_i^j)_{i \in N}, \mathtt{pre}_j \rangle$ as follows:

- $E_j := \{e \in \Sigma \mid$ there is a history of the form $he \in H$ with $len(h) = j\}$
- For each $i \in N$, define $a \preceq_i^j b$ iff either (a) there are $ha, h'b \in H$ such that $len(h) = len(h) = j$ and $ha \leq_i h'b$, or (b) [a new case] a and b are accommodating, and we put $a \simeq b$ (i.e. $a \leq b$ and $b \leq a$).
- For each $e \in E_j$, let $\mathtt{pre}_j(e)$ be the formula that characterizes the set $\{h \mid he \in H$ and $len(h) = j\}$. *Bisimulation invariance* guarantees that there is always such a formula (maybe involving an infinitary syntax).

Again we show that the construction is correct in the following sense:

Claim 4.4 (Correctness). *Let $\leq$ be the plausibility relation in the doxastic temporal model. Let $\preccurlyeq^F_{DDL}$ be the plausibility relation in the forest induced by successive priority updates of the plausibility model by the sequence of event models we constructed. We have:*

$$h \leq h' \text{ iff } h \preccurlyeq^F_{DDL} h'.$$

Proof of the claim We proceed by induction on the length of histories. The base case is clear from our construction of the initial model $\mathcal{M}$.
Now for the induction step, with the same simplified notation as earlier.

From *DoTL* **to** *Forest*(*DEL*) We distinguish two cases.

Case 1. $ha \leq h'b, h \leq h'$. By the inductive hypothesis, $h \leq h'$ implies $h \preccurlyeq^F_{DDL} h'$ (1). Since $ha \leq h'b$, it follows by construction that $a \leq b$ (2). It follows from (1) and (2) that by priority update $ha \preccurlyeq^F_{DDL} h'b$.

Case 2. $ha \leq h'b, h \nleq h'$. Clearly, then, a and b are not *accommodating* and thus the special clause has not been used to build the event model, though we do have $a \leq b$ (1). By the contrapositive of Preference Revelation, we also conclude that for all $ja, j'b \in H$, we have $j'b \nleq ja$ (2). Therefore, our construction gives $b \nleq a$ (3), and we conclude that $a < b$ (4). But then by priority update, we get $ha \preccurlyeq^F_{DDL} h'b$.

From *Forest*(*DEL*) **to** *DoTL* We distinguish again two cases.

Case 1. $ha \preccurlyeq^F_{DDL} h'b, h \preccurlyeq^F_{DDL} h'$. By definition of priority update, $ha \preccurlyeq^F_{DDL} h'b$ implies that $a \leq b$ (1). There are two possibilities. Case 1: The special clause of the construction has been used, and a, b are accommodating (2). By the inductive hypothesis, $h \preccurlyeq^F_{DDL} h'$ implies $h \leq h'$ (3). But (2) and (3) imply that $ha \leq h'b$. Case 2: Clause (1) holds because for some $ja, j'b \in H$, in the *DoTL* model, $ja \leq j'b$ (4). By the inductive hypothesis, $h \preccurlyeq^F_{DDL} h'$ implies $h \leq h'$ (5). Now, it follows from (4), (5) and Preference Propagation that $ha \leq h'b$.

Case 2. $ha \preccurlyeq^F_{DDL} h'b, h \not\preccurlyeq^F_{DDL} h'$. Here is where we put our new accommodation clause to work. Let us label our assertions: $h \not\preccurlyeq^F_{DDL} h'$ (1) and $ha \preccurlyeq^F_{DDL} h'b$ (2). It follows from (1) and (2) by the definition of priority update that $a < b$ (3), and hence, by definition $b \nleq a$ (4). Clearly, a and b are not accommodating (5): for otherwise, we would have had $a \simeq b$, and hence $b \leq a$, contradicting (4). Therefore, (3) implies that there are $ja, j'b \in H$ with $ja \leq j'b$ (6). Now assume for *contradictio* that (in the *DoTL* model) $ha \nleq h'b$ (7). It follows from (6) and (7) by Accommodation that a and b *are* accommodating, contradicting (5). Thus we have $ha \leq h'b$. □

Given a doxastic temporal model describing the evolution of the beliefs of a group of agents, we have determined whether it could have been generated by successive 'local' priority updates of a plausibility model. Of course, further scenarios are possible, e.g., bringing in knowledge as well. We discuss some extensions in the next subsection.

4.3 Extensions and variations of the theorem

Unified plausibility models

There are two roads to merging epistemic indistinguishability and doxastic plausibility. The first works with a plausibility order *and* an epistemic indistinguishability relation, explaining the notion of *belief* with a mixture of the two. Baltag and Smets Baltag and Smets (2006) apply product update to epistemic indistinguishability and priority update to the plausibility relation. A characterization for the doxastic epistemic temporal models induced in this way follows from van Benthem et al. Benthem et al. (2008) Theorem 1.1 plus Theorem 4.3 of previous subsection (or its simpler counterpart for total orders).

All this has the flavor of working with *prior* beliefs and information partitions, taking the *posteriors* to be computed from them.

However there are also reasons for working with (*posterior*) beliefs only (see e.g. Morris (1995)). Indeed, Baltag and Smets Baltag and Smets (2006) take this second road, using *unified* 'local' plausibility models with just one explicit relation $\trianglelefteq$. We briefly show how our earlier results transform to this setting. In what follows, we write $a \cong b$ iff $a \trianglelefteq b$ and $b \trianglelefteq a$.

Definition 4.1. The priority update of a *unified* plausibility model $\mathcal{M} = \langle W, (\trianglelefteq_i)_{i\in N}, V\rangle$ and a $\trianglelefteq$-event model $\epsilon = \langle E, (\trianglelefteq_i)_{i\in N}, \mathtt{pre}\rangle$ is the unified plausibility model $\mathcal{M} \otimes \epsilon = \langle W', (\trianglelefteq'_i)_{i\in N}, V'\rangle$ constructed as follows:

- $W' = \{(w, e) \in W \times E \mid \mathcal{M}, w \Vdash \mathtt{pre}(e)\}$,
- $(w, a) \trianglelefteq'_i (w', b)$ iff either 1. $a \trianglelefteq_i b$, $b \not\trianglelefteq a$ and $w \trianglelefteq w' \ \vee\ w' \trianglelefteq w$ or 2. $a \trianglelefteq_i b$, $b \trianglelefteq a$ and $w \trianglelefteq w'$,
- $V'((s, e)) = V(s)$.

Here are our, by now familiar, key properties in this setting.

Agent revision properties in terms of $\trianglelefteq_i$

- $\trianglelefteq$-**PerfectRecall** whenever $ha \trianglelefteq h'b$ we have $h \trianglelefteq h' \ \vee\ h' \trianglelefteq h$.
- $\trianglelefteq$-**PreferencePropagation** whenever $h \trianglelefteq h'$ and $ja \trianglelefteq j'b$ then $ha \trianglelefteq h'b$.
- $\trianglelefteq$-**PreferenceRevelation**, whenever $ha \trianglelefteq h'b$ and $jb \trianglelefteq j'a$, also $h \trianglelefteq h'$.
- $\trianglelefteq$-**Accommodation** if, whenever ($ja \trianglelefteq j'b$, $h' \trianglelefteq h$ and $ha \not\trianglelefteq h'b$), for all $ga, g'b \in H$ ($g \trianglelefteq g' \leftrightarrow ga \trianglelefteq g'b$), and for all $g'a, gb \in H$ ($g \trianglelefteq g' \leftrightarrow gb \trianglelefteq g'a$).

The last axiom is slightly weaker than Accommodation. The following result is proved in the extended version of this paper.

Theorem 4.5. *Let $\mathcal{H}$ be a unified doxastic-temporal model, $\mathcal{M}$ a unified plausibility model, $\vec{\epsilon}$ be a sequence of unified event models, while $\otimes$ is* priority *update. The following assertions are equivalent:*

- *$\mathcal{H}$ is isomorphic to the forest generated by $\mathcal{M} \otimes \vec{\epsilon}$,*
- *$\mathcal{H}$ satisfies bisimulation invariance, propositional stability, synchronicity, $\trianglelefteq$-Perfect Recall, $\trianglelefteq$-Preference Propagation, $\trianglelefteq$-Preference Revelation and $\trianglelefteq$-Accommodation.*

Proof. For the sake of space we just give the idea of how the core of the argument goes. The construction is the same as in the case for general pre-orders. Necessity goes smoothly. For sufficiency and keeping the notation of the previous case.

From *DoTL* **to** *Forest*(*DDL*)

Case 1. $ha \leq h'b, h \leq h'$**.** Follows from the standard clause of the construction, IH and the update rule.

Case 2. $ha \leq h'b, h \not\leq h'$**.** Using again the standard clause of the construction, $\trianglelefteq$-Perfect Recall, the contrapositive of $\trianglelefteq$-Preference Revelation and the update rule.

From *Forest*(*DEL*) **to** *DoTL*

Case 1. $ha \preccurlyeq^F_{DDL} h'b, h \preccurlyeq^F_{DDL} h'$**.** Distinguishing between clauses of the construction. If a, b are accommodating the result is immediate from IH. If they are not then we have $\exists j$, j' $ja \leq j'b$ (1). And the result follows from IH, (1) and Preference Propagation.

Case 2. $ha \preccurlyeq^F_{DDL} h'b, h \not\preccurlyeq^F_{DDL} h'$**.** For this case. It follows by the update rule and construction that a, b are not accomodating in the DoTL model. But then by construction we have $\exists j$, j' $ja \leq j'b$ (2). Thus by accomodation the assumption that $ha \not\leq h'b$ (in the DTL model) would mean that a, b are accomodating which would contradict (2). The result follows by reductio. □

Our next source of variation is an issue that we have left open throughout our analysis so far, which may have bothered some readers.

Bisimulations and pre-condition languages

Our definition of event models presupposed a language for the preconditions, and correspondingly, the right notion of bisimulation in our representation results should matching (at least, on finite models) the precondition language used. For instance, if the precondition language contains a belief operator scanning the *intersection* of a plausibility $\leq_i$ relation and an epistemic indistinguishability relation $\sim$, then the *zig* and *zag* clauses should not only apply to $\leq_i$ and $\sim_i$ separately, but also to $\leq_i \cap \sim_i$. And things get even more complicated if we allow temporal operators in our languages (cf. Benthem et al. (2007)). We do not commit to any specific choice here, since the choice of a language seems orthogonal to our main concerns. But we will discuss formal languages in the next section, taking definability of our major structural constraints as a guide.

Finally, our results can be generalized by including one more parameter.

Protocols

So far we have assumed that the same sequences of events were executable uniformly anywhere in the initial doxastic model, provided the worlds fulfilled the preconditions. This strong assumption is lifted in Benthem et al. (2007; 2008), who allow the *protocol*, i.e., the set of executable sequences of events forming our current informational process, to vary from state to state. Initially, they still take the protocol to be *common knowledge*, but eventually, they allow for scenarios where agents need not know which protocol is running. These variations change the complete dynamic-epistemic logic of the system. It would be of interest to extend this work to our extended doxastic setting.

5 Dynamic and Temporal Doxastic Languages

Our emphasis so far has been on structural properties of models. To conclude, we turn to the logical languages that can express these, and hence also, the type of doxastic reasoning our agents can be involved with.

5.1 Dynamic doxastic language

We first look at a core language that matches dynamic belief update.

Syntax

Definition 5.1 (Dynamic Doxastic-Epistemic language)**.** The language of dynamic doxastic language $DDE\mathcal{L}$ is defined as follows:

$$\phi := p \mid \neg\phi \mid \phi \vee \phi \mid \langle\leq_i\rangle\phi \mid \langle i\rangle\phi \mid \mathtt{E}\phi \mid \langle\epsilon, \mathbf{e}\rangle\phi$$

where i ranges over over N, p over a countable set of proposition letters $Prop$, and $(\epsilon, \mathbf{e})$ ranges over a suitable set of symbols for event models.

All our dynamic doxastic logics will be interpreted on the following models.

Models

Definition 5.2 (Epistemic Plausibility Models)**.** An *epistemic plausibility model* $\mathcal{M} = \langle W, (\leq_i)_{i\in N}, (\sim_i)_{i\in N}, V\rangle$ has $W \neq \emptyset$, and for each $i \in N$, $\leq_i$ is a pre-order on W, and $\sim_i$ any relation, while $V : Prop \to \wp H$.

Definition 5.3 ($\sim$,$\leq$-event model)**.** An *epistemic plausibility event model* ($\sim$,$\leq$-event model for short) ϵ is of the form $\langle E, (\leq_i)_{i\in N}, (\sim_i)_{i\in N}, \mathtt{pre}\rangle$ where $E \neq \emptyset$, for each $i \in N$, $\leq_i$ is a pre-order on E and $\sim_i$ is a relation on W. Also, there is a precondition function $\mathtt{pre} : E \to DDE\mathcal{L}$

Definition 5.4 (Priority update)**.** The *priority update* of an epistemic plausibility model $\mathcal{M} = \langle W, (\leq_i)_{i\in N}, (\sim_i)_{i\in N}, V\rangle$ and a $\sim$,<-event model $\epsilon = \langle E, (\leq_i)_{i\in N}, (\sim_i)_{i\in N}, \mathtt{pre}\rangle$ is the plausibility model $\mathcal{M} \otimes \epsilon = \langle W', (\leq'_i)_{i\in N}, V'\rangle$ whose structure is defined as follows:

- $W' = \{(w, e) \in W \times E \mid \mathcal{M}, w \Vdash \mathtt{pre}(e)\}$
- $(w, e) \leq'_i (w', e')$ iff $e <_i e'$, or $e \simeq_i e'$ and $w \leq_i w'$
- $(w, e) \sim'_i (w', e')$ iff $e \sim_i e'$ and $w \sim_i w'$
- $V'((s, e)) = V(s)$.

Semantics

Here is how we interpret the $DDE(L)$ language. A pointed event model is an event model plus an element of its domain. To economize on notation we use event symbols in the semantic clause. We write $\mathtt{pre}(e)$ for $\mathtt{pre}_\epsilon(e)$ when it is clear from context.

Definition 5.5 (Truth definition)**.** Let $K_i[w] = \{v \mid w \sim_i v\}$.

$$\begin{array}{lll}
\mathcal{M}, w \Vdash p & \text{iff} & w \in V(p) \\
\mathcal{M}, w \Vdash \neg\phi & \text{iff} & \mathcal{M}, w \nVdash \phi \\
\mathcal{M}, w \Vdash \phi \vee \psi & \text{iff} & \mathcal{M}, w \Vdash \phi \text{ or } \mathcal{M}, w \Vdash \psi \\
\mathcal{M}, w \Vdash \langle \leq_i \rangle \phi & \text{iff} & \exists v \text{ such that } w \leq_i v \text{ and } \mathcal{M}, v \Vdash \phi \\
\mathcal{M}, w \Vdash \langle i \rangle \phi & \text{iff} & \exists v \text{ such that } v \in K_i[w] \text{ and } \mathcal{M}, v \Vdash \phi \\
\mathcal{M}, w \Vdash \mathtt{E}\phi & \text{iff} & \exists v \in W \text{ such that } \mathcal{M}, v \Vdash \phi \\
\mathcal{M}, w \Vdash \langle \epsilon, \mathbf{e} \rangle \phi & \text{iff} & \mathcal{M}, w \Vdash \mathtt{pre}(e) \text{ and } \mathcal{M} \times \epsilon, (w, e) \Vdash \phi
\end{array}$$

Knowledge K_i and the universal modality $\mathtt{A}$ are defined as usual.

Reduction axioms

The methodology of dynamic epistemic and dynamic doxastic logics revolves around *reduction* axioms. On top of some complete static base logic, these fully describe the dynamic component. Here is well-known *Action* – *Knowledge* reduction axiom of Baltag et al. (1998):

$$[\epsilon, \mathbf{e}] K_i \phi \leftrightarrow (\mathtt{pre}(e) \rightarrow \bigwedge \{K_i [\epsilon, \mathbf{f}] \phi \ : \ e \sim_i f\}) \tag{1}$$

Similarly here is the key reduction axiom for $\langle \epsilon, \mathbf{e} \rangle \langle \leq_i \rangle$ with priority update:

Proposition 5.1. *The following dynamic-doxastic principle is sound for plausibility change:*

$$\langle \epsilon, \mathbf{e} \rangle \langle \leq_i \rangle \phi \leftrightarrow (\mathtt{pre}(e) \wedge (\langle \leq_i \rangle \bigvee \{\langle \mathbf{f} \rangle \phi \ : \ e \simeq_i f\} \vee \mathtt{E} \bigvee \{\langle \mathbf{g} \rangle \phi \ : \ e <_i g\})) \tag{2}$$

The crucial feature of such a dynamic 'recursion step' is that the order between *action* and *belief* is reversed. This works because, conceptually, the current beliefs already *pre-encode* the beliefs after some specified event. In the epistemic setting, principles like this also reflect agent properties of Perfect Recall and No Miracles Benthem and Pacuit (2006). Here, they rather encode radically 'event-oriented' revision policies, and the same point applies to the principles we will find later in a doxastic temporal setting.

Finally for the existential modality $\langle \epsilon, \mathbf{e} \rangle \mathtt{E}$ we note the following fact:

Proposition 5.2. *The following axiom is valid for the existential modality:*

$$\langle \epsilon, \mathbf{e} \rangle \mathtt{E} \phi \leftrightarrow (\mathtt{pre}(e) \wedge (\mathtt{E} \bigvee \{\langle \mathbf{f} \rangle \phi \ : \ f \in Dom(\epsilon)\})) \tag{3}$$

We do not pursue further issues of axiomatic completeness here, since we are just after the model theory of our dynamic and temporal structures.

5.2 Doxastic epistemic temporal language

Next *epistemic-doxastic temporal models* are simply our old doxastic temporal models $\mathcal{H}$ extended with epistemic accessibility relations $\sim_i$.

Syntax

Definition 5.6 (Doxastic Epistemic Temporal Languages)**.** The language of $DET\mathcal{L}$ is defined by the following inductive syntax:

$$\phi := p \mid \neg\phi \mid \phi \vee \phi \mid \langle e\rangle\phi \mid \langle e^{-1}\rangle\phi \mid \langle\leq_i\rangle\phi \mid \langle i\rangle\phi \mid \mathtt{E}\phi$$

where i ranges over N, e over Σ, and p over proposition letters *Prop*.

Semantics

The language $DET\mathcal{L}$ is interpreted over nodes h in our trees (cf. Benthem and Pacuit (2006)):

Definition 5.7 (Truth definition)**.** Let $K_i[h] = \{h' \mid h \sim_i h'\}$.

$\mathcal{H}, h \Vdash p$	iff	$h \in V(p)$
$\mathcal{H}, h \Vdash \neg\phi$	iff	$\mathcal{H}, h \nVdash \phi$
$\mathcal{H}, h \Vdash \phi \vee \psi$	iff	$\mathcal{H}, h \Vdash \phi$ or $\mathcal{H}, h \Vdash \psi$
$\mathcal{H}, h \Vdash \langle e\rangle\phi$	iff	$\exists h' \in H$ such that $h' = he$ and $\mathcal{H}, h' \Vdash \phi$
$\mathcal{H}, h \Vdash \langle e^{-1}\rangle\phi$	iff	$\exists h' \in H$ such that $h'e = h$ and $\mathcal{H}, h' \Vdash \phi$
$\mathcal{H}, h \Vdash \langle\leq_i\rangle\phi$	iff	$\exists h'$ such that $h \leq_i h'$ and $\mathcal{H}, h' \Vdash \phi$
$\mathcal{H}, h \Vdash \langle i\rangle\phi$	iff	$\exists h'$ such that $h' \in K_i[h]$ and $\mathcal{H}, h' \Vdash \phi$
$\mathcal{H}, h \Vdash \mathtt{E}\phi$	iff	$\exists h' \in H$ such that $\mathcal{H}, h' \Vdash \phi$

Now we have the right syntax to analyze our earlier structural conditions.

5.3 Defining the frame conditions

We will prove semantic *correspondence results* (cf. Blackburn et al. (2001)) for our crucial properties using somewhat technical axioms that simplify the argument. Afterwards, we present some reformulations whose meaning for belief-revising agents is more intuitive.

The key correspondence result

Theorem 5.3 (Definability)**.** *Preference Propagation, Preference Revelation and Accommodation are definable in the doxastic-epistemic temporal language $DET\mathcal{L}$.*

- *$\mathcal{H}$ satisfies Preference Propagation iff the following axiom is valid:*

$$\mathtt{E}\langle a\rangle\langle\leq_i\rangle\langle b^{-1}\rangle\top \rightarrow (((\langle\leq_i\rangle\langle b\rangle p \wedge \langle a\rangle q) \rightarrow \langle a\rangle(q \wedge \langle\leq_i\rangle p) \qquad (PP)$$

- *$\mathcal{H}$ satisfies Preference Revelation iff the following axiom is valid:*

$$\mathtt{E}\langle b\rangle\langle\leq_i\rangle\langle a^{-1}\rangle\top \rightarrow (\langle a\rangle\langle\leq_i\rangle(p \wedge \langle b^{-1}\rangle\top) \rightarrow \langle\leq_i\rangle\langle b\rangle p) \qquad (PR)$$

- *$\mathcal{H}$ satisfies Accommodation iff the following axiom is valid:*

$$\begin{aligned}
&\mathtt{E}\langle a\rangle\langle\leq_i\rangle\langle b^{-1}\rangle\top \\
&\quad \wedge \mathtt{E}\,[\langle a\rangle\,(p_1 \wedge \mathtt{E}\,(p_2 \wedge \langle b^{-1}\rangle\top)\,) \wedge [a]\,(p_1 \rightarrow [\leq_i]\neg p_2)] \\
&\qquad \rightarrow (\,(\langle\leq_i\rangle\langle b\rangle q \rightarrow [a]\langle\leq_i\rangle q) \\
&\qquad\quad \wedge (\langle a\rangle\langle\leq_i\rangle(r \wedge \langle b^{-1}\rangle\top) \rightarrow \langle\leq_i\rangle\langle b\rangle r) \qquad (AC)
\end{aligned}$$

Proof. We only prove the case of *Preference Propagation*, the other two are in the extended version of the paper. We drop agent labels for convenience.

(PP) **characterizes Preference Propagation** We first show that (PP) is valid on all models $\mathcal{H}$ based on preference-propagating frames. Assume that $\mathcal{H}, h \Vdash \mathtt{E}\langle a\rangle\langle\leq_i\rangle\langle b^{-1}\rangle\top$ (1). Then there are $ja, j'b \in H$ such that $ja \leq j'b$ (2). Now let $\mathcal{H}, h \Vdash (\langle\leq\rangle\langle b\rangle p \wedge \langle a\rangle q)$ (3). Then there is $h' \in H$ such that $h \leq h'$ (4) and $\mathcal{H}, h' \Vdash \langle b\rangle p$ (5), while also $\mathcal{H}, ha \Vdash q$ (6). We must show that $\mathcal{H}, h \vDash \langle a\rangle(q \wedge \langle\leq_i\rangle p)$ (7). But, from (2),(4),(6) and *Preference Propagation*, we get $ha \leq h'b$, and the conclusion follows by the truth definition.

Next, we assume that axiom (PP) is valid on a doxastic temporal frame, that is, true under any interpretation of its proposition letters. So, assume that $ja \leq j'b$ (1), and also $h \leq h'$ (2). Moreover, let $ha, h'b \in H$ (3). First note that (1) automatically verifies the antecedent of (PP) in any node of the tree. Next, we make the antecedent of the second implication in (PP) true at h by interpreting the proposition letter p as just the singleton set of nodes $h'b$, and q as just ha (4). Since (PP) is valid, its consequent will also hold under this particular valuation V. Explicitly we have $\mathcal{H}, V, h \Vdash \langle a\rangle(q \wedge \langle\leq_i\rangle p)$. But spelling out what p, q mean there, we get just the desired conclusion that $ha \leq h'b$. □

The preceding correspondence argument is really just a *Sahlqvist* substitution case (cf. Blackburn et al. (2001)), and so are the other two. (See Benthem and Dégremont (2008) for the proofs.)

(PR) **characterizes Preference Revelation**

(AC) **charachterizes Accommodation** We do not prove a further completeness result, but will show one nice derivation, as a syntactic counterpart to our earlier Fact 3.1.

$$\begin{aligned} &\mathtt{E}\,[\langle a\rangle\,(\psi \;\wedge\; \mathtt{E}\,(\phi \wedge \langle b^{-1}\rangle\top)\,) \;\wedge\; [a]\,(\psi \;\rightarrow\; [\leq_i]\neg\phi)] \\ &\quad\rightarrow\; ((\langle a\rangle\langle\leq_i\rangle(\phi \wedge \langle b^{-1}\rangle\top) \rightarrow \langle\leq_i\rangle\langle b\rangle\phi) \end{aligned} \tag{F}$$

Here is an auxiliary correspondence observation:

Fact 5.4. *On total doxastic temporal models the following axiom is valid:*

$$\begin{aligned} &\langle a\rangle(\psi \;\wedge\; \mathtt{E}\,(\phi \wedge \langle b^{-1}\rangle\top)) \;\;\rightarrow \\ &\qquad(\,\langle a\rangle(\psi \;\wedge\; \langle\leq_i\rangle\phi) \;\vee\; \mathtt{E}\langle b\rangle(\phi \;\wedge\; \langle\leq_i\rangle(\psi \wedge \langle a^{-1}\rangle\top)) \end{aligned} \tag{Tot}$$

Now we can state an earlier semantic fact in terms of axiomatic derivability in some obvious minimal system for the language $DET\mathcal{L}$:

Fact 5.5. $\bullet \vdash ((PP) \wedge (F)) \rightarrow (\mathrm{AC})$ $\bullet \vdash ((PR) \wedge (Tot)) \rightarrow (F)$

Benthem and Dégremont (2008) has the details. We now get an immediate counterpart to Fact 3.1:

Corollary 5.6.

$$\vdash ((PP) \wedge (PR) \wedge (Tot)) \rightarrow (\mathrm{AC}) \tag{4}$$

Two intuitive explanations

Here are two ways to grasp the intuitive meaning of our technical axioms.

Reformulation with safe belief. An intermediate notion of knowledge first considered by Stalnaker (1981) has been argued for doxastically as *safe belief* by Baltag and Smets (2006) as describing those beliefs we do not give up under true new information. The safe belief modality $\Box^{\geq}$ is just the universal dual of the existential modality $\langle\geq\rangle$ scanning the converse of $\leq$. Without going into details of its logic (e.g., safe belief is positively, but not negatively introspective), here is how we can rephrase our earlier axiom:
• $\mathcal{H}$ satisfies Preference Propagation iff the following axiom is valid on $\mathcal{H}$:

$$\mathtt{E}\langle a\rangle\langle\geq\rangle\langle b^{-1}\rangle\top \rightarrow (\langle a\rangle\Box^{\geq_i}p \rightarrow \Box^{\geq_i}[b]p) \qquad (PP')$$

A similar reformulation is easy to give for Preference Revelation. These principles reverse action modalities and safe belief much like the better-known Knowledge-Action interchange laws in the epistemic-temporal case. We invite the reader to check their intuitive meaning in terms of acquired safe beliefs as informative events happen.

Analogies with reduction axioms Another way to understand the above axioms in their original format with existential modalities is their clear analogy with the reduction axiom for priority update. Here are two cases juxtaposed:

$$\langle\epsilon,\mathbf{e}\rangle\langle\leq_i\rangle p \leftrightarrow (\mathtt{pre}(e) \wedge (\langle\leq_i\rangle \bigvee\{\langle f\rangle p \ : \ e\simeq_i f\} \vee \mathtt{E}\bigvee\{\langle g\rangle p \ : \ e <_i g\})) \qquad (2)$$

$$\mathtt{E}\langle a\rangle\langle\leq_i\rangle\langle b^{-1}\rangle\top \rightarrow ((\langle\leq_{\mathtt{i}}\rangle\langle\mathtt{b}\rangle p \rightarrow [\mathtt{a}]\langle\leq_{\mathtt{i}}\rangle p) \qquad (PP)$$

$$\mathtt{E}\langle b\rangle\langle\leq_i\rangle\langle a^{-1}\rangle\top \rightarrow ((\langle\mathtt{a}\rangle\langle\leq_{\mathtt{i}}\rangle(p \wedge \langle b^{-1}\rangle\top) \rightarrow \langle\leq_{\mathtt{i}}\rangle\langle\mathtt{b}\rangle p) \qquad (PR)$$

Family resemblance is obvious, and indeed, (*PP*) and (*PR*) may be viewed as the two halves of the reduction axiom, transposed to the more general setting of arbitrary doxastic-temporal models.

5.4 Variations and extensions of the language

Weaker languages

The above doxastic-temporal language is by no means the only reasonable one. Weaker forward-looking modal fragments also make sense, dropping both converse and the existential modality. But they do not suffice for the purpose of our correspondence.

Proposition 5.7 (Undefinability)**.**
Preference propagation, Preference Revelation and Accommodation are not *definable in the forward looking fragment of* **DET**$\mathcal{L}$

Proof. The reason is the same in all cases: we show that these properties are not preserved under taking *bounded p-morphic images*. The Figure gives an indication how this works concretely. □

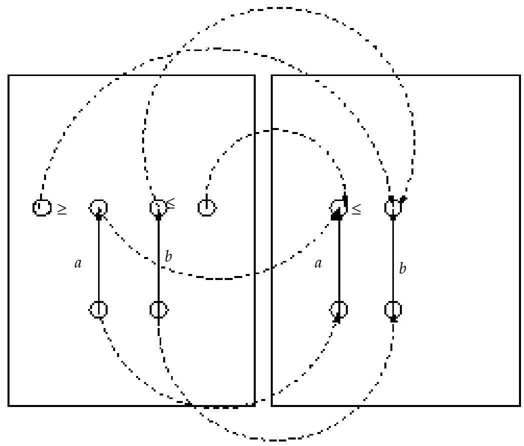


Figure 1: Propagation is not preserved in the p-morphic images

Richer languages

But there is also a case to be made for richer languages. For instance, if we want to define the frame property of *synchronicity*, we must introduce an *equilevel relation* in our models, with a corresponding modality for it. While expressing synchronicity then becomes easy, this move is dangerous in principle. Van Benthem and Pacuit Benthem and Pacuit (2006) point at the generally high complexity of tree logics when enriched with this expressive power.

Likewise, finer epistemic and doxastic process descriptions will require further temporal modalities, such as "Since" and "Until", beyond the basic operators that we used for matching the needs of dynamic doxastic logic directly.

Finally, there may be even more urgent language extensions for doxastic temporal logic, having to do with our very notion of belief. We have emphasized the notion of *safe belief*, which scans the plausibility relation $\geq$ as an ordinary modality. This notion can be used to define the more standard notion of belief as truth in all most plausible worlds: cf. Boutilier (1994). But it has been argued recently by Baltag and Smets (2006), and also by de Jongh and Liu (2006) that we really want a more 'entangled' version of the latter notion as well, referring to the most plausible worlds *inside the epistemically accessible ones*. Such a notion of 'posterior belief' has the following semantics:

$$\mathcal{H}, h \Vdash B_i\phi \quad \text{iff} \quad \forall h' \in Min(K_i[h], \leq_i) \text{ we have } \mathcal{H}, h' \Vdash \phi$$

Technically, expressing this requires an additional intersection modality. While this extension loses some typical modal properties, it does satisfy reduction axioms in the format discussed here: cf. Liu (2008).

6 Axiomatization of protocols-based dynamic logics of belief revision

While PAL is an interesting special case of DEL, Dynamic belief revision (Lexicographic Upgrade) is an interesting special case of DDL. For the interested reader the precise analogy will be given as the end of the section.

The aim of this section is to consider protocols-based version of logics of belief revision. First of all one will therefore be interested in a particular cases of soft priority update, namely whenever informally events can be identified with their preconditions, i.e. with formulas of a doxastic language. Secondly not all events can be executed everywhere, i.e. certain formulas will be declared/observed in some states but not in some other. This information is precisely what a protocol is about: indicating what can be "said" (what sequences of formulas) where or when (in which doxastic state). Precisely:

Definition 6.1. Dynamic Belief Revision Protocol

Given a doxastic language $\mathcal{L}$ we let $Ptcl(\mathcal{L}) = \{P | P \subseteq \mathcal{L}^*$ and P is closed under initial segments$\}$. Given a doxastic model $\mathcal{M} = \langle W, (\leq_i)_{i\in N}, V\rangle$, a dynamic belief revision protocol is a mapping $\mathtt{p} : W \to Ptcl(\mathcal{L})$.

For each n, let h_n stand for the initial segment of h of length n. We will be interested in a particular kind of forests. Namely:

Definition 6.2 (*DoTL* model generated by a sequence of protocol-based lexicographic upgrade)**.** Each initial plausibility model $\mathcal{M} = \langle W, (\leq_i)_{i\in N}, V\rangle$ and each dynamic belief revision protocol $\mathtt{p} : W \to Ptcl(\mathcal{L})$ yields a yields a *generated DoTL plausibility model* $\mathcal{H} = \langle \Sigma, H, (\leq_i)_{i\in N}, \mathbf{V}\rangle$ as follows:

- Let $\Sigma := \mathcal{L}$.
- Let $H_1 := W$ and for any $1 < n$ let $H_{n+1} := \{(w\phi_1 \ldots \phi_n) | (w\phi_1 \ldots \phi_{n-1}) \in H_n$ such that $\phi_1 \ldots \phi_n \in \mathtt{p}(w)\}$

  Finally let $H = \bigcup_{1\leq k} H_k$.
- If $h, h' \in H_1$, then $h \leq_i h'$ iff $h \leq_i^{\mathcal{M}} h'$.
- For $1 < k$ and for $h = w\phi_1 \ldots \phi_{k-1}$, $h' = w'\phi_1 \ldots \phi_{k-1}$, $h \leq_i h'$ iff one of the following holds:
  1. $\mathcal{H}, (w\phi_1 \ldots \phi_{k-2}) \Vdash \phi_{k-1}$ while $\mathcal{H}, (w'\phi_1 \ldots \phi_{k-2}) \nVdash \phi_{k-1}$
  2. $\mathcal{H}, (w\phi_1 \ldots \phi_{k-2}) \Vdash \phi_{k-1}$ iff $\mathcal{H}, (w'\phi_1 \ldots \phi_{k-2}) \Vdash \phi_{k-1}$, and $(w\phi_1 \ldots \phi_{k-2}) \leq_i (w'\phi_1 \ldots \phi_{k-2})$.
- For each $1 \leq k$, for each $h, h' \in H_k$, let $h \equiv h'$ iff $h_k = h'k$.
- Let $wh \in \mathbf{V}(p)$ iff $w \in V(p)$.

One might be interested in different natural doxastic language. Depending on some cardinality assumptions about the model some of the fragments of this logic have the needed expressive power. Moreover some of them are easier to axiomatize than other. In general one would like our language to be able to define (static) conditional belief in the static part of our language. We will look at axiomatizations of protocol-based lexicographic updgrade for different static doxastic languages. In general there are two strategies to have our axiomatization include conditional belief. One is to take is as a defined notion. The other is to handle it directly, giving in particular a pseudo reduction axiom for it. Of course one would expect the latter to be derivable from the axioms for the modalities with which it can be defined.

On finite models, it is known (see (Girard 2008, 3.3.6)) that conditional beliefs can defined using a modality for the plausibility pre-order and the universal modality. In the general case more expressive power is needed. One way to go is to allow for state variable and a hybrid state binder, binding a variable to the current state. We first discuss our first option in the next subsection. We discuss different languages in the one after.

6.1 Axiomatization of the dynamic logic of protocols-based belief revision.

We first present the static doxastic language $\mathcal{L}_{Dox}(\mathtt{E}, \geq)$.

Definition 6.3 (Protocol-constrained Multi-agent Belief Revision Language)**.** The purely doxastic language $\mathcal{L}_{Dox}(\mathtt{E}, \geq)$ is:

$$\phi := p \mid \neg\phi \mid \phi \vee \phi \mid \langle\leq_i\rangle\phi \mid \langle\geq_i\rangle\phi \mid \mathtt{E}_{\equiv}\phi \mid$$

where i ranges over N and p over a countable set of proposition letters *Prop*

Its dynamic extension is defined as expected:

The language of $TBR(\mathcal{L}_{Dox}(\mathtt{E}, \geq))$ is defined by the following inductive syntax:

$$\phi := p \mid \neg\phi \mid \phi \vee \phi \mid \langle\leq_i\rangle\phi \mid \langle\geq_i\rangle\phi \mid \mathtt{E}_{\equiv}\phi \mid \langle\Uparrow \psi\rangle\phi \mid$$

where i ranges over N, ψ over $\mathcal{L}_{Dox}(\mathtt{E}, \geq)$ and p over a countable set of proposition letters *Prop*. We make use of the usual shortcuts. In particular the universal modality $\mathtt{A}_{\equiv}$ is defined as usual. Moreover (Girard 2008, 3.3.6) has shown that conditional belief can be defined as follows:

$$B_i^{\psi}\phi \leftrightarrow A(\psi \rightarrow \langle\geq_i\rangle(\psi \wedge [\geq_i](\psi \rightarrow \phi)))$$

Definition 6.4 (Truth definition)**.** The language $TBR(\mathcal{L}_{Dox}(\mathtt{E}, \geq))$ is interpreted over nodes wh in our trees (cf. Benthem and Pacuit (2006)) - where w is a sequence of length 1 and h is possibly the empty sequence - together with an assignment $g : \mathtt{SVAR} \rightarrow H_1$ (reminding to the reader that H_1 is the set of sequences of length 1.)

$\mathcal{H}, wh, g \Vdash p$	iff	$wh \in V(p)$
$\mathcal{H}, wh, g \Vdash \neg\phi$	iff	$\mathcal{H}, wh, g \nVdash \phi$
$\mathcal{H}, wh, g \Vdash \phi \vee \psi$	iff	$\mathcal{H}, wh, g \Vdash \phi$ or $\mathcal{H}, wh, g \Vdash \psi$
$\mathcal{H}, wh, g \Vdash \langle\leq_i\rangle\phi$	iff	$\exists h'$ such that $wh \leq_i h'$ and $\mathcal{H}, h', g \Vdash \phi$
$\mathcal{H}, wh, g \Vdash \langle\geq_i\rangle\phi$	iff	$\exists h'$ such that $h' \leq_i wh$ and $\mathcal{H}, h', g \Vdash \phi$
$\mathcal{H}, wh, g \Vdash \mathtt{E}_{\equiv}\phi$	iff	$\exists h' \equiv wh$ such that $\mathcal{H}, h', g \Vdash \phi$
$\mathcal{H}, wh, g \Vdash \langle\Uparrow \phi\rangle\psi$	iff	$\exists h' \in H$ such that $h' = wh\phi$ and $\mathcal{H}, h', g \Vdash \psi$

To be sure, on finite state spaces, the syntactic definition of $B_i^{\phi}\psi$ is sound with respect to the following definition: For each ϕ, let $B_i^{\phi}[wh] = Min(\{h' \mid wh \equiv h'$ and $\mathcal{H}, h' \vDash \phi\}, \leq_i)$.

$$\mathcal{H}, wh, g \Vdash B_i^{\phi}\psi \quad \text{iff} \quad \forall h' \text{ such that } h' \in B_i^{\phi}[wh] \text{ we have } \mathcal{H}, h', g \Vdash \psi$$

Let us now turn to the axiomatization of this fragment.
We will call it $\mathbf{Dox}(\mathtt{E}, \langle \geq \rangle)$.

Definition 6.5 (Axiomatization).

Ax1 **S4** for $\langle \leq_i \rangle$ and $\langle \geq_i \rangle$

Ax2 **S5** for $\mathtt{E}_\equiv$

Ax3 $(p \rightarrow [\leq_i]\langle \geq_i \rangle p) \wedge (p \rightarrow [\geq_i]\langle \leq_i \rangle p)$

Ax4 $B_i^\psi \phi \leftrightarrow A(\psi \rightarrow \langle \geq_i \rangle(\psi \wedge [\geq_i](\psi \rightarrow \phi)))$
For any $\phi \in \mathcal{L}_{Dox}$ we have also:

Ax5 **K** for $\langle \Uparrow \phi \rangle$

Ax6 For p atomic we have $\langle \Uparrow \phi \rangle p \leftrightarrow \langle \Uparrow \phi \rangle \top \wedge p$.

Ax7 $\langle \Uparrow \phi \rangle \neg \psi \leftrightarrow \langle \Uparrow \phi \rangle \top \wedge \neg \langle \Uparrow \phi \rangle \psi$

Ax8 $\langle \Uparrow \phi \rangle \psi \wedge \chi \leftrightarrow \langle \Uparrow \phi \rangle \psi \wedge \langle \Uparrow \phi \rangle \chi$

Ax9

$$\begin{aligned} \langle \Uparrow \phi \rangle \langle \leq_i \rangle \psi \leftrightarrow \langle \Uparrow \phi \rangle \top \wedge & \\ & [\, (\phi \wedge \mathtt{E}_\equiv(\neg\phi \wedge \langle \Uparrow \phi \rangle \psi)) \\ & \vee\ (\phi \wedge \langle \leq_i \rangle \langle \Uparrow \phi \rangle \psi) \\ & \vee\ (\neg\phi \wedge \langle \leq_i \rangle(\neg\phi \wedge \langle \Uparrow \phi \rangle \psi))\,] \end{aligned} \tag{5}$$

Ax10 $\langle \Uparrow \phi \rangle \mathtt{E}_\equiv \psi \leftrightarrow \langle \Uparrow \phi \rangle \top \wedge \mathtt{E}_\equiv \langle \Uparrow \phi \rangle \psi$

It is easily checked that the following are derivable.

$$\langle \Uparrow \phi \rangle \mathtt{A}_\equiv \psi \leftrightarrow \langle \Uparrow \phi \rangle \top \wedge \mathtt{A}_\equiv(\langle \Uparrow \phi \rangle \top \rightarrow \langle \Uparrow \phi \rangle \psi) \tag{Th1a}$$

$$\begin{aligned} \langle \Uparrow \phi \rangle [\leq_i] \psi \leftrightarrow \langle \Uparrow \phi \rangle \top \wedge & \\ & [(\phi \rightarrow \mathtt{A}_\equiv(\neg\phi \rightarrow \neg\langle \Uparrow \phi \rangle \neg\psi)) \\ & \wedge (\phi \rightarrow [\leq_i]\neg\langle \Uparrow \phi \rangle \neg\psi) \\ & \wedge [\leq_i](\neg\phi \rightarrow \neg\langle \Uparrow \phi \rangle \neg\psi)] \end{aligned} \tag{Th1b}$$

Completeness

Since conditional belief can be cashed out in terms of the existential and the $\langle \geq_i \rangle$ modalities we will consider the conditional belief free fragment of $TBR\mathcal{L}$. For the completeness proof we start by defining a `Stbr`-canonical initial model and prove that it can be unfolded in a satisfactory belief revision forest.

Definition 6.6 (Canonical initial doxastic model).
`Stbr` canonical initial model $\mathcal{M}_0^\Sigma = \langle W^0, \leq_i^0, \geq_i^0, \equiv^0, V^0 \rangle$ defined as follows:

- W^0 is the set of `Stbr`-MCSs

- For each $w, v \in W^0$, define $w \leq_i^0 v$ iff $\{\phi \mid [\leq_i]\phi \in w\} \subseteq v$
- For each $w, v \in W^0$, define $w \geq_i^0 v$ iff $\{\phi \mid [\geq_i]\phi \in w\} \subseteq v$
- For each $w, v \in W^0$, define $w \equiv^0 v$ iff $\{\phi \mid \mathtt{A}\phi \in w\} \subseteq v$
- Finally define $V^0(p) = \{w \in H^0 \mid p \in w\}$.

A piece of notation $h_{(n)}$ stand for the n-th element of h. We can now define a canonical doxastic forest by unfolding the canonical initial model:

Definition 6.7 (Canonical doxastic forest)**.**
`Stbr` canonical forest $\mathcal{H}^\Sigma = \langle H^\Sigma, \lambda, \leq_i^\Sigma, \geq_i^\Sigma, V^\Sigma \rangle$ is defined as follows:

- $H^0 = W^0$
- For each $h \in H^0$ let $\lambda(h) = h$
- $H_{n+1} = \{h\phi \mid h \in H_n \text{ and } \langle \Uparrow \phi \rangle \top \in \lambda(h)\}$
- For each $k > 0$ and $h\phi \in H_k$ let $\lambda(h\phi) = \{\psi \mid \langle \Uparrow \phi \rangle \psi \in \lambda(h)\}$
- $H^\Sigma = \bigcup_{k \geq 0} H^k$
- For each $h, h' \in H^0$, define $h \leq_i^\Sigma h'$ iff $\{\phi \mid [\leq_i]\phi \in \lambda(h)\} \subseteq \lambda(h')$
- For each $h, h' \in H^0$, define $h \geq_i^\Sigma h'$ iff $\{\phi \mid [\geq_i]\phi \in \lambda(h)\} \subseteq \lambda(h')$
- For each $h, h' \in H^0$, define $h \equiv^\Sigma h'$ iff $\{\phi \mid \mathtt{A}_\equiv\phi \in \lambda(h)\} \subseteq \lambda(h')$
- For each $k > 0$ and $h\phi, h'\psi \in H^k$, define $h\phi \leq^\Sigma h'\psi$ iff $h = h'$, $\phi = \psi$ and one of the following holds:
  1. $\phi \in \lambda(h)$ while $\phi \notin \lambda(h')$
  2. $\phi \in \lambda(h)$ iff $\phi \in \lambda(h')$, and $h \leq_i h'$.
- For each $k > 0$ and $h\phi, h'\psi \in H^k$, define $h\phi \geq^\Sigma h'\psi$ iff $h'\psi \leq^\Sigma h\phi$
- For each $k > 0$ and for each $h\phi, h'\psi \in H^k$, define $h\phi \equiv^\Sigma h'\psi$ iff $h \equiv^\Sigma h'$ and $\phi = \psi$.
- For each $h \in H^\Sigma$, define $V^\sigma(p) = \{h \in H^\Sigma \mid p \in \lambda(h_{(1)})\}$.

Lemma 6.1. *For each $k \geq 0$, for each $h \in H^k$, $\lambda(h)$ is* `Stbr`*-MCSs.*

Proof. The proof is by induction on k. The base case holds by definition. Assume that the claim holds for $k = n$. Now assume that $h\phi \in H^{n+1}$. By IH $\lambda(h)$ is a MCSs. Moreover by construction we have $\langle \Uparrow \phi \rangle \top \in \lambda(h)$ (1). Let $\phi \in TBR\mathcal{L}$. Since $\lambda(h)$ is a MCSs we have either $\langle \Uparrow \phi \rangle \psi \in \lambda(h)$ or $\neg \langle \Uparrow \phi \rangle \psi \in \lambda(h)$. If $\langle \Uparrow \phi \rangle \psi \in \lambda(h)$, then by construction $\psi \in \lambda(h\phi)$. If instead $\neg \langle \Uparrow \phi \rangle \psi \in \lambda(h)$ then by (1) and Ax7 we have $\langle \Uparrow \phi \rangle \neg \psi \in \lambda(h)$. It follows by construction that $\neg \psi \in \lambda(h\phi)$. Therefore for each $\phi \in TBR\mathcal{L}$ we have either $\neg \psi \in \lambda(h\phi)$ or $\neg \psi \in \lambda(h\phi)$.

Now we have to prove $\lambda(h\phi)$ is consistent. Assume for *contradictio* that it is not. Then by definition we have a finite set of formulas $\{\phi_1, \ldots, \phi_m\} \subseteq \lambda(h\phi)$

such that $\vdash (\bigwedge_{i=1}^{m} \phi_i) \rightarrow \perp$ (2). It follows from (2) by standard modal reasoning that $\vdash \langle \Uparrow \phi \rangle \top \rightarrow \langle \Uparrow \phi \rangle (\bigvee_{i=1}^{m} \neg \phi_i)$. It follows by Ax5 again that $\vdash \langle \Uparrow \phi \rangle \top \rightarrow (\bigvee_{i=1}^{m} \langle \Uparrow \phi \rangle \neg \phi_i)$ (3). By (1) and (3) it follows that $(\bigvee_{i=1}^{m} \langle \Uparrow \phi \rangle \neg \phi_i) \in \lambda(h)$ (4). But since by IH $\lambda(h)$ is a MCSs, there is some j such that $1 \leq j \leq m$ and $\langle \Uparrow \phi \rangle \neg \phi_j \in \lambda(h)$ (5). From (5) and Ax7 we have $\neg \langle \Uparrow \phi \rangle \phi_j \in \lambda(h)$ (6). From (2) it follows by contruction that $\langle \Uparrow \phi \rangle \phi_i \in \lambda h$ for each i such that $1 \leq i \leq m$ (7). But (6) and (7) together contradicts the fact that $\lambda(h)$ is consistent. It follows by *reductio* that $\lambda(h\phi)$ is consistent. □

Now we need to prove a Truth Lemma.

Lemma 6.2 (Truth Lemma)**.** *For every $\phi \in TBR\mathcal{L}$, for each $h \in H^{\Sigma}$ we have:*

$$\phi \in \lambda(h) \text{ iff } \mathcal{H}^{\Sigma}, h \Vdash \phi$$

Proof. The proof is by induction on the complexity of ϕ. Base case (for atomic formulas) and boolean cases are easy.

[$\mathtt{A}_{\equiv}$-modality.] From left to right. Assume that $\mathtt{A}_{\equiv}\psi \in \lambda(h)$ (0). There are two cases. Either $h \in H^0$ (1) or $h \in (H^{\Sigma} - H^0)$ (2).

Let us consider the first case. Assume that $h, h' \in H^0$ and that $h \equiv^{\Sigma} h'$ (3). It follows from (3) by construction that $\{\phi \mid \mathtt{A}_{\equiv}\phi \in \lambda(h)\} \subseteq \lambda(h')$ (4). From (4) and (0) we know in particular that $\psi \in \lambda(h')$ (5). By (5) and the IH of the main induction on formulas it follows that $\mathcal{H}^{\Sigma}, h' \Vdash \psi$ (6). Since h' was arbitrary, it follows therefore from (6) and truth definition of $\mathtt{A}_{\equiv}$ that $\mathcal{H}^{\Sigma}, h \Vdash \mathtt{A}_{\equiv}\psi$ (7).

Let us now the consider the second case: $h \in (H^{\Sigma} - H^0)$ (2). Without loss of generality let us assume that h is of the form $w\phi_1 \dots \phi_{n+1}$ (8). From (8) and (0) it follows by construction that $\langle \Uparrow \phi_{n+1} \rangle \mathtt{A}_{\equiv}\psi \in \lambda(w\phi_1 \dots \phi_n)$ (9). Since by Lemma 6.1 $\lambda(w\phi_1 \dots \phi_n)$ is a $\mathtt{Stbr}$-MCS, it follows from (9) and (Th1*a*) that $\langle \Uparrow \phi_{n+1} \rangle \top \in \lambda(w\phi_1 \dots \phi_n)$ (10) and $\mathtt{A}_{\equiv}(\langle \Uparrow \phi_{n+1} \rangle \top \rightarrow \langle \Uparrow \phi_{n+1} \rangle \psi) \in \lambda(w\phi_1 \dots \phi_n)$ (11). Iterating the same argument we find that:

$$\mathtt{A}_{\equiv}(\langle \Uparrow \phi_1 \rangle \top \rightarrow (\langle \Uparrow \phi_2 \rangle \top \rightarrow \dots (\langle \Uparrow \phi_n \rangle \top \rightarrow (\langle \Uparrow \phi_{n+1} \rangle \top \rightarrow \langle \Uparrow \phi_{n+1} \rangle \psi)) \dots)) \in \lambda(w) \tag{12}$$

and that for each i, $1 \leq k \leq n$

$$\langle \Uparrow \phi_{k+1} \rangle \top \in \lambda(w\phi_1 \dots \phi_k) \tag{13}$$

and that

$$\langle \Uparrow \phi_1 \rangle \top \in \lambda(w) \tag{14}$$

Now assume that $h \equiv^{\Sigma} h'$ (15). It follows from (15) by construction and some easy induction that h' is of the form $v\phi_1 \dots \phi_{n+1}$ (16). Similarly it follows from (15) by construction and some induction that $w \equiv^{\Sigma} v$ (17). It follows by construction from (17) and (12) that $(\langle \Uparrow \phi_1 \rangle \top \rightarrow (\langle \Uparrow \phi_2 \rangle \top \rightarrow \dots (\langle \Uparrow \phi_n \rangle \top \rightarrow (\langle \Uparrow \phi_{n+1} \rangle \top \rightarrow \langle \Uparrow \phi_{n+1} \rangle \psi)) \dots)) \in \lambda(v)$ (18). But it is then easy to check that from (18), (13) and (14) we have $\psi \in h'$ (19). It follows from (19) by the main IH that $\mathcal{H}, h' \Vdash \psi$. But since h' was arbitrary, we have: $\mathcal{H}^{\Sigma}, h \Vdash \mathtt{A}_{\equiv}\psi$.

[$\mathtt{A}_\equiv$-modality.] From right to left. Assume that $\mathtt{A}_\equiv\psi \notin \lambda(h)$ (0). There are two cases. Either $h \in H^0$ (1) or $h \in (H^\Sigma - H^0)$ (2).

Let us consider the first case. Assume that $\mathtt{A}_\equiv\psi \notin \lambda(h)$ (0). We have to prove that $\mathcal{H}^\Sigma, h \nVdash \mathtt{A}_\equiv\psi$. To prove (0) it is sufficient to find a MCSs $\lambda(h')$ such that $\psi \notin \lambda(h')$, but $\{\phi \mid \mathtt{A}_\equiv\phi \in \lambda(h)\} \subseteq \lambda(h')$. By Lindenbaum Lemma, it is sufficient to show that $v_0 = \{\neg\psi\} \cup \{\phi \mid \mathtt{A}_\equiv\phi \in \lambda(h)\}$ is consistent. Assume for *contradictio* that it is not. Then we have a finite set of formulas $\phi_1 \dots \phi_m \in \{\phi \mid \mathtt{A}_\equiv\phi \in \lambda(h)\}$ such that $(\bigwedge_{i=1}^m \phi_i) \rightarrow \psi$ (3). But then it follows by standard modal reasoning that $(\bigwedge_{i=1}^m \mathtt{A}_\equiv\phi_i) \rightarrow \mathtt{A}_\equiv\psi$ (4). But since $\lambda(h)$ is a MCS, it follows from (4) that $\mathtt{A}_\equiv\psi \in \lambda(h)$ which contradicts (0).

Let us now consider the other case. Assume WLOG that $h = w\phi_1 \dots \phi_{n+1}$. Now assume that $\mathtt{A}_\equiv\psi \notin \lambda(h)$ (0). It follows from (0) by maximality of $\lambda(h)$ and construction that $\langle\Uparrow \phi_{n+1}\rangle\neg\mathtt{A}_\equiv\psi \in \lambda(w\phi_1 \dots \phi_n)$ (1). It follows from (1) by Ax7 that $\langle\Uparrow \phi_{n+1}\rangle\top \in \lambda(w\phi_1 \dots \phi_n)$ (2) and $\neg\langle\Uparrow \phi_{n+1}\rangle\mathtt{A}_\equiv\psi \in \lambda(w\phi_1 \dots \phi_n)$ (3). It follows from (2), (3), an easy argument give us by Th1 that $\neg\mathtt{A}_\equiv(\langle\Uparrow \phi_{n+1}\rangle\top \rightarrow \langle\Uparrow \phi_{n+1}\rangle\psi) \in \lambda(w\phi_1 \dots \phi_n)$ (4). Repeting this argument we find that $\neg\mathtt{A}_\equiv(\langle\Uparrow \phi_1\rangle\top \rightarrow \langle\Uparrow \phi_1\rangle(\langle\Uparrow \phi_2\rangle\top \rightarrow (\dots\langle\Uparrow \phi_n\rangle(\langle\Uparrow \phi_{n+1}\rangle\top \rightarrow \langle\Uparrow \phi_{n+1}\rangle\psi)\dots))) \in \lambda(w)$ (5).

We will know prove that there is a v such that $w\phi_1 \dots \phi_{n+1} \equiv v\phi_1 \dots \phi_{n+1}$ and $\psi \notin \lambda(v\phi_1 \dots \phi_{n+1})$ (7). First take the following set $v_0 = \{\chi \mid A \equiv \chi \in \lambda(w)\} \cup \{\neg\langle\Uparrow \phi_1\rangle\top \rightarrow \langle\Uparrow \phi_1\rangle(\langle\Uparrow \phi_2\rangle\top \rightarrow (\dots\langle\Uparrow \phi_n\rangle(\langle\Uparrow \phi_{n+1}\rangle\top \rightarrow \langle\Uparrow \phi_{n+1}\rangle\psi)\dots))\}$. Assume for contradictio that v_0 is inconsistent (8). It follows from (8) that there is a finite set of formumas $\{\chi_1, \dots \chi_m\} \subseteq \{\chi \mid A \equiv \chi \in \lambda(w)\}$ such that $\vdash (\bigwedge_{i=1}^m \chi_i) \rightarrow ((\langle\Uparrow \phi_1\rangle\top \rightarrow \langle\Uparrow \phi_1\rangle(\langle\Uparrow \phi_2\rangle\top \rightarrow (\dots\langle\Uparrow \phi_n\rangle(\langle\Uparrow \phi_{n+1}\rangle\top \rightarrow \langle\Uparrow \phi_{n+1}\rangle\psi)\dots)))$ (9). But from (9) and standard modal reasoning we find that $\vdash (\bigwedge_{i=1}^m \mathtt{A}_\equiv\chi_i) \rightarrow \mathtt{A}_\equiv(\langle\Uparrow \phi_1\rangle\top \rightarrow \langle\Uparrow \phi_1\rangle(\langle\Uparrow \phi_2\rangle\top \rightarrow (\dots\langle\Uparrow \phi_n\rangle(\langle\Uparrow \phi_{n+1}\rangle\top \rightarrow \langle\Uparrow \phi_{n+1}\rangle\psi)\dots)))$ (10). But since $\lambda(h)$ is a MCS, it follows from (10) that $\mathtt{A}_\equiv(\langle\Uparrow \phi_1\rangle\top \rightarrow \langle\Uparrow \phi_1\rangle(\langle\Uparrow \phi_2\rangle\top \rightarrow (\dots\langle\Uparrow \phi_n\rangle(\langle\Uparrow \phi_{n+1}\rangle\top \rightarrow \langle\Uparrow \phi_{n+1}\rangle\psi)\dots))) \in \lambda(h)$ which contradicts (5). Thus by *reductio* v_0 is consistent. By Lindenbaum Lemma we can extend v_0 to a maximally consistent v. But by construction $v \in H^\Sigma$ (11). Since $\{\chi \mid A \equiv \chi \in \lambda(w)\} \subseteq v_0 \subseteq v$ it follows by construction that $w \equiv v$. But then an easy induction shows that for every $1 \leq j \leq n+1$ we have:

$$w\phi_1 \dots \phi_j \equiv v\phi_1 \dots \phi_j \qquad (A_\equiv)$$

Since by construction $\neg\langle\Uparrow \phi_1\rangle\top \rightarrow \langle\Uparrow \phi_1\rangle(\langle\Uparrow \phi_2\rangle\top \rightarrow (\dots\langle\Uparrow \phi_n\rangle(\langle\Uparrow \phi_{n+1}\rangle\top \rightarrow \langle\Uparrow \phi_{n+1}\rangle\psi)\dots)) \in \lambda(v)$ it follows that $\langle\Uparrow \phi_1\rangle\top \in \lambda(v)$ (12) and $\neg\langle\Uparrow \phi_1\rangle(\langle\Uparrow \phi_2\rangle\top \rightarrow (\dots\langle\Uparrow \phi_n\rangle(\langle\Uparrow \phi_{n+1}\rangle\top \rightarrow \langle\Uparrow \phi_{n+1}\rangle\psi)\dots))$ (13). But it follows from (12), (13) and Ax7, that $\langle\Uparrow \phi_1\rangle\neg(\langle\Uparrow \phi_2\rangle\top \rightarrow (\dots\langle\Uparrow \phi_n\rangle(\langle\Uparrow \phi_{n+1}\rangle\top \rightarrow \langle\Uparrow \phi_{n+1}\rangle\psi)\dots)) \in \lambda(v)$ (14). But then by construction $\neg(\langle\Uparrow \phi_2\rangle\top \rightarrow (\dots\langle\Uparrow \phi_n\rangle(\langle\Uparrow \phi_{n+1}\rangle\top \rightarrow \langle\Uparrow \phi_{n+1}\rangle\psi)\dots)) \in \lambda(v\phi_1)$ (15). Repeting this argument we find that $\neg\psi \in \lambda(v\phi_1 \dots \phi_{n+1})$ (16). By Lemma 6.2 $\lambda(v\phi_1 \dots \phi_{n+1})$ is consistent, (16) therefore implies that $\psi \notin \lambda(v\phi_1 \dots \phi_{n+1})$ (17). But $(A_\equiv)$, (11) and (17) is all we need to prove (7). We can now apply the main IH on formulas to get $\mathcal{H}^\equiv, v\phi_1 \dots \phi_{n+1} \nVdash \psi$ (18). By (18) and the truth condition of $A_\equiv$ it follows that $\mathcal{H}^\Sigma, h \Vdash \mathtt{A}_\equiv\psi$. Concluding the proof for this direction of for the $A_\equiv$-subcase.

[[≤]-modality.] From left to right. Assume that $\mathtt{A}_\equiv\psi \in \lambda(h)$ (0). There are two cases. Either $h \in H^0$ (1) or $h \in (H^\Sigma - H^0)$ (2).

Let us consider the first case. Assume that $h, h' \in H^0$ and that $h \leq^\Sigma h'$ (3). It follows from (3) by construction that $\{\phi \mid [\leq]\phi \in \lambda(h)\} \subseteq \lambda(h')$ (4). From (4) and (0) we know in particular that $\psi \in \lambda(h')$ (5). By (5) and the IH of the main induction on formulas it follows that $\mathcal{H}^{\Sigma}, h' \Vdash \psi$ (6). Since h' was arbitrary, it follows therefore from (6) and truth definition of $[\leq]$ that $\mathcal{H}^{\Sigma}, h \Vdash [\leq]\psi$ (7).

Let us now the consider the second case: $h \in (H^{\Sigma} - H^{0})$ (2). For simplicity we assume that h is of the form $w\phi$ (8). The proof can be generalized along the lines of the $\mathtt{A}_{\equiv}$-case. From (8) and (0) it follows by construction that $\langle\Uparrow \phi\rangle[\leq]\psi \in \lambda(w\phi)$ (9). Since by Lemma 6.1 $\lambda(w)$ is a $\mathtt{Stbr}$-MCS, it follows from (9) and (Th1*b*) that $\langle\Uparrow \phi_{n+1}\rangle\top \in \lambda(w)$ (10) and $[(\phi \rightarrow \mathtt{A}_{\equiv}(\neg\phi \rightarrow \neg\langle\Uparrow \phi\rangle\neg\psi)) \wedge (\phi \rightarrow [\leq_i]\neg\langle\Uparrow \phi\rangle\neg\psi) \wedge [\leq_i](\neg\phi \rightarrow \neg\langle\Uparrow \phi\rangle\neg\psi)] \in \lambda(w)$ (11).

Now assume that $h \leq^{\Sigma} h'$ (15). It follows from (15) by construction h' is of the form $v\phi$ (16). By (15) we know that we are in one of the following cases:

1. $\phi \in \lambda(w)$ while $\phi \notin \lambda(v)$
2. $\phi \in \lambda(w)$, $\phi \in \lambda(v)$, and $w \leq_i v$
3. $\phi \notin \lambda(w)$, $\phi \notin \lambda(v)$, and $w \leq_i v$.

Case 1: $\phi \in \lambda(w)$ (17.1) while $\phi \notin \lambda(v)$ (17.2). By (11) we have: $(\phi \rightarrow \mathtt{A}_{\equiv}(\neg\phi \rightarrow \neg\langle\Uparrow \phi\rangle\neg\psi)) \in w$ (17.3). It follows from (17.1), (17.3) and Lemma 6.2 that $(\mathtt{A}_{\equiv}(\neg\phi \rightarrow \neg\langle\Uparrow \phi\rangle\neg\psi)) \in w$ (17.4). From (17.4) we have by construction $(\neg\phi \rightarrow \neg\langle\Uparrow \phi\rangle\neg\psi) \in v$ (17.5). But it is easy to check that $\neg\phi \in v$ (17.6). Thus by (17.5), (17.6) and Lemma 6.2 we have $\neg\langle\Uparrow \phi\rangle\neg\psi \in v$ (17.7). Again it is easy to check that $\langle\Uparrow \phi\rangle\top \in v$ (17.8). But (17.7), (17.8), Ax7 and Lemma 6.2 gives us $\langle\Uparrow \phi\rangle\psi \in v$ (17.9). By construction (17.9) gives us $\psi \in v\phi$ (17.10). (17.10) gives us by IH $h' \Vdash \psi$ (17.11). But since h' was arbitrary, we have: $\mathcal{H}^{\Sigma}, h \Vdash [\leq]\psi$ (17.12).

Case 2: $\phi \in \lambda(w)$ (18.1), $\phi \in \lambda(v)$ (18.2) and $w \leq v$ (18.3). By (11) we have: $(\phi \rightarrow [\leq_i]\neg\langle\Uparrow \phi\rangle\neg\psi) \in w$ (17.3). An easy argument gives us $[\leq_i]\neg\langle\Uparrow \phi\rangle\neg\psi \in w$ (17.4). Thus by construction $\neg\langle\Uparrow \phi\rangle\neg\psi \in v$ (17.5). By construction we have also $\langle\Uparrow \phi\rangle\top \in v$ which together with (17.5) and Ax7 gives us $\langle\Uparrow \phi\rangle\psi \in v$ (17.6). (17.6) implies by construction that $\psi \in v$. The usual argument concludes the proof.

Case 3: $\phi \notin \lambda(w)$ (19.1), $\phi \notin \lambda(v)$ (19.2) and $w \leq v$ (19.3). By (11) we have: $[\leq_i](\neg\phi \rightarrow \neg\langle\Uparrow \phi\rangle\neg\psi) \in w$ (19.4). An easy argument gives us $[\leq_i]\neg\langle\Uparrow \phi\rangle\neg\psi \in w$ (17.4). Thus by construction $\neg\langle\Uparrow \phi\rangle\neg\psi \in v$ (17.5). By construction we have also $\langle\Uparrow \phi\rangle\top \in v$ which together with (17.5) and Ax7 gives us $\langle\Uparrow \phi\rangle\psi \in v$ (17.6). (17.6) implies by construction that $\psi \in v$. The usual argument concludes the proof for this case and this direction.

[$[\leq_i]$-modality.] From right to left. Assume that $[\leq_i]\psi \notin \lambda(h)$ (0). There are two cases. Either $h \in H^0$ (1) or $h \in (H^{\Sigma} - H^0)$ (2).

The first case is along line the lines of the proof in the previous section.

Let us now consider the other case. For the sake of simplicity we assume that $h = w\phi$. The proof can be generalized along the lines of the $A_{\equiv}$-case. Now assume that $[\leq_i]\psi \notin \lambda(h)$ (0). It follows from (0) by maximality of $\lambda(h)$ and construction that $\langle\Uparrow \phi\rangle\neg[\leq_i]\psi \in \lambda(w)$ (1). It follows from (1) by Ax7 that $\langle\Uparrow \phi\rangle\top \in \lambda(w)$ (2) and $\neg\langle\Uparrow \phi\rangle[\leq_i]\psi \in \lambda(w)$ (3).

Given that (2) and (3), an easy argument give us by Th1*b* that $\neg[(\phi \rightarrow \mathtt{A}_{\equiv}(\neg\phi \rightarrow \neg\langle\Uparrow \phi\rangle\neg\psi)) \wedge (\phi \rightarrow [\leq_i]\neg\langle\Uparrow \phi\rangle\neg\psi) \wedge [\leq_i](\neg\phi \rightarrow \neg\langle\Uparrow \phi\rangle\neg\psi)] \notin w$ and thus that $\neg(\phi \rightarrow \mathtt{A}_{\equiv}(\neg\phi \rightarrow \neg\langle\Uparrow \phi\rangle\neg\psi)) \vee \neg(\phi \rightarrow [\leq_i]\neg\langle\Uparrow \phi\rangle\neg\psi) \vee \neg[\leq_i](\neg\phi \rightarrow \neg\langle\Uparrow \phi\rangle\neg\psi)]$ (4). The preceding disjunction naturally displays three cases.

Case 1: $\neg(\phi \rightarrow \mathtt{A}_{\equiv}(\neg\phi \rightarrow \neg\langle\Uparrow \phi\rangle\neg\psi)) \in w$ (4.1). Since by Lemma 6.2 $\phi \in w$ (4.2) and $\neg\mathtt{A}_{\equiv}(\neg\phi \rightarrow \neg\langle\Uparrow \phi\rangle\neg\psi) \in w$ (4.3). We will now prove that there is a v such that $w \equiv v$, $\phi \notin v$ and $\psi \notin \lambda(v\phi)$. First take the following set $v_0 = \{\chi \mid A \equiv \chi \in \lambda(w)\} \cup \{\neg\phi \wedge \langle\Uparrow \phi\rangle\neg\psi\}$ (4.5). Assume for *contradictio* that v_0 is inconsistent (4.6). It follows from (4.6) that there is a finite set of formulas $\{\chi_1, \dots \chi_m\} \subseteq \{\chi \mid A \equiv \chi \in \lambda(w)\}$ such that $\vdash (\bigwedge_{i=1}^{m} \chi_i) \rightarrow (\neg\phi \rightarrow \neg\langle\Uparrow \phi\rangle\neg\psi)$ (4.7). But from (4.7) and standard modal reasoning we find that $\vdash (\bigwedge_{i=1}^{m} \mathtt{A}_{\equiv}\chi_i) \rightarrow \mathtt{A}_{\equiv}(\neg\phi \rightarrow \neg\langle\Uparrow \phi\rangle\neg\psi)$ (4.8). But from (4.8), (4.2) and Lemma 6.2 we have $\mathtt{A}_{\equiv}(\neg\phi \rightarrow \neg\langle\Uparrow \phi\rangle\neg\psi)$ (4.9). But (4.9) contradicts (4.3), thus by *reductio* v_0 is consistent. By Lindenbaum Lemma it can be extended to a maximal consistent set v^+. But by construction $v^+ \in H^\Sigma$ (4.10). Since $\{\chi \mid A \equiv \chi \in \lambda(w)\} \subseteq v_0 \subseteq v^+$ it follows by construction that $w \equiv v^+$ (4.11). By construction of v_0 we have also $\neg\phi \in v_0 \subseteq v^+$ (4.12). But it follows from (4.2), (4.12) and (4.11) by construction that $w\phi \leq v\phi$ (4.13). But since by (4.5) and a simple argument $\langle\Uparrow \phi\rangle\neg\psi \in v$ too, we have $\psi \notin v\phi$ (4.14). But (4.13) and (4.14) gives us $w\phi \nVdash [\leq_i]\psi$. Concluding our proof for this subcase.

Case 2: $\neg(\phi \rightarrow [\leq_i]\neg\langle\Uparrow \phi\rangle\neg\psi) \in w$ (5.1). It follows from (5.1) by Lemma 6.2 that $\phi \in w$ (5.2) and that $\neg[\leq]\neg\langle\Uparrow \phi\rangle\neg\psi \in w$ (5.3). By Ax7 it follows that $\neg[\leq](\langle\Uparrow \phi\rangle\top \rightarrow \langle\Uparrow \phi\rangle\psi) \in w$. The right to left argument of the $\mathtt{A}_{\equiv}$-case applies replacing $\equiv$ by $\leq$, i.e. we construct v^+ such that by construction $w \leq v$ (5.4) and $\langle\Uparrow \phi\rangle\neg\psi \in v^+$ (5.5). We find by (5.5) and construction that $\psi \notin v\phi$ (5.6). But since by construction (5.4) and (5.2) give us $w\phi \leq v\phi$ (5.7). The rest of the argument is as usual.

Case 3: $\neg[\leq_i](\neg\phi \rightarrow \neg\langle\Uparrow \phi\rangle\neg\psi) \in w$ (6.1). An easy argument show that by (6.1) we have $\neg[\leq_i](\neg\phi \rightarrow (\langle\Uparrow \phi\rangle\top \rightarrow \langle\Uparrow \phi\rangle\psi)) \in w$ (6.2). We construct v_0 such that $\neg\phi \in v_0$ (6.3), $\langle\Uparrow \phi\rangle\top \in v_0$ (6.4) but $\langle\Uparrow \phi\rangle\psi \notin v_0$ (6.5) that can be extended it to a MCS v^+ (6.6) such that $w \leq v^+$ (6.7). By (6.6), (6.7) and (6.3) we have by construction $w\phi \leq v\phi$ (6.8). But from (6.3) and (6.8) an easy argument gives the desired result by the main IH. Concluding the proof for this subcase, this direction and the case.

The case of $[\geq]$ is similar. □

We have shown that for every consistent set of formulae, there was a initial state in our canonical forest such that all formulae in the set where satisfied there. We finally need to show that our canonical forest can be generated by an initial plausibility model and a dynamic belief revision protocol.

Lemma 6.3 (The canonical forest is isomorphic to a forest generated by an initial plausibility model and a dynamic belief revision protocol)**.**

Proof. We start by proving that the initial part of the canonical forest is isomorphic to a plausibility model $\mathcal{M}$ (1). This proof follows by a classical Sahlqvist correspondence argument (Blackburn et al. 2001, ch. 4). It remains to prove that that there is a dynamic belief revision protocol $\mathtt{p} : W \rightarrow Ptcl(\mathcal{L})$ such that the forest generated by the initial part of the canonical forest according the protocol $\mathtt{p}$. Pick $\mathtt{p}$ such that $\mathtt{p}(w) = \{\sigma \mid w\sigma \in H^\Sigma\}$. We claim that a history $w\sigma \in H^\Sigma$ iff $h\sigma \in H(\mathcal{M}, p) = Dom(\mathcal{H}(\mathcal{M}, p))$. (The rest of the Lemma – i.e. the clauses for the plausibility relation and for the valuation function – follows by construction as the reader can check by inspecting clauses in Definition 6.2 and Definition 6.7). The proof is by induction on the length of $w\sigma$. The base case is immediate from (1). Now assume that that the equivalence holds for histories of length n. Assume that $w\sigma\phi$ is of length $n+1$ (2) and $w\sigma\phi \in H^\Sigma$ (3). It follows by construction of the canonical model that $w\sigma \in H^\Sigma$ (4). But then by IH and(4) we have $w\sigma \in H(\mathcal{M}, p)$ (5). It follows by construction of $\mathtt{p}$ and (3) that $\sigma\phi \in \mathtt{p}$ (6). But then by Definition 6.2 , (5) and (6) we have $w\sigma\phi \in H(\mathcal{M}, p)$ (7). The other direction is similar. □

Theorem 6.4 (Completeness)**.** `Stbr` *is sound and strongly complete with respect to the class of DoTL models generated by a sequence of protocol-based lexicographic upgrade.*

Proof. The proof follows from Truth Lemma and preceding Lemma by a standard argument Blackburn et al. (2001). □

6.2 More languages.

The preceding language is somewhat natural in terms of expressive power and pleasant to work with due to its close connection to the model-theory. One might be however interested in other languages. First of all do we have a direct pseudo-reduction axiom for conditional belief. The answer is yes.

Proposition 6.5. *The following axiom is sound with respect to the class of DoTL models generated by a sequence of protocol-based lexicographic upgrade.*

$$\begin{aligned} \langle \Uparrow \phi \rangle B_i^{\psi} \chi \ \leftrightarrow \ \langle \Uparrow \phi \rangle \top \wedge & \\ & [(\mathtt{E}(\phi \ \wedge \ \langle \Uparrow \phi \rangle \psi) \ \wedge \ B_i^{\phi \wedge \langle \Uparrow \phi \rangle \psi} \langle \Uparrow \phi \rangle \chi) \\ & \vee \ (\neg \mathtt{E}(\phi \ \wedge \ \langle \Uparrow \phi \rangle \psi) \ \wedge \ B_i^{\langle \Uparrow \phi \rangle \psi} \langle \Uparrow \phi \rangle \chi)] \end{aligned} \tag{6}$$

It is also natural to be willing to work both with information sets and a plausibility pre-order. Agents' beliefs being then given as the best state (according to the plausibility order) within the current information set. Both public announcement and soft update can in such a setting happily and fruitfully live together. Here is an example of such a *doxastic-epistemic* language.

$$\phi := p \mid \neg \phi \mid \phi \ \vee \ \phi \mid K_i \phi \mid \langle \leq_i \rangle \phi \mid \langle \geq_i \rangle \phi \mid \langle \sim_i \cap \geq_i \rangle \phi \mid \mathtt{E}_{\equiv} \phi \mid B_i^K \phi \mid \langle \Uparrow \phi \rangle \phi \mid$$

where i ranges over N, ψ over $\mathcal{L}_{Dox}(\mathtt{E}, \geq)$ and p over a countable set of proposition letters *Prop*. We make use of the usual shortcuts. In particular the dual of the knowledge operator will be written $\hat{K}$ and defined as usual.

By lack of space we are not introducing of the definition of doxastic-epistemic models generated by a sequence of protocol-based lexicographic upgrade, but the key idea is that information sets are never changed by lexicographic upgrades, the rest of the definition is just as for purely doxastic models. We turn to the new clauses of the truth definition.

Definition 6.8 (Truth definition)**.** Assuming that the floor of our forests is finite, the language are interpreted as previously. We display only the new cases. Let $K_i[h] = \{h' \mid h \sim_i h'\}$.

$\mathcal{H}, wh, g \Vdash K_i \phi$	iff	$\forall h'$ such that $wh \sim_i h'$ we have $\mathcal{H}, h', g \Vdash \phi$
$\mathcal{H}, wh, g \Vdash \langle \sim_i \cap \geq_i \rangle \phi$	iff	$\exists h'$ such that $h' \leq_i wh$, $wh \sim_i h'$ and $\mathcal{H}, h', g \Vdash \phi$
$\mathcal{H}, wh, g \Vdash B_i^K \phi$	iff	$\forall h'$ such that $h' \in Min(K_i[wh], \leq_i)$ we have $\mathcal{H}, h', g \Vdash \phi$

We now have a syntactic look at their behavior with respect to protocol-based lexicographic update.

Proposition 6.6. *The following axioms is sound with respect to the class of DoTL models generated by a sequence of protocol-based lexicographic upgrade.*

$$\langle\Uparrow\phi\rangle K_i^{\psi} \leftrightarrow \langle\Uparrow\phi\rangle\top \wedge K_i(\langle\Uparrow\phi\rangle\top \rightarrow \langle\Uparrow\phi\rangle\psi) \tag{7}$$

$$\begin{aligned}\langle\Uparrow\phi\rangle\langle\sim_i \cap \geq_i\rangle\psi \leftrightarrow \langle\Uparrow\phi\rangle\top \wedge & \\ & \neg\phi \wedge \hat{K}_i(\phi \wedge \langle\Uparrow\phi\rangle\psi) \\ & \neg\phi \wedge \langle\sim_i \cap \geq_i\rangle(\neg\phi \wedge \langle\Uparrow\phi\rangle\psi) \\ & \phi \wedge \langle\sim_i \cap \geq_i\rangle(\phi \wedge \langle\Uparrow\phi\rangle\psi)\end{aligned} \tag{8}$$

$$\begin{aligned}\langle\Uparrow\phi\rangle B_i^K\psi \leftrightarrow \langle\Uparrow\phi\rangle\top \wedge & \\ & [(\hat{K}_i(\phi \wedge \langle\Uparrow\phi\rangle\top) \wedge K_i(\langle\sim_i \cap \geq_i\rangle((\phi \wedge \langle\Uparrow\phi\rangle\top) \wedge \\ & [\sim_i \cap \geq_i]((\phi \wedge \langle\Uparrow\phi\rangle\top) \rightarrow \langle\Uparrow\phi\rangle\psi)))) \\ & \vee (K_i(\phi \rightarrow \neg\langle\Uparrow\phi\rangle\top) \wedge K_i(\langle\sim_i \cap \geq_i\rangle((\langle\Uparrow\phi\rangle\top) \wedge \\ & [\sim_i \cap \geq_i]((\langle\Uparrow\phi\rangle\top \rightarrow \langle\Uparrow\phi\rangle\psi))))]\end{aligned} \tag{9}$$

7 Conclusion

Agents that update their knowledge and revise their beliefs can behave very differently over time. We have determined the special constraints that capture agents operating with the 'local updates' of dynamic doxastic logic. This took the form of some representation theorems that state just when a general doxastic temporal model is equivalent to the forest model generated by successive priority updates of an initial doxastic model by a protocol sequence of event models. We have also shown how these conditions can be defined in an appropriate extended modal language, making it possible to reason formally about agents engaged in such updates and revisions. Finally we have developed a systematic "protocol logic" of axiomatic completeness for constrained revision processes, analogous to the purely epistemic theory of observation and conversation protocols initiated in Benthem et al. (2008),

Our methods are like those of existing epistemic work, but the doxastic case came with some interesting new notions.

As for open problems, the paper has indicated several technical issues along the way, e.g., concerning the expressive power of different languages over our models and their complexity effects (cf. Benthem and Pacuit (2006) for the epistemic case). In particular, we have completely omitted issues of common knowledge and common belief, even though these are known to generate complications Benthem et al. (2006).

But from where we are standing now, we see several larger directions to pursue:

- A comparison of our 'constructive' *DDL*-inspired approach to *DTL* universes with the more abstract *AGM*-style postulational approach of Bonanno (2006),

- A theory of variation for different sorts of agents with different abilities and tendencies, as initiated in Liu (2008),
- An analysis of knowledge and belief dynamics in games Benthem (2007), Dégremont and Zvesper (2007), Baltag et al. (2008)
- Connections with formal learning theory over epistemic-doxastic temporal universes (cf. Kelly (2008)).

Acknowledgements.

The authors would like to thank Tomohiro Hoshi and Eric Pacuit for very useful comments and discussions. The authors would also like to thank Olivier Roy, Hans van Ditmarsch, Nina Gierasimczuk, Valentin Goranko, Ramaswamy Ramanujam and Jonathan A. Zvesper for their comments on earlier versions of this paper.

References

A. Baltag and L. S. Moss. Logics for Epistemic Programs. *Synthese*, 139(2): 165–224, 2004.

A. Baltag and S. Smets. Dynamic belief revision over multi-agent plausibility models. In G. Bonanno, W. van der Hoek, and M. Wooldridge, editors, *Logic and the Foundations of Game and Decision Theory: Proceedings of LOFT'06*, Texts in Logic and Games, pages 11–24. Amsterdam University Press, 2006.

A. Baltag and S. Smets. A qualitative theory of dynamic interactive belief revision. In G. Bonanno, W. van der Hoek, and M. Wooldridge, editors, *Logic and the Foundation of Game and Decision Theory (LOFT7)*, volume 3 of *Texts in Logic and Games*, pages 13–60. Amsterdam University Press, 2008.

A. Baltag, L. S. Moss, and S. Solecki. The logic of public announcements, common knowledge, and private suspicions. In *TARK '98: Proceedings of the 7th conference on Theoretical aspects of rationality and knowledge*, pages 43–56, San Francisco, CA, USA, 1998. Morgan Kaufmann Publishers Inc. ISBN 1558605630. URL `http://portal.acm.org/citation.cfm?id=671885`.

A. Baltag, S. Smets, and J. A. Zvesper. Dynamic Rationality in Extensive Games. In J. v. Benthem and E. Pacuit, editors, *Proceedings of Workshop on Logic and Intelligent Interaction, ESSLLI*, 2008.

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

J. v. Benthem. Games in Dynamic Epistemic Logic. *Bulletin of Economic Research*, 53(4):219–248, 2001.

J. v. Benthem. Priority product update as social choice. *ILLC Prepublications Series*, PP-2008-09, 2008.

J. v. Benthem. Rational dynamics and epistemic logic in games. *International Game Theory Review*, 9(1):377–409, 2007. Erratum reprint, Volume 9(2), 2007, 377 - 409.

J. v. Benthem and C. Dégremont. Multi-agent belief dynamics: bridges between dynamic doxastic and doxastic temporal logics. Technical report, ILLC, University of Amsterdam, 2008. http://staff.science.uva.nl/~cdegremo/.

J. v. Benthem and E. Pacuit. The Tree of Knowledge in Action: Towards a Common Perspective. In I. H. G. Governatori and Y. Venema, editors, *Advances in Modal Logic*, volume 6. College Publications, 2006.

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

J. v. Benthem, J. Gerbrandy, and E. Pacuit. Merging frameworks for interaction: DEL and ETL. In D. Samet, editor, *Proceedings of Theoretical Aspects of Rationality and Knowledge (TARK 2007)*, 2007.

J. v. Benthem, J. Gerbrandy, T. Hoshi, and E. Pacuit. Merging frameworks for interaction: DEL and ETL. Working paper, 2008.

P. Blackburn, M. de Rijke, and Y. Venema. *Modal Logic*, volume 53 of *Cambridge Tracts in Theoretical Computer Science*. Cambridge University Press, 2001.

O. Board. Dynamic interactive epistemology. *Games and Economic Behavior*, 49: 49–80, 2004.

G. Bonanno. Axiomatic characterization of the agm theory of belief revision in a temporal logic. *Artif. Intell.*, 171(2-3):144–160, 2007.

G. Bonanno. Belief revision in a temporal framework. In G. Bonanno, W. van der Hoek, and M. Wooldridge, editors, *Logic and the Foundations of Game and Decision Theory: Proceedings of LOFT'06*, Texts in Logic and Games, pages 43–50. Amsterdam University Press, 2006.

C. Boutilier. Conditional logics of normality: A modal approach. *Artificial Intelligence*, 68(1):87–154, 1994.

D. de Jongh and F. Liu. Optimality, belief and preference. In S. Artemov and R. Parikh, editors, *Proceedings of the Workshop on Rationality and Knowledge*. ESSLLI, Malaga, 2006.

C. Dégremont and J. A. Zvesper. Logique dynamique pour le raisonnement stratégique dans les jeux extensifs. In *Journées Francophones sur les Modèles Formels de l'Interaction (MFI'07)*, pages 61–73, 2007.

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

K. Eliaz and E. A. Ok. Indifference or indecisiveness? Choice-theoretic foundations of incomplete preferences. *Games and Economic Behavior*, 56(1):61–86, July 2006. URL http://ideas.repec.org/a/eee/gamebe/v56y2006i1p61-86.html.

R. Fagin, J. Y. Halpern, Y. Moses, and M. Y. Vardi. *Reasoning About Knowledge*. MIT Press, Cambridge, 1995. ISBN 0262061627.

P. Girard. *Modal Logic for Preference Change*. PhD thesis, Stanford, 2008.

K. Kelly. Ockham's razor, truth, and information. In J. van Behthem and P. Adriaans, editors, *Handbook of the Philosophy of Information*. 2008.

F. Liu. *Changing for the Better: Preference Dynamics and Agent Diversity*. Phd dissertation, ILLC Amsterdam, 2008.

S. Morris. The common prior assumption in economic theory. *Economics and Philosophy*, 11:227–253, 1995.

E. Pacuit. Some comments on history based structures. *Journal of Applied Logic*, 5(4):613–624, 2007.

R. Parikh and R. Ramanujam. A knowledge based semantics of messages. *Journal of Logic, Language and Information*, 12(4):453–467, 2003. ISSN 0925-8531. doi: http://dx.doi.org/10.1023/A:1025007018583.

R. Stalnaker. A theory of conditionals. In R. Stalnaker, W. L. Harper, and G. Pearce, editors, *Ifs: Conditionals, Belief, Decision, Chance and Time*. D. Reidel, Dordrecht, 1981.

Public Announcement Logic with Protocol Constraints

Tomohiro Hoshi
Department of Philosophy, Stanford University.
thoshi@stanford.edu

Abstract

There are two important ingredients in describing multi-agent intelligent interaction: one concerns how agents' epistemic states change over informational events; the other concerns what informational events can take place in the course of agents' interaction. As a version of *Dynamic Epistemic Logic*, *Public Announcement Logic* (PAL) is suitable for describing the former but less so for capturing the later. On the other hand, *Epistemic Temporal Logic* (ETL) is suitable for the second but less so for the first. The purpose of this paper is to merge PAL and ETL to provide a framework that well represents both of the two aspects. To achieve this goal, we assign to each point of a given epistemic model a set of sequences of public announcements, which we call a *protocol*. Then, by successively applying model relativizations based on the assigned protocols, we generate tree structures of ETL that represent all possible evolutions of the original epistemic model under the constraints represented by the protocol. We will study the logic of public announcement over the class of ETL-models generated by the method and provide a complete axiomatization.

1 Introduction

There are two important ingredients in describing multi-agent intelligent interaction. One concerns how agents' epistemic states change over informational events. As we can observe in the literature, informational events of the simplest kind could affect agents' knowledge in a very sensitive manner. Thus it is crucial to get a good grasp on informational events and their epistemic effects. The other aspect concerns what informational events can take place in the course of agents' interaction. For instance, our everyday conversations are (supposedly) governed by a certain kind of "courtesy protocols", such as "Do not blurt out everything at the beginning," "Let the others speak in turn", etc. Also, puzzles and games have concrete rules on how players should communicate their information. It is important to capture the kind of communication constraints in order to properly deal with some of the main questions in the literature, e.g.

whether or how epistemic states of our interest can be reached, etc.

Among the variety of multi-agent systems, *Dynamic Epistemic Logic* (DEL, e.g. Baltag et al. (1998), Gerbrandy (1999), van Ditmarsch et al. (2007)) is suitable for describing the first aspect, but less so for the second. For instance, *Public Announcement Logic* (PAL, e.g. Plaza (1989), van Benthem (2002)), the prime example of DEL, describes the dynamics of agents' epistemic states when true information is *publicly announced*. In PAL, the public announcement of a true statement φ is captured by the relativization of epistemic models to the set of states where φ is true. With such a framework, PAL models a general aspect of informational events that can be suitably interpreted as model relativization, such as observation, learning, etc. On the other hand, PAL is based on the assumption that whatever is true can be publicly announced. Thus it does not give us a way to represent communication constraints beyond truth on true information obtainable through public announcements. This feature of PAL is a general feature of all systems in DEL: although DEL provides systematic methods for performing model transformations in order to represent informational events, it does not adequately capture the constraints on what informational events can happen.

In contrast, *Epistemic Temporal Logic* (ETL, e.g. Parikh and Ramanujam (2003)) is suitable for the second aspect but less so for the first. ETL represents temporal evolutions of agents' epistemic states by branching-time tree structures consisting of sequences of events. In ETL-models, each branch of a given node represents what sequence of events can take place after the temporal point represented by the node. Thus ETL can straightforwardly capture constraints present in various communications. However, ETL does not specify the epistemic effect of each event independently from given ETL-models. Instead, the epistemic effects of events must be specified in each ETL-model by imposing desired properties on agents' indistinguishability relations. Thus ETL does not provide a systematic way of representing informational events and their epistemic effects as in the case of DEL.

The purpose of this paper is to merge PAL and ETL to provide a system that captures well both of the two important aspects of multi-agent intelligent interaction. To put the two frameworks together, we assign to each state in a given epistemic model a set of sequences of public announcements, which we call a *protocol*. Then we successively apply to the epistemic model sequences of public announcements based on the assigned protocols. This will generate ETL-tree structures that represent all possible evolutions of the original epistemic states under the constraints represented by the protocol. The basic ideas of this construction have been provided in van Benthem et al. (2008) with the restriction that public announcements in the protocols must be built only from epistemic formulas. The current paper generalizes the original method to remove the restriction.

After presenting the merged framework, we will study the logic of public announcement over the class of ETL-models generated by the above method. The language of our logic, which we call *TAPAL*,[1] will include the operators that quantify over public announcements, e.g. "There is some public announcement after which…", "There is some sequence of public announcements after

[1] The name is given to indicate that the system merges *Temporal Public Announcement Logic*, TPAL, in van Benthem et al. (2008) and *Arbitrary Public Announcement Logic*, APAL, in Balbiani et al. (2008).

which…”, etc. as well as the standard public announcement operators, “The public announcement that φ can be made after which…”. We include these “generalized” operators, since they enable us to express the question whether there are some (sequence of) public announcements after which epistemic states of interest will be reached. That is, with the generalized operators, we can formulate such a reachability question, which motivated our protocol-based semantic framework in the first place. Furthermore, the generalized operators of the kind have been considered in the original setting of PAL by Balbiani et al. (2008). Our study of TAPAL will show that the ideas in Balbiani et al. (2008) can be combined with our new semantic framework.

We proceed as follows. We start out by briefly reviewing the systems, PAL and ETL (§2). Next we present the method of generating ETL-models from epistemic models by assigning a protocol to each state (§3) and define the logic TAPAL over the class of ETL-models generated by the method (§4). Then after studying semantic results in TAPAL (§5), we will provide the axiomatization TAPAL (§6) and prove the soundness and completeness of TAPAL (§7, 8). We conclude the paper by discussing some open problems (§9).

2 PAL and ETL

We start by reviewing PAL and ETL. We only give the bare necessities for our presentation. For further technical details, the readers are referred to the sources mentioned below. Fix $\mathcal{A}$ as a finite set of agents and At, as a countable set of propositional letters.

2.1 PAL

Definition 2.1. Language of PAL The language $\mathcal{L}_{pal}$ of *PAL* is inductively defined as follows:

$$\varphi \ ::= \top \mid p \mid \neg\varphi \mid \varphi \wedge \varphi \mid \langle i \rangle \varphi \mid \langle !\varphi \rangle \varphi$$

where $p \in \mathsf{At}$ and $i \in \mathcal{A}$. The duals, $[i]$ and $[!\varphi]$, of $\langle i \rangle$ and $\langle !\varphi \rangle$, and the other boolean operators are defined in the standard way. $\mathcal{L}_{el}$ is the fragment of $\mathcal{L}_{pal}$ without the operator $\langle !\varphi \rangle$.

The intended readings of $\langle i \rangle \varphi$ and $\langle !\psi \rangle \varphi$ are respectively “i considers φ possible” and “The public announcement that ψ can be made after which φ.” The intended readings of $[i]\varphi$ and $[!\psi]\varphi$ are respectively “i knows that φ” and “After the public announcement that ψ, φ.”

Definition 2.2. Epistemic Models An *epistemic model* is a triple $(W, \sim, V)$, where (i) W is a nonempty set, (ii) $\sim : \mathcal{A} \to \wp(W \times W)$, and (iii) $V : \mathsf{At} \to \wp(W)$. When $(w, v) \in \sim (i)$, we write $w \sim_i v$ by convention.

Definition 2.3. Truth Let $\mathcal{M} = (W, \sim, V)$ be an epistemic model. The truth of $\varphi \in \mathcal{L}_{pal}$ at w in W is inductively defined as follows:

$$
\begin{array}{lll}
\mathcal{M}, w \models p & \text{iff} & w \in V(p) \quad (\text{with } p \in \mathsf{At}) \\
\mathcal{M}, w \models \neg\varphi & \text{iff} & \mathcal{M}, w \not\models \varphi \\
\mathcal{M}, w \models \varphi \wedge \psi & \text{iff} & \mathcal{M}, w \models \varphi \text{ and } \mathcal{M}, w \models \psi \\
\mathcal{M}, w \models \langle i \rangle\varphi & \text{iff} & \exists w' \in W : (w, w') \in \sim(i) \text{ and } \mathcal{M}, w' \models \varphi \\
\mathcal{M}, w \models \langle !\varphi \rangle\psi & \text{iff} & \mathcal{M}, w \models \varphi \text{ and } \mathcal{M}|_\varphi, w \models \psi
\end{array}
$$

where $\mathcal{M}|_\varphi = (W|_\varphi, \sim|_\varphi, V|_\varphi)$ is defined by:

$$
\begin{array}{lll}
W|_\varphi & = & \{v \in W \mid \mathcal{M}, v \models \varphi\} \\
\sim|_\varphi(i) & = & \sim(i) \cap (W|_\varphi \times W|_\varphi) \\
V|_\varphi(p) & = & V(p) \cap W|_\varphi.
\end{array}
$$

In the framework of PAL, Balbiani et al. (2008) consider the operator $\Diamond$, where the intended reading of $\Diamond\varphi$ is "Some public announcement can be made after which φ." The semantics of the operator is given by:

$$
\mathcal{M}, w \models \Diamond\varphi \quad \text{iff} \quad \exists\psi \in \mathcal{L}_{pal} : \mathcal{M}, w \models \langle !\psi \rangle\varphi.
$$

The extension of PAL with the operator is denoted by *APAL*. The language of APAL is denoted by $\mathcal{L}_{apal}$.

2.2 ETL

There are different languages for ETL (See e.g. Fagin et al. (1995)). We present the minimal language that is needed for our presentation. Fix the set Σ of events. The language $\mathcal{L}_{etl}$ extends $\mathcal{L}_{el}$ (as defined in Definition 2.1) by the operator $\langle e \rangle$ where $e \in \Sigma$, $\Diamond$, and $\Diamond^*$. The intended readings of $\langle e \rangle\varphi$, $\Diamond\varphi$, and $\Diamond^*\varphi$ are respectively "The event e can happen after which φ.", "Some event can happen after which φ" and "Some sequence of events can happen after which φ." The duals are defined and read in the standard way.

We write Σ^* for the set of finite sequences of elements in Σ. A *history* is an element in Σ^*. Given $h \in \Sigma^*$ and $e \in \Sigma$, we write he for the history h followed by the event e. Also given $h, h' \in \Sigma^*$, we write $h \preceq h'$ if h is a finite prefix of h'. If $h \prec h'$ and $h \neq h'$, we write $h \prec h'$. We denote the length of a given sequence σ by $\mathsf{len}(\sigma)$. Also we denote the empty sequence by λ.

Definition 2.4. ETL Models An *ETL-model* is a tuple $(H, \sim, V)$ where (i) H is a subset of Σ^* that is closed under finite prefixes,[2] (ii) $\sim: \mathcal{A} \to \wp(H \times H)$, and (iii) $V : \mathsf{At} \to \wp(H)$.

Definition 2.5. Truth Let $\mathcal{H} = (H, \sim, V)$ be an ETL-model. The truth definitions of the operators $\langle e \rangle$, $\Diamond$, and $\Diamond^*$ are given as follows. (The definitions of the other operators are similar to the ones given in Definition 2.3.)

$$
\begin{array}{lll}
\mathcal{H}, h \models \langle e \rangle\varphi & \text{iff} & he \in H \text{ and } \mathcal{H}, he \models \varphi \\
\mathcal{H}, h \models \Diamond\varphi & \text{iff} & \exists e \in \Sigma : he \in H \text{ and } \mathcal{H}, he \models \varphi \\
\mathcal{H}, h \models \Diamond^*\varphi & \text{iff} & \exists h' \in \Sigma^* : hh' \in H \text{ and } \mathcal{H}, hh' \models \varphi
\end{array}
$$

[2] I.e. for every h, g, if $h \in H$ and $g \prec h$, then $g \in H$.

3 PAL-Generated ETL-Modes

To provide the desired semantic framework, we first assign to each state in a given epistemic model a set of sequences of public announcements, which represents the set of "permissible" sequences of announcements. We call those assigned sets *protocols*.

Definition 3.1. *PAL*-Protocols Let Σ_{pal} be the set of public announcements $!\varphi$ where $\varphi \in \mathcal{L}_{pal}$. A *protocol* is a subset of the set of finite sequences of public announcements in Σ_{pal} that is closed under finite prefix. We denote the set of protocols by *Ptcl*. A *state-dependent* protocol (abbreviated as *sd*-protocol) p on an epistemic model $\mathcal{M}$ is a function from $Dom(\mathcal{M})$ to *Ptcl*. We denote the class of *sd*-protocols by $\mathbb{PAL}$.

Given an epistemic model $\mathcal{M}$ and an *sd*-protocol p on $\mathcal{M}$, we construct from $\mathcal{M}$ the ETL-model that represents all the possible temporal evolutions of the epistemic state given the protocol information encoded in p. The basic intuition of the construction can be illustrated by the following example. Let $\mathcal{M}$ be an epistemic model consisting of the three points, w, v, u, indistinguishable[3] for an agent i, where p is true only at w, v and q is true only at w. Let p be an *sd*-protocol on $\mathcal{M}$ such that $\mathsf{p}(w) = \{!p![i]q\}$, $\mathsf{p}(v) = \{!p![i]q, !\neg q\}$, $\mathsf{p}(u) = \{!p, !\neg q!\top\}$. The ETL-model we construct from $\mathcal{M}$ and p can be visualized as in Figure 1. The basic procedure to construct the model is to (i) check what is permitted by p as a public announcement at each stage, (ii) create a new node if what is permitted is in fact true at the stage and (iii) carry over the indistinguishability relation in the previous stage for the new nodes. Thus, we start from the first stage $\mathcal{M}$ (indicated by the solid line enclosing the three points). In all states in $\mathcal{M}$, $!p$ is assigned by p. Since p is true at w, v, we create the nodes $w!p$ and $v!p$, while we do not create the node "$u!p$" since p is false at u. Also we connect $w!p$ and $v!p$ by the indistinguishability relation (indicated by the horizontal dashed line), since they are indistinguishable in $\mathcal{M}$. Note that the created nodes constitute the model obtained by applying $!p$ to $\mathcal{M}$, i.e. the model $\mathcal{M}|_p$. In this second stage (indicated by the circle enclosing the two nodes), $![i]q$ is permitted and true at both nodes. Thus we produces the third stage consisting of $w!p![i]p$ and $v!p![i]p$. Similarly the nodes $v!\neg q$ and $u!\neg q$ are created since $\neg q$ are permitted and true at v, u while $w!\neg q$ is not, since $!\neg q$ is neither permitted nor true at u. Furthermore, the node $u!\neg q!\top$ is created but the node "$v!\neg q!\top$" is not present, since $!\top$ is only permitted at $u!\neg q$ though $\top$ is true.

Constructing such ETL-models from epistemic models by *sd*-protocols, we now interpret the operators $\langle !\varphi \rangle$ as the ETL-operators. That is, given a public announcement $!\varphi$, we consider it as an event in ETL and the operator $\langle !\varphi \rangle$ as an ETL-operator. This will give us the following new semantic definition for the public announcement operators:

TPAL $\quad \mathcal{H}, h \models \langle !\varphi \rangle \psi \quad$ iff $h!\varphi$ is in $\mathcal{H}$ and $\quad \mathcal{H}, h!\varphi \models \psi$.

Thus, when we denote the above model by $\mathcal{H}$, we will have, for instance, $\mathcal{H}, w \models \langle !p \rangle [i]p$, $\mathcal{H}, u!\neg q \models \langle !\top \rangle \top$, etc. The generalized operators, $\Diamond$ and $\Diamond^*$, will

[3]To simplify our explication, we assume that the indistinguishability relation is an equivalence relation in the examples below. However, our framework does not assume this in general.

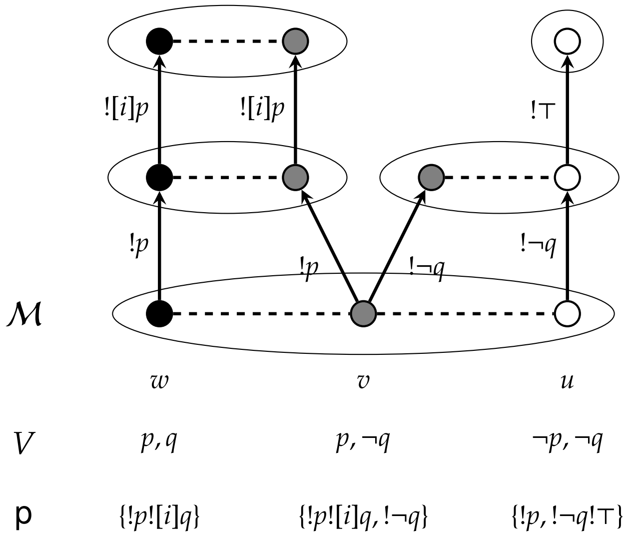


Figure 1: Basic Construction of PAL-Generated ETL-Models

be also redefined accordingly.

However, this illustration gives only a partial story. Note that the *sd*-protocol p in the above example does not contain formulas with public announcement operators. Once we have such formulas in protocols, we cannot simply generate *ETL*-trees straight up from the bottom epistemic model, as we did in the above example. To see the problem, suppose a protocol p allows the formula $\langle !A\rangle\langle !B\rangle\top$ to be announced at w. To determine whether the formula is indeed announceable, we need to know whether the formula is true at w. However, to determine whether the formula is true at w, we need to know in advance whether A is true at w and whether p allows $!A$ at w. Moreover, we need to know whether B is true at $w!A$ and whether p allows $!B$ after $!A$ at w (if the node $w!A$ is generated). Otherwise, we cannot determine the announceability of $\langle !A\rangle\langle !B\rangle\top$.

This example illustrates the following points. First, if φ contains announcement operators, thus making $!\varphi$ a "higher-order" public announcement about the "lower-order" public announcements contained in $!\varphi$, then we need to know in advance about the announceability of the lower-order announcements in order to determine the announceability of $!\varphi$. Second, $!\varphi$ may "refer" to lower-order public announcements after some sequence of public announcements. In such cases, we need to know in advance the announceability of the relevant sequences of announcements in order to determine the announceability of $!\varphi$. That is, we need to know the structure of the tree above the current node at least concerning the lower-order announcements mentioned in φ, while we only needed to know the structure of the current stage in the above example of tree constructions.

Therefore, we need to generalize our construction in the above example to take into account higher order announcements. The key idea is to construct ETL-structures by induction on the *orders* of announcements occurring in the announcement sequences. Having an epistemic model $\mathcal{M}$ and a protocol p, we first construct ETL-trees from $\mathcal{M}$ by the above construction method, based on the initial segments of the sequences given by p that consist only of epistemic formulas. That is, we begin the construction by dealing with only the "first-order" announcements. Then we add nodes to the resulted trees, based on the second-order announcements that refer to the first-order announcements. This second construction process can be carried out, since the truth values of the formulas we need to know to make the second-order announcements, i.e. the ones of first-order announcements, will have been determined at this point after the first-order construction process. We then continue this way by constructing nodes of second-order announcements after nodes for first-order announcements, until all first- and second-order announcements are taken care of. Then we next goes on to the third-order announcement process and continue similarly. And so forth. Below we make this idea more precise.

Definition 3.2. Order of Formulas *The order* $o(!\varphi)$ of a public announcement $!\varphi \in \Sigma_{pal}$ is defined inductively as follows:

- $o(!p) = 1$ with $p \in \mathsf{At}$
- $o(!(\varphi \wedge \psi)) = max(o(!\varphi), o(!\psi))$
- $o(!\neg\varphi) = o(!\varphi)$
- $o(![i]\varphi) = o(!\varphi)$
- $o(!\langle !\varphi\rangle\psi) = max(o(!\varphi) + 1, o(!\psi))$

For example, $o(!(\langle !p\rangle\top)) = 2$, $o(!(\langle !q\rangle\langle !\langle !p\rangle\top\rangle\top)) = 3$ etc.[4] The order of a given public announcement indicates the greatest number of nested "!" operators. Given a sequence $\sigma = \varphi_0 \ldots \varphi_{n-1} \in \Sigma_{pal}$, we define the order $o(\sigma)$ of σ by

$$o(\sigma) = (o(!\varphi_0), \ldots o(!\varphi_{n-1})).$$

We denote the set of the orders of sequences by $\mathbb{O}$.

Definition 3.3. Lexicographic Ordering on Orders We define the ordering $\ll$ on the set of orders $\mathbb{O}$ lexicographically as follows. For every pair of sequences in Σ_{pal}, $\sigma = \sigma_0 \ldots \sigma_{n-1}$ and $\tau = \tau_0 \ldots \tau_{m-1}$, $o(\sigma) \ll o(\tau)$ if

1. $\sigma \prec \tau$ (σ is a proper initial segment of τ as defined above) or
2. There is some $i \in \mathbb{N}$ such that
   - for all $j \in \mathbb{N}$, $j < i \to o(\sigma_j) = o(\tau_j)$), and
   - $o(\sigma_i) < o(\tau_i)$.

Definition 3.4. Union of Models Let $\mathbb{F} = \{\mathcal{H}_k\}_{k\in I}$ be a family of *ETL*-models $\mathcal{H}_k = (H_k, \sim_k, V_k)$. The union $\bigcup_{k\in I} \mathcal{H}_k$ of *ETL*-models in $\mathbb{F}$ is a triple $(H, \sim, V)$:

- $H = \bigcup_{k\in I} H_k$
- $\sim (i) \;\; = \bigcup_{k\in I} \sim_k (i)$

[4] Parentheses are added for clarification here. We will use additional parentheses below when we need clarification or emphasis.

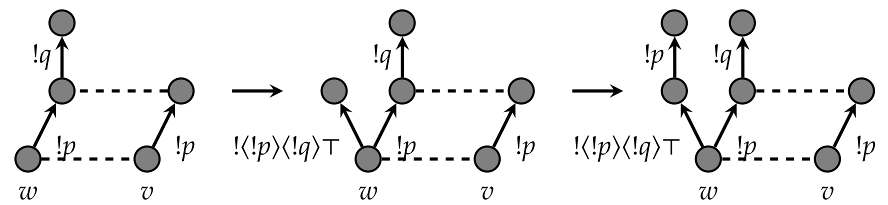


Figure 2: PAL-Generated ETL-Models

- For all $p \in \mathsf{At}$, $V(p) = \bigcup_{k \in I} V_k(p)$.

Now, as we mentioned above, given an ETL-model, we interpret the truth of formulas of $\mathcal{L}_{pal}$ in ETL by taking the operator $\langle !\varphi \rangle$ as the labeled future operator in ETL. The semantic definition for the $\langle !\varphi \rangle$ is as given in **TPAL** above. The definitions for other operators are as given in Definition 2.5.

Given a sequence σ, we denote by $\sigma_{(k)}$ $(0 \leq k \leq m)$ the initial segment of σ of length k and by σ_k $(1 \leq k \leq m)$ the k-th element of σ.

Definition 3.5. σ-Generated Models Let $\mathcal{M} = (W, \sim, V)$ and p be an epistemic model and an *sd*-protocol on $\mathcal{M}$ respectively. For every sequence $\sigma \in \Sigma^*_{pal}$ and every order $x \in \mathbb{O}$, we define *the σ-generated model* $\mathcal{H}^{\sigma,\mathsf{p}} = (H^{\sigma,\mathsf{p}}, \sim^{\sigma,\mathsf{p}}, V^{\sigma,\mathsf{p}})$ and *the order-x-fragment model* $\mathcal{H}^{\mathsf{p}}_x = (H^{\mathsf{p}}_x, \sim^{\mathsf{p}}_x, V^{\mathsf{p}}_x)$ by simultaneous induction as follows:

1. $\mathcal{H}^{\lambda,\mathsf{p}} = \mathcal{M}$, $\mathcal{H}^{\mathsf{p}}_{\lambda} = \mathcal{M}$
2. $\mathcal{H}^{\mathsf{p}}_{o(\tau)} = \bigcup \{\mathcal{H}^{\tau',\mathsf{p}} \mid o(\tau') \ll o(\tau)\}$
3. $H^{\sigma_{(n+1)},\mathsf{p}} = H^{\mathsf{p}}_{o(\sigma_{(n+1)})} \cup \{w\sigma_{(n+1)} \mid \mathcal{H}^{\mathsf{p}}_{o(\sigma_{(n+1)})}, w\sigma_{(n)} \models \sigma_{n+1}$ and $\sigma_{(n+1)} \in \mathsf{p}(w)\}$
4. $(w\tau, v\tau') \in \sim^{\sigma_{(n+1)},\mathsf{p}} (i)$ iff $(w, v) \in \sim (i)$ and $\tau = \tau'$
5. $V^{\sigma_{(n+1)},\mathsf{p}}(p) = \{w\tau \in H^{\sigma_{(n+1)},\mathsf{p}} \mid w \in V(p)\}$

Definition 3.6. PAL-Generated ETL-Models An *ETL*-model *Forest*$(\mathcal{M}, \mathsf{p})$ generated from an epistemic model $\mathcal{M} = (W, \sim, V)$ based on a *sd*-protocol p is defined by:

$$Forest(\mathcal{M}, \mathsf{p}) := \bigcup_{w \in W} \bigcup_{\sigma \in \mathsf{p}(w)} \mathcal{H}^{\sigma,\mathsf{p}}$$

We call a *PAL-generated ETL-model* an ETL-model generated this way. We denote by $\mathbb{F}(\mathbb{PAL})$ the class of PAL-generated ETL-models.

Example 1. Example Let $\mathcal{M}$ be an epistemic model consisting of two indistinguishable points (for an agent) w, v, in which p is true at both w and v and q is true only at w. Define a protocol p so that $\mathsf{p}(w) = \{!p!q, !\langle !p \rangle \langle !q \rangle \top\}$ and $\mathsf{p}(v) = \{!p!q, !\langle !p \rangle \langle !q \rangle \top\}$.

Figure 2 illustrate the construction process. The model on the left is obtained by calculating the first-order public announcement. The model on the right is obtained by calculating the second-order public announcements. The model on the right is the ETL-model generated from $\mathcal{M}$ by p as specified.

4 Logic: TAPAL

Definition 4.1. Language of TAPAL The language $\mathcal{L}_{tapal}$ of TAPAL extends $\mathcal{L}_{pal}$ with the operators $\Diamond$ and $\Diamond^*$. The formulas in $\mathcal{L}_{tapal}$ is inductively defined by:

$$\varphi \ ::= \top \mid p \mid \neg\varphi \mid \varphi \wedge \varphi \mid \langle i \rangle \varphi \mid \langle !\theta \rangle \varphi \mid \Diamond\varphi \mid \Diamond^*\varphi$$

where $p \in \mathsf{At}$, $i \in \mathcal{A}$ and $\theta \in \mathcal{L}_{pal}$. The duals of the modal operators, $\Box$ and $\Box^*$, are defined in the standard way. The other operators are defined as mentioned in Definition 2.1. We call *TPAL* the fragment without $\Diamond$ and $\Diamond^*$, and denote the language by $\mathcal{L}_{tpal}$. Note that $\mathcal{L}_{tpal} = \mathcal{L}_{pal}$.

The intended interpretations of $\Diamond\varphi$ and $\Box\varphi$ are "Some public announcement can be made after which φ is true" and "After every public announcement, φ is true." respectively. Also the intended interpretations of $\Diamond^*\varphi$ and $\Box^*\varphi$ are "Some sequences of public announcement can be made after which φ is true." and "After every sequence of public announcement, φ is true." respectively. (Sequences here are possibly empty.)

Definition 4.2. Truth Given an *ETL*-model $\mathcal{H} = (H, \sim, V) \in \mathbb{F}(\mathbb{PAL})$ and a history $h \in H$, the truth of a *TAPAL*-formula φ at h is inductively defined as follows. We only give the definitions for $\langle !\varphi \rangle$, $\Diamond$, and $\Diamond^*$. The other definitions are as given in Definition 2.5:

$$\begin{array}{lll}
\mathcal{H}, h \models \langle !\psi \rangle \varphi & \text{iff} & h!\psi \in H \text{ and } \mathcal{H}, h!\psi \models \varphi \\
\mathcal{H}, h \models \Diamond\varphi & \text{iff} & \exists !\psi \in \Sigma_{pal} : h!\psi \in H \text{ and } \mathcal{H}, h!\psi \models \varphi \\
\mathcal{H}, h \models \Diamond^*\varphi & \text{iff} & \exists \sigma \in \Sigma_{pal}^* : h\sigma \in H \text{ and } \mathcal{H}, h\sigma \models \varphi
\end{array}$$

Consistency, satisfiability, validity etc. are defined in a familiar way.

Some remarks are in order concerning the language of TAPAL. First note that *sd*-protocols defined in Definition 3.1 are restricted to the PAL-formulas. This is to make the definition of the generalized operators $\Diamond$ and $\Diamond^*$ well-defined, as in the case of *APAL* in Balbiani et al. (2008). For suppose all *TAPAL* formulas are allowed in *sd*-protocols. Assume further that $\Box\varphi \in \mathsf{p}(w)$ for some w in a given model. By the truth definition in Definition 4.2, to determine the truth value of $\Box\varphi$, we need to know the truth value of $\Box\varphi$. Second, given this restriction on *sd*-protocols, we also needed to defined the public announcement operators (Definition 4.1) to be formed only from the formulas in *PAL*. By this, we do not allow formulas such as $\langle !\Diamond\psi \rangle \varphi$, which is in fact allowed in *APAL*.

5 Semantic Results

Next we see some semantic features of the system TAPAL. First, we see some basic results of TAPAL in comparison with those in PAL and APAL. Then, we consider the results concerning two kinds of model transformations, which we call *normalization* and *grafting*.

5.1 Semantic Comparison

The first result states that our semantics framework generalizes that of PAL. If we permit all formulas to be publicly announced, then the truth of a formula in our framework corresponds to that in the framework of PAL.

Proposition 1. *Let $\mathcal{M}$ be an epistemic model. Let p_{pal} be the sd-protocol on $\mathcal{M}$ such that for every w in $\mathcal{M}$, $\mathsf{p}_{pal}(w) = \Sigma^*_{pal}$. Then, for any formula φ in $\mathcal{L}_{apal}$,*

$$\mathcal{M}, w \models \varphi \quad \textit{iff} \quad \mathsf{Forest}(\mathcal{M}, \mathsf{p}_{pal}), w \models \varphi$$

Proof. By a straightforward induction on φ. □

Next we give a semantic comparison between the public announcement operators $\langle !\varphi \rangle$ in TPAL (or TAPAL) and that in PAL (or APAL).

Proposition 2 (Public Announcement Operators)**.** *The following properties hold in PAL but not in TPAL.*

(A) $\models \langle !\alpha \rangle \langle !\beta \rangle \varphi \leftrightarrow \langle ! \langle !\alpha \rangle \beta \rangle \varphi.$

(B) $\models \langle !\varphi \rangle \leftrightarrow \varphi$

For PAL, **A** and **B** follows straightforwardly from the semantic definition of $\langle !\varphi \rangle$, as given Definition 2.3. The readers are invited to give counterexamples against **A** and **B** in TAPAL.

The validity of **A** in PAL shows that sequences of public announcements are identified with some single announcements in PAL. On the other hand, it is invalid in TAPAL. Even if it allows given sequence of announcements, protocols may not allow the corresponding single announcements. The validity of **B** in PAL reflects the assumption that every true formula is announceable. TAPAL removes this assumption and invalidates the principle, while it assumes the truthfulness of announcements and validates the left-to-right direction. Because the invalidity of the principle, the standard reduction axioms in PAL do not hold. See the axiomatization of TAPAL in §6.

Next consider the following properties:

1. $\models \Box\varphi \to \varphi$	2. $\models \Box\varphi \to \Box\Box\varphi$
3. $\models \Box\Diamond\varphi \to \Diamond\Box\varphi$	4. $\models \Diamond\Box\varphi \to \Box\Diamond\varphi$

Proposition 3. *Generalized Operators*

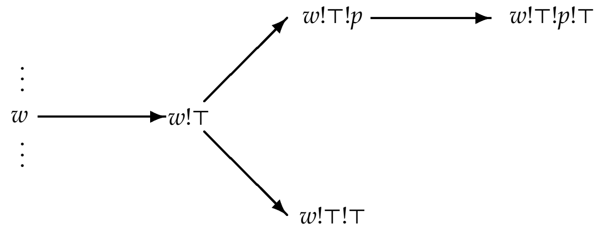
(A) *All of the properties 1-4 hold in APAL.*

(B) *None of the properties 1-4 holds in TAPAL.*

(C) *The properties 1-2 hold, but 3 and 4 don't in TAPAL, when $\Diamond$ and $\Box$ are replaced with $\Diamond^*$ and $\Box^*$ respectively.*

Proof. The proofs of the properties A1-4 in *APAL* are in Balbiani et al. (2008). We only do B3-4 and C3-4. The counterexamples are as follows:

B3 Let $\mathcal{M}, w \models p$. Define $\mathsf{p}(w) = \{!\top, !\top!\top, !\top!p, !\top!p!\top\}$. The model $\mathcal{H} = \mathit{Forest}(\mathcal{M}, \mathsf{p})$ can be represented by the figure below. Here we have $\mathcal{H}, w!\top!p \models \langle !\top \rangle \top$, but $\mathcal{H}, w!\top!\top \not\models \langle !\top \rangle \top$. Therefore, we have $\mathcal{H}, w \models \Box\Diamond\langle !\top \rangle \top$, but $\mathcal{H}, w \models \Box\Diamond\neg\langle !\top \rangle \top$, i.e. $\mathcal{H}, w \not\models \Diamond\Box\langle !\top \rangle \top$.



B4 In the above model, $\mathcal{H}, w!\top!p \models \Box\top$, which yields $\mathcal{H}, w!\top \models \Diamond\Box\top$, but $\mathcal{H}, w!\top!\top \not\models \Diamond\top$, which yields $\mathcal{H}, w!\top \not\models \Box\Diamond\top$.

C3 Let $\mathcal{M}, w \models p$. Define $\mathsf{p}(w) = \{!\top, !\top!p, !\top!p!\top, !\top!p!\top!p, ...\}$. Let $\mathcal{H}$ be $\mathit{Forest}(\mathcal{M}, \mathsf{p})$. We claim that, for every h in $\mathcal{H}$, there exists $\sigma, \sigma' \in \mathsf{p}(w)$ such that $\mathcal{H}, h\sigma \models \langle !\top \rangle \top$ and $\mathcal{H}, h\sigma' \not\models \langle !\top \rangle \top$. To see this, note that every h ends with either $\top$ or p. If h ends with $!\top$, then put $\sigma =!p$ and $\sigma' = \emptyset$; if h ends with $!p$, then put $\sigma = \emptyset$ and $\sigma' =!\top$. This fact implies $\mathcal{H}, w \models \Box\Diamond\langle !\top \rangle \top$ and $\mathcal{H}, w \models \Box\Diamond\neg\langle !\top \rangle \top$. Thus, this model is a counterexample against 3.

C4 The models for B4 similarly works. □

Even though formulas in TAPAL and APAL behave quite differently, there is some property they share. That is, the truth of the formulas at most depend on the current and future epistemic states, not on the past. We will state this fact formally in TAPAL. Let p be an *sd*-protocol on $\mathcal{M}$. Also let $\mathcal{H} = \mathsf{Forest}(\mathcal{M}, \mathsf{p}) = (H, \sim, V)$ and $\sigma \in \Sigma^*_{pal}$.

Definition 5.1. Epistemic Model after σ *The epistemic model after σ in $\mathcal{M}$ based on* p, $\mathcal{M}^{\mathsf{p},\sigma} = (W^{\mathsf{p},\sigma}, \sim^{\mathsf{p},\sigma}, V^{\mathsf{p},\sigma})$, is defined by:

- $W^{\mathsf{p},\sigma} := \{w\sigma \mid w\sigma \in H\}$
- $\sim^{\mathsf{p},\sigma} (i) := \sim (i) \cap (W^{\mathsf{p},\sigma} \times W^{\mathsf{p},\sigma})$
- $V^{\mathsf{p},\sigma}(p) := V(p) \cap W^{\mathsf{p},\sigma}$

Definition 5.2. Protocol after σ *The sd-protocol after σ on* $\mathcal{M}^{\mathsf{p},\sigma}$, $\mathsf{p}^{\sigma<}$, so that, for all $w\sigma$ in $\mathcal{M}^{\mathsf{p},\sigma}$ with w in $\mathcal{M}$,

$$\mathsf{p}^{\sigma<}(w\sigma) = \{\tau \mid w\sigma\tau \text{ in } \mathsf{Forest}(\mathcal{M}, \mathsf{p})\}.$$

Definition 5.3. ETL-Model after σ *The ETL-model after σ in* $\mathcal{H} = \mathsf{Forest}(\mathcal{M}, \mathsf{p})$, $\mathcal{H}^{\sigma\leq}$, is defined by:

$$\mathcal{H}^{\sigma\leq} := \mathsf{Forest}(\mathcal{M}^{\mathsf{p},\sigma}, \mathsf{p}^{\sigma<})$$

Proposition 4. *Let* $\mathcal{H} = \mathsf{Forest}(\mathcal{M}, \mathsf{p}) \in \mathbb{F}(\mathbb{PAL})$. *Also let w be in $\mathcal{M}$ and σ in Σ_{pal}. Then, for every $\varphi \in \mathcal{L}_{tapal}$,*

$$\mathcal{H}, w\sigma \models \varphi \quad \Leftrightarrow \quad \mathcal{H}^{\sigma\leq}, w\sigma \models \varphi.$$

◁

Proof. Straightforward induction on φ. □

Furthermore, the expressive power of $\Diamond$ and $\Diamond^*$ renders the systems non-compact, as in the case of APAL (see Balbiani et al. (2008)).

Proposition 5. *TAPAL is not compact.* ◁

Proof. Straightforward by considering the set $\Gamma = \{\neg\langle !\theta\rangle p | \theta \in \mathcal{L}_{tpal}\} \cup \{\Diamond p\}$ or the set $\bigcup_{i=0}^{\infty} \Gamma_i \cup \{\Diamond^* p\}$, where $\Gamma_i = \{\neg\langle !\theta_0\rangle ... \langle !\theta_i\rangle p | \theta_j \in \mathcal{L}_{tpal}\ (0 \leq j \leq i)\}$. □

5.2 Model Normalization

Next, we consider a model transformation, which we call *normalization*. Let $\varphi_0, \varphi_1, \ldots$ and $\top_0, \top_1, \ldots$ be a pair of (possibly infinite) sequences of *TPAL*-formulas such that (i) $\top_i$ is a tautologous formula in *TPAL* and (ii) $\varphi_i \neq \varphi_j$ and $\top_i \neq \top_j$ for all $i, j \geq 0$.

Definition 5.4. Normalization of Sequences Given a sequence $\sigma \in \Sigma^*_{pal}$, we define $\sigma[!\top_0/!\varphi_0, !\top_1/!\varphi_1, \ldots]$ to be the result of replacing all occurrences of $!\varphi_i$ in σ with $!\top_i$ for all i.

Definition 5.5. Normalization of Models Let p be an *sd*-protocol on $\mathcal{M}$. Also let $\mathcal{H} = \mathsf{Forest}(\mathcal{M}, \mathsf{p}) = (H, \sim, V)$. Define $\mathcal{H}[!\top_0/!\varphi_0, !\top_1/!\varphi_1, \ldots] = (H', \sim', V')$ by:

- $H' := \{h[!\top_0/!\varphi_0, !\top_1/!\varphi_1, \ldots] \mid h \in H\}$
- $(h[!\top_0/!\varphi_0, !\top_1/!\varphi_1, \ldots], g[!\top_0/!\varphi_0, !\top_1/!\varphi_1, \ldots]) \in \sim' (i)$ iff $(h, g) \in \sim (i)$
- $V'(p) := \{h[!\top_0/!\varphi_0, !\top_1/!\varphi_1, \ldots], \mid h \in V(p)\}$

Now note that, given that $\mathcal{H}$ is in $\mathbb{F}(\mathbb{PAL})$, $\mathcal{H}[!\top_0/!\varphi_0, !\top_1/!\varphi_1, \ldots]$ is also in $\mathbb{F}(\mathbb{PAL})$. Indeed, when $h\varphi_i$ is in $\mathcal{H}$, $h!\top_i$ must be in $\mathcal{H}[!\top_0/!\varphi_0, !\top_1/!\varphi_1, \ldots]$ since the tautologous formula $!\top_i$ is guaranteed to be true at h. Also if $(h, g) \in \sim (i)$, the corresponding nodes for h and g will be indistinguishable by construction. We state this fact more precisely as follows.

Definition 5.6. Protocol above σ in $\mathcal{H}$ Let $\mathcal{H} = \mathsf{Forest}(\mathcal{M}, \mathsf{p}) \in \mathbb{F}(\mathbb{PAL})$. Let $\sigma \in \Sigma^*_{pal}$. Then define $\mathsf{p}^{\mathcal{H},\sigma<}$ on $\mathcal{M}^{\mathsf{p},\sigma}$ so that

$$\mathsf{p}^{\mathcal{H},\sigma<}(w\sigma) = \{\tau \mid w\sigma\tau \text{ in } \mathcal{H}\}$$

Observation 1. Let $\varphi_0, \varphi_1, \ldots$ be a sequence of formulas in $\mathcal{L}_{pal}$ and $\top_0, \top_1, \ldots$, a sequence of tautologous formulas in $\mathcal{L}_{pal}$. Suppose, for every $i, j \leq 0$, if $i \neq j$, then $\varphi_i \neq \varphi_j$ and $\top_i \neq \top_j$. Let $\mathcal{H} = \mathsf{Forest}(\mathcal{M}, \mathsf{p})$. Put $\mathcal{G} = \mathcal{H}[!\varphi_0/!\top_0, !\varphi_1/!\top_1, \ldots]$.

$$\mathcal{G} = \mathsf{Forest}(\mathcal{M}, \mathsf{p}^{\mathcal{G},\lambda})$$

where λ is the empty sequence. ◁

Now we prove a distinct property of PAL-generated ETL-models. Given a formula $\varphi \in \mathcal{L}_{tapal}$, even if we replace the announcements in a given *sd*-protocol that do not occur in φ with "new" tautologous formulas, we can preserve the truth of φ. To do this, we need some definitions.

Definition 5.7. Announcement Occurrence Set The *announcement occurrence set* $AOC(\varphi)$ of a $TAPAL$-formula φ is defined inductively as follows:

- $AOC(p) = \emptyset$ with $p \in P$
- $AOC(\neg\varphi) = AOC(\varphi)$
- $AOC(\varphi \wedge \psi) = AOC(\varphi) \cup AOC(\psi)$
- $AOC([i]\varphi) = AOC(\varphi)$
- $AOC(\langle !\psi \rangle\varphi) = \{\psi\} \cup AOC(\psi) \cup AOC(\varphi)$
- $AOC(\Diamond\varphi) = AOC(\varphi)$
- $AOC(\Diamond^*\varphi) = AOC(\varphi)$

Given a sequence $\sigma = !\varphi_1 \dots !\varphi_n \in \Sigma^*_{pal}$, we define

$$AOC(\sigma) := AOC(\varphi_1) \cup \dots \cup AOC(\varphi_n).$$

Furthermore, given an sd-PAL-protocol p on $\mathcal{M} = (W, \sim, V)$, we define

$$AOC(\mathsf{p}) := \bigcup_{\{\sigma \mid \exists w \in W: \sigma \in \mathsf{p}(w)\}} AOC(\sigma).$$

Proposition 6 (Normalization). *Let* $\mathcal{H} = \mathsf{Forest}(\mathcal{M}, \mathsf{p}) \in \mathbb{F}(\mathbb{PAL})$. *Let X be a finite subset of $\mathcal{L}_{pal}$. Furthermore, let $\varphi_0, \varphi_1 \dots$ be an enumeration of the formulas in $\mathcal{L}_{pal} \backslash X$ without repetition, and $\top_0, \top_1, \dots$ be an enumeration of tautologous formulas in $\mathcal{L}_{pal} \backslash X$ without repetition. Then, for every h and $TAPAL$-formula φ such that $AOC(\varphi) \subseteq X$,*

$$\mathcal{H}, h \models \varphi \quad \Leftrightarrow \quad \mathcal{H}[!\top_0/!\varphi_0, !\top_1/!\varphi_1, \dots], h[!\top_0/!\varphi_0, !\top_1/!\varphi_1, \dots] \models \varphi$$

◁

Proof. By a straightforward induction on φ. We only do the case for $\langle !A \rangle$. Let $h = w\sigma$ with w in $\mathcal{M}$ and $\sigma \in \mathsf{p}(w)$. First, assume φ is of the form $\langle !A \rangle\psi$. Assume LHS. Then we have $\mathsf{Forest}(\mathcal{M}, \mathsf{p}), h!A \models \psi$. Since $AOC(\psi) \subseteq AOC(\langle !A \rangle\psi) \subseteq X$, we can by IH obtain

$$\mathcal{H}[!\top_0/!\varphi_0, !\top_1/!\varphi_1, \dots], h!A[!\top_0/!\varphi_0, !\top_1/!\varphi_1, \dots] \models \psi.$$

Since $A \in AOC(\langle !A \rangle\psi)$, we have

$$\mathcal{H}[!\top_0/!\varphi_0, !\top_1/!\varphi_1, \dots], h[!\top_0/!\varphi_0, !\top_1/!\varphi_1, \dots]!A \models \psi.$$

Therefore, we have $\mathcal{H}[!\top_0/!\varphi_0, !\top_1/!\varphi_1, \dots], h[!\top_0/!\varphi_0, !\top_1/!\varphi_1, \dots] \models \langle !A \rangle\psi$. □

5.3 Grafting

Next, we introduce another model transformation, which we call *grafting*. We will need the results concerning the model transformation, when we prove the soundness of the axiomatic system for TAPAL. Let $\mathcal{H} = \mathsf{Forest}(\mathcal{M}, \mathsf{p}) = (H, \sim, V) \in \mathbb{F}(\mathbb{PAL})$. Let σ, τ be a pair of finite sequences in Σ^*_{pal} and $\top_0$ a tautologous formula in $\mathcal{L}_{pal}$ such that $\top_0 \notin AOC(\mathsf{p})$.

Definition 5.8. Grafting *The model $\mathcal{H}^{[\sigma\tau \mapsto \sigma!\top_0]}$ obtained by grafting $\mathcal{H}$ with respect to $\sigma\tau \mapsto \sigma!\top_0$* is a triple $(H^{[\sigma\tau\mapsto\sigma!\top_0]}, \sim^{[\sigma\tau\mapsto\sigma!\top_0]}, V^{[\sigma\tau\mapsto\sigma!\top_0]})$ defined by:

- $H^{[\sigma\tau\mapsto\sigma!\top_0]} := H \cup \{w\sigma!\top_0 v \mid \exists v \in \mathsf{PAL}^* : \ w\sigma\tau v \in H \text{ and } w \text{ in } \mathcal{M}\}$
- $(h, h') \in \sim^{[\sigma\tau\mapsto\sigma!\top]} (i)$ iff
  - $(h, h') \in \sim (i)$, or
  - $h = w\sigma!\top_0 v$, $h' = v\sigma!\top_0 v'$ and $(\sigma\tau v, v\sigma\tau v') \in \sim (i)$.
- $h \in V^{[\sigma\tau\mapsto\sigma!\top_0]}(p)$ iff
  - $h \in V(p)$
  - $h = w\sigma!\top_0 v$ and $w\sigma\tau v \in V(p)$.

The idea of grafting is as follows. Given a sequence $\sigma\tau$, we "take branches" in the *ETL*-model above $\sigma\tau$ in $\mathcal{H}$, i.e. $\mathcal{H}^{\sigma\tau\leq}$. Then we concatenate the "new" tautologous formula $!\top_0$ at the bottom of the branches and "graft" the branches to the corresponding nodes of the form $w\sigma$ with w in the base epistemic model.

Observation 2. Let $\mathcal{G} = \mathcal{H}^{[\sigma\tau\mapsto\sigma!\top_0]}$. Then

$$\mathcal{G} = \mathsf{Forest}(\mathcal{M}, \mathsf{p}^{\mathcal{G},\lambda<})$$

where λ is the empty sequence. ◁

Proof. By the similar reasoning given to obtain Observation 1. □

Proposition 7 (Preservation at Grafted Branches). *For every $\varphi \in \mathcal{L}_{tapal}$,*

$$\mathcal{H}, w\sigma\tau \models \varphi \quad \Leftrightarrow \quad \mathcal{H}^{[\sigma\tau\mapsto\sigma!\top_0]}, w\sigma!\top_0 \models \varphi$$

◁

Proof. The proof is straightforward by Proposition 4 and the construction of $\mathcal{H}^{[\sigma\tau\mapsto\sigma!\top_0]}$. □

The last result gives one truth-preservation result in grafted models, i.e. the truth of formulas is preserved at the bottom of newly grafted branches. However, unfortunately, grafting does not in general preserve the truth of formulas. We must be careful when transforming models by grafting to preserve the truth of formulas of our interest. We will see more on this below when we prove the soundness theorem of our axiomatization.

6 Axiomatization

Now we present the complete axiomatization for TAPAL. Let $\Diamond^n$ and $\Box^n$ be the sequences of n $\Diamond$'s and $\Box$'s respectively. When $n = 0$, $\Diamond^n$ and $\Box^n$ denote φ. Also given $\sigma = \sigma_0 \dots \sigma_{n-1} \in \Sigma^*_{pal}$, denote the sequences $\langle\sigma_0\rangle \dots \langle\sigma_{n-1}\rangle$ and $[\sigma_0] \dots [\sigma_{n-1}]$ by $\langle\sigma\rangle$ and $[\sigma]$ respectively. When $n = 0$, $\langle\sigma\rangle\varphi$ and $[\sigma]\varphi$ denote φ. Now we can present the axiomatization of TAPAL. Finally, we define the *complexity* $|\varphi|$ of a TAPAL-formula φ by:

- $|p| = 0$ with p propositional.
- $|\neg\varphi| = |\Diamond\varphi| = |\Diamond^*\varphi| = |\varphi| + 1$
- $|\varphi \wedge \psi| = |\langle!\varphi\rangle\psi| = |\varphi| + |\psi| + 1$.

Definition 6.1. Axiomatization of *TAPAL* The axiomatization TAPAL of TAPAL consists of the following axiom schemas and inference rules:

Axiom Schema

PC Propositional validities

M1 $[i](\varphi \to \psi) \to ([i]\varphi \to [i]\psi)$

M2 $[!\theta](\varphi \to \psi) \to ([!\theta]\varphi \to [!\theta]\psi)$

R1 $\langle!\psi\rangle p \leftrightarrow \langle!\psi\rangle\top \wedge p$

R2 $\langle!\psi\rangle\neg\varphi \leftrightarrow \langle!\psi\rangle\top \wedge \neg\langle!\psi\rangle\varphi$

R3 $\langle!\psi\rangle[i]\varphi \leftrightarrow \langle!\psi\rangle\top \wedge [i](\langle!\psi\rangle\top \to \langle!\psi\rangle\varphi)$

A1 $\langle!\varphi\rangle\top \to \varphi$

A2 $\langle!\chi\rangle\varphi \to \Diamond\varphi$ for any $\chi \in \mathcal{L}_{pal}$

A3 $\Diamond^*\varphi \leftrightarrow \varphi \vee \Diamond\Diamond^*\varphi$

Inference Rules

[i]-Nec If $\vdash \varphi$, then $\vdash [i]\varphi$.

[!A]-Nec If $\vdash \varphi$, then $\vdash [!A]\varphi$ where $!A \in \Sigma_{pal}$

$R(\Box)$ If $\vdash \varphi \to [\sigma][!\top_0]\psi$, then $\vdash \varphi \to [\sigma]\Box\psi$, where $\top_0$ is a tautologous formula in $\mathcal{L}_{pal}$ such that $\top_0 \notin AOC(\varphi) \cup AOC(\sigma) \cup AOC(\psi)$.

$R(\Box^*)$ If $\vdash \varphi \to [\sigma]\Box^k\psi$ for every k such that $0 \leq k \leq |\varphi| + 1$, then $\vdash \varphi \to [!A_1] \dots [!A_n]\Box^*\psi$.

First, note that R1-3 are similar to the reduction axioms in PAL. The only difference is that they have $\langle!\psi\rangle\top$ in the place of ψ in the reduction axioms (as mentioned in Proposition 2). Second, A1 implies that public announcements are truthful. Since our framework lifts the assumption in PAL that whatever is true is announceable, the other direction of A1 does not obtain. Third, the purpose of A2 is clear given the semantic definitions. Fourth, A3 plays the role of Fixed Point Axiom as in PDL (See e.g. Blackburn et al. (2001)), as can be seen by their schematic similarity.

Fifth, we will be able to show below (Corollary 1) that $R(\Box)$ is in fact equivalent to the following sound rule:

$R'(\Box)$ If $\vdash \varphi \to [\sigma][!p]\psi$ where p does not occur in φ, σ, or ψ, then $\vdash \varphi \to [\sigma]\Box\psi$.

This form of the rule clarifies what the rule $R(\Box)$ is for. Observe the similarity between $R'(\Box)$ and the first-order rule:

FOQ If $\vdash \varphi \to \psi$ with no occurrence of x in φ, then $\vdash \varphi \to \forall x\psi$.

In fact, as we will see below in the completeness proof of TAPAL, the use of $R(\Box)$ is very similar to the use of this first-order rule in the completeness proof of first-order logic. Nonetheless, we chose $R(\Box)$ instead of $R'(\Box)$, since it extracts from the property of PAL-generated ETL-models that they preserve truth over *model normalization*. More on this in §7.1.

Finally, to see the role of $R(\Box^*)$, consider the following rule:

$R'(\Box^*)$ If $\vdash \varphi \to [\sigma]\Box^n\psi$ for all $n \geq 0$, then $\vdash \varphi \to [\sigma]\Box^*\psi$.

Given the semantic definition, it is straightforward to see that this infinitary rule is sound. The idea of our rule $R(\Box^*)$ is that we can extract a bound on n in the infinitary rule from the complexity of the formula φ. (We will in fact use a more complicated notion of complexity, but it is bounded by the above simple notion of complexity. More on this in §7.2.)

7 Soundness

Now we show the soundness of TAPAL. The soundness of the axiom schemas and the necessitation rules are straightforward. Thus leaving the details of the proofs to the reader, we go on to show that $R(\Box)$ and $R(\Box^*)$ are sound. For our proofs, the model transformations investigated in Section 5.2 and 5.3, *normalization* and *grafting* will be used as essential tools.

7.1 The Soundness of $R(\Box)$

The idea of the proof can be sketched as follows. Suppose that $\varphi \wedge \langle\sigma\rangle\Diamond\psi$ is true at some h in $\mathcal{H}$. Then, φ is true at h. Also there is some $!\theta$ such that $h\sigma!\theta$ is in $\mathcal{H}$ and ψ is true at $h\sigma!\theta$. This situation is visualized in the left figure in Figure 3. We modify the model $\mathcal{H}$ by (i) taking the subtree starting from $h\sigma!\theta$ (the node labeled with ψ in the figure), (ii) obtaining the subtree with a new branch $!\top_0$ attached to its bottom, (iii) and grafting the new subtree to $h\sigma$ and the nodes connected to $h\sigma$ by indistinguishability relations in which $!\theta$ can happen. Let us denote the model obtained this way by $\mathcal{H}'$. $\mathcal{H}'$ is visualized in the right figure in Figure3. We claim that the formula $\varphi \wedge \langle\sigma\rangle\Diamond\psi$ is true at h in $\mathcal{H}'$. First, since TAPAL-formulas are 'future-looking', ψ is true at $h\sigma!\top_0$ by Proposition 4, since the structure of the new subtree is the same as the old subtree. Therefore, $\Diamond\psi$ is true at $h\sigma$, which implies that $\langle\sigma\rangle\Diamond\psi$ is true at h. Furthermore, the truth of φ is preserved over this transformation, since φ cannot distinguish the new

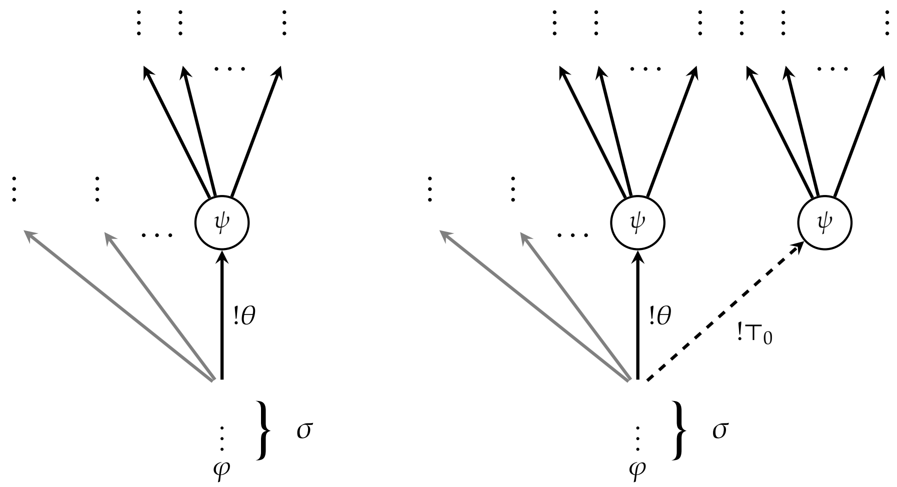


Figure 3: Grafting

and old subtrees by the assumption that $!\top_0$ does not occur in φ. Therefore, $\varphi \wedge \langle\sigma\rangle\Diamond\psi$ is indeed satisfiable.

Below, we make the above idea precise. We start by observing the following fact stating that grafting with respect to $\sigma\tau \mapsto \sigma!\top_0$ where $\mathsf{len}(\tau) = 1$ preserves the truth of TAPAL-formulas.

Lemma 1 (Grafting with $\mathsf{len}(\tau) = 1$)**.** *Let* p *be an sd-protocol on* $\mathcal{M} = (W, \sim, V)$*. Let* $\mathcal{H} = \mathsf{Forest}(\mathcal{M}, \mathsf{p})$ *and* $w\tau\sigma!\psi$ *in* $\mathcal{H}$ *where* $w \in W$*. For every* φ*, if a tautologous formula* $\top \in \mathcal{L}_{pal}$ *is not in* $AOC(\varphi) \cup AOC(\mathsf{p})$*, then*

$$\mathcal{H}, w\tau \models \varphi \quad \Leftrightarrow \quad \mathcal{H}^{[\tau\sigma!\psi \mapsto \tau\sigma!\top_0]}, w\tau \models \varphi.$$

Proof. Straightforward by induction on φ. □

Theorem 1 (Soundness of $R(\Box)$)**.** *If* $\varphi \wedge \langle\sigma\rangle\Diamond\psi$ *is satisfiable in* $\mathbb{F}(\mathbb{PAL})$*, then* $\varphi \wedge \langle\sigma\rangle\langle!\top_0\rangle\psi$ *with* $\top_0 \notin AOC(\varphi) \cup AOC(\sigma) \cup AOC(\psi)$ *is satisfiable in* $\mathbb{F}(\mathbb{PAL})$. ◁

Proof. Assume that $\varphi\wedge\langle\sigma\rangle\Diamond\psi$ is satisfiable. Thus, let $\mathsf{Forest}(\mathcal{M}, \mathsf{p}), h \models \varphi\wedge\langle\sigma\rangle\Diamond\psi$. This implies $\mathsf{Forest}(\mathcal{M}, \mathsf{p}), h \models \varphi \wedge \langle\sigma\rangle\langle\alpha\rangle\psi$ for some $\alpha \in \Sigma_{pal}$. Now take

$$X := AOC(\varphi \wedge \langle\sigma\rangle\langle\alpha\rangle\psi).$$

Also let *Taut* be the set of tautologous formulas in *TPAL*. Take $Taut' := Taut \backslash X$. Then enumerate the elements in $Taut'$ and let $\top'_0, \top'_1, \dots$ be the result of the enumeration. Also take an enumeration of $\mathcal{L}_{pal} \backslash X$ without repetition so that $\top'_0$ comes as the first element. We write the enumeration as $\top'_0, \varphi'_1, \varphi'_2, \dots$. Then apply Proposition 6 by taking the following parameters:

- $X := AOC(\varphi \wedge \langle\alpha\rangle\psi)$
- $\varphi_0 := \top'_0, \varphi_1 := \varphi'_1, \dots, \varphi_i := \top'_i, \dots$
- $\top_0 := \top'_1, \top_1 := \top'_2, \dots, \top_i := \top'_{i+1}, \dots$

Then, by this application of Proposition 6 together with Observation 1, we obtain

$$\mathsf{Forest}(\mathcal{M}, \mathsf{p}'), h' \models \varphi \wedge \langle\sigma\rangle\langle\alpha\rangle\psi$$

for some p' such that $\top_0' \notin AOC(\mathsf{p}')$. Now, since this implies $\mathsf{Forest}(\mathcal{M}, \mathsf{p}'), h'\sigma\alpha \models \psi$, we can apply Lemma 1 (or Proposition 7) to obtain

$$\mathsf{Forest}(\mathcal{M}, \mathsf{p}')^{[\tau\sigma\alpha \mapsto \tau\sigma!\top_0']}, h'\sigma!\top_0' \models \psi$$

Similarly, by applying Lemma 1 to $\mathsf{Forest}(\mathcal{M}, \mathsf{p}'), h' \models \varphi$, we can obtain

$$\mathsf{Forest}(\mathcal{M}, \mathsf{p}')^{[\tau\sigma\alpha \mapsto \tau\sigma!\top_0']}, h' \models \varphi.$$

By Observation 2, the model $\mathsf{Forest}(\mathcal{M}, \mathsf{p}')^{[\tau\sigma\alpha \mapsto \tau\sigma!\top_0]}$ is in $\mathbb{F}(\mathbb{PAL})$ and, therefore, $\varphi \wedge \langle\sigma\rangle\langle!\top_0\rangle\psi$ is satisfiable in $\mathbb{F}(\mathbb{PAL})$. □

Corollary 1. *Let all $\varphi, \psi \in \mathcal{L}_{tapal}$ and $\sigma \in \Sigma^*_{pal}$. Also let $p, \top_0$ be respectively a propositional letter and a tautologous formula in $\mathcal{L}_{pal}$ such that $p, \top_0 \notin AOC(\varphi) \cup AOC(\psi) \cup AOC(\sigma)$. Then*

$$\vdash \varphi \to [\sigma][!p]\psi \;\Leftrightarrow \vdash \varphi \to [\sigma][!\top_0]\psi \;\Leftrightarrow \vdash \varphi \to [\sigma]\Box\varphi$$

⊲

Proof. This follows immediatlye from the soundness of the rule $R'(\Box)$ given in Hoshi (2008) and Theorem 1 via the semantic definition of □. □

7.2 The Soundness of $R(\Box^*)$

Now to deal with the soundness of $R(\Box^*)$, let us start by observing the following fact.

Proposition 8 (Reduction of $\Diamond^*$ to $\Diamond$)**.** *For every $\varphi \in \mathcal{L}_{tapal}$, if $\Diamond^*\varphi$ is satisfiable in $\mathbb{F}(\mathbb{PAL})$, then $\Diamond^n\varphi$ is satisfiable in $\mathbb{F}(\mathbb{PAL})$ for $n = 0$ or $n = 1$.*

Proof. If $\mathcal{H}, h \models \Diamond^*\varphi$ with $h = w\tau$, then there is some $\sigma \in \Sigma^*_{pal}$ such that $\mathcal{H}, h \models \langle\sigma\rangle\varphi$. If σ is empty, we are done. Thus, assume that σ is not empty. By the method used in the proof of Theorem 1, obtain a tautologous formula $\top_0$, a model $\mathcal{H}'$, and a history h' in $\mathcal{H}'$ such that $\top_0$ does not occur in $\mathcal{H}'$ and $\mathcal{H}', h' \models \Diamond^*\varphi$. Then by a similar argument given in the proof of Theorem 1, we obtain

$$(\mathcal{H}')^{[\tau\sigma \mapsto \tau!\top_0]}, h'!\top_0 \models \varphi.$$

This implies the satisfiability of $\Diamond\varphi$. □

Corollary 2. *For every $\varphi \in \mathcal{L}_{tapal}$,*

$$\vdash \Box\varphi \quad\Leftrightarrow\quad \vdash \Box^*\varphi.$$

⊲

In the light of this result, it might be expected that the following claim holds:

Claim If $\varphi \wedge \langle\sigma\rangle\Diamond^*\psi$ is satisfiable in $\mathbb{F}(\mathbb{PAL})$, then $\varphi \wedge \langle\sigma\rangle\Diamond^n\psi$ is satisfiable in $\mathbb{F}(\mathbb{PAL})$ for $n = 0$ or $n = 1$.

Unfortunately this claim does not hold, due to the semantics of the $\Box$-operator. For simplicity, consider the case where σ is empty. Take $\Box\theta \wedge \Diamond^*\psi$. If this formula is satisfiable, then there will be a sequence τ after which ψ is satisfied. Here even if we appeal to the grafting method as in Proposition 8, we may not obtain the satisfiability of the whole formula $\Box\theta \wedge \Diamond^*\psi$. For the new node added to the model as a result of grafting must be quantified by $\Box$ in $\Box\theta$ and there is no guarantee that the node satisfies θ.

What this example illustrates is that, in general, the formula φ in $\varphi \wedge \langle\sigma\rangle\Diamond^*\psi$ may 'refer' to the nodes between the current node h and $h\tau$, where τ is a sequence of announcements, whose existence is claimed by $\Diamond^*$ in the formula. When this 'reference' is made by $\Box$, we cannot safely graft as we did for Proposition 8.

However how 'high up' in the tree φ can 'refer' can be read off from the complexity of φ. In particular, what is problematic is the occurrences of $\Box$ in φ and we need to know the highest number of nexted occurrences of $\Box$ in φ. Once we know such a number for φ, we can safely graft above the height that φ can refer to, as we did for Proposition 8.

Thus we first introduce a measure to indicate the number of the occurrences of the $\Box$-operator that must be taken care of in the relevant sense here.

Observation 3. Every *TAPAL*-formula is equivalent to some formula of *TAPAL* built up by the following inductive definition:

$$\varphi \ ::= \top \mid p \mid \neg p \mid \varphi \wedge \varphi \mid \varphi \wedge \varphi \mid \langle i\rangle\varphi \mid [i]\varphi \mid \langle !A\rangle\varphi \mid [!\theta]\varphi \mid \Diamond\varphi \mid \Box\varphi \mid \Diamond^*\varphi \mid \Box^*\varphi.$$

where $p \in \mathsf{At}$, $i \in \mathcal{A}$ and $\theta \in \mathcal{L}_{pal}$.

Proof. Immediate by the definitions of the dual operators and the standard boolean equivalences. $\Box$

Thus, we can interchangeably use the inductive definition in Definition 4.1 and the one given here.

Definition 7.1. Initial Box Iteration

The initial box iteration $ibi(\varphi)$ of a *TAPAL*-formula φ is defined inductively as follows:

- $ibi(p) = ibi(\neg p) = 0$ for p propositional
- $ibi(\varphi \wedge \psi) = ibi(\varphi \vee \psi) = max(ibi(\varphi), ibi(\psi))$
- $ibi(\langle i\rangle\varphi) = ibi([i]\varphi) = ibi(\varphi)$
- $ibi(\langle !A\rangle\varphi) = ibi([!A]\varphi) = 0$
- $ibi(\Diamond\varphi) = 0$
- $ibi(\Box) = ibi(\varphi) + 1$
- $ibi(\Diamond^*\varphi) = ibi(\Box^*\varphi) = ibi(\varphi)$

Now we can explain the basic idea of the soundness proof for $R(\Box^*)$ as follows. Suppose $\varphi \wedge \langle\sigma\rangle\Diamond^*\psi$ is true at $w\tau$ in $\mathcal{H}$ (w in the base epistemic model $\mathcal{M}$). Then, ψ is true at $w\tau\sigma v$ for some v. Now we graft the model $\mathcal{H}$ with respect to $\tau\sigma v_0 \mapsto \tau\sigma!\top_0$. This will preserve the truth of $\varphi \wedge \langle\sigma\rangle\Diamond^*\psi$ by Lemma 1. After this, we again graft with respect to $\tau\sigma!\top_0 v_1 \mapsto \tau\sigma!\top_1$ similarly by preserving the truth of $\varphi \wedge \langle\sigma\rangle\Diamond^*\psi$. We repeat grafting this way as many times as $ibi(\varphi)$, i.e. the number of the $\Box$-operators that must be taken care of. Once we graft $ibi(\varphi)$ times and go high enough along the tree, we can safely apply the grafting method for $\Diamond^*$-operator as we did in Proposition 8. This process of iterated grafting preserves the truth of $\varphi \wedge \langle\sigma\rangle\Diamond^*\psi$ and thus we can put the desired bound k for the satisfiability of the formula $\varphi \wedge \langle\sigma\rangle\Diamond^k\psi$ given the satisfiability of the formula $\varphi \wedge \langle\sigma\rangle\Diamond^*\psi$. Below we make this idea more precise.

Lemma 2 (Grafting for $\Box$)**.** *Let* p *be an sd-PAL-protocol on* $\mathcal{M} = (W, \sim, V)$ *and* φ *a TAPAL-formula. Let* $w\sigma \in \mathsf{p}(w)$*. Put* $\mathcal{H} = \mathsf{Forest}(\mathcal{H}, \mathsf{p})$ *and* $ibi(\varphi) = m$*. Also let* τ *be a sequence of TAPAL-formula such that* $\mathsf{len}(\tau) \geq m$*. Finally let* $\top_0 \notin AOC(\varphi) \cup AOC(\mathsf{p})$*. Then, for every* $v \in \Sigma^*_{pal}$ *and* $w \in W$*, if* $w\sigma\tau v$ *is in* $\mathcal{H}$*,*

$$\mathcal{H}, w\sigma \models \varphi \quad \Rightarrow \quad \mathcal{H}^{[\sigma\tau v \mapsto \sigma\tau!\top_0]}, w\sigma \models \varphi.$$

Proof. The proof can be given by straightforward induction on φ in terms of the equivalent formulation of the formulas in *TAPAL* as in Observation 3. $\Box$

Theorem 2. *If* $\varphi \wedge \langle\sigma\rangle\Diamond^*\psi$ *is satisfiable in* $\mathbb{F}(\mathbb{PAL})$*, then* $\varphi \wedge \langle\sigma\rangle\Diamond^k\psi$ *is satisfiable in* $\mathbb{F}(\mathbb{PAL}^+)$ *for some* k *such that* $0 \leq k \leq ibi(\varphi) \dot{-} \mathsf{len}(\sigma) + 1$*, where* $a \dot{-} b = a - b$ *if* $a - b > 0$*;* $a \dot{-} b = 0$ *otherwise.*

Proof. Let $\mathcal{H} = \mathsf{Forest}(\mathcal{M}, \mathsf{p})$ and $w\tau$ in $\mathcal{H}$ with w in $\mathcal{M}$. Assume that $\mathcal{H}, w\tau \models \varphi \wedge \langle\sigma\rangle\Diamond^*\psi$. By the semantics of $\Diamond^*$, there is some $v = v_0 \dots v_{n-1}$ such that

$$\mathcal{H}, w\tau \models \varphi \wedge \langle\sigma\rangle\langle v\rangle\psi. \tag{1}$$

If $ibi(\varphi) \dot{-} \mathsf{len}(\sigma) \geq \mathsf{len}(v)$, we are done since we have $\mathcal{H}, w\tau \models \varphi \wedge \langle\sigma\rangle\Diamond^k\psi$ for some $k \leq ibi(\varphi) \dot{-} \mathsf{len}(\sigma) + 1$.

Thus suppose $ibi(\varphi) \dot{-} \mathsf{len}(\sigma) < \mathsf{len}(v)$. Let $a = \mathsf{len}(\sigma)$ and $b = ibi(\varphi)$. Then take a sequence of distinct tautologous formulas in $\mathcal{L}_{pal}$, $\top_0, \dots, \top_{[(b-a)-1]+1}$. By a similar argument given in the proof of Theorem 1, we can assume that $\top_0, \dots, \top_{[(b-a)-1]+1} \notin AOC(\mathsf{p})$. Then, define

$$\mathcal{H}' = (\dots(\mathcal{H}^{[w\tau\sigma v_0 \mapsto w\tau\sigma!\top_0]})\dots)^{[w\tau\sigma!\top_0 \dots !\top_{(b-a)-2} v_{(b-a)-1} \mapsto w\tau\sigma!\top_0 \dots !\top_{(b-a)-2}!\top_{(b-a)-1}]}$$

By repeatedly applying Lemma 1, we have

$$\mathcal{H}', w\tau \models \varphi \tag{2}$$

Also since (1) implies

$$\mathcal{H}', w\tau\sigma v_0 \dots v_{(b-a)-1} \models \langle v_{(b-1)} \dots v_{n-1}\rangle\psi,$$

by repeatedly applying Lemma 7, we have

$$\mathcal{H}', w\tau\sigma!\top_0 \dots !\top_{(b-a)-1} \models \langle v_{(b-a)} \dots v_{n-1}\rangle\psi.$$

and thus $\mathcal{H}', w\tau\sigma!\top_0 \dots !\top_{(b-a)-1} \models \Diamond^*\psi$. Here consider the model

$$\mathcal{H}'' := (\mathcal{H}')^{[w\tau!\top_0 \dots !\top_{(b-a)-1} v_{b-a} \dots v_{n-1} \mapsto w\tau!\top_0 \dots !\top_{(b-a)-1}!\top_{b-a}]}$$

By the argument given in the proof of Proposition 8, this implies

$$\mathcal{H}'', w\tau!\top_0 \ldots !\top_{(b-a)-1}!\top_{b-a} \models \psi.$$

This gives us

$$\mathcal{H}'', w\tau \models \Diamond^{b-a+1}\psi.$$

In addition, (2) together with Lemma 2 implies

$$\mathcal{H}'', w\tau \models \varphi.$$

Therefore, we have $\varphi \wedge \Diamond^{b-a+1}$ is satisfied in $\mathcal{H}''$, which is clearly in $\mathbb{F}(\mathbb{PAL})$ by construction (and Observation 1 and 2). □

Corollary 3. *Soundness of* $R(\Box^*)$ $R(\Box^*)$ *is sound with respect to the class* $\mathbb{F}(\mathbb{PAL})$.

Proof. Immediate from the above theorem and the fact that

$$ibi(\varphi)\dot{-}\mathsf{len}(\sigma) + 1 \leq ibi(\varphi) + 1 \leq |\varphi| + 1.$$

□

8 Completeness

Finally we prove the weak completeness of TAPAL. For this, we construct the canonical model by following the method in van Benthem et al. (2008). The completeness proof of TAPAL differs from that of the system given in van Benthem et al. (2008), since the construction is not from the set of all the maximal consistent set, but from the set of the maximal consistent sets Σ with the following special properties.

Definition 8.1. Saturation wrt $\Diamond$ A set Σ of formula is *saturated with respect to* $\Diamond$, if, for every sentence of the form $\langle\sigma\rangle\Diamond\varphi$ with $\sigma \in \Sigma^*_{pal}$, $\langle\sigma\rangle\Diamond\varphi \in \Sigma$ implies that there is some formula θ such that $\langle\sigma\rangle\langle!\theta\rangle\varphi \in \Sigma$.

Definition 8.2. Saturation wrt $\Diamond^*$ A set Σ of formulas is *saturated with respect to* $\Diamond^*$, if, for every formula of the form $\langle\sigma\rangle\Diamond^*\varphi$ with $\sigma \in \Sigma^*_{pal}$, $\langle\sigma\rangle\Diamond^*\varphi \in \Sigma$ implies that there is some n such that $\langle\sigma\rangle\Diamond^n\varphi \in \Sigma$.

The motivation for these properties is to make sure that there are formulas that "witness" $\Diamond$ and $\Diamond^*$ in every formula in a given maximally consistent set. Here the analogy mentioned in the above remark between $R(\Box)$ and the first-order rule comes back again. In the proof below, when we construct a maximal consistent set from a consistent formula, we add witnessing formulas for the formulas of the above form. The consistency of the resulting set with witnessing formulas will be guaranteed by the rule $R(\Box)$, and this is very similar to the way that the first-order rule in question (or its equivalent) is used in the completeness proof of first-order logic. Similarly, $R(\Box^*)$ gives a witness for $\Diamond^*\varphi$ by finding an appropriate n for $\Diamond^n\varphi$ to be added, consistently, to a set, when we construct maximally consistent sets. These roles of the two rules are clear in the proof of the following lemma.

Lemma 3 (Lindenbaum Lemma)**.** *Every consistent TAPAL-formula φ can be expanded to a maximal consistent set saturated with respect to $\Diamond$ and $\Diamond^*$.*

Proof. Let $\alpha_0, .\alpha_1...$ be an enumeration of the TAPAL-formulas such that $\alpha_0 = \varphi$. We construct a sequence $\Sigma_0, \Sigma_1, ...$ of sets as follows:

- $\Sigma_0 = \emptyset$
- If $\Sigma_n \cup \{\alpha_n\}$ is inconsistent, then $\Sigma_{n+1} = \Sigma_n$.
- If $\Sigma_n \cup \{\alpha_n\}$ is consistent and α_n is neither of the form $\langle\sigma\rangle\Diamond\psi$ nor of the form $\langle\sigma\rangle\Diamond^*\psi$, then $\Sigma_{n+1} = \Sigma_n \cup \{\alpha_n\}$.
- If $\Sigma_n \cup \{\alpha_n\}$ is consistent and α_n is of the form $\langle\sigma\rangle\Diamond\psi$, then $\Sigma_{n+1} = \Sigma_n \cup \{\langle\sigma\rangle\Diamond\psi, \langle\sigma\rangle\langle !\top_0\rangle\psi\}$ for a tautologous formula $\top_0$ in $\mathcal{L}_{pal}$ such that $\top_0 \notin AOC(\psi) \cup AOC(\sigma) \cup \bigcup_{\theta\in\Sigma_n} AOC(\theta)$. Such a tautologous formula exists since Σ_n is finite and we have a countable number of tautologous formulas in $\mathcal{L}_{pal}$.
- If $\Sigma_n \cup \{\alpha_n\}$ is consistent and α_n is of the form $\langle\sigma\rangle\Diamond^*\psi$, then take k such that $\Sigma_n \cup \{\langle\sigma\rangle\Diamond^*\psi, \langle\sigma\rangle\Diamond^k\psi\}$ is consistent and put $\Sigma_{n+1} = \Sigma_n \cup \{\langle\sigma\rangle\Diamond^*\psi, \langle\sigma\rangle\Diamond^k\psi\}$.

We show by induction that Σ_n is consistent for $n \geq 1$. The base case is given by the assumption that φ is consistent. Assume that Σ_n is consistent for an arbitrary n. Clearly it suffices to show the following claims:

Claim 1: Σ_{n+1} is consistent, if α_n is of the form $\langle\sigma\rangle\Diamond\psi$.

Claim 2: If $\Sigma_n \cup \{\alpha_n\}$ is consistent and α_n is of the form $\langle\sigma\rangle\Diamond^*\psi$, there is some m such that $\Sigma_n \cup \{\alpha_n, \langle\sigma\rangle\Diamond^m\psi\}$ is consistent.

Proof of Claim 1 Suppose Σ_{n+1} is inconsistent. Then, there must be some formulas $\psi_1, \psi_2, ..., \psi_l \in \Sigma_m \cup \{\langle\sigma\rangle\Diamond\psi\}$ such that

$$\vdash (\psi_1 \wedge ... \wedge \psi_l) \rightarrow \neg\langle\sigma\rangle\langle !\top_0\rangle\psi.$$

However, this implies

$$\vdash (\psi_1 \wedge ... \wedge \psi_l) \rightarrow [\sigma][!\top_0]\neg\psi.$$

Since $\top_0$ is chosen so that $\top_0 \notin AOC(\psi) \cup AOC(\sigma) \cup \bigcup_{\theta\in\Sigma_n} AOC(\theta)$, we can apply $R(\Box)$ to obtain

$$\vdash (\psi_1 \wedge ... \wedge \psi_l) \rightarrow [\sigma]\Box\neg\psi$$

This gives us $\Sigma_m \vdash [\sigma]\Box\neg\psi$ and $\Sigma_m \vdash \neg\langle\sigma\rangle\Diamond\psi$. However this contradicts the assumption that $\Sigma_n \cup \{\alpha_n\}$ is consistent.

Proof of Claim 2: Suppose toward contradiction that there is no such m. Then, for all $m \geq 0$, we have:

$$\vdash \bigwedge \Sigma_n \rightarrow \neg\langle\sigma\rangle\Diamond^m\psi.$$

where $\bigwedge \Sigma_n$ is a conjunction of the formulas in Σ_{m-1}. This implies that, for all m,

$$\vdash \bigwedge \Sigma_n \cup \{\alpha_n\} \to [\sigma]\Box^m \neg\psi$$

and by $R(\Box^*)$

$$\vdash \bigwedge \Sigma_n \cup \{\alpha_n\} \to [\sigma]\Box^* \neg\psi.$$

Therefore, we have $\Sigma_n \cup \{\alpha_n\} \vdash [\sigma]\Box^*\neg\psi$ and thus $\Sigma_n \cup \{\alpha_n\} \vdash \neg\langle\sigma\rangle\Diamond^*\psi$. This contradicts our assumption that $\Sigma_n \cup \{\langle\sigma\rangle\Diamond^*\psi\}$ is consistent.

Now take $\Sigma' = \bigcup_{i=0}^{\infty} \Sigma_i$. The maximality and saturation with respect to $\Diamond$ and $\Diamond^*$ is clear by the construction. The consistency is shown in the standard way by the consistency of Σ_n for $n \geq 1$. □

Definition 8.3. Base Epistemic Model We define $\mathcal{M}_0 = (W_0, \sim_0, V_0)$ as follows:

- W_0 : the set of the maximal consistent sets in $\mathcal{L}_{tapal}$ saturated with respect to $\Diamond$ and $\Diamond^*$.
- $\sim_0$: $\mathcal{A} \to \wp(W_0 \times W_0)$: for $a, b \in W_0, (a, b) \in\sim_0 (i)$ iff $\{\varphi | [i]\varphi \in a\} \subseteq b$.
- $V_0(p) = \{w \in W_0 \mid p \in w\}$

Definition 8.4. Canonical Histories Let W_0 be the set of the maximal consistent sets of $\mathcal{L}_{tapal}$. We define λ_n and G_n are defined as follows:

- Define $G_0 = W_0$ and for each $w \in G_0$, $\lambda_0(w) = w$.
- Let $G_{n+1} = \{h!\theta | h \in G_n$ and $\langle !\theta\rangle\top \in \lambda_n(h)\}$. For every $h = h'!\theta \in G_{n+1}$, define $\lambda_{n+1}(h) = \{\varphi | \langle !\theta\rangle\varphi \in \lambda_n(h')\}$.

Lemma 4. *For each $n \geq 0$, for each $\sigma \in G_n$, $\lambda_n(\sigma)$ is a maximally consistent set that is saturated with respect to $\Diamond$ and $\Diamond^*$.*

Proof. Straightforward induction on n. □

We define *the canonical history* G_{can} as $\cup^\infty G_i$ and a function $\lambda : G_{can} \to W_0$ as $\lambda(\sigma) = \lambda_n(\sigma)$ for each $n = \mathsf{len}(\sigma)$.

Definition 8.5. Canonical Model The canonical model of TAPAL is $\mathcal{G}_{can} = (G_{can}, \sim_{can}, V_{can})$, where each element is defined as follows:

- $G_{can} := \cup^\infty G_i$.
- $\sim_{can}$: $\mathcal{A} \mapsto G_{can} \times G_{can}$: for $g, h \in G_{can}$ where $h = w\sigma$ and $g = v\tau$, $(h, g) \in\sim_{can} (i)$ iff $(w, v) \in\sim_0 (i)$ and $\sigma = \tau$.
- For every $p \in \mathsf{At}$, $h \in V_{can}(p)$ iff $w \in V_0(p)$ where $h = wh'$.

Proposition 9. *Let $\sigma \in \Sigma^*_{pal}$ and $\mathsf{len}(\sigma) = n$. Then,*

1. $\vdash \langle\sigma\rangle\varphi \to \Diamond^n\varphi$.
2. $\vdash \Diamond^n\varphi \to \Diamond^*\varphi$.

Proof. Straightforward. The proof for the second appeals to the axiom A3. □

Lemma 5. *(Truth Lemma) For every formula $\varphi \in \mathcal{L}_{tapal}$,*

$$\varphi \in \lambda(h) \quad \textit{iff} \quad \mathcal{G}_{can}, h \models \varphi.$$

Proof. : The proof is by induction on φ. For the cases other than $\Diamond$ and $\Diamond^*$, the argument is the same as in van Benthem et al. (2008).

Assume that φ is of the form $\Diamond\psi$. First assume that $\Diamond\psi \in \lambda(h)$. Since $\lambda(h)$ is saturated with respect to $\Diamond$, we have $\langle !\theta\rangle\psi \in \lambda(h)$ for some θ. By the construction of $\mathcal{G}_{can}$, we have $\psi \in \lambda(h!\theta)$. By IH, we obtain $\mathcal{G}_{can}, h!\theta \models \psi$. Therefore, we have $\mathcal{G}_{can}, h \models \Diamond\psi$ by truth definition.

For the other direction, assume that $\mathcal{G}_{can}, h \models \Diamond\psi$. By definition, there is some θ such that $h!\theta \in G_{can}$ and $\mathcal{G}_{can}, h!\theta \models \psi$. By IH, we have $\psi \in \lambda(h!\theta)$, which, by the construction of $\mathcal{G}_{can}$, implies $\langle !\theta\rangle\varphi \in \lambda(h)$. This implies by A2 that $\Diamond\varphi \in \lambda(h)$.

Next, assume that $\Diamond^*\psi \in \lambda(h)$. Since $\lambda(h)$ is a maximally consistent set saturated with respect to $\Diamond^*$, there is some $k \geq 0$ such that $\Diamond^k\psi \in \lambda(h)$ Now, since $\lambda(h)$ is also saturated with respect to $\Diamond$, we have $\langle !\theta_1\rangle \ldots \langle !\theta_k\rangle\psi \in \lambda(h)$. Thus, by the construction of canonical model, we have $\psi \in \lambda(h!\theta_1 \ldots !\theta_k)$, which implies by IH that $\mathcal{G}_{can}, h!\theta_1 \ldots !\theta_k \models \psi$. This gives us $\mathcal{G}_{can}, \sigma \models \Diamond^*\psi$.

Assume that $\mathcal{G}_{can}, h \models \Diamond^*\psi$. By definition, this is equivalent to saying that there is some σ such that $h\sigma \in \mathcal{G}_{can}$ and $\mathcal{G}_{can}, h\sigma \models \psi$. By IH, we have $\psi \in \lambda(h\sigma)$, which, by the construction of λ, implies $\langle\sigma\rangle\psi \in \lambda(h)$. By Proposition 9, we have that $\Diamond^*\psi \in \lambda(h)$. □

To obtain completeness, we still need to show that the canonical model is in the class $\mathbb{F}(\mathbb{PAL})$. For this, we need to prove some properties of PAL-generated models. Given two sequences σ and τ, we denote by $\sigma\tau$ the sequence obtained by concatenating σ and τ in order.

Proposition 10. *Given an epistemic model $\mathcal{M}$ and an sd-protocol p on $\mathcal{M} = (W, \sim, V)$, define $\mathcal{H}^{\mathsf{p}}_x$ for every $x \in \mathbb{N}^*$ as defined in 3.5. Let $y, z \in \mathbb{O}$, $n \geq 1$. Further, suppose $yn \ll z$. For every $h \in H^{\mathsf{p}}_{yn}$ and every $\varphi \in \mathcal{L}_{pal}$ with $o(\varphi) \leq n$,*

$$\mathcal{H}^{\mathsf{p}}_{yn}, h \models \varphi \Leftrightarrow \mathcal{H}^{\mathsf{p}}_z, h \models \varphi.$$

Proof. First, observe that, by Definition 3.5, $h \in H^{\mathsf{p}}_{yn}$ implies that $h \in H^{\mathsf{p}}_z$ (by the assumption that $yn \ll z$). Thus, on the assumption that $h \in H^{\mathsf{p}}_{yn}$, we have $h \in H^{\mathsf{p}}_{yn} \Leftrightarrow h \in H^{\mathsf{p}}_z$. Denote this fact by (i). We show the claim by induction on φ. The base and boolean cases are clear. Suppose φ is of the form $[i]\varphi$. Assume LHS. Let $(h, h') \in \sim^{\mathsf{p}}_{yn} (i)$. Then, by IH, $\mathcal{H}^{\mathsf{p}}_z, h' \models \psi$. Here, by the construction in Definition 3.5 and the fact (i), it follows that $(h, h') \in \sim^{\mathsf{p}}_{yn} (i) \Leftrightarrow (h, h') \in \sim^{\mathsf{p}}_z (i)$. Thus we have $\mathcal{H}^{\mathsf{p}}_z, h \models [i]\psi$. The other way is similar.

Next suppose φ is of the form $\langle !\theta\rangle\psi$. First, LHS is equivalent to $\mathcal{H}^{\mathsf{p}}_{yn}, h!\theta \models \psi$. Furthermore, since $o(!\varphi) \leq n$, we have $o(!\theta) \leq n$ by the definition of o. By this fact and the construction in Definition 3.5, $h \in H^{\mathsf{p}}_{yn}$ implies that $h!\theta \in H^{\mathsf{p}}_{yn} \Leftrightarrow h!\theta \in H^{\mathsf{p}}_z$ (by the same reasoning as for the fact (i)). Thus, we can apply IH and obtain $\mathcal{H}^{\mathsf{p}}_{yn}, h!\theta \models \psi \Leftrightarrow \mathcal{H}^{\mathsf{p}}_z, h!\theta \models \psi$. This gives us the equivalence between LHS and RHS. □

Let $\mathcal{M}_0 = (W_0, \sim_0, V_0)$ be the base epistemic model, from which the canonical *ETL*-model is constructed. Also, let $\mathcal{G} = (G, \approx, U)$ be the canonical model.

Define p_0 on $\mathcal{G}$ so that $\mathsf{p}_0(w) = \{\sigma | w\sigma \in G\}$ for all $w \in W_0$. Given $\mathcal{M}_0$ and p_0, generate $\mathcal{H}^{\sigma,\mathsf{p}_0}$ and $\mathcal{H}_x^{\mathsf{p}_0}$ for $\sigma \in (\Sigma_{pal})^*$ and $x \in \mathbb{O}$, as defined in Definition 3.5. For simplicity, we write $\mathcal{H}^\sigma$ and $\mathcal{H}_x$ respectively for $\mathcal{H}^{\sigma,\mathsf{p}_0}$ and $\mathcal{H}_x^{\mathsf{p}_0}$. Also let $\mathcal{H} = (H, \sim, V) = Forest(\mathcal{M}_0, \mathsf{p}_0)$.

Proposition 11. *Let $w \in W_0$ and $\sigma \in \Sigma^*_{pal}$. Assume $v\sigma \in G \Leftrightarrow v\sigma \in H^\sigma$ for every $v \in W_0$ (Denote by "Assumption 1"). Then, for every $\varphi \in \mathcal{L}_{pal}$,*

$$\mathcal{G}, w\sigma \models \varphi \Leftrightarrow \mathcal{H}_{o(\sigma)o(\varphi)}, w\sigma \models \varphi.$$

Proof. We go by induction on φ. The base and boolean cases are straightforward. Suppose that φ is of the form $[i]\psi$. Assume $\mathcal{G}, w\sigma \models [i]\psi$. Let w' be such that $(w, w') \in\sim_0 (i)$. Then we have $(w\sigma, w'\sigma) \in\approx (i)$ by construction, and thus $\mathcal{G}, w'\sigma \models \psi$. Thus, by IH, $\mathcal{H}_{o(\sigma)o(!\varphi)}, w'\sigma \models \psi$. Put $\mathcal{H}_{o(\sigma)o(!\psi)} = \mathcal{H}' = (H', \sim', V')$. Here, note, for every $u \in W_0$, $u\sigma \in H^\sigma \Leftrightarrow u\sigma \in H'$, by the construction in Definition 3.5. Therefore, Assumption 1 implies that, for any u, $(w\sigma, u\sigma) \in\sim' (i) \Leftrightarrow (w\sigma, u\sigma) \in\approx (i)$. This gives us $\mathcal{H}', w\sigma \models [i]\psi$. Here $\mathcal{H}' = \mathcal{H}_{o(\sigma)o(\psi)} = \mathcal{H}_{o(\sigma)o([i]\psi)}$, since $o(\psi) = o([i]\psi)$ by the definition of o. Thus, we obtain the LHS-RHS direction. The other direction is similar.

Next, suppose φ is of the form $\langle !\theta\rangle\psi$. First, we claim that Assumption 1 implies $v\sigma!\theta \in G \Leftrightarrow v\sigma!\theta \in H^{\sigma!\theta}$ for all $v \in W_0$.

Proof of the claim: $v\sigma!\theta \in G$ implies $\langle !\theta\rangle\top \in \lambda(w\sigma)$, and by A1, $\theta \in \lambda(w\sigma)$. By truth lemma, we have $\mathcal{G}, v\sigma \models \theta$. Thus, by IH, $\mathcal{H}_{o(\sigma)o(!\theta)}, v\sigma \models \theta$. Since we have $\sigma!\theta \in \mathsf{p}_0(v)$ by the construction of p_0, we have $v\sigma!\theta \in H^{\sigma!\theta}$. The other direction is similar.

Now, assume the LHS of the biconditional. It implies that $\mathcal{G}, w\sigma!\theta \models \psi$. By the claim, we can apply IH and obtain $\mathcal{H}_{o(\sigma!\theta)o(!\psi)}, w\sigma!\theta \models \psi$. Here, note that $o(\sigma!\theta)o(!\psi) = o(\sigma)o(!\theta)o(!\psi) \ll o(\sigma)o(!\varphi)$ since $o(!\theta), o(!\psi) < o(!\langle !\theta\rangle\psi)$. Thus, applying Proposition 10, we obtain $\mathcal{H}_{o(\sigma)o(!\langle !\theta\rangle\psi)}, w\sigma!\theta \models \psi$. Therefore, we have $\mathcal{H}_{o(\sigma)o(!\langle !\theta\rangle\psi)}, w\sigma \models \langle !\theta\rangle\psi$, as desired. The RHS-LHS direction is similar. □

Lemma 6 (Canonicity)**.** *The canonical model $\mathcal{G}$ is in $\mathbb{F}(\mathbb{PAL})$.*

Proof. It suffices to show the following claim:

Claim 1: For every $w \in W_0$ and every $\sigma \in (\mathsf{PAL}^+)^*$, $w\sigma \in G \Leftrightarrow w\sigma \in H^\sigma$.

For this implies $G = H$ and then, by inspecting the constructions of *PAL*-generated *ETL*-models and the canonical model, we see that $\mathcal{G} = \mathcal{H}$.

We go by complete induction on the order of σ. The base case ($o(\sigma) = \lambda$) is clear by Definition 8.5 and 3.5. Assume that the claim holds for every τ such that $o(\tau) \ll o(\sigma)$. Let $\sigma = \sigma_1 \ldots \sigma_k$. Suppose that $w\sigma_1 \ldots \sigma_k \in G$. This implies $\mathcal{G}, w\sigma_1 \ldots \sigma_{k-1} \models \sigma_k$ (by truth lemma) and $\sigma_1 \ldots \sigma_k \in \mathsf{p}_0(w)$. Also IH implies that, for every $v \in W_0$, $w\sigma_1 \ldots \sigma_{k-1} \in G \Leftrightarrow w\sigma_1 \ldots \sigma_{k-1} \in H^{\sigma_1 \ldots \sigma_{k-1}}$. By the construction in Definition 3.5, this is equivalent to:

For every $v \in W_0$, $w\sigma_1 \ldots \sigma_{k-1} \in G \Leftrightarrow w\sigma_1 \ldots \sigma_{k-1} \in H_{o(\sigma_1 \ldots \sigma_{k-1})}$.

Thus, we can apply Proposition 11 and obtain $\mathcal{H}_{o(\sigma_1\ldots\sigma_{k-1})o(\sigma_k)}, w\sigma_1\ldots\sigma_{k-1} \models \sigma_k$. Given $\sigma_1\ldots\sigma_k \in \mathsf{p}_0(w)$, we have $w\sigma_1\ldots\sigma_k \in H^{\sigma_1\ldots\sigma_k}$ by Definition 3.5. The other way is similar. □

Theorem 3 (Completeness)**.** *TAPAL is weakly complete with respect to* $\mathbb{F}(\mathbb{PAL})$.

Proof. By a standard argument with Lemma 5 and 6 □

9 Conclusion and Discussion

We have provided a framework for PAL that can represent protocol constraints on what is announceable. We did this by assigning a protocol to each point of a given epistemic model and constructing ETL-models by successively applying the sequences of public announcements based on the assigned protocols. In the resulting framework TAPAL, the public announcement operators $\langle !\varphi\rangle$, $\Diamond$, and $\Diamond^*$, exhibit semantic properties distinct from those in the framework of PAL. Also by appealing to some distinct properties of our PAL-generated ETL-models, we have provided the complete axiomatization of TAPAL.

We now conclude by discussing some open problems concerning our framework for future research.

Complexity

We did not discuss the complexity of TAPAL. For the fragment TPAL, we can finitize the completeness proof (by a method similar to the one in van Benthem et al. (2008)) and show its decidability. In addition, we can prove that PAL is faithfully embeddable to TPAL. Given that the satisfiability problem for PAL is PSPACE-complete, this result tells us that the satisfiability problem for TPAL is at least PSPACE hard. However, the decidability and complexity of TAPAL is unknown.

Other Kinds of Operators

We can extend TAPAL by various interesting operators that have been considered in the literature. One important operator is the *common knowledge* operator. However, in the presence of $\Diamond^*$, we expect that we obtain the undecidability result by Miller and Moss (2005). Another kind of operators are past operators, such as "The public announcement of φ has been made before which...". Operators of the kind have been considered in the framework of DEL by Yap (2007). Tree structures of our framework seem to accommodate the semantic intuition of those operators. See Hoshi and Yap (2009).

Extension to DEL

Our framework only deals with a specific kind of informational events, i.e. public announcements. Thus it is natural to ask whether we can provide a similar framework for the full class of informational events in DEL. In fact, we can redefine protocols so that they consist of sequences of event models, and provide a method to generate ETL-models. For such a class of ETL-models, we can give a complete axiomatization in the language of DEL by making minor modifications to our proofs above. However, we have not obtained the axiomatization for the language of DEL with the generalized operators, "There is some single event after which…" and "There is some sequence of events after which…".

Acknowledgements I acknowledge my gratitude to Johan van Benthem, Cédric Dégremont, Yu Junhua, and Eric Pacuit for their helpful comments on earlier drafts of this paper.

References

P. Balbiani, A. Baltag, H. van Ditmarsch, A. Herzig, T. Hoshi, and de Lima. Arbitrary announcement logic. *Review of Symbolic Logic*, 2008. To appear.

A. Baltag, L. Moss, and S. Solecki. The logic of public announcements, common knowledge and private suspicions. In I. Gilboa, editor, *TARK 1998*, number 43-56, 1998.

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

R. Fagin, J. Halpern, Y. Moses, and M. Vardi. *Reasoning about Knowledge*. Synthese Library. MIT Press, Boston, 1995.

J. Gerbrandy. *Bisimulations on Planet Kripke*. PhD thesis, ILLC, 1999.

T. Hoshi. Logics of public announcement with constrained protocols. In *LOFT*, 2008.

T. Hoshi and A. Yap. Dynamic epistemic logic with branching temporal structures. To appear in Knowledge, Rationality and Action, 2009.

J. Miller and L. Moss. The undecidability of iterated modal relativization. *Studia Logica*, 79(3):373–407, 2005.

R. Parikh and R. Ramanujam. A knowledge based semantics of messages. *Journal of Logic, Language, and Information*, 12:453–467, 2003.

J. Plaza. Logics of public announcements. In *Proceedings 4th International Symposium on Methodologies for Intelligent Systems: Poster Session Program*, pages 201–216, 1989.

J. van Benthem. One is lonely number: On the logic of communication. Technical Report PP-2002-27, University of Amsterdam, 2002.

J. van Benthem, J. Gerbrandy, T. Hoshi, and E. Pacuit. Merging frameworks for interaction. *Journal of Philosophical Logic*, 2008. To appear.

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

A. Yap. Dynamic epistemic logic and temporal modality. University of Victoria, 2007.

Inference and Update

Fernando R. Velázquez-Quesada
Institute for Logic, Language and Computation, Universiteit van Amsterdam.
`F.R.VelazquezQuesada@uva.nl`

Abstract

We look at two fundamental logical processes, often intertwined in planning and problem solving: inference and update. Inference is an internal process with which we uncover what is implicit in the information we already have. Update, on the other hand, is produced by external communication, usually in the form of announcements and in general in the form of observations, giving us information that might not have been available (even implicitly) before. Both processes have received attention from the logic community, usually separately. In this work, we develop a logical language that allows us to describe them together. We present syntax, semantics and a complete axiom system; we discuss similarities and differences with other approaches and mention how the work can be extended.

1 Introduction

Consider the following situation, from van Benthem (2008a):

> You are in a restaurant with your parents, and you have ordered fish, meat, and vegetarian, for you, your father and your mother, respectively. Now a new waiter comes from the kitchen with the three dishes. What can he do to get to know which dish corresponds to which person?

The waiter can ask *"Who has the fish?"*; then he can ask *"Who has the meat?"*. Now he does not have to ask anymore: "two questions plus one inference are all that is needed" (van Benthem 2008a).

The present work looks at these two fundamental logical processes, often intertwined in real-life activities. Inference is an *internal* process: the agent revises her own information in search of what can be derived from it. Update, on the other hand, is produced by *external* communication: the agent gets new information via observations. Both are logical processes, both describe dynamics of information, both are used in every day situations and still, they have been studied separately.

Inference has been the main subject of study of logic, allowing us to extract new information from what we already have. Among the most important branches, we can mention Hilbert-style proof systems, natural deduction and tableaux. Recent works, like Duc (Duc 1995; 1997) and Jago (Jago 2006a;b) have incorporated modal logics to the field, representing inference as a non-deterministic step-by-step process.

Update, on the other hand, has been the main subject of *Dynamic Epistemic Logic* (van Ditmarsch et al. 2007). Works like Plaza (1989) and Gerbrandy (1999) turned attention to the effect public announcements have on the knowledge of an agent. Many works have followed them, including the study of more complex actions (Baltag et al. 1999, van Ditmarsch 2000) and the effect of announcements on more propositional attitudes (the soft/hard facts of van Benthem (2007); the knowledge/belief of Baltag and Smets (2008)).

In van Benthem (2008c), the author shows how these two phenomena fall directly within the scope of modern logic. As he emphasizes, "asking a question and giving an answer is just as 'logical' as drawing a conclusion!". Here, we propose a merging of the two traditions. We consider that both processes are equally important, but so is their interaction. In this work, we develop a logical language that allows us to express inference and update together.

We start in Section 2 by providing a framework for representing *implicit* and *explicit* information, and isolate the case where this information is true. Section 3 provides a representation of truth-preserving *inference*; moreover, we show how dynamics of the inference process itself can be represented. Section 4 introduces the other logical process: *update*. Then we compare our proposal with other approaches (Section 5) and present a summary and further work (Section 6). We focus in the single-agent case, leaving group-information concepts for future work.

2 Implicit and explicit information

The *Epistemic Logic* (EL) framework with Kripke models (Hintikka 1962) is one of the most widely used for representing and reasoning about an agent's information. Nevertheless, it is not fine enough to represent the restaurant example above. Agents whose information is represented with this framework suffer from what Hintikka called the *logical omniscience* problem[1]: they are informed of all validities and their information is closed under truth-preserving inference.

This feature, useful in some applications, is too much in some others. More importantly for us, it *hides* the inference process. If we represent our example within such framework, the answer to the second question tells the waiter not only that your father will get the meat but also that your mother will get the vegetarian dish. In this case, the inference is very simple, but in general this is not the case: proving a theorem, for example, consists on applications of deductive inference steps to show that the conclusion indeed follows from the premises. Some theorems may be straightforward, but some are not.

As argued in van Benthem (2006), we can give the modal operator another interpretation, reading the formula $\Box\varphi$ as *"the agent is implicitly informed about*

[1]See Sim (1997) for a survey about the logical omniscience problem.

φ". We follow that idea, and we extend EL to also represent *explicit* information; moreover, we also provide a mechanism with which the agent can increase it. The work of this section resembles those presented in Fagin and Halpern (1988), Duc (Duc 1995; 1997) and Jago (Jago 2006a;b); the precise relation will be clarified in Section 5.

2.1 Formulas, rules and the explicit/implicit information language

The agent's explicit information is given by a set of formulas; the mechanism that allows her to increase it is given by syntactic rules. We start by defining the language to represent explicit information and by indicating what a rule in that language is.

Definition 2.1 (Formulas and rules)**.** Let $\mathtt{P}$ be a set of atomic propositions. The *internal language* $\mathcal{I}$ is given by the propositional language over $\mathtt{P}$. A *rule* based on $\mathcal{I}$ is a pair (Γ, γ) (sometimes represented as $\Gamma \Rightarrow \gamma$) where Γ is a finite set of formulas and γ is a formula, all of them in $\mathcal{I}$. Given a rule $\rho = (\Gamma, \gamma)$, we call Γ the *set of premises of* ρ ($\text{prem}(\rho)$) and γ the *conclusion of* ρ ($\text{conc}(\rho)$). We denote by $\mathcal{R}$ the set of rules based on formulas of $\mathcal{I}$.

With this internal language, the agent can have explicit information about the real world but not about her own information. This is indeed a limitation, but it makes the update definition of Section 4 possible. In Section 6 we briefly discuss the reasons for this limitation, leaving a deep analysis for further work.

Our language extends that of EL by adding two kinds of formulas: one for expressing the agent's explicit information ($I\,\gamma$) and another expressing the rules she can apply ($L\,\rho$).

Definition 2.2 (Explicit/implicit information language $\mathcal{EI}$)**.** Let $\mathtt{P}$ be a set of atomic propositions. The formulas of the *explicit/implicit information language* $\mathcal{EI}$ are given by

$$\varphi ::= \top \mid p \mid I\,\gamma \mid L\,\rho \mid \neg\varphi \mid \varphi \vee \psi \mid \Diamond\varphi$$

with $p \in \mathtt{P}$, $\gamma \in \mathcal{I}$ and $\rho \in \mathcal{R}$. The boolean connectives $\wedge$, $\rightarrow$ and $\leftrightarrow$ as well as the modal operator $\Box$ are defined as usual.

The semantic model extends a Kripke model by assigning two new sets to each possible world: one indicating the *formulas* the agent is explicitly informed about, and other indicating the *rules* she can apply.

Definition 2.3 (Explicit/implicit information model)**.** Let $\mathtt{P}$ be a set of atomic propositions. An *explicit/implicit information model* is a tuple $M = \langle W, R, V, Y, Z\rangle$ where:

- $\langle W, R, V\rangle$ is a standard Kripke model with W the non-empty set of worlds, $R \subseteq W \times W$ the *accessibility relation* and $V : W \to \wp(\mathtt{P})$ the *atomic valuation function*.
- $Y : W \to \wp(\mathcal{I})$ is the *information set function*, satisfying ***coherence for formulas*** (if $\gamma \in Y(w)$ and Rwu, then $\gamma \in Y(u)$).
- $Z : W \to \wp(\mathcal{R})$ is the *rule set function* satisfying ***coherence for rules*** (if $\rho \in Z(w)$ and Rwu, then $\rho \in Z(u)$).

We denote by **EI** the class of all explicit/implicit information models.

Our restrictions reflect the following idea. The sets $Y(w)$ and $Z(w)$ represent the formulas and rules the agent is explicitly informed about; if while staying in w the agent considers u possible, it is natural to ask for u to preserve the agent's explicit information at w.

Definition 2.4 (Semantics for $\mathcal{EI}$)**.** Given a model $M = \langle W, R, V, Y, Z\rangle$ in **EI** and a world $w \in W$, the *satisfaction* relation $\Vdash$ between (M, w) and formulas of $\mathcal{EI}$ is given by:

$$
\begin{array}{lll}
(M,w) \Vdash p & \textit{iff} & p \in V(w) \\
(M,w) \Vdash \neg\varphi & \textit{iff} & (M,w) \nVdash \varphi \\
(M,w) \Vdash \varphi \vee \psi & \textit{iff} & (M,w) \Vdash \varphi \text{ or } (M,w) \Vdash \psi \\
(M,w) \Vdash \Diamond\varphi & \textit{iff} & \text{there is } u \in W \text{ such that } Rwu \text{ and } (M,u) \Vdash \varphi \\
(M,w) \Vdash I\,\gamma & \textit{iff} & \gamma \in Y(w) \\
(M,w) \Vdash L\,\rho & \textit{iff} & \rho \in Z(w)
\end{array}
$$

Note how the operators I and L just look into the correspondent sets.

(P)	All propositional tautologies	$(Coh_{\mathcal{I}})$	$\vdash I\,\gamma \to \Box I\,\gamma$
(K)	$\vdash \Box(\varphi \to \psi) \to (\Box\varphi \to \Box\phi)$	$(Coh_{\mathcal{R}})$	$\vdash L\,\rho \to \Box L\,\rho$
$(Dual)$	$\vdash \Diamond\varphi \leftrightarrow \neg\Box\neg\varphi$		
(MP)	From $\vdash \varphi$ and $\vdash \varphi \to \psi$ infer $\vdash \psi$	(Gen)	From $\vdash \varphi$ infer $\vdash \Box\varphi$

Table 1: Axioms and inference rules for the logic EI.

We provide a syntactic characterization of formulas of $\mathcal{EI}$ that are valid in the class of models **EI**.

Theorem 1 (Soundness and completeness of EI w.r.t. **EI**)**.** *The logic* EI*, built from the axioms and rules shown in Table 1 is sound and strongly complete with respect to the class* **EI**.

Sketch. Soundness follows from the validity of the axioms and the validity-preserving property of the rules. Completeness can be proved by defining the canonical model in the standard way with the information set and rule set functions given by $Y^{\mathsf{EI}}(w) := \{\gamma \in \mathcal{I} \mid I\,\gamma \in w\}$ and $Z^{\mathsf{EI}}(w) := \{\rho \in \mathcal{R} \mid L\,\rho \in w\}$, and by showing that it satisfies the two coherence properties. $\Box$

Note how the agent's explicit information does not suffer from the *logical omniscience* problem: the validity of γ does not imply the validity of $I\,\gamma$, and $I\,(\gamma \to \delta) \to (I\,\gamma \to I\,\delta)$ is not valid.

2.2 When information is knowledge

An explicit/implicit information model does not impose any restriction on the information of the agent, but we can distinguish those models in **EI** where implicit and explicit information are true and the rules are truth-preserving. For implicit information, we consider models with equivalence accessibility relations, as it is usually done in EL. For explicit information, we ask for every formula in the information set to be true in the corresponding world. For the rules, we define a translation TR that maps each rule in $\mathcal{R}$ to an implication whose antecedent is the conjunction of the premises and whose consequent is the conclusion, and we ask for this translation to be true in the correspondent world.

$$\mathrm{TR}(\rho) := \bigwedge \mathrm{prem}(\rho) \rightarrow \mathrm{conc}(\rho)$$

Definition 2.5 (The class $\mathbf{EI}_K$)**.** We denote by $\mathbf{EI}_K$ the class of explicit/implicit information models satisfying ***equivalence*** (R is an equivalence relation), ***truth for formulas*** (for every world w, if $\gamma \in Y(w)$, then $(M, w) \Vdash \gamma$) and ***truth for rules*** (for every world w, if $\rho \in Z(w)$, then $(M, w) \Vdash \mathrm{TR}(\rho)$).

(T)	$\vdash \varphi \rightarrow \Diamond\varphi$	$(Tth_{\mathcal{I}})$	$\vdash I\,\gamma \rightarrow \gamma$
(4)	$\vdash \Diamond\Diamond\varphi \rightarrow \Diamond\varphi$	$(Tth_{\mathcal{R}})$	$\vdash L\,\rho \rightarrow \mathrm{TR}(\rho)$
(B)	$\vdash \Diamond\Box\varphi \rightarrow \varphi$		

Table 2: Extra axioms for the logic EI_K.

Formulas valid in $\mathbf{EI}_K$ models are characterized by the logic EI_K.

Theorem 2 (Soundness and completeness of EI_K w.r.t. $\mathbf{EI}_K$)**.** *The logic EI_K, extending EI by adding the axioms of Table 2, is sound and strongly complete with respect to the class $\mathbf{EI}_K$.*

Proof. Soundness follows from the validity of the new axioms. Completeness follows from the fact that the canonical model for the logic EI_K satisfies *equivalence* (because of axioms T, 4, B), *truth for formulas* (because of $Tth_{\mathcal{I}}$) and *truth for rules* (because of $Tth_{\mathcal{R}}$). □

For models in $\mathbf{EI}_K$, the coherence properties put the same information and rule set in worlds belonging to the same equivalence class. Note also that in these models, from axiom $Coh_{\mathcal{I}}$ ($I\,\gamma \rightarrow \Box I\,\gamma$) and axiom $Tth_{\mathcal{I}}$ ($I\,\gamma \rightarrow \gamma$) we get $I\,\gamma \rightarrow \Box\gamma$: whatever is part of the agent's explicit information belongs to her implicit information too.

It is now time to turn our attention into the *dynamics* of information. In the following sections, we extend the language to describe inference and update. With the first, the agent will extract information that is entailed by her explicit information; with the second, she will get information that may not be available (even implicitly) to her before.

3 Inference

The agent can extend her explicit knowledge by using rules. A rule (Γ, γ) intuitively indicates that if every $\delta \in \Gamma$ is true, so it is γ. However, so far, we have not stated any restriction on how the agent can use a rule. She can use it to get the conclusion without having all the premises, or even deriving the premises whenever she has the conclusion. In the previous section we focused on true-information models; in the same spirit, this section deals with truth-preserving inference.

3.1 A particular case: truth-preserving inference

The inference process adds formulas to the information set. In order to preserve truth, we restrict the way in which the rule can be applied. The *deduction* operation over a model M is defined as follows.

Definition 3.1 (Deduction operation)**.** Let $M = \langle W, R, V, Y, Z \rangle$ be a model in the class **EI**, and let σ be a rule in $\mathcal{R}$. The model $M_\sigma = \langle W, R, V, Y', Z \rangle$ has the same worlds, accessibility relation, valuation and rule set function as M. In the case of information sets, we have

$$Y'(w) := \begin{cases} Y(w) \cup \{\mathrm{conc}(\sigma)\} & \text{if } \mathrm{prem}(\sigma) \subseteq Y(w) \text{ and } \sigma \in Z(w) \\ Y(w) & \text{otherwise} \end{cases}$$

Note how the conclusion of the rule is added to a world just when all the premises and the rule are already there. This allows us to prove that, in particular, the deduction operation preserves models in $\mathbf{EI}_K$.

Proposition 1. *Let σ be a rule. If M is a model in $\mathbf{EI}_K$, so is M_σ.*

Proof. Equivalence and both properties of rules are immediate since neither the accessibility relation nor the rule set function are modified. The properties of formulas can be verified easily. □

The language $\mathcal{EID}$ extends $\mathcal{EI}$ by closing it under deduction modalities $\langle \mathrm{D}_\sigma \rangle$ for σ a rule: if φ is a formula in $\mathcal{EID}$, so is $\langle \mathrm{D}_\sigma \rangle\, \varphi$. These new formulas are read as *"there is a way of deductively applying σ after which φ is the case"*. Define the abbreviation $\mathrm{Pre}_\sigma \equiv I\ \mathrm{prem}(\sigma) \wedge L\,\sigma$ where, given Γ a finite set of formulas in $\mathcal{I}$, we write $I\,\Gamma$ for $\bigwedge_{\gamma \in \Gamma} I\,\gamma$. The semantics for deduction formulas is given as follows.

Definition 3.2. Let M be a model in **EI**, and take a world w in it.

$$(M, w) \Vdash \langle \mathrm{D}_\sigma \rangle\, \varphi \quad \textit{iff} \quad (M, w) \Vdash \mathrm{Pre}_\sigma \text{ and } (M_\sigma, w) \Vdash \varphi$$

The formula $[\mathrm{D}_\sigma]\,\varphi$ is given by $[\mathrm{D}_\sigma]\,\varphi \leftrightarrow \neg\langle \mathrm{D}_\sigma \rangle\, \neg\varphi$, as usual.

For an axiom system, Proposition 1 tells us that the deduction operation preserves models in $\mathbf{EI}_K$, so we can rely on the logic EI_K: all we need to do is give a set of reduction axioms for formulas of the form $\langle \mathrm{D}_\sigma \rangle\, \varphi$, expressing how the truth value of formulas *after* the deduction operation depends on the truth value of formulas *before* it.

$\vdash \langle \mathrm{D}_\sigma \rangle \top \leftrightarrow \mathrm{Pre}_\sigma$	$\vdash \langle \mathrm{D}_\sigma \rangle\, p \leftrightarrow (\mathrm{Pre}_\sigma \wedge p)$
$\vdash \langle \mathrm{D}_\sigma \rangle \neg\varphi \leftrightarrow (\mathrm{Pre}_\sigma \wedge \neg\langle \mathrm{D}_\sigma \rangle\, \varphi)$	$\vdash \langle \mathrm{D}_\sigma \rangle\, (\varphi \vee \psi) \leftrightarrow (\langle \mathrm{D}_\sigma \rangle\, \varphi \vee \langle \mathrm{D}_\sigma \rangle\, \psi)$
$\vdash \langle \mathrm{D}_\sigma \rangle \Diamond\varphi \leftrightarrow (\mathrm{Pre}_\sigma \wedge \Diamond\langle \mathrm{D}_\sigma \rangle\, \varphi)$	
$\vdash \langle \mathrm{D}_\sigma \rangle\, I\ \mathrm{conc}(\sigma) \leftrightarrow \mathrm{Pre}_\sigma$	$\vdash \langle \mathrm{D}_\sigma \rangle\, L\,\rho \leftrightarrow (\mathrm{Pre}_\sigma \wedge L\,\rho)$
$\vdash \langle \mathrm{D}_\sigma \rangle\, I\,\gamma \leftrightarrow (\mathrm{Pre}_\sigma \wedge I\,\gamma)$ for $\gamma \neq \mathrm{conc}(\sigma)$	
From $\vdash \varphi$, infer $\vdash [\mathrm{D}_\sigma]\,\varphi$	

Table 3: Axioms and rules for deduction operation formulas.

Theorem 3 (Soundness and completeness of EI_{KD} w.r.t. $\mathbf{EI}_K$). *The logic* EI_{KD}*, built from axioms and rules of* EI_K *(Table 2) plus axioms and rules in Table 3, is sound and strongly complete w.r.t.* $\mathbf{EI}_K$*.*

Proof. Soundness comes from the validity of the new axioms and the validity-preserving property of the new rule. Strong completeness comes from the fact that, by a repetitive application of such axioms, any deduction formula can be reduced to a formula in $\mathcal{EI}$, for which EI_K is strongly complete with respect to $\mathbf{EI}_K$. □

Our model-operation treatment of inference gives us nice properties; for example, inference preserves atomic valuation. But not only that; the formulas **(1)** $\langle \mathrm{D}_\sigma \rangle \top \rightarrow \mathrm{Pre}_\sigma$, **(2)** $I\gamma \rightarrow [\mathrm{D}_\sigma] I\gamma$, **(3)** $[\mathrm{D}_\sigma] I \operatorname{conc}(\sigma)$ and **(4)** $\langle \mathrm{D}_\sigma \rangle I\gamma \rightarrow I\gamma$ (for $\gamma \neq \operatorname{conc}(\sigma)$) are valid. The first expresses that in order to apply a rule we need the premises and the rule; the second says that after applying a rule we preserve the explicit information we had before; the third and the fourth indicate that the explicit information is increased only by the rule's conclusion.

3.2 Dynamics of deduction

Just as the agent's explicit information changes, her inferential abilities can also change. This may be because she is informed about another rule (as with the *updates* of Section 4), but it may be also because she *builds* new rules from the ones she already has. For example, from the rules $\{p\} \Rightarrow q$ and $\{q\} \Rightarrow r$, we can derive the rule $\{p\} \Rightarrow r$. It takes one step to derive it, but it will save intermediate steps later.

In fact the example, a kind of *transitivity*, represents the application of *Cut* over the mentioned rules. In general, inference relations can be characterized by *structural rules*, indicating how to derive new rules from the ones already present. In the case of deduction, we have

Reflexivity: $\dfrac{}{\varphi \Rightarrow \varphi}$ *Contraction:* $\dfrac{\psi, \chi, \xi, \chi, \phi \Rightarrow \varphi}{\psi, \chi, \xi, \phi \Rightarrow \varphi}$

Permutation: $\dfrac{\psi, \chi, \xi, \phi \Rightarrow \varphi}{\psi, \xi, \chi, \phi \Rightarrow \varphi}$ *Monotonicity:* $\dfrac{\psi, \phi \Rightarrow \varphi}{\psi, \chi, \phi \Rightarrow \varphi}$

Cut: $\dfrac{\chi \Rightarrow \xi \quad \psi, \xi, \phi \Rightarrow \varphi}{\psi, \chi, \phi \Rightarrow \varphi}$

Each time a structural rule is applied, we get a new rule that can be added to the rule set. *Contraction* and *Permutation* are not so interesting for us, since we are already considering the premises of a rule as a set. On the other hand, *Reflexivity*, *Monotonicity* and *Cut* can increase the agent's inferential abilities.

Definition 3.3 (Structural operations). Let $M = \langle W, R, V, Y, Z \rangle$ be a model in **EI**. The structural operations, $(\cdot)_{\mathbf{Ref}(\delta)}$, $(\cdot)_{\mathbf{Mon}(\delta,\varsigma)}$ and $(\cdot)_{\mathbf{Cut}(\varsigma_1,\varsigma_2)}$, return a model that differs from M just in the new rule set function Z'.

Reflexivity. Let δ be a formula of the internal language, and consider the rule $\varsigma_\delta = (\{\delta\}, \delta)$. The new rule set function is given by

$$Z'(w) := Z(w) \cup \{\varsigma_\delta\}$$

Monotonicity. Let δ be a formula in $\mathcal{I}$, and let ς be a rule over $\mathcal{I}$. Consider the rule $\varsigma' = (\mathrm{prem}(\varsigma) \cup \{\delta\}, \mathrm{conc}(\varsigma))$, extending ς by adding δ to its premises. The new rule set function is given by

$$Z'(w) := \begin{cases} Z(w) \cup \{\varsigma'\} & \text{if } \varsigma \in Z(w) \\ Z(w) & \text{otherwise} \end{cases}$$

Cut. Let ς_1, ς_2 be rules over the internal language, such that the conclusion of ς_1 is contained in the premises of ς_2. Consider the rule ς' with $(\mathrm{prem}(\varsigma_2) - \{\mathrm{conc}(\varsigma_1)\}) \cup \mathrm{prem}(\varsigma_1)$ as premises and $\mathrm{conc}(\varsigma_2)$ as conclusion. The new rule set function is given by

$$Z'(w) := \begin{cases} Z(w) \cup \{\varsigma'\} & \text{if } \{\varsigma_1, \varsigma_2\} \subseteq Z(w) \\ Z(w) & \text{otherwise} \end{cases}$$

The three structural operations preserve models in $\mathbf{EI}_K$.

Proposition 2. *If M is a model in $\mathbf{EI}_K$, then $M_{\mathbf{Ref}(\delta)}$, $M_{\mathbf{Mon}(\delta,\varsigma)}$ and $M_{\mathbf{Cut}(\varsigma_1,\varsigma_2)}$ are also in $\mathbf{EI}_K$.*

Proof. Equivalence and both properties of formulas are immediate. Coherence for rules follows from the definitions and coherence for rules of M, so we just have to prove truth for rules property for the three operations. Note that in the three cases it is enough to show that the rules are truth-preserving in M because the truth-value of the translation depends just on the valuation, which is preserved by the operations.

- ***Reflexivity.*** Recall that $\varsigma_\delta = (\{\delta\}, \delta)$ and pick any $\rho \in Z'(w)$. If ρ is already in $Z(w)$, we have $(M, w) \Vdash \mathrm{TR}(\rho)$ since M is in $\mathbf{EI}_K$. Otherwise, ρ is ς_δ, but we obviously have $(M, w) \Vdash \delta \to \delta$.
- ***Monotonicity.*** Recall that $\varsigma' = (\mathrm{prem}(\varsigma) \cup \{\delta\}, \mathrm{conc}(\varsigma))$ and pick any $\rho \in Z'(w)$. If ρ is already in $Z(w)$, we have $(M, w) \Vdash \mathrm{TR}(\rho)$. Otherwise, ρ is ς' and we should have $\varsigma \in Z(w)$; therefore, we have $(M, w) \Vdash \bigwedge \mathrm{prem}(\varsigma) \to \mathrm{conc}(\varsigma)$ and hence $(M, w) \Vdash (\bigwedge \mathrm{prem}(\varsigma) \wedge \delta) \to \mathrm{conc}(\varsigma)$.
- ***Cut.*** Recall that $\varsigma' = ((\mathrm{prem}(\varsigma_2) - \{\mathrm{conc}(\varsigma_1)\}) \cup \mathrm{prem}(\varsigma_1), \mathrm{conc}(\varsigma_2))$ and pick any $\rho \in Z'(w)$. If $\rho \in Z(w)$, we have $(M, w) \Vdash \mathrm{TR}(\rho)$ since M is in $\mathbf{EI}_K$. Otherwise, ρ is ς' and we have $\{\varsigma_1, \varsigma_2\} \subseteq Z(w)$.

  Suppose $(M, w) \Vdash \bigwedge \mathrm{prem}(\varsigma')$, so every premise of ς' is true at w in M. This includes every premise of ς_1 and every premise of ς_2 except $\mathrm{conc}(\varsigma_1)$. But since every premise of ς_1 is true at w in M and ς_1 is in $Z(w)$, truth for rules of M tells us that $\mathrm{conc}(\varsigma_1)$ is true at w in M and hence every premise of ς_2 is true at w in M. Now, since ς_2 is in $Z(w)$, truth for rules of M tell us that $\mathrm{conc}(\varsigma_2)$, that is, $\mathrm{conc}(\varsigma')$, is true at w in M. Then we have $(M, w) \Vdash \mathrm{TR}(\varsigma')$.

□

The language $\mathcal{EID}^*$ extends $\mathcal{EID}$ by making it closed under formulas representing structural operations: if φ is in $\mathcal{EID}^*$, so are $\langle Ref_\delta \rangle \varphi$, $\langle Mon_{\delta,\varsigma} \rangle \varphi$ and $\langle Cut_{\varsigma_1,\varsigma_2} \rangle \varphi$. Each one of the formulas are read as *"there is a way of applying the structural operation after which φ is the case"*. With $\mathrm{Pre}_{\mathbf{Mon}(\delta,\varsigma)} \equiv L\,\varsigma$ and $\mathrm{Pre}_{\mathbf{Cut}(\varsigma_1,\varsigma_2)} \equiv L\,\varsigma_1 \wedge L\,\varsigma_2 \wedge (I\ \mathrm{prem}(\varsigma_2) \to I\ \mathrm{conc}(\varsigma_1))$, the semantics of the new formulas is given as follows.

Definition 3.4. Let M be a model in **EI**, and take a world w in it.

$$\begin{array}{lll}
(M, w) \Vdash \langle Ref_\delta \rangle \varphi & \textit{iff} & (M_{\mathbf{Ref}(\delta)}, w) \Vdash \varphi \\
(M, w) \Vdash \langle Mon_{\delta,\varsigma} \rangle \varphi & \textit{iff} & (M, w) \Vdash \mathrm{Pre}_{\mathbf{Mon}(\delta,\varsigma)} \text{ and } (M_{\mathbf{Mon}(\delta,\varsigma)}, w) \Vdash \varphi \\
(M, w) \Vdash \langle Cut_{\varsigma_1,\varsigma_2} \rangle \varphi & \textit{iff} & (M, w) \Vdash \mathrm{Pre}_{\mathbf{Cut}(\varsigma_1,\varsigma_2)} \text{ and } (M_{\mathbf{Cut}(\varsigma_1,\varsigma_2)}, w) \Vdash \varphi
\end{array}$$

Just as before, the *boxed* versions of the structural operation formulas is defined as the dual of their correspondent *diamond* versions.

To provide axioms for the new formulas, Proposition 2 allows us to rely on the logic EI_K once again.

Theorem 4 (Soundness and completeness of EI_{KDS} w.r.t. $\mathbf{EI}_K$)**.** *For uniformity, define* $\mathrm{Pre}_{\mathbf{Ref}(\delta)} \equiv \top$. *The logic* EI_{KDS}*, extending* EI_{KD} *with axioms and rule of Table 4 (where STR stands for either Ref, Mon or Cut and* ς' *is the correspondent new rule), is sound and strongly complete w.r.t. models in* $\mathbf{EI}_K$*.*

$\vdash \langle STR \rangle \top \leftrightarrow \mathrm{Pre}_{\mathbf{STR}}$	$\vdash \langle STR \rangle p \leftrightarrow (\mathrm{Pre}_{\mathbf{STR}} \wedge p)$
$\vdash \langle STR \rangle \neg\varphi \leftrightarrow (\mathrm{Pre}_{\mathbf{STR}} \wedge \neg\langle STR \rangle \varphi)$	
$\vdash \langle STR \rangle (\varphi \vee \psi) \leftrightarrow (\langle STR \rangle \varphi \vee \langle STR \rangle \psi)$	
$\vdash \langle STR \rangle \Diamond\varphi \leftrightarrow (\mathrm{Pre}_{\mathbf{STR}} \wedge \Diamond\langle STR \rangle \varphi)$	
$\vdash \langle STR \rangle L\,\varsigma' \leftrightarrow \mathrm{Pre}_{\mathbf{STR}}$	$\vdash \langle STR \rangle I\,\gamma \leftrightarrow (\mathrm{Pre}_{\mathbf{STR}} \wedge I\,\gamma)$
$\vdash \langle STR \rangle L\,\rho \leftrightarrow (\mathrm{Pre}_{\mathbf{STR}} \wedge L\,\rho)$ for $\rho \neq \varsigma'$	
From $\vdash \varphi$, infer $\vdash [STR]\,\varphi$	

Table 4: Axioms and rules for *reflexivity*, *monotonicity* and *cut* formulas.

We finish this section presenting some validities (Table 5), expressing how deduction after structural operations is related to deduction before them at models in $\mathbf{EI}_K$.

Theorem 5. *The formulas in Table 5 are valid in* $\mathbf{EI}_K$ *models.*

Reflexivity with ς_δ the rule $(\{\delta\}, \delta)$	
• $\langle Ref_\delta \rangle \langle \mathrm{D}_\sigma \rangle \varphi \leftrightarrow \langle \mathrm{D}_\sigma \rangle \langle Ref_\delta \rangle \varphi$	for $\sigma \neq \varsigma_\delta$
• $\langle Ref_\delta \rangle \langle \mathrm{D}_{\varsigma_\delta} \rangle \varphi \leftrightarrow (\langle \mathrm{D}_{\varsigma_\delta} \rangle \varphi \vee (I\,\delta \wedge \langle Ref_\delta \rangle \varphi))$	
Monotonicity with ς' the rule $(\mathrm{prem}(\varsigma) \cup \{\delta\}, \mathrm{conc}(\varsigma))$	
• $\langle Mon_{\delta,\varsigma} \rangle \langle \mathrm{D}_\sigma \rangle \varphi \leftrightarrow \langle \mathrm{D}_\sigma \rangle \langle Mon_{\delta,\varsigma} \rangle \varphi$	for $\sigma \neq \varsigma'$
• $\langle Mon_{\delta,\varsigma} \rangle \langle \mathrm{D}_{\varsigma'} \rangle \varphi \leftrightarrow (\langle \mathrm{D}_{\varsigma'} \rangle \varphi \vee (I\,\delta \wedge L\,\varsigma \wedge \langle \mathrm{D}_\varsigma \rangle \langle Mon_{\delta,\varsigma} \rangle \varphi))$	
Cut with ς' the rule $(\,(\mathrm{prem}(\varsigma_2) - \{\mathrm{conc}(\varsigma_1)\}) \cup \mathrm{prem}(\varsigma_1),\ \mathrm{conc}(\varsigma_2)\,)$	
• $\langle Cut_{\varsigma_1,\varsigma_2} \rangle \langle \mathrm{D}_\sigma \rangle \varphi \leftrightarrow \langle \mathrm{D}_\sigma \rangle \langle Cut_{\varsigma_1,\varsigma_2} \rangle \varphi$	for $\sigma \neq \varsigma'$
• $\langle Cut_{\varsigma_1,\varsigma_2} \rangle \langle \mathrm{D}_{\varsigma'} \rangle \varphi \leftrightarrow$ $(\langle \mathrm{D}_{\varsigma'} \rangle \varphi \vee (I\ \mathrm{prem}(\varsigma_1) \wedge L\,\varsigma_1 \wedge (I\ \mathrm{conc}(\varsigma_1) \rightarrow \langle \mathrm{D}_{\varsigma_2} \rangle \langle Cut_{\varsigma_1,\varsigma_2} \rangle \varphi)))$	

Table 5: Formulas relating structural operations and deduction

Proof. The validity of the formulas follows from the bisimilarities between models stated below. In our case, the bisimulation concept extends the standard one by asking for related worlds to have the same information and rule set: given two models $M_1 = \langle W_1, R_1, V_1, Y_1, Z_1 \rangle$ and $M_2 = \langle W_2, R_2, V_2, Y_2, Z_2 \rangle$, a non empty relation $B \subseteq (W_1 \times W_2)$ is a bisimulation if and only if it is a standard bisimulation between $\langle W_1, R_1, V_1\rangle$ and $\langle W_2, R_2, V_2\rangle$ and, if Bw_1w_2, then $Y_1(w_1) = Y_2(w_2)$ and $Z_1(w_1) = Z_2(w_2)$.

Let $M = \langle W, R, V, Y, Z\rangle$ be a model in $\mathbf{EI}_K$, and take $w \in W$. Models of the form $M_{\mathbf{STR}_\sigma}$ are the result of applying first the structural operation STR and then the deduction operation with rule σ, and analogously for models of the form $M_{\sigma\mathbf{STR}}$. In all cases, the bisimulation is the identity relation over worlds reachable from w.

Reflexivity. Let ς_δ be the rule $(\{\delta\}, \delta)$:
- If $\sigma \neq \varsigma_\delta$, then $(M_{\mathbf{Ref}(\delta)_\sigma}, w) \leftrightarroweq (M_{\sigma\mathbf{Ref}(\delta)}, w)$.
- If $\varsigma_\delta \in Z(w)$, then $(M_{\mathbf{Ref}(\delta)_{\varsigma_\delta}}, w) \leftrightarroweq (M_{\varsigma_\delta}, w)$.
- If $\delta \in Y(w)$, then $(M_{\mathbf{Ref}(\delta)_{\varsigma_\delta}}, w) \leftrightarroweq (M_{\varsigma_\delta\mathbf{Ref}(\delta)}, w)$.

Monotonicity. Let ς' be the rule $(\text{prem}(\varsigma) \cup \{\delta\}, \text{conc}(\varsigma))$:
- If $\sigma \neq \varsigma'$, then $(M_{\mathbf{Mon}(\delta,\varsigma)_\sigma}, w) \leftrightarroweq (M_{\sigma\mathbf{Mon}(\delta,\varsigma)}, w)$.
- If $\varsigma' \in Z(w)$, then $(M_{\mathbf{Mon}(\delta,\varsigma)_{\varsigma'}}, w) \leftrightarroweq (M_{\varsigma'}, w)$.
- If $\delta \in Y(w)$ and $\varsigma \in Z(w)$, then $(M_{\mathbf{Mon}(\delta,\varsigma)_{\varsigma'}}, w) \leftrightarroweq (M_{\varsigma\mathbf{Mon}(\delta,\varsigma)}, w)$.

Cut. Let ς' be the rule $(\,(\text{prem}(\varsigma_2) - \{\text{conc}(\varsigma_1)\}) \cup \text{prem}(\varsigma_1),\ \text{conc}(\varsigma_2)\,)$:
- If $\sigma \neq \varsigma'$, then $(M_{\mathbf{Cut}(\varsigma_1,\varsigma_2)_\sigma}, w) \leftrightarroweq (M_{\sigma\mathbf{Cut}(\varsigma_1,\varsigma_2)}, w)$.
- If $\varsigma' \in Z(w)$, then $(M_{\mathbf{Cut}(\varsigma_1,\varsigma_2)_{\varsigma'}}, w) \leftrightarroweq (M_{\varsigma'}, w)$.
- If $(\text{prem}(\varsigma_1) \cup \{\text{conc}(\varsigma_1)\}) \in Y(w)$ and $\varsigma_1 \in Z(w)$, then $(M_{\mathbf{Cut}(\varsigma_1,\varsigma_2)_{\varsigma'}}, w) \leftrightarroweq (M_{\varsigma_2\mathbf{Cut}(\varsigma_1,\varsigma_2)}, w)$.

The involved operations (structural ones and deduction) preserve worlds, accessibility relations and valuations. To show that the identity relation over worlds reachable from w is a bisimulation, we just need to show that they have the same information and rule set in both models.

Consider as an example the third bisimilarity for monotonicity. For information sets, take any γ in the information set of w at $M_{\mathbf{Mon}(\delta,\varsigma)_{\varsigma'}}$; by definition, either it was already in that of w at $M_{\mathbf{Mon}(\delta,\varsigma)}$ or else it was added by the deduction operation. In the first case, it is in w at M (structural operations do not modify information sets); then it is also in w at M_ς and hence it is in w at $M_{\varsigma\mathbf{Mon}(\delta,\varsigma)}$. In the second case, γ should be $\text{conc}(\varsigma')$, but then we have the premises of ς' (and hence those of ς) in w at $M_{\mathbf{Mon}(\delta,\varsigma)}$. Then, they are already in w at M and, by hypothesis, we have ς in w at M, so $\text{conc}(\varsigma) = \text{conc}(\varsigma')$ is in w at M_ς and hence it is in w at $M_{\varsigma\mathbf{Mon}(\delta,\varsigma)}$.

For the other direction, take γ in w at $M_{\varsigma\mathbf{Mon}(\delta,\varsigma)}$. Then it is in w at M_ς and therefore either it was already in w at M or else it was added by the deduction operation. In the first case, γ is preserved through the monotonicity and the deduction operations, and therefore it is in w at $M_{\mathbf{Mon}(\delta,\varsigma)_{\varsigma'}}$. In the second case, γ should be $\text{conc}(\varsigma)$, and then we should have $\text{prem}(\varsigma)$ and ς in the correspondent sets of w at M. By hypothesis we have δ in w at M, so we have all the premises of ς' in w at M and therefore they are also in w at $M_{\mathbf{Mon}(\delta,\varsigma)}$. Since we have ς in w at M, we have ς' in w at $M_{\mathbf{Mon}(\delta,\varsigma)}$ too. Hence, we have $\text{conc}(\varsigma') = \text{conc}(\varsigma)$ in w at $M_{\mathbf{Mon}(\delta,\varsigma)_{\varsigma'}}$. The case for rules is similar.

Now suppose a world u is reachable from w through the accessibility relation at $M_{\mathbf{Mon}(\delta,\varsigma)_{\varsigma'}}$. Since neither the relations nor the worlds are modified by the operations, u is reachable from w at M and therefore u is reachable from w at $M_{\varsigma\mathbf{Mon}(\delta,\varsigma)}$, too. Now we use the coherence properties: since $\delta \in Y(w)$ and $\varsigma \in Z(w)$, we have δ and ς in the corresponding sets of u, and then we can apply the argument used for w to show that u has the same information and rule set on both models. □

4 Update

So far, our language can express the agent's *internal* dynamics, but it cannot express *external* ones. We can express how deductive steps modify explicit knowledge, and even how structural operations extends the rules the agent can apply, but we cannot express how both explicit and implicit knowledge are affected by external observations. Here we add the other fundamental source of information; in this section, we extend the language to express updates.

Updates are the result of the agent's social nature. We get new information because of the interaction with our environment, information that does not have to follow from what we explicitly know. In *Public Announcement Logic* (PAL), an announcement is interpreted as an operation that removes the worlds where the announced formula does not hold, restricting the epistemic relation to those that are not deleted. In our semantic model, we have a finer representation of the agent's knowledge: we have implicit one (given by the accessibility relation) but we also have explicit knowledge (the information sets). We can extend PAL by defining operations affecting explicit and implicit knowledge in different forms, and therefore expressing different ways the agent processes the incoming information. Here, we present one of the possible definitions, what we have called *explicit observations*.

4.1 Explicit observations

The previously defined operations just add formulas or rules to the corresponding sets, but do not modify the accessibility relation and therefore do not affect implicit knowledge. Explicit observations, on the other hand, do modify the accessibility relation because they remove worlds where the observation does not hold. With respect to explicit information, they add arbitrary *true* information (a formula or a rule), no matter if it was implicitly available or not.

Definition 4.1 (Explicit observation operation)**.** Let $M = \langle W, R, V, Y, Z \rangle$ be a model in **EI** and let χ be a formula of (a rule based on) $\mathcal{I}$. The model $M_{\chi!} = \langle W', R', V', Y', Z' \rangle$ is given by

- $W' := \{ w \in W \mid (M, w) \Vdash \chi \}$ $\quad (W' := \{ w \in W \mid (M, w) \Vdash \mathrm{TR}(\chi) \})$,
- $R' := R \cap (W' \times W')$,
- $V'(w) := V(w)$ for every $w \in W'$,
- $Y'(w) := Y(w) \cup \{\chi\}$ $\quad (Y'(w) := Y(w))$ $\quad$ for every $w \in W'$,
- $Z'(w) := Z(w)$ $\quad (Z'(w) := Z(w) \cup \{\chi\})$ $\quad$ for every $w \in W'$.

The operation preserves models in $\mathbf{EI}_K$ too.

Proposition 3. *Let M be a model in $\mathbf{EI}_K$ and let χ be a formula (a rule). If M is in $\mathbf{EI}_K$, so is $M_{\chi!}$.*

Proof. Suppose χ is a formula. Equivalence follows immediately, as well as the properties for rules since Z is not affected in the remaining worlds. Coherence for formulas holds because χ is added uniformly, and truth for formulas holds because of the definition of W'. Suppose χ is a rule. Equivalence and the properties for formulas are just as before. Coherence for rules holds because χ is added uniformly, and truth for rules holds because of the definition of W'. $\square$

The language $\mathcal{EID}^{*!}$ extends $\mathcal{EID}^{*}$ by closing it under explicit observations: if φ is in $\mathcal{EID}^{*!}$, so is $\langle\chi!\rangle\,\varphi$. These formulas are read as *"there is a way of explicitly observing χ after which φ is the case"*. In case χ is a formula, define $\text{Pre}_{\chi!} \equiv \chi$; in case χ is a rule, define $\text{Pre}_{\chi!} \equiv \text{TR}(\chi)$. The semantics for the new formulas is given as follows.

Definition 4.2. Let M be a model in **EI**, and take a world w in it.

$$(M, w) \Vdash \langle\chi!\rangle\,\varphi \quad \textit{iff} \quad (M, w) \Vdash \text{Pre}_{\chi!} \text{ and } (M_{\chi!}, w) \Vdash \varphi$$

The formula $[\chi!]\,\varphi$ is defined as the dual of $\langle\chi!\rangle\,\varphi$, as usual.

Theorem 6 (Soundness and completeness of EI_{KDSO} w.r.t. $\mathbf{EI}_K$)**.** *The logic EI_{KDSO}, built from EI_{KDS} plus axioms and rule in Table 6, is sound and strongly complete w.r.t. the class of models $\mathbf{EI}_K$.*

$\vdash \langle\chi!\rangle\,\top \;\leftrightarrow\; \text{Pre}_{\chi!}$	$\vdash \langle\chi!\rangle\,p \;\leftrightarrow\; (\text{Pre}_{\chi!} \wedge\; p)$
$\vdash \langle\chi!\rangle\,\neg\varphi \;\leftrightarrow\; (\text{Pre}_{\chi!} \wedge \neg\langle\chi!\rangle\,\varphi)$	$\vdash \langle\chi!\rangle\,(\varphi \vee \psi) \;\leftrightarrow\; (\langle\chi!\rangle\,\varphi \vee \langle\chi!\rangle\,\psi)$
$\vdash \langle\chi!\rangle\,\Diamond\varphi \;\leftrightarrow\; (\text{Pre}_{\chi!} \wedge \Diamond\langle\chi!\rangle\,\varphi)$	
If χ is a formula:	
$\vdash \langle\chi!\rangle\,I\,\chi \;\leftrightarrow\; \text{Pre}_{\chi!}$	$\vdash \langle\chi!\rangle\,L\,\rho \;\leftrightarrow\; (\text{Pre}_{\chi!} \wedge\; L\,\rho)$
$\vdash \langle\chi!\rangle\,I\,\gamma \;\leftrightarrow\; (\text{Pre}_{\chi!} \wedge I\,\gamma)$ for $\gamma \neq \chi$	
If χ is a rule:	
$\vdash \langle\chi!\rangle\,L\,\chi \;\leftrightarrow\; \text{Pre}_{\chi!}$	$\vdash \langle\chi!\rangle\,I\,\gamma \;\leftrightarrow\; (\text{Pre}_{\chi!} \wedge\; I\,\gamma)$
$\vdash \langle\chi!\rangle\,L\,\rho \;\leftrightarrow\; (\text{Pre}_{\chi!} \wedge\; L\,\rho)$ for $\rho \neq \chi$	
From $\vdash \varphi$, infer $\vdash [\chi!]\,\varphi$	

Table 6: Axioms and rules for *explicit observation* formulas.

We finish this section presenting some validities (Table 7) expressing interaction between explicit observations and deduction in models of $\mathbf{EI}_K$.

Theorem 7. *The formulas in Table 5 are valid in $\mathbf{EI}_K$ models.*

If χ is a formula	
• $\langle\chi!\rangle\,\langle\text{D}_\sigma\rangle\,\varphi \;\leftrightarrow\; \langle\text{D}_\sigma\rangle\,\langle\chi!\rangle\,\varphi$	if $\chi \notin \text{prem}(\sigma)$
• $\langle\chi!\rangle\,\langle\text{D}_\sigma\rangle\,\varphi \;\leftrightarrow\; (\langle\text{D}_\sigma\rangle\,\langle\chi!\rangle\,\varphi \vee (I\,\delta \rightarrow \langle\text{D}_\sigma\rangle\,\langle\chi!\rangle\,\varphi))$	if $\chi \in \text{prem}(\sigma)$
If χ is a rule	
• $\langle\chi!\rangle\,\langle\text{D}_\sigma\rangle\,\varphi \;\leftrightarrow\; \langle\text{D}_\sigma\rangle\,\langle\chi!\rangle\,\varphi$	if $\chi \neq \sigma$
• $\langle\chi!\rangle\,\langle\text{D}_\chi\rangle\,\varphi \;\leftrightarrow\; (\langle\text{D}_\chi\rangle\,\langle\chi!\rangle\,\varphi \vee (L\,\chi \rightarrow \langle\text{D}_\chi\rangle\,\langle\chi!\rangle\,\varphi))$	if $\chi = \sigma$

Table 7: Formulas relating explicit observations and deduction

Proof. Just as the case of structural operations and deduction, the validity of the formulas follows from the bisimilarities stated below.

If χ is a formula:
- If $\chi \notin \text{prem}(\sigma)$, then $(M_{\chi!_\sigma}, w) \leftrightarroweq (M_{\sigma\chi!}, w)$.
- If $\chi \in \text{prem}(\sigma)$ and $\chi \in Y(w)$, then $(M_{\chi!_\sigma}, w) \leftrightarroweq (M_{\sigma\chi!}, w)$.

If χ is a rule:
- If $\chi \neq \sigma$, then $(M_{\chi!_\sigma}, w) \leftrightarroweq (M_{\sigma\chi!}, w)$.
- If $\chi = \sigma$ and $\chi \in Z(w)$, then $(M_{\chi!_\sigma}, w) \leftrightarroweq (M_{\sigma\chi!}, w)$.

The proof is similar to the case of structural operations and deduction, keeping in mind that observations remove worlds. □

5 Comparison with other works

The present work explores a representation of explicit/implicit information, allowing us to describe the way different process affects them. Several other works have proposed similar frameworks, and this section provides a brief comparison between some of them and our proposal.

5.1 Fagin-Halpern's logics of awareness

Fagin and Halpern presented in 1988 a *logic of general awareness* (L_A). The language is a set of atomic propositions P closed under negation, conjunction and the operators A_i and L_i (for an agent i). Formulas of the form $A_i\varphi$ are read as *"the agent i is aware of φ"*, and formulas of the form $L_i\varphi$ are read as *"the agent i implicitly believes that φ"*. The operator B_i for explicit beliefs is defined as $B_i\varphi := A_i\varphi \wedge L_i\varphi$.

An *structure for general awareness* is a tuple $M = (W, \mathfrak{A}_i, \mathfrak{L}_i, V)$, where $W \neq \emptyset$ is the set of possible worlds, $\mathfrak{A}_i : W \to \wp(L_A)$ is a function that assigns a set of formulas of L_A to agent i in each world (her awareness set), $\mathfrak{L}_i$ is a serial, transitive and Euclidean relation over W for each agent i (L_A deals with beliefs) and $V : \mathtt{P} \to \wp(W)$ is a valuation. Semantics for atomic propositions, negations and conjunctions are standard; for formulas of the form $A_i\,\varphi$ we look into the awareness set, and L_i is a box modal operator.

The main difference between this *logic of general awareness* and our approach is in the dynamics. First, we include in the semantic model our rule set function Z, indicating the processes the agent can use to increase her explicit information. She has not only facts about the world, but also rules that allow her to infer new facts. It is not that the agent knows that after a rule application her information set will change; it is that she knows the *process* that leads the change. Second, the L_A does not express changes in the agent's awareness set, though later in the same paper the authors add a binary relation over W to represent steps in time. Our approach uses inference as the process that transform explicit information, and this process is represented not as relation between worlds, but as a model operation that adds formulas to information sets. Moreover, we also consider dynamics of the inference process itself, witness the Structural operations. Third, the language of our information sets is less expressive than the awareness sets, but that allows us to define the update operation for representing external dynamics, a process not considered in L_A.

5.2 Duc's dynamic epistemic logic

Duc proposed in 1995, 1997, 2001 a dynamic epistemic logic to reason about agents that are neither logically omniscient nor logically ignorant. He defined the language L_{BDE}, based on formulas of the form $K\gamma$ (for γ a propositional formula) and closed under negation, conjunction and the modal operator $\langle F \rangle$. Formulas of the form $K\gamma$ are read as *"γ is known"*, and formulas of the form $\langle F \rangle \varphi$ are read as *"φ is true after some course of thought"*. The language does not provide formulas to talk about the real world.

A *BDE*-model M is a tuple (W, R, Y), with R a transitive relation over W and Y a function assigning a set of propositional formulas to each possible world. The definition asks for properties guaranteeing that the set of formulas will grow as the agent reasons, and that her information will be closed under modus ponens and will contain all tautologies at some point. Semantics for negation and conjunctions are standard. For formulas of the form $K\gamma$ we look into $Y(w)$ for w the correspondent world and the operator $\langle F \rangle$ is interpreted as a diamond with R.

Duc's framework does not consider implicit information and, while it does express changes in the sets of formulas, this mechanism is represented as a relation between worlds, different from our model operation approach. Also, his language is restricted to express what the agent can infer through some *"course of thought"*, but it does not express external dynamics, as our explicit observations do.

5.3 Jago's logic for resource-bounded agents

In 2006a, 2006b, Jago presented a logic for resource-bounded agents. He considered a semantic model similar to ours, extending Kripke models with a set of formulas of some internal language for every agent in each possible world to describe explicit information. He also considered rule-based inference as the mechanism through which the agent can increase her information. Similar to Duc's work, inference is represented as a relation between worlds, with a relation R_ρ linking w and u *iff* w contains the rule ρ and its premises and u extends w by ρ's conclusions.

There are two main differences in the approaches. The first one is again the treatment of the mechanism to increase explicit information. Extending what we have said before, our model-operation representation facilitates the work by giving us a functional treatment of inference, while the modal representation forces us to ask for properties of the relation in order for inference to behave in a functional way. Those properties may need a more powerful language to be expressed (the uniqueness of the result of a rule application needs nominals) and some of them may be not preserved after updates (the existence of a world resulting from an available rule application is not preserved since new information may turn applicable more rules). The second one is our updates, not considered in Jago's work.

5.4 van Benthem's acts of realization

In van Benthem (2008b), the author considers a language based on atomic propositions and formulas of the form $I\gamma$ (γ a factual formula), and closed under boolean connectives and the modal operator K. The semantic model is of the form $(W, W^{acc}, \sim, V)$ where $(W, \sim, V)$ is a standard Kripke model and W^{acc} is a set of *access worlds*: pairs (w, X) with $w \in W$ and X a set of factual formulas (the access set). There are two restrictions: for every (w, X), all formulas in X should be true at w and epistemically indistinguishable worlds should have the same access set. Given a model and an access world (w, X), atomic propositions are interpreted according to the valuation at w, boolean connectives and K (a box) are interpreted as usual, and $I\gamma$ is true at (w, X) *iff* $\gamma \in X$.

Given a factual formula, he defines two model operations. An *implicit observation* removes the worlds that do not satisfy the formula, just as an announcement in PAL. An *explicit observation* removes worlds but also adds the formula to the access sets of the remaining ones. So far the setup is similar to ours, except for the set of rules we consider at each world and its syntactic counterpart $L\,\rho$.

But here van Benthem notices that the two operations overlap in their effects on the model. He proposes two more "orthogonal" operations: one simply removing worlds (a *"bare observation"*) and another one simply adding true formulas to the access sets (an *"act of realization"*). An implicit observation is then a *bare observation* while an explicit one composes a *bare observation* and an *act of realization*.

Note that an *act of realization* is more general than our deduction. As it is shown, *any* formula that is part of the implicit information can be added to the access set; in particular, validities, can be added at any point. Our framework, on the other hand, allows us to add a formula only if it is the conclusion of an applicable rule, that is, we have the rule and its premises.

6 Final remarks and further work

Let us describe the restaurant example with our framework. The initial setting can be given by a model M with six possible worlds, each one of them indicating a possible distribution of the dishes, and all of them indistinguishable from each other.

For explicit information, consider a set of atomic propositions of the form $\mathsf{p_d}$ where p stands for a person (**f**ather, **m**other or **y**ou) and d stands for some dish (**m**eat, **f**ish or **v**egetarian). The waiter explicitly knows each person will get only one dish, so we can put the rules

$$\rho_1 : \{\mathsf{y_f}\} \Rightarrow \neg\mathsf{y_v}, \qquad \rho_2 : \{\mathsf{f_m}\} \Rightarrow \neg\mathsf{f_v}$$

and similar ones in each world. Moreover, he explicitly knows that each dish corresponds to one person, so the rule

$$\sigma : \{\neg\mathsf{y_v}, \neg\mathsf{f_v}\} \Rightarrow \mathsf{m_v}$$

can be also added, among many others. Let w be the real world, where $\mathsf{y_f}$, $\mathsf{f_m}$ and $\mathsf{m_v}$ are true. The formula $\neg I\,\mathsf{m_v} \wedge \neg\Box\mathsf{m_v}$, indicating that the waiter does not know (neither explicitly nor implicitly) that your mother has the vegetarian, is true at (M, w).

While approaching the table, the waiter can increase the rules he knows. This does not give him new explicit facts, but it will allow him to infer faster later. From *Cut* over ρ_1 and σ, he gets

$$\varsigma_1 : \{\mathsf{y_f}, \neg \mathsf{f_v}\} \Rightarrow \mathsf{m_v}$$

Then, the formulas

$$\langle Cut_{\rho_1,\sigma} \rangle \neg I\, \mathsf{m_v} \qquad\qquad \langle Cut_{\rho_1,\sigma} \rangle\, L\, \varsigma_1$$

are also true at (M, w). Moreover, he can apply *Cut* again, this time with ρ_2 and ς_1, obtaining the rule

$$\varsigma_2 : \{\mathsf{y_f}, \mathsf{f_m}\} \Rightarrow \mathsf{m_v}$$

and making

$$\langle Cut_{\rho_1,\sigma} \rangle \langle Cut_{\rho_2,\varsigma_1} \rangle \neg I\, \mathsf{m_v} \qquad\qquad \langle Cut_{\rho_1,\sigma} \rangle \langle Cut_{\rho_2,\varsigma_1} \rangle\, L\, \varsigma_2$$

true at (M, w).

After the answer to the question *"Who has the fish?"*, the waiter explicitly knows that you have the fish. Four possible worlds are removed, but he still does not know that your mother has the vegetarian. We have

$$(M, w) \Vdash \langle Cut_{\rho_1,\sigma} \rangle \langle Cut_{\rho_2,\varsigma_1} \rangle \langle \mathsf{y_f}! \rangle (\neg I\, \mathsf{m_v} \wedge \neg \Box \mathsf{m_v})$$

Then he asks *"Who has the meat?"*, and the answer removes one of the remaining worlds. Now he knows implicitly that your mother has the vegetarian dish and, moreover, he is able to infer it and add it to his explicit information:

$$(M, w) \Vdash \langle Cut_{\rho_1,\sigma} \rangle \langle Cut_{\rho_2,\varsigma_1} \rangle \langle \mathsf{y_f}! \rangle \langle \mathsf{f_m}! \rangle (\Box \mathsf{m_v} \wedge \langle \mathrm{D}_{\varsigma_2} \rangle\, I\, \mathsf{m_v})$$

Two structural operations, two explicit observations and one inference are all that the waiter needs.

The proposal can be extended in several ways. The first one is by extending the internal language beyond the propositional one. As we mentioned, we choose it because it makes the definition of updates of Section 4 possible. In general, a true observation in the full explicit/implicit information language cannot be simply added to an information set, since it may become false after being observed (witness Moore sentences, like $p \wedge \neg \Box p$). A first attempt would be to keep in the new information set those formulas that are true in the new model, but we would face circularity: we define the new information set by keeping those formulas that are still true, but in order to decide whether an explicit information formula $I\,\gamma$ is true or not, we need this new information set. A further analysis providing a solution to this limitation will greatly increase the expressivity of the framework.

We have analyzed the case in which the information is true, but this is not the general situation. By removing such restriction we can talk not only about knowledge but also about *beliefs*. Some recent works combine these two notions, giving us a nice way of studying these two propositional attitudes together. Moreover, we have analyzed the case where inference preserves truth, but there are other inference processes, like *default reasoning*, *abduction* or *belief revision*, which are widely used, particularly in incomplete information situations. Within the proposed framework, we can represent different inference processes, and we can study how all of them work together.

For the external dynamics, we mentioned that this finer representation of knowledge allows us to define different kinds of observations. So far we can represent observations that do not affect explicit information (like van Benthem's *bare observations*) and our already defined explicit observations. With a more expressive internal language, we could represent more kinds of observations, all differing between them in how introspective is the agent about the observed fact.

In the context of agent diversity (Liu 2006, 2008), our framework allows us to represent agents having different rules and therefore having different reasoning abilities. The idea works also for external dynamics: agents may have different observational powers. It will be interesting to explore how agents that differ in their reasoning and observational abilities interact with each other.

Acknowledgements The author thanks Johan van Benthem for his invaluable ideas and suggestions that led to the present work. He also thanks the group of the ILLC's *Seminar on Logics for Dynamics of Information and Preferences*, as well as the organizers and participants of the *Second Workshop in Decisions, Games & Logic (DGL'08)* and the *Workshop on Logic and Intelligent Interaction* in ESSLLI 2008; their comments have improved earlier versions.

References

T. Agotnes and N. Alechina, editors. *Proceedings of the Workshop on Logics for Resource-Bounded Agents, organized as part of the 18th European Summer School on Logic, Language and Information (ESSLLI)*, Malaga, Spain, August 2006.

A. Baltag and S. Smets. A qualitative theory of dynamic interactive belief revision. In G. Bonanno, W. van der Hoek, and M. Wooldridge, editors, *Logic and the Foundations of Game and Decision Theory (LOFT7)*, volume 3 of *Texts in Logic and Games*, pages 13–60. Amsterdam University Press, 2008.

A. Baltag, L. S. Moss, and S. Solecki. The logic of public announcements, common knowledge and private suspicious. Technical Report SEN-R9922, CWI, Amsterdam, 1999.

H. N. Duc. Logical omniscience vs. logical ignorance on a dilemma of epistemic logic. In C. A. Pinto-Ferreira and N. J. Mamede, editors, *EPIA 1995*, volume 990 of *Lecture Notes in Computer Science*, pages 237–248, Heilderberg, 1995. Springer. ISBN 3-540-60428-6.

H. N. Duc. Reasoning about rational, but not logically omniscient, agents. *Journal of Logic and Computation*, 7(5):633–648, 1997.

H. N. Duc. *Resource-Bounded Reasoning about Knowledge*. PhD thesis, Institut für Informatik, Universität Leipzig, Leipzig, Germany, 2001.

R. Fagin and J. Y. Halpern. Belief, awareness, and limited reasoning. *Artificial Intelligence*, 34(1):39–76, 1988. ISSN 0004-3702. doi: http://dx.doi.org/10.1016/0004-3702(87)90003-8.

J. Gerbrandy. Dynamic epistemic logic. In L. S. Moss, J. Ginzburg, and M. de Rijke, editors, *Logic, Language and Computation*, volume 2, pages 67–84. CSLI Publications, Stanford, CA, USA, 1999. ISBN 1-57586-181-X.

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

M. Jago. Rule-based and resource-bounded: A new look at epistemic logic. In Agotnes and Alechina (2006), pages 63–77.

M. Jago. *Logics for Resource-Bounded Agents*. PhD thesis, Department of Philosophy, University of Nottingham, Nottingham, UK, July 2006b.

F. Liu. Diversity of agents. In Agotnes and Alechina (2006), pages 88–98.

F. Liu. *Changing for the Better. Preference Dynamics and Agent Diversity*. PhD thesis, Institute for logic, Language and Computation, Universiteit van Amsterdam, Amsterdam, The Netherlands, February 2008. ILLC Dissertation series DS-2008-02.

J. A. Plaza. Logics of public communications. In M. L. Emrich, M. S. Pfeifer, M. Hadzikadic, and Z. W. Ras, editors, *Proceedings of the 4th International Symposium on Methodologies for Intelligent Systems*, pages 201–216, 1989.

K. M. Sim. Epistemic logic and logical omniscience. A survey. *International Journal of Intelligent Systems*, 12(1):57–81, January 1997.

J. van Benthem. Epistemic logic and epistemology: The state of their affairs. *Philosophical Studies*, 128:49–76, March 2006. doi: doi:10.1007/s11098-005-4052-0.

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

J. van Benthem. Logic and reasoning: Do the facts matter? *Studia Logica*, 88 (1):67–84, February 2008a. ISSN 0039-3215. doi: http://dx.doi.org/10.1007/s11225-008-9101-1.

J. van Benthem. Merging observation and access in dynamic logic. *Journal of Logic Studies*, 1(1), 2008b.

J. van Benthem. Tell it like it is: Information flow in logic. *Journal of Peking University (Humanities and Social Science Edition)*, 1:80–90, 2008c.

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

H. P. van Ditmarsch. *Knowledge games*. PhD thesis, University of Groningen, Groningen, The Netherlands, 2000. ILLC Dissertation Series DS-2000-06.

Many-Valued Hybrid Logic

Jens Ulrik Hansen, Thomas Bolander and Torben Braüner
Programming, Logic and Intelligent Systems Research Group and Science Studies Research Group, Roskilde University;
Informatics and Mathematical Modelling, Technical University of Denmark;
Programming, Logic and Intelligent Systems Research Group, Roskilde University.
`jensuh@ruc.dk, tb@imm.dtu.dk, torben@ruc.dk`

Abstract

In this paper we define a many-valued semantics for hybrid logic and we give a sound and complete tableau system which is proof-theoretically well-behaved, in particular, it gives rise to a decision procedure for the logic. This shows that many-valued hybrid logics is a natural enterprise and opens up the way for future applications.[1]

1 Introduction

Classical hybrid logic is obtained by adding to ordinary, classical modal logic further expressive power in the form of a second sort of propositional symbols called nominals, and moreover, by adding so-called satisfaction operators. A nominal is assumed to be true at exactly one world, so a nominal can be considered the name of a world. Thus, in hybrid logic a name is a particular sort of propositional symbol whereas in first-order logic it is an argument to a predicate. If i is a nominal and ϕ is an arbitrary formula, then a new formula $@_i\phi$ called a satisfaction statement can be formed. The part $@_i$ of $@_i\phi$ is called a satisfaction operator. The satisfaction statement $@_i\phi$ expresses that the formula ϕ is true at one particular world, namely the world at which the nominal i is true. Hybrid logic is proof-theoretically well-behaved, which is documented in the forthcoming book Braüner (2008). Hybrid-logical proof-theory includes a long line of work on tableau systems for hybrid logic, see Blackburn (2000), Blackburn and Marx (2002), Bolander and Braüner (2006), Bolander and Blackburn (2007a), Hansen (2007), Bolander and Blackburn (2007b).

Now, classical hybrid logic can be viewed as a combination of two logics, namely classical, two-valued logic (where the standard propositional connectives are interpreted in terms of the truth-values true and false) and hybrid modal logic (where modal operators, nominals, and satisfaction operators are

[1]This article also appears in the proceeding of the workshop *Advances in Modal Logic 2008*, Nancy, France.

interpreted in terms of a set of possible worlds equipped with an accessibility relation). The present paper concerns many-valued hybrid logic, that is, hybrid logic where the two-valued logic basis has been generalized to a many-valued logic basis. To be more precise, we shall define a many-valued semantics for hybrid logic, and we shall give a tableau system that is sound and complete with respect to the semantics. Not only is the many-valued semantics a generalization of the two-valued semantics, but if we chose a two-valued version of the many-valued tableau system, then modulo minor reformulations and the deletion of superfluous rules, the tableau system obtained is identical to an already known tableau systems for hybrid logic. Our many-valued semantics is a hybridized version of a many-valued semantics for modal logic given in the papers Fitting (1992a;b; 1995). A notable feature of this semantics is that it allows the accessibility relation as well as formulas to take on many truth-values (in other many-valued modal logics it is only formulas that can take on many truth-values).

A leading idea behind our work is that we distinguish between the way of reasoning and what the reasoning is about, and in accordance with this idea, we generalize the way of reasoning from two-valued logic to many-valued logic such that we reason in a many-valued way about time, space, knowledge, states in a computer, or whatever the subject-matter is. Given our distinction between the way of reasoning and what the reasoning is about, we take it that the concerns of hybrid logic basically are orthogonal to as whether the logic basis is two-valued or many-valued. Thus, it is expectable that the already known proof-theoretically well-behaved tableau systems for two-valued hybrid logic can be generalized to proof-theoretically well-behaved tableau systems for many-valued hybrid logic. Accordingly, if we define a many-valued hybrid logic and give a tableau system that satisfies standard proof-theoretic requirements (it is cut-free, it satisfies a version of the subformula property, and it gives rise to a decision procedure), then we learn more about hybrid logic and we provide more evidence that hybrid logic and hybrid-logical proof-theory is a natural enterprise.

This paper is structured as follows. In the second section of the paper we define the many-valued semantics for hybrid logic and we make some remarks on the relation to intuitionistic hybrid logic. In the third section we introduce a tableau system, in the fourth section we prove termination, and in the fifth section we prove completeness.

2 A Many-Valued Hybrid Logic language

In this section a Many-Valued Hybrid Logic language (denoted by **MVHL**) is presented and a semantics for the language is given. We have included global modalities, one reason being that they are used in our motivation for our choice of semantics for the nominals, but our termination and completeness proofs later in the paper do not include global modalities. In the following let $\mathcal{T}$ denote a fixed finite Heyting algebra. That is, $\mathcal{T}$ is a finite lattice such that for all a and b in $\mathcal{T}$ there is a greatest element x of $\mathcal{T}$ satisfying $a \wedge x \leq b$. The element x is called the *relative pseudo-complement* of a with respect to b (denoted $a \Rightarrow b$). To avoid notational ambiguity in relation to the syntax of our hybrid logic, we will in the following use the symbol $\Rightarrow$ for relative pseudo-complement, and

$\sqcup$ and $\sqcap$ for meet and join, respectively. The largest and smallest elements of $\mathcal{T}$ are denoted $\top$ and $\bot$, respectively. The elements of the Heyting algebra $\mathcal{T}$ are going to be used as truth values for our many-valued logic. Thus, in the following, we will often refer to the elements of $\mathcal{T}$ as *truth values*.[2]

2.1 Syntax for MVHL

Let a countable infinite set of propositional variables PROP and a countable infinite set of nominals NOM be given. In addition to the usual connectives of propositional model logic, we include the global modalities E and A, and for every $i \in \mathsf{NOM}$, a satisfaction operator $@_i$.

Definition 2.1 (**MVHL**-formulas)**.** The set of **MVHL**-formulas is given by the following grammar:

$$\varphi \;::=\; p \mid a \mid i \mid (\psi_1 \wedge \psi_2) \mid (\psi_1 \vee \psi_2) \mid (\psi_1 \rightarrow \psi_2) \mid \Box\psi \mid \Diamond\psi \mid @_i\psi \mid E\psi \mid A\psi\ ,$$

where $p \in \mathsf{PROP}$, $a \in \mathcal{T}$, and $i \in \mathsf{NOM}$.

In general we will use i, j, k and so on for nominals and a, b, c for elements of $\mathcal{T}$.

2.2 Semantics for MVHL

The semantics for **MVHL** is a Kripke semantics in which the accessibility relation is allowed to take values in $\mathcal{T}$. This is inspired by Fitting (1995). A model $\mathcal{M}$ is a tuple $\mathcal{M} = \langle W, R, \mathbf{n}, v\rangle$, where W is the set of worlds, and R a mapping $R : W \times W \rightarrow \mathcal{T}$ called the accessibility relation. $\mathbf{n}$ is a function interpreting the nominals, i.e. $\mathbf{n} : \mathsf{NOM} \rightarrow W$. Finally the valuation $v : W \times \mathsf{PROP} \rightarrow \mathcal{T}$ assigns truth values to the propositional variables at each world.

Now given a model $\mathcal{M} = \langle W, R, \mathbf{n}, v\rangle$, we can extend the valuation v to all formulas in the following inductive way, where $w \in W$:

[2] In order to give reasonable semantics for $\wedge$ and $\vee$ a Lattice structure is needed. A complete Lattice would be enough if the accessibility relation was only allowed to have two values, but since we also allows for the accessibility relation to take values in $\mathcal{T}$, the structure of a Heyting algebra is needed. For further discussions of the choice of a finite Heyting algebra as the set of truth values see Fitting (1992b; 1995).

$$
\begin{aligned}
\nu(w,a) &:= a \quad \text{for } a \in \mathcal{T} \\
\nu(w,i) &:= \begin{cases} \top & \text{, if } \mathbf{n}(i) = w \\ \bot & \text{, else} \end{cases} \\
\nu(w,\varphi \wedge \psi) &:= \nu(w,\varphi) \sqcap \nu(w,\psi) \\
\nu(w,\varphi \vee \psi) &:= \nu(w,\varphi) \sqcup \nu(w,\psi) \\
\nu(w,\varphi \to \psi) &:= \nu(w,\varphi) \Rightarrow \nu(w,\psi) \\
\nu(w,\Box\varphi) &:= \bigsqcap\{R(w,v) \Rightarrow \nu(v,\varphi) \mid v \in W\} \\
\nu(w,\Diamond\varphi) &:= \bigsqcup\{R(w,v) \sqcap \nu(v,\varphi) \mid v \in W\} \\
\nu(w,@_i\varphi) &:= \nu(\mathbf{n}(i),\varphi) \\
\nu(w,A\varphi) &:= \bigsqcap\{\nu(v,\varphi) \mid v \in W\} \\
\nu(w,E\varphi) &:= \bigsqcup\{\nu(v,\varphi) \mid v \in W\}
\end{aligned}
$$

The semantics chosen for the hybrid logical constructions is discussed in the following. The semantics for $@_i\varphi$ is obvious, its truth value is simply the truth value of φ at the world i denotes. The semantics chosen for the global modalities A and E reflect the fact that these modalities are simply the global versions of the modalities $\Box$ and $\Diamond$. The choice of semantics for nominals is less obvious. In this paper we have chosen to assign each nominal i the value $\top$ in exactly one world, and $\bot$ in all other worlds. This is in agreement with the the standard semantics for hybrid logic in which a nominal "points to a unique world". It would probably also be possible to allow nominals to take values outside the set $\{\top, \bot\}$, but at least a nominal should receive the value $\top$ in one and only one world in order for the semantics to be in accordance with classical, two-valued, hybrid logic (and for nominals to be semantically different from ordinary propositional symbols). Our decision of making the semantics of nominals two-valued rests primarily on the fact that it allows us to preserve the following well-known logical equivalence from classical, two-valued, hybrid logic:

$$
\begin{aligned}
@_i\varphi &\leftrightarrow E(i \wedge \varphi) \\
@_i\varphi &\leftrightarrow A(i \to \varphi)
\end{aligned}
$$

With the chosen semantics, these equivalences also hold in MVHL:

$$
\begin{aligned}
\nu(w,@_i\varphi) = \nu(\mathbf{n}(i),\varphi) &= \bigsqcup\{\nu(v,i) \sqcap \nu(v,\varphi) \mid v \in W\} = \nu(w,E(i \wedge \varphi)) \\
\nu(w,@_i\varphi) = \nu(\mathbf{n}(i),\varphi) &= \bigsqcap\{\nu(v,i) \Rightarrow \nu(v,\varphi) \mid v \in W\} = \nu(w,A(i \to \varphi)).
\end{aligned}
$$

Here we have been using that the following holds in a Heyting algebra: $\top \sqcap a = a$, $\bot \sqcap a = \bot$, $a \sqcup \bot = a$, $\top \Rightarrow a = a$ and $\bot \Rightarrow a = \top$. Another pleasant property resulting from the choice of semantics for nominals is the following:

$$
\nu(w,@_i\Diamond j) = \nu(\mathbf{n}(i),\Diamond j) = \bigsqcup\{R(\mathbf{n}(i),v) \sqcap \nu(v,j) \mid v \in W\} = R(\mathbf{n}(i),\mathbf{n}(j)).
$$

This identity expresses that the reachability of the world denoted by j from the world denoted by i is described by the formula $@_i\Diamond j$. This property also holds in classical hybrid logic. Identity between worlds denoted by nominals can also be expressed as usual, since we have:

$$\nu(w, @_i j) = \top \text{ iff } \mathbf{n}(i) = \mathbf{n}(j).$$

2.3 The relation to intuitionistic hybrid logic

As pointed out in the paper Fitting (1992b), there is a close relation between the many-valued modal logic given in that paper and intuitionistic modal logic. We shall in this subsection consider the relation between many-valued hybrid logic and a variant of the intuitionistic hybrid logic given in the paper Braüner and de Paiva (2006) (which in turn is a hybridization of an intuitionistic modal logic introduced in a tense-logical version in Ewald (1986)). In the present subsection we do not assume that a finite Heyting algebra has been fixed in advance, so the only atomic formulas we consider are ordinary propositional symbols, nominals, and the symbol $\bot$. We first define an appropriate notion of an intuitionistic model, which can be seen as a restricted variant of the notion of a model given in Braüner and de Paiva (2006)[3].

Definition 2.2. A *restricted model* for intuitionistic hybrid logic is a tuple

$$(W, \leq, D, \{R_w\}_{w \in W}, \{\nu_w\}_{w \in W})$$

where

1. W is a non-empty finite set partially ordered by $\leq$;
2. D is a non-empty set;
3. for each w, R_w is a binary relation on D such that $w \leq v$ implies $R_w \subseteq R_v$; and
4. for each w, ν_w is a function that to each ordinary propositional symbol p assigns a subset of D such that $w \leq v$ implies $\nu_w(p) \subseteq \nu_v(p)$.

The elements of the set W are states of knowledge and for any such state w, the relation R_w is the set of known relationships between possible worlds and the set $\nu_w(p)$ is the set of possible worlds at which p is known to be true. Note that the definition requires that the epistemic partial order $\leq$ preserves these kinds of knowledge, that is, if an advance to a greater state of knowledge is made, then what is known is preserved.

Given a restricted model $\mathfrak{M} = (W, \leq, D, \{R_w\}_{w \in W}, \{\nu_w\}_{w \in W})$, an *assignment* is a function $\mathbf{n}$ that to each nominal assigns an element of D. The relation

[3] Compare to Definition 2, p. 237, of the paper Braüner and de Paiva (2006). The differences are the following: i) In Braüner and de Paiva (2006) the set W need not be finite. ii) Instead of D there is a family $\{D_w\}_{w \in W}$ of non-empty sets such that $w \leq v$ implies $D_w \subseteq D_v$, R_w is a binary relation on D_w, and $\nu_w(p)$ is a subset of D_w. iii) There is a family $\{\sim_w\}_{w \in W}$ where $\sim_w$ is an equivalence relation on D_w such that $w \leq v$ implies $\sim_w \subseteq \sim_v$ and such that if $d \sim_w d'$, $e \sim_w e'$, and $dR_w e$, then $d' R_w e'$, and similarly, if $d \sim_w d'$ and $d \in \nu_w(p)$, then $d' \in \nu_w(p)$. The equivalence relations are used for the interpretation of nominals. Such a model for intuitionistic hybrid logic corresponds to a standard model for intuitionistic first-order logic with equality where equality is interpreted using the equivalence relations, cf. Troelstra and van Dalen (1988).

$\mathfrak{M}, \mathbf{n}, w, d \models \phi$ is defined by induction, where w is an element of W, $\mathbf{n}$ is an assignment, d is an element of D, and ϕ is a formula.

$$\begin{array}{rcl}
\mathfrak{M}, \mathbf{n}, w, d \models p & \text{iff} & d \in v_w(p) \\
\mathfrak{M}, \mathbf{n}, w, d \models i & \text{iff} & d = \mathbf{n}(i) \\
\mathfrak{M}, \mathbf{n}, w, d \models \phi \wedge \psi & \text{iff} & \mathfrak{M}, \mathbf{n}, w, d \models \phi \text{ and } \mathfrak{M}, \mathbf{n}, w, d \models \psi \\
\mathfrak{M}, \mathbf{n}, w, d \models \phi \vee \psi & \text{iff} & \mathfrak{M}, \mathbf{n}, w, d \models \phi \text{ or } \mathfrak{M}, \mathbf{n}, w, d \models \psi \\
\mathfrak{M}, \mathbf{n}, w, d \models \phi \rightarrow \psi & \text{iff} & \text{for all } v \geq w, \\
 & & \mathfrak{M}, \mathbf{n}, v, d \models \phi \text{ implies } \mathfrak{M}, \mathbf{n}, v, d \models \psi \\
\mathfrak{M}, \mathbf{n}, w, d \models \bot & \text{iff} & \text{falsum} \\
\mathfrak{M}, \mathbf{n}, w, d \models \Box\phi & \text{iff} & \text{for all } v \geq w, \text{ for all } e \in D, \\
 & & dR_v e \text{ implies } \mathfrak{M}, \mathbf{n}, v, e \models \phi \\
\mathfrak{M}, \mathbf{n}, w, d \models \Diamond\phi & \text{iff} & \text{for some } e \in D, dR_w e \text{ and } \mathfrak{M}, \mathbf{n}, w, e \models \phi \\
\mathfrak{M}, \mathbf{n}, w, d \models @_i\phi & \text{iff} & \mathfrak{M}, \mathbf{n}, w, \mathbf{n}(i) \models \phi \\
\mathfrak{M}, \mathbf{n}, w, d \models A\phi & \text{iff} & \text{for all } v \geq w, \text{ for all } e \in D, \mathfrak{M}, \mathbf{n}, v, e \models \phi \\
\mathfrak{M}, \mathbf{n}, w, d \models E\phi & \text{iff} & \text{for some } e \in D, \mathfrak{M}, \mathbf{n}, w, e \models \phi
\end{array}$$

This semantics can be looked upon in two different ways: As indicated above, it can be seen as a restricted variant of the semantics given in Braüner and de Paiva (2006), but it can also be seen as a hybridized version of a semantics given in the paper Fitting (1992b). In the latter paper, the epistemic worlds of the semantics are thought of as experts and the epistemic partial order is thought of as a relation of dominance between experts: One expert dominates another one if whatever the first expert says is true is also said to be true by the second expert.

As pointed out in Fitting (1992b), the intuitionistic semantics for modal logic is in a certain sense equivalent to the many-valued semantics. This also holds in the hybrid-logical case. In what follows, we outline this equivalence. It can be shown that given a restricted model $\mathfrak{M} = (W, \leq, D, \{R_w\}_{w \in W}, \{v_w\}_{w \in W})$, cf. Definition 2.2, and an assignment $\mathbf{n}$, the $\leq$-closed subsets of W ordered by $\subseteq$ constitute a finite Heyting algebra, and moreover, a many-valued model $(D, R^*, \mathbf{n}, v^*)$ can be defined by letting

- $R^*(d, e) = \{w \in W \mid dR_w e\}$ and
- $v^*(d, p) = \{w \in W \mid d \in v_w(p)\}$.

By a straightforward extension of the corresponding proof in Fitting (1992b), it can be proved that for any formula ϕ, it is the case that $v^*(d, \phi) = \{w \in W \mid \mathfrak{M}, \mathbf{n}, w, d \models \phi\}$. Conversely, given a finite Heyting algebra $\mathcal{T}$ and a many-valued model $(D, R, \mathbf{n}, v)$, a restricted model $\mathfrak{M} = (W, \subseteq, D, \{R^*_w\}_{w \in W}, \{v^*_w\}_{w \in W})$ can be defined by letting

- $W = \{w \mid w \text{ is a proper prime filter in } \mathcal{T}\}$,
- $dR^*_w e$ if and only if $R(d, e) \in w$, and
- $d \in v^*_w(p)$ if and only if $v(d, p) \in w$.

Details can be found in the paper Fitting (1992b). Again, by a straightforward extension of the corresponding proof in that paper, it can be proved that for any formula ϕ, it is the case that $\mathfrak{M}, \mathbf{n}, w, d \models \phi$ if and only if $v(d, \phi) \in w$.

Thus, in the above sense the intuitionistic semantics for hybrid logic is equivalent to the many-valued semantics for hybrid logic. It is an interesting question whether there is such an equivalence if instead of the restricted models of Definition 2.2 one considers the more general models for intuitionistic hybrid logic given in the paper Braüner and de Paiva (2006)[4]. We shall leave this to further work.

3 A tableau calculus for MVHL

In the following we will present a tableau calculus for MVHL. The basic notions for tableaux are defined as usual (see e.g. Fitting (1983)). The formulas occurring in our tableaux will all be of the form $@_i(a \to \varphi)$ or $@_i(\varphi \to a)$ prefixed either a T or an F, where $i \in \mathsf{NOM}$ and $a \in \mathcal{T}$. That is, the formulas occurring in our tableaux will be *signed formulas* of hybrid logic. A signed formula of the form $T@_i(a \to \varphi)$ is used to express that the formula $a \to \varphi$ is true at i, that is, receives the value $\top$ at i. If $v(\mathbf{n}(i), a \to \varphi) = \top$ then, by definition of v, $a \Rightarrow v(\mathbf{n}(i), \varphi) = \top$. By definition of relative pseudo-complement we then get that $\top$ is the greatest element of $\mathcal{T}$ satisfying $a \wedge \top \leq v(\mathbf{n}(i), \varphi)$. In other words, we simply have $a \leq v(\mathbf{n}(i), \varphi)$. Thus what is expressed by a formula $T@_i(a \to \varphi)$ is that the truth value of φ at i is greater than or equal to a. Symmetrically, a signed formula of the formula $T@_i(\varphi \to a)$ expresses that the truth value of φ at i is less than or equal to a. Dually, a signed formula of the form $F@_i(a \to \varphi)$ ($F@_i(\varphi \to a)$) expresses that the truth value of φ at i is *not* greater than or equal to (less than or equal to) a.

The tableau rules are divided into four classes; Branch Closing Rules, Non-modal Rules, Modal Rules and Hybrid Rules. The Branch Closing Rules and Propositional Rules are direct translations of Fitting's corresponding rules for the pure modal case Fitting (1995).

Branch Closing Rules:

A tableau branch Θ is said to be *closed* if one of the following holds:

1. $T@_i(a \to b) \in \Theta$, for some a, b with $a \nleq b$.
2. $F@_i(a \to b) \in \Theta$, for some a, b with $a \leq b$, $a \neq \bot$, and $b \neq \top$.
3. $F@_i(\bot \to \varphi) \in \Theta$, for some formula φ.
4. $F@_i(\varphi \to \top) \in \Theta$, for some formula φ.
5. $T@_i(b \to \varphi), F@_i(a \to \varphi) \in \Theta$, for some a, b with $a \leq b$.
6. $T@_j(a \to i), F@_i(b \to j) \in \Theta$, for some $a, b \neq \bot$.
7. $T@_i(i \to a) \in \Theta$, for some nominal i and truth value a with $a \neq \top$.

[4]As indicated in the previous footnote, in the intuitionistic semantics of Braüner and de Paiva (2006), nominals are interpreted using a family $\{\sim_w\}_{w \in W}$ of equivalence relations, not identity. This seems to imply that in an equivalent many-valued semantics, nominals should be allowed to take on arbitrary truth-values, not just top and bottom.

$$\frac{T@_i(a \to (\varphi \wedge \psi))}{\begin{array}{c} T@_i(a \to \varphi) \\ T@_i(a \to \psi) \end{array}}\,(\mathbf{T}\wedge)^1 \qquad \frac{F@_i(a \to (\varphi \wedge \psi))}{F@_i(a \to \varphi) \mid F@_i(a \to \psi)}\,(\mathbf{F}\wedge)^1$$

$$\frac{T@_i((\varphi \vee \psi) \to a)}{\begin{array}{c} T@_i(\varphi \to a) \\ T@_i(\psi \to a) \end{array}}\,(\mathbf{T}\vee)^2 \qquad \frac{F@_i((\varphi \vee \psi) \to a)}{F@_i(\varphi \to a) \mid F@_i(\psi \to a)}\,(\mathbf{F}\vee)^2$$

$$\frac{F@_i(a \to (\varphi \to \psi))}{\begin{array}{c|c|c} T@_i(b_1 \to \varphi) & \cdots & T@_i(b_n \to \varphi) \\ F@_i(b_1 \to \psi) & \cdots & F@_i(b_n \to \psi) \end{array}}\,(\mathbf{F}\!\to)^3 \qquad \frac{T@_i(a \to (\varphi \to \psi))}{F@_i(b \to \varphi) \mid T@_i(b \to \psi)}\,(\mathbf{T}\!\to)^4$$

[1] Where $a \neq \bot$.
[2] Where $a \neq \top$.
[3] Where $a \neq \bot$ and $b_1, ..., b_n$ are all the members of $\mathcal{T}$ with $b_i \leq a$ except $\bot$.
[4] Where $a \neq \bot$ and b is any member of $\mathcal{T}$ with $b \leq a$ except $\bot$.

Figure 1: Propositional Rules for **MVHL**.

The two last conditions, 6 and 7, have no counterpart in Fitting's system, but are required in ours to deal with the semantics chosen for nominals. Note that if a formula $F@_i(a \to i)$ with $a \neq \top$ occurs on a branch then the branch can also be closed: In case $a = \bot$, condition 3 immediately implies closure. If $a \neq \bot$ then using the reversal rule ($\mathbf{F} \geq$) (see below), we can add a formula $T@_i(i \to b)$ to the branch, where b is one of the maximal members of $\mathcal{T}$ not above a. Because b is not above a, b cannot be $\top$. Thus condition 7 implies closure.

Non-modal Rules:

The tableau rules for the propositional connectives and the rules capturing the properties of the Heyting algebra are given in Figure 1 and Figure 2, respectively. The rules of Figure 2 are called *reversal rules*, as in Fitting (1995). The reversal rules together with the closure rules ensure that no formula can be assigned more than one truth value (relative to a given world and a given branch).

Modal Rules:

These modal rules, presented in Figure 3, differ from the ones of Fitting and heavily employs the hybrid logic machinery. Note that the tableau rules contain formulas of the form $T@_i(a \leftrightarrow \Diamond j)$. Such formulas are simply used as shorthand notation for the occurrence of both the formulas $T@_i(a \to \Diamond j)$ and $T@_i(\Diamond j \to a)$. In each of the rules of our calculus, the leftmost premise is called the *principal premise*. If α is a signed formula on one of the forms $T@_i(a \to \varphi)$, $T@_i(\varphi \to a)$, $F@_i(a \to \varphi)$ or $F@_i(\varphi \to a)$, we call φ the *body* of α and i its *prefix*. If α and β are two signed formulas such that the body of α is a subformula of the body of β, then α is said to be a *quasi-subformula* of β.

$$\frac{F@_i(a \to \varphi)}{T@_i(\varphi \to b_1) \mid \cdots \mid T@_i(\varphi \to b_n)}\ \textbf{(F}\geq\textbf{)}^{1,2} \qquad \frac{T@_i(a \to \varphi)}{F@_i(\varphi \to b)}\ \textbf{(T}\geq\textbf{)}^{1,3}$$

$$\frac{F@_i(\varphi \to a)}{T@_i(b_1 \to \varphi) \mid \cdots \mid T@_i(b_n \to \varphi)}\ \textbf{(F}\leq\textbf{)}^{1,4} \qquad \frac{T@_i(\varphi \to a)}{F@_i(b \to \varphi)}\ \textbf{(T}\leq\textbf{)}^{1,5}$$

[1] φ is a formula other than a propositional constant from $\mathcal{T}$.
[2] Where $b_1, ..., b_n$ are all maximal members of $\mathcal{T}$ with $a \nleq b_i$ and $a \neq \bot$.
[3] Where b is any maximal member of $\mathcal{T}$ with $a \nleq b$ and $a \neq \bot$.
[4] Where $b_1, ..., b_n$ are all minimal members of $\mathcal{T}$ with $b_i \nleq a$ and $a \neq \top$.
[5] Where b is any minimal member of $\mathcal{T}$ with $b \nleq a$ and $a \neq \top$.

Figure 2: Reversal Rules for **MVHL**.

$$\frac{F@_i(a \to \Box\varphi)}{\begin{array}{c} T@_i(b_1 \leftrightarrow \Diamond j) \\ F@_j((a \sqcap b_1) \to \varphi) \end{array} \left| \begin{array}{c} \cdots \\ \cdots \end{array} \right| \begin{array}{c} T@_i(b_n \leftrightarrow \Diamond j) \\ F@_j((a \sqcap b_n) \to \varphi) \end{array}}\ \textbf{(F}\Box\textbf{)}^{1} \qquad \frac{T@_i(a \to \Box\varphi) \qquad T@_i(b \to \Diamond j)}{T@_j((a \sqcap b) \to \varphi)}\ \textbf{(T}\Box\textbf{)}$$

$$\frac{F@_i(\Diamond\varphi \to a)}{\begin{array}{c} T@_i(b_1 \leftrightarrow \Diamond j) \\ F@_j(\varphi \to (b_1 \Rightarrow a)) \end{array} \left| \begin{array}{c} \cdots \\ \cdots \end{array} \right| \begin{array}{c} T@_i(b_n \leftrightarrow \Diamond j) \\ F@_j(\varphi \to (b_n \Rightarrow a)) \end{array}}\ \textbf{(F}\Diamond\textbf{)}^{1,2} \qquad \frac{T@_i(\Diamond\varphi \to a) \qquad T@_i(b \to \Diamond j)}{T@_j(\varphi \to (b \Rightarrow a))}\ \textbf{(T}\Diamond\textbf{)}^{2}$$

$$\frac{F@_i(E\varphi \to a)}{F@_j(\varphi \to a)}\ \textbf{(FE)}^{3} \qquad \frac{T@_i(E\varphi \to a)}{T@_j(\varphi \to a)}\ \textbf{(TE)}^{4}$$

$$\frac{T@_i(a \to A\varphi)}{T@_j(a \to \varphi)}\ \textbf{(TA)}^{4} \qquad \frac{F@_i(a \to A\varphi)}{F@_j(a \to \varphi)}\ \textbf{(FA)}^{3}$$

[1] Where $\mathcal{T} = \{b_1, ..., b_n\}$ and j is a nominal new to the branch.
[2] Where the principal premise is a quasi-subformula of the root formula.
[3] Where j is a nominal new to the branch.
[4] Where j is a nominal already occurring on the branch.

Figure 3: Modal Rules for **MVHL**.

$$\frac{T@_i(@_j\varphi \to a)}{T@_j(\varphi \to a)}\,(@_L) \qquad \frac{T@_i(a \to @_j\varphi)}{T@_j(a \to \varphi)}\,(@_R)$$

$$\frac{F@_i\varphi \qquad T@_i(a \to j)}{F@_j\varphi}\,\textbf{(F-NOM)}^{1,2} \qquad \frac{T@_i\varphi \qquad T@_i(a \to j)}{T@_j\varphi}\,\textbf{(T-NOM)}^{1,2}$$

$$\frac{T@_k(\Diamond i \to b) \qquad T@_i(a \to j)}{T@_k(\Diamond j \to b)}\,(\textbf{BRIDGE}_L)^1 \qquad \frac{T@_k(b \to \Diamond i) \qquad T@_i(a \to j)}{T@_k(b \to \Diamond j)}\,(\textbf{BRIDGE}_R)^1$$

$$\frac{T@_i(\top \to j) \qquad T@_j(\top \to k)}{T@_i(\top \to k)}\,(\textbf{TRANS})$$

$$\frac{T@_i(a \to j)}{T@_i(\top \to j)}\,\textbf{(NOM EQ)}^1$$

[1] Where $a \neq \bot$.
[2] Where the principal premise is a quasi-subformula of the root formula.

Figure 4: Hybrid Rules for **MVHL**.

Hybrid Rules:

These hybrid rules, presented in Figure 4, are inspired by the standard rules from classical hybrid logic (see Blackburn (2000), Bolander and Braüner (2006), Bolander and Blackburn (2007a)). Note that for the **(NOM)** rule, two versions are needed. Furthermore a new rule is needed due to the fact that we are in a many-valued setting, this is the rule **(NOM EQ)**, which ensures our semantic definition of nominals as being $\top$ in exactly one world.

A *tableau proof* of a formula ϕ is a closed tableau with root $F@_i(\top \to \phi)$, where i is an arbitrary nominal not occurring in ϕ. The intuition here is that the root formula $F@_i(\top \to \phi)$ asserts that ϕ does *not* have the value $\top$, and if the tableau closes, this assertion is refuted. If i is a nominal occurring in the root formula of a tableau then i is called a *root nominal* of the tableau. Other nominals occurring on the tableau are called *non-root nominals*.

4 Termination

The tableau calculus presented above is not terminating. This is due to the rules (**TA**) and (**FA**) for the global modality A. If the rules for the global modalities—(**FE**), (**TE**), (**TA**) and (**FA**)—are all removed, we obtain a tableau calculus for the many-valued hybrid logic with these modalities removed. We will refer to this calculus as the *basic calculus*, and refer to its tableaux as *basic tableaux*. In the following we will prove that the basic calculus terminates. The proof closely follows the method introduced in Bolander and Blackburn (2007a).

If α and β are signed formulas on a tableau branch, then β is said to be *produced* by α if β is one of the conclusions of a rule application with principal premise α. The signed formula β is said to be *indirectly produced* by α if there exists a sequence of signed formulas $\alpha, \alpha_1, \alpha_2, \ldots, \alpha_n, \beta$ in which each formula is produced by its predecessor. We now have the following result.

Lemma 1 (Quasi-subformula Property). *Let $\mathcal{T}$ be a basic tableau. For any signed formula α occurring on $\mathcal{T}$, one of the following holds:*

1. *α is a quasi-subformula of the root formula of $\mathcal{T}$.*
2. *α is a formula of one of the forms $T@_i(a \rightarrow \Diamond j)$, $T@_i(\Diamond j \rightarrow a)$, $F@_i(a \rightarrow \Diamond j)$ or $F@_i(\Diamond j \rightarrow a)$, for which one of the following holds:*
   - *(a) j is a root nominal.*
   - *(b) α is indirectly produced by ($\mathbf{F}\Box$) or ($\mathbf{F}\Diamond$) by a number of applications of the reversal rules.*

Proof. The proof goes by induction on the construction of $\mathcal{T}$. In the basic case α is just the root formula, which of course is of type 1. Now assume that α have been introduced by one of the propositional rules. These rules does not take premises of type 2 and thus by induction they must be of type 1. But then the conclusions produced by these rules must also be of type 1, thus α must be of type 1. If α have been produced by once of the reversal rules by a formula of type 1, then α will also by of type 1 and if α is produced by a formula of type 2, α is also of type 2. Now the modal rules. If α have been produced by the rule ($\mathbf{T}\Box$) then the principal premise can not be a formula of type 2 and thus by induction it must be of type 1. But then so is α. Similar for the rule ($\mathbf{T}\Diamond$) where the side condition insures that the principal premise is of type 1. If α is introduced by on of the rules ($\mathbf{F}\Box$) or ($\mathbf{F}\Diamond$) again the premise must be of type 1. These rules produce two formulas, the first one is by definition of type 2b and the second must be of type 1 since the premise is. Thus in this case α is either of type 1 or type 2b. Finally for the hybrid rules. In the rules (**TRANS**), (**NOM EQ**), ($@_L$) or ($@_R$) the premises can not be of type 2 and thus by induction they must be of type 1. But then the conclusions will also be of type 1. Now if the rule used is (**T-NOM**) or (**F-NOM**) then the side condition insures that the principal premise are of type 1. But then the conclusion will also be of type 1. Now assume that one of the rules ($\mathbf{BRIDGE}_L$) or ($\mathbf{BRIDGE}_R$) have been applied to produce α. Then the non-principal premise can not be of type 1 and thus must be of type 2, which implies that j is a root nominal. Thus the conclusion α must be of type 2a. This completes the proof. $\Box$

Note that in the basic calculus the only rules that can introduce new nominals to a tableau are ($\mathbf{F}\Box$) and ($\mathbf{F}\Diamond$).

Definition 4.1. Let Θ be a branch of a basic tableau. If a nominal j has been introduced to the branch by applying either ($\mathbf{F}\Box$) or ($\mathbf{F}\Diamond$) to a premise with prefix i then we say that j is *generated* by i on Θ, and we write $i \prec_\Theta j$.

Lemma 2. *Let Θ be a branch of a basic tableau. The graph $G = (N^\Theta, \prec_\Theta)$, where N^Θ is the set of nominals occurring on Θ, is a finite set of wellfounded, finitely branching trees.*

Proof. That G is wellfounded follows from the observation that if $i \prec_\Theta j$, then the first occurrence of i on Θ is before the first occurrence of j. That G is finitely branching is shown as follows. For any given nominal i the number of nominals j satisfying $i \prec_\Theta j$ is bounded by the number of applications of (**F**$\Box$) and (**F**$\Diamond$) to premises of the form $F@_i(a \to \Box\varphi)$ and $F@_i(\Diamond\varphi \to a)$. So to prove that G is finitely branching, we only need to prove that for any given i the number of such premises is finite. However, this follows immediately from the fact that all such premises must be quasi-subformulas of the root formula (cf. Lemma 1 and the condition on applications of (**F**$\Diamond$)). What is left is to prove that G is a finite set of trees. This follows from the fact that each nominal in N^Θ can be generated by at most one other nominal, and the fact that each nominal in N^Θ must have one of the finitely many root nominals of Θ as an ancestor. $\Box$

Lemma 3. *Let Θ be a branch of a basic tableau. Then Θ is infinite if and only if there exists an infinite chain of nominals*

$$i_1 \prec_\Theta i_2 \prec_\Theta i_3 \prec_\Theta \cdots .$$

Proof. The 'if' direction is trivial. To prove the 'only if' direction, let Θ be any infinite tableau branch. Θ must contain infinitely many distinct nominals, since it follows immediately from Lemma 1 that a tableau with finitely many nominals can only contain finitely many distinct formulas. This implies that the graph $G = (N^\Theta, \prec_\Theta)$ defined as in Lemma 2 must be infinite. Since by Lemma 2, G is a finite set of wellfounded, finitely branching trees, G must then contain an infinite path $(i_1, i_2, i_3, \dots)$. Thus we get an infinite chain $i_1 \prec_\Theta i_2 \prec_\Theta i_3 \prec_\Theta \cdots$. $\Box$

Definition 4.2. Let Θ be a branch of a basic tableau, and let i be a nominal occurring on Θ. We define $m_\Theta(i)$ to be the maximal length of any formula with prefix i occurring on Θ.

Lemma 4 (**Decreasing length**)**.** *Let Θ be a branch of a basic tableau. If $i \prec_\Theta j$ then $m_\Theta(i) > m_\Theta(j)$.*

Proof. For any signed formula α, we will use $|\alpha|$ to denote the length of α. Assume $i \prec_\Theta j$. Let α be a signed formula satisfying: 1) α has maximal length among the formulas on Θ with prefix j; 2) α is the earliest occurring formula on Θ with this property. We need to prove $m_\Theta(i) > |\alpha|$. The formula α can not have been introduced on Θ by applying any of the propositional rules (Figure 1), since this would contradict maximality of α. It can not have been directly produced by any of the reversal rules (Figure 2) either, since this would contradict the choice of α as the earliest possible on Θ of maximal length with prefix j. By the same argument, α can not have been directly produced by any of the rules (**BRIDGE**$_L$), (**BRIDGE**$_R$), (**TRANS**) or (**NOM EQ**). Assume now α has been introduced by applying ($@_L$) or ($@_R$) to a premise of the form $T@_k(@_j\varphi \to a)$ or $T@_k(a \to @_j\varphi)$. By Lemma 1, the premise must be a quasi-subformula of the root formula. Thus j must be a root nominal. However, this is a contradiction, since by assumption j is generated by i, and can thus not be a root nominal. Thus neither ($@_L$) nor ($@_R$) can have been the rule producing α. Now assume that α has been produced by an application of either (**F-NOM**) or (**T-NOM**). Since α has index j, the non-principal premise used in this rule application must have the form $T@_i(a \to j)$. By Lemma 1, this premise must be a quasi-subformula of the root formula, and thus j is again a root nominal, which is a contradiction. Thus α can not have been produced by (**F-NOM**) or (**T-NOM**) either. Thus α must have been introduced by one of the rules (**F**$\Box$), (**T**$\Box$), (**F**$\Diamond$) or (**T**$\Diamond$). Consider first the case of the (**F**$\Box$) and (**F**$\Diamond$) rules. If an instance of one of these produced α, then this instance must have been applied to a premise β with prefix i, since we have assumed $i \prec_\Theta j$ and by Lemma 2 there cannot be an $i' \neq i$ satisfying $i' \prec_\Theta j$. (Note that if α is of the form $T@_j(b \to \Diamond k)$ or $T@_j(\Diamond k \to b)$ produced by a formula $F@_j(a \to \Box\varphi)$ or $F@_j(\Diamond\varphi \to a)$, this would lead to a contradiction with the assumption that α has maximal length with prefix j and is the earliest occurring formula with this property.) Since the rules in question always produce conclusions that are shorter than their premises, β must be longer than α. Since β is a formula with prefix i we then get:

$$m_\Theta(i) \geq \left|\beta\right| > |\alpha|\,, \tag{1}$$

as required. Now consider finally the case where α has been produced by either (**T**$\Box$) or (**T**$\Diamond$). Then α has been produced by a rule instance with non-principal premise of the form $T@_k(b \to \Diamond j)$. Since j is not a root nominal, this premise can not be a quasi-subformula of the root formula. Neither can it be of the tybe ($2a$) mentioned in lemma 1. It must thus be of type ($2b$), that is, it must be indirectly produced by formulas of the form $T@_k(b_m \to \Diamond j')$ or $T@_k(\Diamond j' \to b_m)$ obtained as conclusion by applications of **(F**$\Box$**)** or **(F**$\Diamond$**)**. Since only reversal rules have been applied in the indirect production from these conclusions, we must have $j = j'$ and thus $k \prec_\Theta j$. Since we already have $i \prec_\Theta j$ we get $k = i$, using Lemma 2. We can conclude that the non-principal premise of the rule instance producing α must have the form $T@_i(b \to \Diamond j)$, and thus the principal premise must be a formula β with index i. Since the rules in question always produce conclusions that are shorter than their premises, β must be longer than α. Since β is a formula with prefix i we then again get the sequence of inequalities (1), as required. $\Box$

We can now finally prove termination of the basic calculus.

Theorem 1 (**Termination of the basic calculus).** *Any tableau in the basic calculus is finite.*

Proof. Assume there exists an infinite basic tableau. Then it must have an infinite branch Θ. By Lemma 3, there exists an infinite chain

$$i_1 \prec_\Theta i_2 \prec_\Theta i_3 \prec_\Theta \cdots .$$

Now by Lemma 4 we have

$$m_\Theta(i_1) > m_\Theta(i_2) > m_\Theta(i_3) > \cdots$$

which is a contradiction, since $m_\Theta(i)$ is a non-negative number for any nominal i. $\Box$

5 Completeness of the basic calculus

In this section we prove completeness of the basic calculus, that is, the calculus without the global modalities. In this connection we remark that we have proved completeness for a calculus including the global modalities similar to the calculus of the present paper. Let Θ be an open saturated branch in the tableau calculus. We will use this branch to construct a model $\mathcal{M}_\Theta = \langle W_\Theta, R_\Theta, \mathbf{n}_\Theta, v_\Theta \rangle$. The set of worlds, W_Θ is simply defined to be the set of nominals occurring on Θ. The definition of the other elements of the model requires a bit more work. First we define the mapping $\mathbf{n}_\Theta$.

Fix a choice function σ that for any given set of nominals on Θ returns one of these nominals. We now define the mapping $\mathbf{n}_\Theta$ in the following way:

$$\mathbf{n}_\Theta(i) = \begin{cases} \sigma\{j \mid T@_i(\top \to j) \in \Theta\} & \text{if } \{j \mid T@_i(\top \to j) \in \Theta\} \neq \emptyset \\ i & \text{otherwise.} \end{cases}$$

A nominal i is called an *urfather* on Θ if $i = \mathbf{n}_\Theta(j)$ for some nominal j.

Lemma 5. *Let Θ be a saturated tableau branch. Then we have the following properties:*

1. *If $T@_i\varphi \in \Theta$ is a quasi-subformula of the root formula then $T@_{\mathbf{n}_\Theta(i)}\varphi \in \Theta$. Similarly, if $F@_i\varphi \in \Theta$ is a quasi-subformula of the root formula then $F@_{\mathbf{n}_\Theta(i)}\varphi \in \Theta$.*
2. *If $T@_i(\top \to j) \in \Theta$ then $\mathbf{n}_\Theta(i) = \mathbf{n}_\Theta(j)$.*
3. *If i is an urfather on Θ then $\mathbf{n}_\Theta(i) = i$.*

Proof. First we prove (i). Assume $T@_i\varphi \in \Theta$ is a quasi-subformula of the root formula. If $\mathbf{n}_\Theta(i) = i$ then there is nothing to prove. So assume $\mathbf{n}_\Theta(i) = \sigma\{j \mid T@_i(\top \to j) \in \Theta\}$. Then $T@_i(\top \to \mathbf{n}_\Theta(i)) \in \Theta$, and by applying (**T-NOM**) to premises $T@_i\varphi$ and $T@_i(\top \to \mathbf{n}_\Theta(i))$ we get $T@_{\mathbf{n}_\Theta(i)}\varphi$, as needed. The case of $F@_i\varphi \in \Theta$ is proved similarly, using (**F-NOM**) instead of (**T-NOM**). We now prove (ii). Assume $T@_i(\top \to j) \in \Theta$. To prove $\mathbf{n}_\Theta(i) = \mathbf{n}_\Theta(j)$ it suffices to prove that for all nominals k, $T@_i(\top \to k) \in \Theta \Leftrightarrow T@_j(\top \to k) \in \Theta$. So let k be an arbitrary nominal. If $T@_i(\top \to k) \in \Theta$ then we can apply (**T-NOM**) (since $T@_i(\top \to k)$ is a quasi-subformula of the root formula by Lemma 1) to premises $T@_i(\top \to k)$ and $T@_i(\top \to j)$ to obtain the conclusion $T@_j(\top \to k)$, as required. If conversely $T@_j(\top \to k) \in \Theta$ then we can apply (**TRANS**) to premises $T@_i(\top \to j)$ and $T@_j(\top \to k)$ to obtain the conclusion $T@_i(\top \to k)$, as required. We finally prove (iii). Assume i is an urfather. Then $i = \mathbf{n}_\Theta(j)$ for some j. If $j = i$ we are done. Otherwise we have $i = \mathbf{n}_\Theta(j) = \sigma\{k \mid T@_j(\top \to k) \in \Theta\}$ and thus $T@_j(\top \to i) \in \Theta$. This implies $i = \mathbf{n}_\Theta(j) = \mathbf{n}_\Theta(i)$, using item (ii). □

We now turn to the definition of v_Θ. As in Fitting (1995) we will not define a particular valuation v of the propositional variables occuring on the branch, but only show that any valuation assigning values between a certain lower and upper bound (both given by the branch Θ) will do. Let us first define these bounds.

Definition 5.1. For a formula φ in the language of MVHL and a nominal i, define:

$$bound^{\Theta,i}(\varphi) = \bigsqcap\{a \mid T@_i(\varphi \to a) \in \Theta\}$$
$$bound_{\Theta,i}(\varphi) = \bigsqcup\{a \mid T@_i(a \to \varphi) \in \Theta\}$$

The intuition is that $bound^{\Theta,i}(\varphi)$ is an upper bound for the truth value of φ at the world i decided by the branch Θ and $bound_{\Theta,i}(\varphi)$ is a lower bound for this truth value.

The following lemma corresponds to Lemma 6.4 of Fitting (1995) and can be proved in the same way. It ensures that we can actually always chose a value between the lower and the upper bounds.

Lemma 6. *For all i on Θ and all formulas φ of MVHL*

$$bound_{\Theta,i}(\varphi) \leq bound^{\Theta,i}(\varphi).$$

Later we will show that any valuation assigning a value to p between $bound_{\Theta,i}(p)$ and $bound^{\Theta,i}(p)$ at the world $\mathbf{n}_\Theta(i)$ will do for the truth value of p at this world.

The following lemma corresponds to Proposition 6.5 in Fitting (1995) and is proven in the same way.

Lemma 7. *Let φ be any formula in the MVHL language other than a propositional constant from $\mathcal{T}$, and let $a \in \mathcal{T}$, then:*

- *(i) If $T@_i(a \to \varphi) \in \Theta$, then $a \leq bound_{\Theta,i}(\varphi)$.*
- *(ii) If $T@_i(\varphi \to a) \in \Theta$, then $bound^{\Theta,i}(\varphi) \leq a$.*
- *(iii) If $F@_i(a \to \varphi) \in \Theta$, then $a \nleq bound^{\Theta,i}(\varphi)$.*
- *(iv) If $F@_i(\varphi \to a) \in \Theta$, then $bound_{\Theta,i}(\varphi) \nleq a$.*

The accessibility relation R_Θ is defined as follows:

$$R_\Theta(i,j) = \bigsqcup\{b \mid T@_i(b \to \Diamond k) \in \Theta, \mathbf{n}_\Theta(k) = j\}.$$

We have the following result, which we are going to use in proving completeness.

Lemma 8. *If $T@_i(c \leftrightarrow \Diamond j) \in \Theta$ then $R_\Theta(i, \mathbf{n}_\Theta(j)) = c$.*

Proof. We will prove $R_\Theta(i, \mathbf{n}_\Theta(j)) \geq c$ and $R_\Theta(i, \mathbf{n}_\Theta(j)) \leq c$. First we prove $R_\Theta(i, \mathbf{n}_\Theta(j)) \geq c$. Since $T@_i(c \leftrightarrow \Diamond j) \in \Theta$ we have $T@_i(c \to \Diamond j) \in \Theta$, and thus

$$\begin{aligned} R_\Theta(i, \mathbf{n}_\Theta(j)) &= \bigsqcup\{b \mid T@_i(b \to \Diamond k) \in \Theta, \mathbf{n}_\Theta(k) = \mathbf{n}_\Theta(j)\} \\ &\geq \bigsqcup\{b \mid T@_i(b \to \Diamond j) \in \Theta\} \\ &\geq c. \end{aligned}$$

We now prove $R_\Theta(i, \mathbf{n}_\Theta(j)) \leq c$. By definition of $\mathbf{n}_\Theta$ we have either $\mathbf{n}_\Theta(j) = j$ or $T@_j(\top \to \mathbf{n}_\Theta(j)) \in \Theta$. If $T@_j(\top \to \mathbf{n}_\Theta(j)) \in \Theta$ then since $T@_i(\Diamond j \to c) \in \Theta$ we get $T@_i(\Diamond \mathbf{n}_\Theta(j) \to c) \in \Theta$, using ($\mathbf{BRIDGE}_L$). If $\mathbf{n}_\Theta(j) = j$ we obviously also have $T@_i(\Diamond \mathbf{n}_\Theta(j) \to c) \in \Theta$. Applying Lemma 7 (ii) we then get $bound^{\Theta,i}(\Diamond \mathbf{n}_\Theta(j)) \leq c$. Thus

$$\begin{aligned} R_\Theta(i, \mathbf{n}_\Theta(j)) &= \bigsqcup\{b \mid T@_i(b \to \Diamond k) \in \Theta, \mathbf{n}_\Theta(k) = \mathbf{n}_\Theta(j)\} \\ &\leq \bigsqcup\{b \mid T@_i(b \to \Diamond \mathbf{n}_\Theta(j)) \in \Theta\} \quad \text{(using } (\mathbf{BRIDGE}_R)) \\ &= bound_{\Theta,i}(\Diamond \mathbf{n}_\Theta(j)) \\ &\leq bound^{\Theta,i}(\Diamond \mathbf{n}_\Theta(j)) \quad \text{(using Lemma 6)} \\ &\leq c, \end{aligned}$$

as required. □

The theorem we need for completeness now may be stated in the following way:

Theorem 2. *Let v be a valuation such that for all propositional variables p and all urfather nominals i*

$$bound_{\Theta,i}(p) \leq v(i,p) \leq bound^{\Theta,i}(p).$$

Then for all subformulas φ of the body of root formula of Θ

$$bound_{\Theta,i}(\varphi) \leq v(i,\varphi) \leq bound^{\Theta,i}(\varphi).$$

Proof. By induction on φ. The base cases are where φ is a propositional variable p, a value $c \in \mathcal{T}$ or a nominal j. The case where φ is p follows directly by the assumption. The case where φ is c is easy: First note that for any truth values a, b, if $T@_i(a \to b) \in \Theta$ then $a \leq b$. This follows from closure rule 1 presented in Section 3. Thus we get:

$$bound_{\Theta,i}(c) = \bigsqcup\{a \mid T@_i(a \to c) \in \Theta\} \leq c \leq \bigsqcap\{a \mid T@_i(c \to a) \in \Theta\} = bound^{\Theta,i}(c).$$

Now assume φ is a nominal j. By definition of v, $v(i,j)$ is $\top$ if $\mathbf{n}_\Theta(j) = i$ and $\bot$ otherwise. Assume first $\mathbf{n}_\Theta(j) = i$. Then $v(i,j)$ is $\top$, so trivially we have $bound_{\Theta,i}(j) \leq v(i,j)$. We thus only need to prove $v(i,j) \leq bound^{\Theta,i}(j)$, that is, we need to prove $\top = bound^{\Theta,i}(j) = \bigsqcap\{a \mid T@_i(j \to a) \in \Theta\}$. This amounts to showing that, for all $a \in \mathcal{T}$, $T@_i(j \to a) \in \Theta$ implies $a = \top$. Assume towards a contradiction that, for some a, $T@_i(j \to a) \in \Theta$ and $a \neq \top$. Since we have assumed $\mathbf{n}_\Theta(j) = i$, by definition of $\mathbf{n}_\Theta$ we get that either $j = i$ or $T@_j(\top \to i) \in \Theta$. If $j = i$ then we have that Θ contains a formula of the form $T@_i(i \to a)$ where $a \neq \top$. This immediately contradicts closure rule 7. Assume instead $T@_j(\top \to i) \in \Theta$. Since we also have $T@_i(j \to a) \in \Theta$ where $a \neq \top$, we can apply ($\mathbf{T} \leq$) to conclude that that Θ must contain a formula of the form $F@_i(t \to j)$ where t is some truth value different from $\bot$. Since Θ then contains both $T@_j(\top \to i)$ and $F@_i(t \to j)$ where $t \neq \bot$, we get a contradiction by closure rule 6. Assume now $\mathbf{n}_\Theta(j) \neq i$. Then $v(i,j) = \bot$, and the inequality $v(i,j) \leq bound^{\Theta,i}(j)$ thus holds trivially. To prove the other inequality, $bound_{\Theta,i}(j) \leq v(i,j)$, we need to show that if $T@_i(a \to j) \in \Theta$ then $a = \bot$. Thus assume toward a contradiction that $T@_i(a \to j) \in \Theta$ and $a \neq \bot$. Then rule (**NOM EQ**) implies $T@_i(\top \to j) \in \Theta$. Thus, by item 2 of Lemma 5, we get $\mathbf{n}_\Theta(i) = \mathbf{n}_\Theta(j)$. Since i is assumed to be an urfather, item 3 of Lemma 5 implies $\mathbf{n}_\Theta(i) = i$. Thus we get $\mathbf{n}_\Theta(j) = \mathbf{n}_\Theta(i) = i$, contradiction the assumption.

Now for the induction step. First the case where φ is $@_j\psi$: Note that $v(i, @_j\psi) = v(\mathbf{n}_\Theta(j), \psi)$ and by induction hypothesis, since $\mathbf{n}_\Theta(j)$ is an urfather,

$$bound_{\Theta,\mathbf{n}_\Theta(j)}(\psi) \leq v(\mathbf{n}_\Theta(j), \psi) \leq bound^{\Theta,\mathbf{n}_\Theta(j)}(\psi).$$

Now by the rule ($@_R$), if $T@_i(a \to @_j\psi) \in \Theta$ then $T@_j(a \to \psi) \in \Theta$, for all $a \in \mathcal{T}$.

Thus we get that

$$\begin{array}{rcl} bound_{\Theta,i}(@_j\psi) & = & \bigsqcup\{a \mid T@_i(a \to @_j\psi) \in \Theta\} \\ & \leq & \bigsqcup\{a \mid T@_j(a \to \psi) \in \Theta\} \\ & \leq & \bigsqcup\{a \mid T@_{\mathbf{n}_\Theta(j)}(a \to \psi) \in \Theta\} \quad \text{(using 1 of Lemma 5)} \\ & = & bound_{\Theta,\mathbf{n}_\Theta(j)}(\psi) \\ & \leq & v(\mathbf{n}_\Theta(j), \psi) \\ & = & v(i, @_j\psi). \end{array}$$

Similar by the ($@_L$) rule, $T@_i(@_j\psi \to a) \in \Theta$ implies that $T@_j(\psi \to a) \in \Theta$, for all $a \in \mathcal{T}$. Hence

$$\begin{array}{rcl} v(i, @_j\psi) & = & v(\mathbf{n}_\Theta(j), \psi) \\ & \leq & bound^{\Theta,\mathbf{n}_\Theta(j)}(\psi) \\ & = & \bigsqcap\{a \mid T@_{\mathbf{n}_\Theta(j)}(\psi \to a) \in \Theta\} \\ & \leq & \bigsqcap\{a \mid T@_j(\psi \to a) \in \Theta\} \quad \text{(using 1 of Lemma 5)} \\ & \leq & \bigsqcap\{a \mid T@_i(@_j\psi \to a) \in \Theta\} \\ & = & bound^{\Theta,i}(@_j\psi), \end{array}$$

and the @-case is done.

In case φ is $\Diamond\psi$, we need to prove that

$$bound_{\Theta,i}(\Diamond\psi) \leq v(i, \Diamond\psi) \leq bound^{\Theta,i}(\Diamond\psi),$$

which by definition amounts to showing that

$$\bigsqcup\{a \mid T@_i(a \to \Diamond\psi) \in \Theta\} \leq \bigsqcup\{R_\Theta(i,j) \sqcap v(j,\psi) \mid j \in \Theta\} \leq \bigsqcap\{a \mid T@_i(\Diamond\psi \to a) \in \Theta\}.$$

Proving the first inequality amounts to showing that if $T@_i(a \to \Diamond\psi) \in \Theta$ then

$$a \leq \bigsqcup\{R_\Theta(i,j) \sqcap v(j,\psi) \mid j \in \Theta\}.$$

To prove this assume toward a contradiction that

$$T@_i(a \to \Diamond\psi) \in \Theta \text{ and } a \nleq \bigsqcup\{R_\Theta(i,j) \sqcap v(j,\psi) \mid j \in \Theta\},$$

for an $a \in \mathcal{T}$. Then choose a $b \in \mathcal{T}$ such that $b \geq \bigsqcup\{R_\Theta(i,j) \sqcap v(j,\psi) \mid j \in \Theta\}$ and b is a maximal member of $\mathcal{T}$ with $a \nleq b$. Then by the reversal rule **(T≥)**, $F@_i(\Diamond\psi \to b) \in \Theta$. Then using the **(F◇)** rule there is a $c \in \mathcal{T}$ and a $j \in \Theta$ such that $T@_i(c \leftrightarrow \Diamond j) \in \Theta$ and $F@_j(\varphi \to (c \Rightarrow b)) \in \Theta$. Since $T@_i(c \leftrightarrow \Diamond j) \in \Theta$, Lemma 8 implies $R_\Theta(i, \mathbf{n}_\Theta(j)) = c$. Applying 1 of Lemma 5 to the formula $F@_j(\varphi \to (c \Rightarrow b)) \in \Theta$ we get $F@_{\mathbf{n}_\Theta(j)}(\varphi \to (c \Rightarrow b)) \in \Theta$. Now (*iv*) of Lemma 7 implies $bound_{\Theta,\mathbf{n}_\Theta(j)}(\psi) \nleq c \Rightarrow b$. This further implies that $(bound_{\Theta,\mathbf{n}_\Theta(j)}(\psi) \sqcap c) \nleq b$. But by the induction hypothesis $bound_{\Theta,\mathbf{n}_\Theta(j)}(\psi) \leq v(\mathbf{n}_\Theta(j), \psi)$ and thus

$$\begin{array}{rcl} bound_{\Theta,\mathbf{n}_\Theta(j)}(\psi) \sqcap c & = & bound_{\Theta,\mathbf{n}_\Theta(j)}(\psi) \sqcap R_\Theta(i, \mathbf{n}_\Theta(j)) \\ & \leq & v(\mathbf{n}_\Theta(j), \psi) \sqcap R_\Theta(i, \mathbf{n}_\Theta(j)) \\ & \leq & \bigsqcup\{R_\Theta(i, \mathbf{n}_\Theta(j)) \sqcap v(\mathbf{n}_\Theta(j), \psi) \mid j \in \Theta\} \\ & \leq & \bigsqcup\{R_\Theta(i,j) \sqcap v(j,\psi) \mid j \in \Theta\} \leq b, \end{array}$$

which of course is a contradiction.

In order to prove that

$$\bigsqcup\{R_\Theta(i,j) \sqcap v(j,\psi) \mid j \in \Theta\} \leq \bigsqcap\{a \mid T@_i(\Diamond\psi \to a) \in \Theta\},$$

we must show that if $T@_i(\Diamond\psi \to a) \in \Theta$, then $R_\Theta(i,j) \sqcap v(j,\psi) \leq a$ for all $j \in \Theta$. Thus assume that $T@_i(\Diamond\psi \to a) \in \Theta$ and that $R_\Theta(i,j) \neq \bot$ (or else it's trivial) for an arbitrary $j \in \Theta$. Since $R_\Theta(i,j) \neq \bot$, the definition of R implies that j must be an urfather. Furthermore,

$$R_\Theta(i,j) = \bigsqcup\{b \mid T@_i(b \to \Diamond k) \in \Theta, \mathbf{n}_\Theta(k) = j\}.$$

Let b and k be chosen arbitrarily such that $T@_i(b \to \Diamond k) \in \Theta$ and $\mathbf{n}_\Theta(k) = j$. Then by the $(\mathbf{T}\Diamond)$ rule, $T@_k(\psi \to (b \Rightarrow a)) \in \Theta$. Using 1 of Lemma 5 we get $T@_{\mathbf{n}_\Theta(k)}(\psi \to (b \Rightarrow a)) \in \Theta$, that is, $T@_j(\psi \to (b \Rightarrow a)) \in \Theta$. Now, by induction hypothesis, since j is an urfather,

$$v(j,\psi) \leq bound^{\Theta,j}(\psi) \leq b \Rightarrow a.$$

Since k and b were chosen arbitrarily with $T@_i(b \to \Diamond k) \in \Theta$ and $\mathbf{n}_\Theta(k) = j$, we get

$$v(j,\psi) \leq \bigsqcap\{b \Rightarrow a \mid T@_i(b \to \Diamond k) \in \Theta, \mathbf{n}_\Theta(k) = j\}.$$

We now get

$$\begin{aligned} R_\Theta(i,j) \sqcap v(j,\psi) &\leq \bigsqcup\{b \mid T@_i(b \to \Diamond k) \in \Theta, \mathbf{n}_\Theta(k) = j\} \\ &\quad \sqcap \bigsqcap\{b \Rightarrow a \mid T@_i(b \to \Diamond k) \in \Theta, \mathbf{n}_\Theta(k) = j\} \\ &\leq \bigsqcup\{b \sqcap (b \Rightarrow a) \mid T@_i(b \to \Diamond k) \in \Theta, \mathbf{n}_\Theta(k) = j\} \\ &\leq \bigsqcup\{a \mid T@_i(b \to \Diamond k) \in \Theta, \mathbf{n}_\Theta(k) = j\} \\ &\leq a. \end{aligned}$$

Because $j \in \Theta$ was arbitrary it follows that it holds for all $j \in \Theta$ and the proof of this case is completed.

In case φ is $\Box\psi$, we need to prove that

$$\bigsqcup\{a \mid T@_i(a \to \Box\psi) \in \Theta\} \leq \bigsqcap\{R_\Theta(i,j) \Rightarrow v(j,\psi) \mid j \in \Theta\} \leq \bigsqcap\{a \mid T@_i(\Box\psi \to a) \in \Theta\}.$$

To prove the first inequality we need to prove that if $j \in \Theta$, then

$$a \leq R_\Theta(i,j) \Rightarrow v(j,\psi), \tag{2}$$

for all $a \in \mathcal{T}$ with $T@_i(a \to \Box\psi) \in \Theta$. So let $a \in \mathcal{T}$ be given arbitrarily such that $T@_i(a \to \Box\psi) \in \Theta$. Note that (2) is equivalent to

$$a \sqcap R_\Theta(i,j) \leq v(j,\psi).$$

By definition of R_Θ we have

$$R_\Theta(i,j) = \bigsqcup\{b \mid T@_i(b \to \Diamond k) \in \Theta, \mathbf{n}_\Theta(k) = j\}.$$

Let b and k be chosen arbitrarily such that $T@_i(b \to \Diamond k) \in \Theta$ and $\mathbf{n}_\Theta(k) = j$. Then by the **(T□)**-rule it follows that $T@_k((a \sqcap b) \to \psi) \in \Theta$. By 1 of Lemma 5 this implies $T@_j((a \sqcap b) \to \psi) \in \Theta$. Thus we get $bound_{\Theta,j}(\psi) \geq (a \sqcap b)$. Since b and k were chosen arbitrarily with the properties $T@_i(b \to \Diamond k) \in \Theta$ and $\mathbf{n}_\Theta(k) = j$ we then get

$$bound_{\Theta,j}(\psi) \geq \bigsqcup\{a \sqcap b \mid T@_i(b \to \Diamond k) \in \Theta, \mathbf{n}_\Theta(k) = j\}.$$

Using this inequality and the induction hypothesis we now get

$$\begin{aligned} a \sqcap R_\Theta(i,j) &= a \sqcap \bigsqcup\{b \mid T@_i(b \to \Diamond k) \in \Theta, \mathbf{n}_\Theta(k) = j\} \\ &= \{a \sqcap b \mid T@_i(b \to \Diamond k) \in \Theta, \mathbf{n}_\Theta(k) = j\} \\ &\leq bound_{\Theta,j}(\psi) \leq v(j,\psi). \end{aligned}$$

Since a was arbitrary this holds for all $a \in \mathcal{T}$ and the inequality have been proven.

To show the other inequality we need to show that

$$\text{if } T@_i(\Box\psi \to a) \in \Theta \text{ then } \bigsqcap\{R_\Theta(i,j) \Rightarrow v(j,\psi) \mid j \in \Theta\} \leq a.$$

If $a = \top$ then this is trivial. Thus assume towards a contradiction that there is an $a \neq \top$ with $T@_i(\Box\psi \to a) \in \Theta$ and $\bigsqcap\{R_\Theta(i,j) \Rightarrow v(j,\psi) \mid j \in \Theta\} \nleq a$. Now let $b \leq \bigsqcap\{R_\Theta(i,j) \Rightarrow v(j,\psi) \mid j \in \Theta\}$ be a minimal member of $\mathcal{T}$ such that $b \nleq a$. Then by the reversal rule **(T≤)**, $F@_i(b \to \Box\psi) \in \Theta$. Hence by the **(F□)**-rule there is a nominal $k \in \Theta$ and a $c \in \mathcal{T}$ such that $T@_i(c \leftrightarrow \Diamond k) \in \Theta$ and $F@_k((b \sqcap c) \to \psi) \in \Theta$. From the first it follows that $R_\Theta(i, \mathbf{n}_\Theta(k)) = c$, using Lemma 8. From the second it follows that $F@_{\mathbf{n}_\Theta(k)}((b \sqcap c) \to \psi) \in \Theta$, using 1 of Lemma 5, and thus, by *(iii)* of Lemma 7, $b \sqcap c \nleq bound^{\Theta,\mathbf{n}_\Theta(k)}(\psi)$. But then from the induction hypothesis it follows that

$$b \sqcap c \nleq v(\mathbf{n}_\Theta(k), \psi) \leq bound^{\Theta,\mathbf{n}_\Theta(k)}(\psi).$$

Hence

$$b \nleq c \Rightarrow v(\mathbf{n}_\Theta(k), \psi) = R_\Theta(i, \mathbf{n}_\Theta(k)) \Rightarrow v(\mathbf{n}_\Theta(k), \psi).$$

But by the assumption on b we also have that

$$b \leq \bigsqcap\{R_\Theta(i,j) \Rightarrow v(j,\psi) \mid j \in \Theta\} \leq R_\Theta(i, \mathbf{n}_\Theta(k)) \Rightarrow v(\mathbf{n}_\Theta(k), \psi),$$

and a contradiction have been reached. This concludes the □ case and thus the entire proof of the theorem. □

Now completeness can easily be proven, in the following sense.

Theorem 3. *If there is no tableau proof of the formula φ, then there is a model $\mathcal{M} = \langle W, R, \mathbf{n}, v\rangle$ and a $w \in W$ such that $v(w,\varphi) \neq \top$.*

Proof. Assume that there is no tableau proof of the formula φ. Then there is an saturated tableau with a open branch Θ starting with the formula $F@_i(\top \to \varphi)$ for a nominal i not in φ. By item 1 of Lemma 5 it follows that also $F@_{\mathbf{n}_\Theta(i)}(\top \to \varphi) \in \Theta$.

The model $\mathcal{M}_\Theta = \langle W_\Theta, R_\Theta, \mathbf{n}_\Theta, v_\Theta \rangle$ can now be constructed such that v_Θ satisfies the assumption of Theorem 2. Since $F@_{\mathbf{n}_\Theta(i)}(\top \to \varphi) \in \Theta$ it follows by Lemma 7 that $\top \not\leq bound^{\Theta,\mathbf{n}_\Theta(i)}(\varphi)$. But by Theorem 2, since φ is a subformula of the root formula and $\mathbf{n}_\Theta(i)$ is an urfather, we know that $v_\Theta(\mathbf{n}_\Theta(i), \varphi) \leq bound^{\Theta,\mathbf{n}_\Theta(i)}(\varphi)$ and it thus follows that $\top \not\leq v_\Theta(\mathbf{n}_\Theta(i), \varphi)$ and the proof is completed. □

Acknowledgements Thanks to the anonymous reviewers for useful comments and corrections. The second and third authors are partially funded by the Danish Natural Science Research Council (the HyLoMOL project).

References

P. Blackburn. Internalizing labelled deduction. *Journal of Logic and Computation*, 10:137–168, 2000.

P. Blackburn and M. Marx. Tableaux for quantified hybrid logic. In U. Egly and C. Fermüller, editors, *Automated Reasoning with Analytic Tableaux and Related Methods, TABLEAUX 2002, volume 2381 of Lecture Notes in Artificial Intelligence*, pages 38–52. Springer-Verlag, 2002.

T. Bolander and P. Blackburn. Termination for hybrid tableaus. *Journal of Logic and Computation*, 17(3):517–554, 2007a.

T. Bolander and P. Blackburn. Terminating tableau calculi for hybrid logics extending K. In C. Areces and S. Demri, editors, *Workshop Proceedings of Methods for Modalities 5*, pages 157–175. École Normale Supérieure de Cachan, France, 2007b.

T. Bolander and T. Braüner. Two tableau-based decision procedures for hybrid logic. In H. Schlingloff, editor, *4th Workshop "Methods for Modalities" (M4M), Informatik-Bericht Nr. 194*, pages 79–96. Humboldt-Universität zu Berlin, 2005.

T. Bolander and T. Braüner. Tableau-based decision procedures for hybrid logic. *Journal of Logic and Computation*, 16:737–763, 2006. Revised and extended version of Bolander and Braüner (2005).

T. Braüner. *Hybrid Logic and Its Proof-Theory*. To appear, 2008.

T. Braüner and V. de Paiva. Towards constructive hybrid logic (extended abstract). In C. Areces and P. Blackburn, editors, *Workshop Proceedings of Methods for Modalities 3*, page 15 pages. 2003.

T. Braüner and V. de Paiva. Intuitionistic hybrid logic. *Journal of Applied Logic*, 4: 231–255, 2006. Revised and extended version of Braüner and de Paiva (2003).

W. B. Ewald. Intuitionistic tense and modal logic. *Journal of Symbolic Logic*, 51: 166–179, 1986.

M. Fitting. Imany-valued modal logics. *Fundamenta Informaticae*, 15:235–254, 1992a.

M. Fitting. Many-valued modal logics, II. *Fundamenta Informaticae*, 17:55–73, 1992b.

M. Fitting. Tableaus for many-valued modal logic. *Studia Logica*, 55:63–87, 1995.

M. Fitting. *Proof methods for modal and intuitionistic logics, volume 169 of* Synthese Library. D. Reidel Publishing Co., Dordrecht, 1983.

J. Hansen. A tableau system for a first-order hybrid logic. In J. Villadsen, T. Bolander, and T. Braüner, editors, *Proceedings of the International Workshop on Hybrid Logic 2007(19th European Summerschool in Logic, Language and Information)*, pages 32–40. 2007.

A. Troelstra and D. van Dalen. *Constructivism in Mathematics: An Introduction*. North-Holland, 1988.

Dynamic Epistemic Logic and Graded Modal Logic

Minghui Ma
The School of Humanities and Social Sciences, Department of Philosophy, Tsinghua University.
mmh05@mails.tsinghua.edu.cn

Abstract

Counting information states in the process of information flow is treated in this paper. Both public announcement logic, and dynamic epistemic logic of production update over event models, are merged with graded modalities which play the role of counting. Recursion axioms are given. And basic model theory of these logics is investigated. The dynamic logic based on epistemic graded S5 modal logic is also explored.

1 Introduction

In the logical dynamics of information flow, some modes of information update are specified in terms of precise logical notions. The public announcement of a proposition ϕ plays the role of moving away all those information states where ϕ is not true. It often happens that the agent knows ϕ after the public announcement of ϕ.[1] One general approach to explore logical dynamics of information flow is presented in the book van Ditmarsch et al. (2007).[2] Baltag et al. (1998) and van Benthem (2008) present another approach to explore dynamic epistemic logic (*DEL*). The technical point of the approach is to combine epistemic models with event models, and determine how to interpret dynamic operators in a mathematically precise way. Dynamics in action is analyzed in terms of some logical notions while the static epistemic logic is inherited. In this paper, logical dynamics are investigated in the following two lines.

- Add graded modalities into the basic epistemic multi-S5 modal logic, and work out the recursion axioms for graded modalities prefixed by dynamic operators.
- Grade the knowledge operator, and take the epistemic graded S5 modal logic as the basic static part of logical dynamics.

[1] See Plaza (1998) and van Benthem (2001) for details of public announcement logic.

[2] It is a text book on dynamics of knowledge and belief. The approach is to explore logical dynamics by propositional dynamic logic.

In the first line, one special ingredient is revealed in both statics and dynamics of information flow: counting information states. Statically, epistemic logic can be combined with counting modalities in such a way that the number of information states in the current range of agent's knowledge can be made explicit. Dynamically, counting behaviors substantially paly a important role in agent's actions. Moreover, in the process of information update, numbers of information states in different stages change dramatically. There may be less states after occurrences of some events. A typical example is public announcement that reduces information states. But this is not always the case. There may be more states after some occurrences of a certain event in certain scenarios. And finally, actions in certain scenarios even may not change the number of information states.

In the second line, we have graded knowledge operators like K_n which says that the agent reckons less than n exceptions. Under the new notions of knowledge, logical dynamics may be reconstructed. In van der Hoek and Meyer (1992), graded epistemic logic has been greatly explored, but no dynamics are considered there.

2 Counting information states

In this section, three examples are taken to show that the number of information states may decrease, increase, or keep stable in the information flows, and that one may reason by counting information states in certain scenarios. All of those examples can be found in van Benthem (2008).

2.1 Semantic information in logical consequence

The traditional slogan about the nature of logic that claimed by most logicians is the following: *Logic studies valid forms of inference.* The conclusion followed from premises does not increase the content or information. This problem has been called *the scandal of deduction*. Given a set of sentences Σ, if $\Sigma \models \phi$ then the semantic content of ϕ or the information conveyed by ϕ is already in the premises Σ. But we certainly obtain something new when we conclude ϕ semantically from Σ by logical consequence. What kind of information is involved in the process of logical inference? The paper van Benthem and Martinez (2008) has discussed some solutions. One answer is *information as code*. The conclusion ϕ is logically encoded in the premises, and the role of inference is to make the conclusion explicit. From the view point of logical dynamics, the conclusion may be achieved by public announcements in some finite steps. For instance, consider the logical consequence relation $\{p \to q, p\} \vDash q$. There are 4 semantic information states or truth-value possibilities: $pq, p\overline{q}, \overline{p}q, \overline{pq}$, where a bar stands for the truth value **FALSE** and the lack of a bar means the truth value **TRUE**.

- After $!(p \to q)$, the state $p\overline{q}$ disappear, and three possibilities remain.
- After $!p$, states $\overline{p}q$ and $\overline{pq}$ are moved away. Only does the possibility pq remain. Here we achieve the conclusion q.

- After arbitrary finite many times of $!q$, the final state pq cannot be removed, and the number of possibilities does not change.

This process matches exactly with the notion of logical consequence: *if the premises are true then the conclusion is also true.* The public announcements match with the assumption that all premises are true, and the conclusion is achieved in the final stable stage.

In each stage, we get a range of information states. Then the question arises: how many states are there in each range? The initial stage has 4 states; after $!(p \to q)$, the range has 3 states; after $!p$, the range has only one state. By using some logical language, we may express those numerical or quantitative facts about those ranges.

2.2 Increasing model size by production update

The production of an epistemic model and an event model can increase information states. Consider the following simple model M: M consists of two states $\bullet$ and $\circ$, where p is true at $\bullet$ and false at $\circ$, and both two agents 1 and 2 cannot distinguish the two states.

Take an event model for the scenario that agent 1 hears $!p$ but agent 2 doubts very much about what has happened. Then we have the following event model $\mathcal{E} = (E, \sim_i, Pre)_{i\in\{1,2\}}$: $E = \{!p, Id\}$ where Id is the trivial event that can happen everywhere; $\sim_1 = \sim_2 = \{(!p, Id), (Id, !p)\}$; $Pre(!p) = p$ and $Pre(Id) = \top$.

Then how many states are there in the product model $M \otimes \mathcal{E}$? Three states: $(\bullet, !p)$, $(\bullet, Id)$, and $(\circ, Id)$. After the event $!p$, agent one can distinguish $(\bullet, !p)$ from other two states, but he still cannot distinguish $(\bullet, Id)$ and $(\circ, Id)$ since no informational events happen there. Agent 2 cannot distinguish all of those three states.

2.3 Reasoning by counting

Imagine the following scenario. Three persons A, B, and C go to a restaurant for dinner. A has ordered fish, B has meat and C has vegetable. Then a new person D comes out of the kitchen with three plates. First of all, D asks: who has the meat?. He gets the answer and puts the plate on the table. And then he asks: who has the fish? He puts the plate on the table again after he hears the answer. Finally, D puts the plate on the table without asking any question! D makes a logical inference here. There are exactly 6 possibilities before he asks any question. After the first answer, D counts the possibilities and finds that there are remaining 2 possibilities. Then he should ask a question. After the second answer, D knows that there are only one possibility left.

2.4 What's the role of counting in action?

Counting information states in action plays two roles:

- It gives us the quantitative knowledge about the current stage in the process of information flow while dynamic operations only provide us with qualitative knowledge of the information update;
- Agents often take actions on counting information states.

Counting is involved in the statics of logic, and we can introduce appropriate operators to talk about numbers of information states in different ranges. Thus the quantitative aspect of agents' epistemic ranges can be described. But observations about numbers of states cannot be described by using only the basic epistemic modal language. The expressive power of standard modal logic has limits. Simple numerical statement cannot be expressed in it. The property that there are two successor states constitutes such a counter example. This property is not invariant for standard bisimulation. On the other hand, counting is involved in the logical dynamics. Note again that the number of information states may change after some certain events. The interaction between counting modalities and dynamic operators emerge in the logical investigations. Moreover, we also glimpse at the role of counting in logical inference in action.

The crucial logical matter here is to find out appropriate formal language to express numerical facts in logical dynamics of information flow. Graded modalities serves the purpose of counting information states. The technical task is to interpret the role of counting modalities prefixed by dynamic operators.

3 Merge graded modal logic and epistemic logic

In this section, we explain graded modal logic (*GML*) and epistemic logic (*EL*) separately, and then merge them together. Let's start from *GML*. The decidable language $\Omega = \{S, 0, +\}$ of Presburger additive arithmetic is needed to form the graded modal language, where S is the successor function symbol, 0 is a constant which designate the natural number **0**, and + is the addition symbol. Ω-terms are given by the following rule:

$$\overline{n} := 0 \mid S\overline{n} \mid \overline{n} + \overline{m}$$

All of the Ω-terms are interpreted in the arithmetic structure $(\omega, \mathbf{0}, S, +)$. In the following, we use number n as the designation of the term $\overline{n}$, $n+1$ the designation of the term $S\overline{n}$, and $m+n$ the designation of the term $\overline{m} + \overline{n}$.

The graded modal language $ML(\tau, \Phi)$ consists of graded modal similarity $\tau = \{\langle\overline{n}\rangle : \overline{n} \in Term(\Omega) \;\&\; \langle\overline{n}\rangle \; unary\}$ and a set Φ of propositional letters. Graded modal formulas are given by the following rule:

$$\phi ::= p \mid \bot \mid \phi \to \psi \mid \langle\overline{n}\rangle\phi$$

Define some more operators as follows: $\Diamond\phi := \langle\overline{1}\rangle\phi$, $\Box\phi := \neg\Diamond\neg\phi$, $[\overline{n}]\phi := \neg\langle\overline{n}\rangle\neg\phi$, and $\langle\overline{n}\rangle!\phi := \langle\overline{n}\rangle\phi \wedge \neg\langle S\overline{n}\rangle\phi$.

In a Kripke model $M = (W, R, V)$, where $w \in W$, the following notations are used throughout this paper: let $R[w] := \{u \in W : Rwu\}$; Let $Card(X)$ denote the cardinality of a set $X \subseteq W$; and let $V(\phi) := \{u \in W : \phi \text{ is true at } u \text{ in } M\}$.

Definition 3.1. Given a model $M = (W, R, V)$ and $w \in W$, define truth $M, w \models \phi$ recursively as follows:

$M, w \models p$ iff $w \in V(p)$;

$M, w \not\models \bot$;

$M, w \models \phi \to \psi$ iff $M, w \models \phi$ implies $M, w \models \psi$;

$M, w \models \langle\overline{n}\rangle\phi$ iff $Card(R[w] \cap V(\phi)) \geq n$.

The formula $[\overline{n}]\phi$ says that there are less than n successor states where ϕ is not true. The formula $\langle\overline{n}\rangle!\top$ is true at state w in a model iff w has exactly n successor states in that model.

Theorem 1. *(Fattorosi-Barnaba and Cerrato (1988)) The minimal normal graded modal logic GK is completely axiomatized by the following schemata and inferential rules:*
(a) All instances of propositional tautologies;
(b) $\langle S\overline{n}\rangle\phi \rightarrow \langle\overline{n}\rangle\phi$;
(c) $\Box(\phi \rightarrow \psi) \rightarrow (\langle\overline{n}\rangle\phi \rightarrow \langle\overline{n}\rangle\psi)$;
(d) $\neg\Diamond(\phi \wedge \psi) \rightarrow (\langle\overline{m}\rangle!\phi \wedge \langle\overline{n}\rangle!\psi \rightarrow \langle\overline{m}+\overline{n}\rangle!(\phi \vee \psi))$;
MP: from ϕ *and* $\phi \rightarrow \psi$ *infer* ψ;
Gen: from ϕ *infer* $\Box\phi$.

The basic epistemic logic *EL* is just multi-S5 modal logic. When graded modalities are introduced into *EL*, we should combine *EL* with *GML*. Extend *GK* with the following two axiomatic schemata and we obtain the graded modal logic *GS*5: (T) $\phi \rightarrow \Diamond\phi$; (E) $\Diamond[\overline{n}]\phi \rightarrow [\overline{n}]\phi$. Note that the set of formulas $\{\Diamond[\overline{n}]\phi \rightarrow [\overline{n}]\phi : n \in \omega\}$ corresponds to the Euclidean property on frames: $\forall x\forall y\forall z(xRy \wedge xRz \rightarrow yRz)$.

Theorem 2. *(Cerrato (1990)) GS5 is complete with respect to the class* Ξ *of frames with equivalence accessibility relations.*

Proof. Let *MCS* be the set of all maximal *GS*5-consistent sets of formulas. Define two functions $\mu : MCS \times MCS \rightarrow \omega + 1$ and $\delta : MCS \rightarrow \omega + 1$ as follows:

- $\mu(u, v) = \omega$, if $\langle\overline{n}\rangle\phi \in u$ for all ϕ in v for all $n \in \omega$.
  $\mu(u, v) = min\{n \in \omega : \langle\overline{n}\rangle!\phi \in u \ for\ some\ \phi \in v\}$, otherwise.
- $\delta(u) = sup\{\mu(w, u) : w \in MCS\}$.

Let $c(u) = \{u\} \times \omega$. Define the canonical model $M = (W, R, V)$ for *GS*5 as follows: consider an ordering of type $\omega \otimes \mu(u, u)$ on $c(u)$, and let

- $W = \bigcup_{u \in MCS} c(u)$;
- if $\mu(u, v) \neq 0$ then $(u_{ni}, v_{nj}) \in R$ for each $n \in \omega$, $i \in \mu(u, u)$, $j \in \mu(v, v)$.
- if $\mu(u, v) = 0$, $(c(u) \times c(u)) \cap R = \emptyset$.

Note that$\mu(u, u) \neq 0$ for each $u \in MCS$. □

Model theory for *GML* has not yet been greatly developed.[3] Completeness theorems like above are important results. The point of the construction of canonical models is that it provides a general method to find enough states in a model. Another important model-theoretic result is the notion of graded bisimulation given in de Rijke (2000). Standard definition of bisimulation for basic modal logic doesn't fit for *GML*. With the new notion of graded bisimulation, *GML* is isomorphic to first order logic with identity module graded bisimulation. Thus graded modal model theory may be explored much more. Here we present another notion of bisimulation.

[3]For details of model theory of standard modal logic, see Blackburn et al. (2001) and de Rijke (1995). Many results may be extended to *GML*.

Definition 3.2. A graded bisimulation Z between two models $M = (W, R, V)$ and $M' = (W', R', V')$ (written: $Z : M \rightleftarrows M'$)) is a non-empty relations $Z \subseteq W \times W'$ satisfying the following conditions:

- (atomic) If wZw' then w and w' satisfy the same propositional letters.
- For each number $k \in \omega$, if wZw' then the following forth and back conditions hold:

  (1) (forth) if $\{u_1, \ldots, u_k\} \subseteq R[w]$ then there are k-different points $v_1, \ldots, v_k$ in W' such that $\{v_1, \ldots, v_k\} \subseteq R'[w']$ and $u_i Z v_i$ for all $1 \leq i \leq k$.

  (2) (back) if $\{v_1, \ldots, v_k\} \subseteq R'[w']$ then there are k-different points $u_1, \ldots, u_k$ in W such that $\{u_1, \ldots, u_k\} \subseteq R[w]$ and $u_i Z v_i$ for all $1 \leq i \leq k$.

We write $M, w \rightleftarrows M', w'$, if there exists a graded bisimulation relation $Z : M \rightleftarrows M'$ with wZw'.

The above definition is more closed to the ordinary notion of bisimulation although it is equivalent to de Rijke's notion which is defined on finite subsets. All graded formulas are invariant for graded bisimulation. With this observation, some model-theoretic definability results on certain classes of pointed models are presented in de Rijke (2000).

Next we combine epistemic logic with graded modal logic. Here is a problem: shall we grade the basic EL? How can we interpret epistemic formulas with graded knowledge operators? This problem is a little bit complex, and we detect it in section 6. Another more conservative approach is to keep EL, add graded modalities, and make the fusion of EL and $GS5$.

Definition 3.3. Fix a set of agents I. The language for EL with graded modalities has propositional letters and boolean operators, plus knowledge operators $K_i (i \in I)$ and graded modalities $\langle \overline{n} \rangle$. The inductive syntactic rule is the following:

$$\phi ::= p \mid \neg\phi \mid \phi \vee \psi \mid K_i\phi \mid \langle \overline{n} \rangle\phi$$

Both EL and $GS5$ are complete with respect to the class Ξ of all frames with equivalence accessibility relations. Then the following technical question arises: is $EL \oplus GS5$ (the minimal graded modal logic with laws of EL and $GS5$) complete with respect to Ξ? Here we go to the fusion of modal logics. Fortunately, the paper Kracht and Wolter (1991) has the following theorem: two consistent disjoint logics L and M are complete if and only if $L \oplus M$ is complete. With this elegant results, Kurucz (2007) shows the following theorem: if modal logics L_1 and L_2 are characterized by classes of frames C_1 and C_2 respectively, and if C_1 and C_2 are closed under the formation of *disjoint union* and *isomorphic copies*, then the fusion $L_1 \oplus L_2$ is characterized by the class $C_1 \oplus C_2 = \{(W, R_1, \ldots, R_n, O_1, \ldots, O_m) : (W, R_1, \ldots, R_n) \in C_1 \textit{ and } (W, O_1, \ldots, O_m) \in C_2\}$. Then we have the following theorem for $EL \oplus GS5$.

Theorem 3. *$EL \oplus GS5$ is complete with respect to Ξ.*

We have obtained a nice basis for the logical dynamics of information flow here. The fusion logic has the power to express numerical facts about models. Moreover, dynamic epistemic logics over the fusion may also prove to be complete by using reduction and the above completeness result. In next section, the role of graded modalities will be presented in the combination of a very simple dynamic epistemic logic and graded modal logic: the public announcement logic (*PAL*) with graded modalities. The basic static part of *PAL* with graded modalities is the fusion $EL \oplus GS5$.

4 PAL with graded modalities

The language for *PAL* with graded modalities (*GPAL*) has all primitive symbols of epistemic language with graded modalities plus public announcement operators. The formulas of *GPAL* are given by the following inductive rule:

$$\phi ::= p \mid \neg\phi \mid \phi \vee \psi \mid K_i\phi \mid \langle\overline{n}\rangle\phi \mid \langle !\phi\rangle\psi$$

Define $[!\phi]\psi := \neg\langle !\phi\rangle\neg\psi$.

The language is interpreted as *EL* and *GML*, and the semantic clause for $\langle !\phi\rangle\psi$ is defined as follows:

$$M, w \models \langle !\phi\rangle\psi \quad \text{if and only if} \quad M, w \models \phi \quad \text{and} \quad M|\phi, w \models \psi$$

where $M|\phi$ is the submodel of M produced by set $\{u \in dom(M) : M, u \models \phi\}$.

To find a complete axiomatization of *GPAL* needs to analyze the dynamic recursion equation in the process of information flow. Those axioms put the outer public announcement operators into the scope of modal operators, and the completeness for the new logic reduces to the basic part.

Theorem 4. *The recursion axiom* $\langle !\phi\rangle\langle\overline{n}\rangle\psi \leftrightarrow \phi \wedge \langle\overline{n}\rangle\langle !\phi\rangle\psi$ *is valid.*

Proof. Assume that $M, w \models \langle !\phi\rangle\langle\overline{n}\rangle\psi$. Then $M, w \models \phi$ and $M|\phi, w \models \langle\overline{n}\rangle\psi$. There exists $u_1, \ldots, u_n$ in $M|\phi$ with $M, u_i \models \phi$ and $M|\phi, u_i \models \psi$ for $1 \leq i \leq n$. Hence there are n successor states of w in M where $\langle !\phi\rangle\psi$ is true. Conversely, assume that $M, w \models \phi \wedge \langle\overline{n}\rangle\langle !\phi\rangle\psi$. Then $M, w \models \phi$ and there are n successor states of w in M where $\langle !\phi\rangle\psi$ is true. Hence all of those successors are in $M|\phi$. □

Theorem 5. *GPAL is completely axiomatized by the axiomatic schemata of* $EL \oplus GS5$ *plus the following recursion axioms:*
(a) $\langle !\phi\rangle p \leftrightarrow \phi \wedge p$;
(b) $\langle !\phi\rangle\neg\psi \leftrightarrow \phi \wedge \neg\langle !\phi\rangle\psi$;
(c) $\langle !\phi\rangle(\psi \vee \chi) \leftrightarrow \langle !\phi\rangle\psi \vee \langle !\phi\rangle\chi$;
(d) $\langle !\phi\rangle K_i\psi \leftrightarrow \phi \wedge K_i\langle !\phi\rangle\psi$;
(e) $\langle !\phi\rangle\langle\overline{n}\rangle\psi \leftrightarrow \phi \wedge \langle\overline{n}\rangle\langle !\phi\rangle\psi$;
(f) $\langle !\phi\rangle\langle !\psi\rangle\chi \leftrightarrow \langle !\langle !\phi\rangle\psi\rangle\chi$.

In the following, one model-theoretic aspect of *GPAL* is explored. By an informative observation on the change of models under public announcement update, those public announcement operators $\langle !\phi\rangle$ induce operations on models. An operation # on models is called to *respect graded bisimulation*, if $\#(M, w) \leftrightarrows \#(M', w')$ whenever $M, w \leftrightarrows M', w'$ for any two models (M, w) and (M', w').

Theorem 6. *All public announcement updates* $|\phi$ *over models respect graded bisimulation.*

Proof. Assume that $M, w \rightleftarrows M', w'$. Then there exists a graded bisimulation relation Z with wZw'. We need to show $(M|\phi, w) \leftrightarrows (M'|\phi, w')$. The point is that the relation sequence $Z' = Z \cap (dom(M|\phi)) \times (dom(M'|\phi))$ is a graded bisimulation between $(M|\phi, w)$ and $(M'|\phi, w')$. We only prove the forth condition. Assume that $\{x_1 \ldots, x_k\} \subseteq R[w]$, then, for each $1 \leq i \leq k$, Rwx_i and $M, x_i \models \phi$. Thus, by using graded bisimulation Z, we find $y_1, \ldots, y_k \in dom(M')$, and by bisimulation invariance, $M', y_i \models \phi$ for each $1 \leq i \leq k$. Hence $\{y_1, \ldots, y_k\} \subseteq (dom(M'|\phi))$. □

Theorem 7. *All formulas of GPAL are invariant for graded bisimulation.*

Proof. Assume that $Z : M, w \leftrightarrows M', w'$. It suffices to show that $M, w \models \psi$ if and only if $M', w' \models \psi$ for any formula ψ. Only prove it for the case $\psi := \langle !\phi \rangle \chi$. Assume that $M, w \models \langle !\phi \rangle \chi$. Then $M, w \models \phi$ and $M|\phi, w \models \chi$. By theorem 6, let $Z' : M|\phi, w \leftrightarrows M'|\phi, w'$. Thus, by inductive hypothesis, $M'|\phi, w' \models \chi$, and $M', w' \models \phi$. Hence $M', w' \models \langle !\phi \rangle \chi$. The other direction is also similar. □

5 DEL with graded modalities

A natural generalization of *PAL* is to input more events other than only a single public announcement in each update. The resulting system *DEL* offers a sharp view of update in epistemic actions. Given more events, agents may or may not distinguish occurrences of different events. Thus event models are introduced and the production of epistemic models and event models is considered. In this section, we explore *DEL* with graded modalities (*GDEL*) on the basis of $EL \oplus GS5$.

Definition 5.1. An event model is a structure $\mathcal{E} = (E, \sim_i, Pre)$ with a set of events E, epistemic uncertainty relation $\sim_i$ for each agent i, and a function Pre from E to formulas in basic epistemic language with graded modalities, which give the precondition for the occurrence of a event at a information state.

The restriction of preconditions of events to basic epistemic formulas is just a tick for convenience. An immediate corollary of this restriction is that all preconditions are invariant for graded bisimulation.

Definition 5.2. For any epistemic model $M = (W, \sim_i, V)$ and event model $\mathcal{E} = (E, \sim_i, Pre)$, the product model $M \otimes \mathcal{E} = (W^+, \sim_i, V^+)$ is an epistemic model defined as follows: $W^+ = \{(w, e) : M, w \models Pre(e)\}$; $(w, e) \sim_i (v, f)$ iff $w \sim_i v$ and $e \sim_i f$; $V^+(p) = \{(w, e) : w \in V(p)\}$.

DEL differs from *PAL* in two features: (1) *DEL* uses event structures to form syntactic operators while *PAL* only uses individual events; (2) product models are used to interpret formulas with event structure. Now, there is a natural question in developing dynamic epistemic logic with graded modalities: how can we interpret graded modalities properly in the product models? A formula $\langle \overline{n} \rangle \phi$ true in a product model requires that there exist at least n different pairs where ϕ is true. In set theory, $(w, e) \neq (v, f)$ iff $w \neq v$ or $e \neq f$. The complex situation about n different pairs arises here.

Definition 5.3. The *GDEL* has the following inductive syntactic rule:

$$\phi ::= p \mid \neg\phi \mid \phi \vee \psi \mid K_i\phi \mid \langle \overline{n} \rangle \phi \mid \langle \mathcal{E}, e \rangle \phi$$

where $(\mathcal{E}, e)$ is an event model. The semantic clauses for *GDEL* are those clauses for *GPAL* except for the dynamic operators. Given any model (M, w), define $M, w \models \langle \mathcal{E}, e\rangle\phi$ iff $M, w \models Pre(e)$ and $M \otimes \mathcal{E}, (w, e) \models \phi$.

The complete axiomatization for *GDEL* will be achieved by using recursion axioms. The recursion axioms are just like those for *GPAL* except the recursion axiom for $\langle \mathcal{E}, e\rangle\phi$. First, we see some special cases.

Case 1. $n = 0$. We have $\langle 0\rangle\phi \leftrightarrow \top$. Hence $\langle \mathcal{E}, e\rangle\langle 0\rangle\phi \leftrightarrow \langle \mathcal{E}, e\rangle\top$. Thus $\langle \mathcal{E}, e\rangle\langle 0\rangle\phi \leftrightarrow Pre(e)$.

Case 2. $n = 1$. We have $\langle S0\rangle\phi \leftrightarrow \Diamond\phi$. Hence $\langle \mathcal{E}, e\rangle\langle S0\rangle\phi \leftrightarrow \langle \mathcal{E}, e\rangle\Diamond\phi$. In this case, $\langle \mathcal{E}, e\rangle\Diamond\phi \leftrightarrow Pre(e) \wedge \Diamond \bigvee_{e\sim f}\langle \mathcal{E}, f\rangle\phi$. This recursion axiom is valid. For any model (M, w), assume that $M, w \models \langle \mathcal{E}, e\rangle\Diamond\phi$. Then $M, w \models Pre(e)$ and $M \otimes \mathcal{E}, (w, e) \models \Diamond\phi$. Thus there exists (v, f) in $M \otimes \mathcal{E}$ such that $(w, e) \sim (v, f)$ and $M \otimes \mathcal{E}, (v, f) \models \phi$. Conversely, assume that $M, w \models Pre(e)$ and there exists v in M and f in $\mathcal{E}$ such that $w \sim v$ and $e \sim f$ and $M, v \models Pre(f) \wedge \langle \mathcal{E}, f\rangle\phi$. Thus $M \otimes \mathcal{E}, (v, f) \models \phi$. Hence $M, w \models \langle \mathcal{E}, e\rangle\Diamond\phi$.

Case 3. $n = 2$. We have the following recursion axiom:

$$\langle \mathcal{E}, e\rangle\langle \overline{2}\rangle\phi \leftrightarrow Pre(e) \wedge (\Diamond \bigvee_{e\sim f_1, e\sim f_2, f_1\neq f_2} \bigwedge_{i\in\{1,2\}} \langle \mathcal{E}, f_i\rangle\phi \vee \langle \overline{2}\rangle \bigvee_{e\sim g}\langle \mathcal{E}, g\rangle\phi)$$

Claim. The above recursion axiom is valid.

Proof. Let $\xi := \Diamond \bigvee_{e\sim f_1, e\sim f_2, f_1\neq f_2} \bigwedge_{i\in\{1,2\}}\langle \mathcal{E}, f_i\rangle\phi$, and $\zeta := \langle \overline{2}\rangle \bigvee_{e\sim g}\langle \mathcal{E}, g\rangle\phi$. Assume that $M, w \models \langle \mathcal{E}, e\rangle\langle \overline{2}\rangle\phi$. Then $M, w \models Pre(e)$ and $M \otimes \mathcal{E}, (w, e) \models \langle \overline{2}\rangle\phi$. Thus there exists $(v_1, f_1) \neq (v_2, f_2)$ with $(w, e) \sim (v_i, f_i)$ and $M \otimes \mathcal{E}, (v_i, f_i) \models \phi$ for $i \in \{1, 2\}$. If $v_1 = v_2 = v$, then there exists v in M with $w \sim v$, and $M, v \models \bigwedge_{i\in\{1,2\}}\langle \mathcal{E}, f_i\rangle\phi$. If $v_1 \neq v_2$, then $M, w \models \langle \overline{2}\rangle \bigvee_{e\sim g}\langle \mathcal{E}, g\rangle\phi$. Conversely, assume that $M, w \models Pre(e) \wedge (\xi \vee \zeta)$. There are two cases. In the case $M, w \models \xi$, there are v, f_1 and f_2 with $e \sim f_1$ and $e \sim f_2$ and $f_1 \neq f_2$ such that $M, v \models \langle \mathcal{E}, f_i\rangle\phi$ for $i \in \{1, 2\}$. Thus $M, v \models Pre(f_i)$ and $M \otimes \mathcal{E}, (v, f_i) \models \phi$ for $i \in \{1, 2\}$. Hence $M, w \models \langle \mathcal{E}, e\rangle\langle \overline{2}\rangle\phi$. In the case $M, w \models \zeta$, there are $v_1 \neq v_2$ in M such that $w \sim v_i$ and $M, v_i \models \bigvee_{e\sim g}\phi$ for $i \in \{1, 2\}$. Let $M, v_1 \models \langle \mathcal{E}, g_1\rangle\phi$ and $M, v_2 \models \langle \mathcal{E}, g_2\rangle\phi$. Thus $M, v_1 \models Pre(g_1)$, $M, v_2 \models Pre(g_2)$, $(w, e) \sim (v_1, g_1)$, $(w, e) \sim (v_2, g_2)$, $M \otimes \mathcal{E}, (w, e) \models \langle \overline{2}\rangle\phi$. □

It can be observed that we use a trick here: fix the event model $\langle \mathcal{E}, e\rangle$ and consider different states in the epistemic model. The conditions on events can be changed for our purpose. We also have the following general recursion axiom for $\langle \overline{n}\rangle\phi$.

Theorem 8. *The following recursion axiom is valid:*

$$\begin{aligned}
&\langle \mathcal{E}, e\rangle\langle \overline{n}\rangle\phi \leftrightarrow Pre(e) \wedge [\Diamond \textstyle\bigvee_{e\sim f_1,\ldots,e\sim f_n, f_i\neq f_j \ for\ 1\leq i\leq n} \bigwedge_{1\leq i\leq n}\langle \mathcal{E}, f_i\rangle\phi \\
&\vee(\Diamond \textstyle\bigvee_{e\sim f_1,\ldots,e\sim f_{n-1}, f_i\neq f_j \ for\ 1\leq i<j<n} \bigwedge_{0<i<n}\langle \mathcal{E}, f_i\rangle\phi \wedge \langle \overline{2}\rangle \bigvee_{e\sim g}\langle \mathcal{E}, g\rangle\phi) \\
&\ldots \\
&\vee(\Diamond \textstyle\bigvee_{e\sim f_1, e\sim f_2, f_1\neq f_2} \bigwedge_{i\in\{1,2\}}\langle \mathcal{E}, f_i\rangle\phi \wedge \langle \overline{n-1}\rangle \bigvee_{e\sim g}\langle \mathcal{E}, g\rangle\phi) \\
&\vee\langle \overline{n}\rangle \textstyle\bigvee_{e\sim g}\langle \mathcal{E}, g\rangle\phi].
\end{aligned}$$

Theorem 9. *GDEL is completely axiomatized by axiom schemata of* $EL \oplus GS5$ *plus the recursion axiom for* $\langle \mathcal{E}, e\rangle\langle \overline{n}\rangle\phi$ *in theorem 8 and the following axioms:*

(a) $\langle \mathcal{E}, e\rangle p \leftrightarrow Pre(e) \wedge p$;

(b) $\langle \mathcal{E}, e\rangle \neg\phi \leftrightarrow Pre(e) \wedge \neg\langle \mathcal{E}, e\rangle\phi$;
(c) $\langle \mathcal{E}, e\rangle(\phi \vee \psi) \leftrightarrow \langle \mathcal{E}, e\rangle\phi \vee \langle \mathcal{E}, e\rangle\psi$;
(d) $\langle \mathcal{E}, e\rangle K_i\phi \leftrightarrow Pre(e) \wedge \bigwedge_{e\sim f} K_i\langle \mathcal{E}, f\rangle\phi$.

Just like what we do for *GPAL*, we look further at the graded bisimulation invariance for *GDEL*-formulas. One proof of bisimulation invariance is by using recursion axioms to reduce *GDEL*-formulas to the basic epistemic formulas with graded modalities. But here we speculate on the update product.

Theorem 10. *The production update $\otimes$ respects graded bisimulation.*

Proof. Assume that $Z : M, w \leftrightarrows M', w'$ is a graded bisimulation. Given any event model $(\mathcal{E}, e)$, we show that $M \otimes \mathcal{E}, (w, e) \leftrightarrows M' \otimes \mathcal{E}, (w', e)$. Define Z' as follows: (1) $(w, e)Z(w', e)$; (2) for $X = \{(w_1, e_1), \ldots, (w_k, e_k)\}$, $M, w_i \models Pre(e_i)$ for $1 \leq i \leq k$. By the graded bisimulation Z, select a set $\{v_1, \ldots, v_k\}$ such that $M', v_i \models Pre(e_i)$ for $1 \leq i \leq k$. Let $Y = \{(v_1, e_1), \ldots, (v_k, e_k)\}$, and $(w_i, e_i)Z(v_i, e_i)$ for all $1 \leq i \leq k$; (3) conversely, also select states by using bisimulation relation Z. It is easy to check that $Z' : M \otimes \mathcal{E}, (w, e) \leftrightarrows M' \otimes \mathcal{E}, (w', e)$. □

Theorem 11. *All formulas in GDEL are invariant for graded bisimulation.*

6 Dynamic graded epistemic logic

The epistemic logic usually used above and in most literatures is the standard multi-S5 modal epistemic logic. The road we take in above two sections is to enrich it with graded modalities. This road doesn't seem to catch the genuine meaning of *grading*. By graded necessity operators, we denote different sorts of necessity. The difference between operators $\Box_i$ and $\Box_j$ is indicated by numbers of successors states that needed to compute the truth values of formulas. Knowledge operators also can be graded in the similar way. In this section, we explore the logical dynamics based on graded epistemic logic which deal with knowledge that is not absolutely true in all worlds but may have exceptions in some possible worlds.

The Plato-formula "*Knowledge = Justified True Belief*" may allow uncertainties in our knowledge. Put aside the story about true beliefs here. Knowledge of a proposition ϕ also depends on the justification of ϕ. Even we have very strong justification for a proposition, the knowledge may have uncertainties. One example is the following:

- The degrees to which a proposition is verified also shows the uncertainties. The notion of *verification* has been widely discussed by logical empiricists. The proposition *Crowns are black* will become knowledge as more and more black crowns are discovered. But the knowledge of this proposition does not exclude exceptional cases. Actually, the knowledge became that *most crowns are black* when white crowns were discovered in Australia.

With graded modalities, notions like *uncertain*, *almost true* knowledge or belief can be treated appropriately in principle. Those modalities can be used to reason with degrees of acceptance. Now, the graded epistemic logic is just the logic *GS*5. We use K_n and its dual M_n to replace $[\overline{n}]$ and $\langle \overline{n} \rangle$ respectively.

$K_n\phi$ is true at a state w in a model M iff there are less than n accessible states of w where ϕ is false, namely, the agent reckons with less than n exceptions for ϕ. The the dual $M_n\phi$ means that the agent considers at least n alternatives where ϕ is true. These graded epistemic operators can be used to make some implicit knowledge explicit. Moreover, the graded epistemic logic $GS5$ is complete. So we may explore logical dynamics on the basis of graded epistemic logic.

It is easy to observe that public announcement logic and dynamic epistemic logic can take the epistemic $GS5$ as a basis. The key recursion axioms in theorem 4 and theorem 8 also work very well.

Theorem 12. *The public announcement logic on epistemic $GS5$ is axiomatized completely by the laws of epistemic $GS5$ and the following recursion axioms:*
(1) recursion axioms (a), (b), (c) and (f) in theorem 5;
(2) $\langle !\phi\rangle M_n\phi \leftrightarrow \phi \wedge M_n\langle !\phi\rangle\psi$.

Theorem 13. *Dynamic epistemic logic on epistemic $GS5$ is axiomatized completely by the laws of epistemic $GS5$ and the following recursion axiom:*
(1) recursion axioms (a), (b) and (c) in theorem 9;
(2) the recursion axiom in theorem 8 with replacing $\langle\overline{n}\rangle$ by M_n.

With epistemic $GS5$, life seems to be easier for us to understand some philosophical issues. (See van der Hoek and Meyer (1992).) The KK-thesis $K\phi \to KK\phi$ may be replaced by the more plausible version $K_n\phi \to KK_n\phi$ which says: if the agent considers that it is possible that there are n alternative possibilities for ϕ's being true then he considers that there are n alternative possibilities for ϕ's being true. Moreover, the notion of knowledge in the logical omniscience formula $K(\phi \to \psi) \to (K\phi \to K\psi)$ is too strong or idealistic. But the following formula is more realistic and plausible: $K(\phi \to \psi) \to (K_m\phi \to K_{m+n}\psi)$. The epistemic $GS5$ seems to provide the notion of knowledge which is more approximate to our real one in ordinary life. We also observe that the dynamic logics of information update which base on the epistemic graded modal logic can also be developed smoothly. A crucial point arises here: it seems to be the case that logical dynamics of information flow are irrelevant to choose which basic static epistemic logic of knowledge. Information update has its own logic!

Acknowledgements I would like to thank professor Johan van Benthem for his encouraging me to combine dynamic epistemic logic with graded modalities. Professor Johan van Benthem gave a nice course on logical dynamics at Tsinghua University (Beijing, China) in Autumn 2008. I am greatly indebted to our discussions.

References

A. Baltag, L. Moss, and S. Solecki. The logic of public announcements: Common knowledge and private suspicions. In I. Gilboa, editor, *Proceedings of the 7th conference on Theoretical aspects of rationality and knowledge*, pages 43–56. Morgan Kaufmann Publishers Inc., Evanston Illinois, 1998.

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

C. Cerrato. General canonical models for graded normal logics (graded modalities iv). *Studia Logica*, 49:241–252, 1990.

M. de Rijke. Modal model theory. Technical Report CS-R9517, Computer Science/Department of Software Technology, Amsterdam, 1995.

M. de Rijke. A note on graded modal logic. *Studia Logica*, 64:271–283, 2000.

M. Fattorosi-Barnaba and C. Cerrato. Graded modalities iii. *Studia Logica*, 47: 99–110, 1988.

M. Kracht and F. Wolter. Properties of independently axiomatizable bimodal logics. *The Journal of Symbolic Logic*, 56(4):1469–1485, 1991.

A. Kurucz. Combining modal logics. In P. Blackburn, J. van Benthem, and F. Wolter, editors, *Handbook of Modal Logic*. Elsevier, 2007.

J. Plaza. Logics of public communications. In M. L. Emrich, M. S. Pfeifer, M. Hadzikadic, and Z. Ras, editors, *Proceedings of the fourth international symposium on methodologies for intelligent systems: Poster session program*, pages 201–216. Oak Ridge National Laboratory, 1998.

J. van Benthem. Logics for information update. In J. van Benthem and M. Kaufmann, editors, *Proceedings of the eith conference on Theoretical aspects of rationality and knowledge*, pages 51–67. Morgan Kaufmann Publishers Inc., San Francisco, 2001.

J. van Benthem. *Logical Dynamics of Information Flow and Interaction*. Unpublished Manuscript, 2008.

J. van Benthem and M. Martinez. Logical theories of information. In *Philosophy of Information*. Elsevier Science Publishers, 2008.

W. van der Hoek and J.-J. C. Meyer. *Graded Modalities in Epistemic Logic*, volume 620 of *Lecture Notes in Computer Science*, pages 503–504. Springer, 1992.

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

Formulaic Events and Local Product Models in Dynamic Epistemic Logic

Junhua Yu
Department of Philosophy, Tsinghua University.
DF7G5036@hotmail.com

Abstract

Formulaic events are introduced into the field of dynamic epistemic logic in this paper. We consider a general class of syntactical methods to employ dynamic modalities into the precondition language without self-references. Within this class of methods, we take the "formulaic event" method, which is in some sense the "minimal" one, as an illustration of "reconstructing" **DEL**. The resulting system, i.e., **feDEL**, is showed to be a nice rewriting of **DEL** by verifying "event model correspondence" results. Having introduced formulas into events, we could define "local product model" by employing knowledge-dependent uncertainty relations between events. It is possible to carry out this on **feDEL**, but we introduce **L-pfeDEL**, which is also in the syntactical repair class, with a view of a natural semantic.

1 Reviews and Motivations

Dynamic Epistemic Logic **DEL** (cf. van Benthem (2008), also Baltag et al. (1998) and van Ditmarsch et al. (2007)) is a logical tool of reasoning about knowledge of agents in informational communications. At first,we give a brief review of some basic concepts of **DEL**, with most of our terminologies being inherited from van Benthem (2008).

Definition 1.1 (Language of **EL**)**.** Suppose that we have a finite number(e.g., $m \in \mathbb{N}$) of agents, which are denoted by $1, 2, \cdots, m$, then the language of **EL** is defined by the following inductive rules:

$$\phi ::= \bot | p | \neg\phi | \phi \vee \phi \vee \cdots | K_i\phi$$

with 'p' standing for any propositional letter.

Note that we employ countably infinite disjunctions here.

Definition 1.2 (Pointed epistemic model)**.** Suppose that we have a set I of agents, then a *pointed epistemic model* is $(\mathfrak{M}, s)$ or $(W, \{R_i\}_{i\in I}, V, s)$, where W is a set of states, with s as the current state, R_i is an uncertainty relation between states for the agent i, V is an evaluation function.

This offers the truth definition of Epistemic Logic (**EL**), van Benthem (2008):

Definition 1.3 (Truth Definition of **EL**)**.** For any pointed epistemic model $(\mathfrak{M}, s) = (W, \{R_i\}_{i\in I}, V, s)$,

$\mathfrak{M}, s \Vdash p$ iff $s \in V(p)$
$\mathfrak{M}, s \Vdash \neg\eta$ iff $\mathfrak{M}, s \nVdash \eta$
$\mathfrak{M}, s \Vdash \xi_1 \vee \xi_2 \vee \cdots$ iff $\mathfrak{M}, s \Vdash \xi_i$ for some i
$\mathfrak{M}, s \Vdash K_i\eta$ iff for all $t \in W$ s.t. $(s, t) \in R_i$, we have $\mathfrak{M}, t \Vdash \eta$

The logic **EL** simulates static epistemic activities of a group of agents. For dynamic epistemic activities, such as public announcements, **EL** is no longer capable enough. To express dynamic activities, we need Dynamic Epistemic Logic, which is initialed in Baltag et al. (1998).

For a pointed Kripke frame over the agent set I, any function which is from the domain of the frame to a language $\mathcal{L}$ is called a *precondition function*, with notation *Pre*. The language $\mathcal{L}$ is then called the *precondition language* of **DEL**.

Other than pointed epistemic models, we also have *pointed event models* in **DEL**.

Definition 1.4 (Pointed event model)**.** Suppose that we have a set I of agents, then a pointed event model is $(\mathcal{E}, e)$, or $(E, \{\sim_i\}_{i\in I}, \{Pre_e\}_{e\in E}, e)$, where E is an at most countable set of events, with e as the current event, $\sim_i$ is an uncertainty relation between events of the agent i, and *Pre* is a precondition function which has E as its domain.

In a pointed event model, the value of e by the precondition function, which is denoted by Pre_e, is called the *precondition* of e.

Having defined pointed event models, we could state the definition of **DEL**-language now.

Definition 1.5 (Language of **DEL**)**.** Suppose that we have a finite number(e.g., $m \in \mathbb{N}$) of agents, which are denoted by $1, 2, \cdots, m$, then the language of **DEL** is defined by the following inductive rules:

$$\phi ::= \bot|p|\neg\phi|\phi \vee \phi \vee \cdots |K_i\phi|[\mathcal{E}, e]\phi$$

with 'p' standing for any propositional letter, and '$\mathcal{E}, e$' standing for any *pointed event model*.

Note that we include countably infinite disjunctions[1].

We have the dynamic modality, i.e., $[\mathcal{E}, e]$, in our language. Before the corresponding truth definition is given, we state the definition of *pointed product model*.

Definition 1.6 (Pointed product model)**.** For any pointed epistemic model $(\mathfrak{M}, s) = (W, \{R_i\}_{i\in I}, V, s)$ and any pointed event model $(\mathcal{E}, e) = (E, \{\sim_i\}_{i\in I}, \{Pre_e\}_{e\in E}, e)$, the pointed product model $\mathfrak{M} \times \mathcal{E}, (s, e)$ is a pointed epistemic model $(W', \{R'_i\}_{i\in I}, V', (s, e))$, where

[1]In van Benthem (2008), we have reduction axioms for **DEL**(cf. (1) in this paper), which indicate that we need to employ countably infinite disjunctions(as $\mathcal{L}_{\omega_1\omega}$ in Ebbinghaus et al. (1994)) to deal with the case that $\{f \in E \,|\, e \sim_i f\}$ is infinite.

$W' = \{(t, f) \in W \times E \mid \mathfrak{M}, t \Vdash Pre_f\}$
$R'_i = \{((t_1, f_1), (t_2, f_2)) \in W' \times W' \mid (t_1, t_2) \in R_i \textit{ and } (f_1, f_2) \in \sim_i\}$
$V' = \{(t, f) \in W' \mid t \in V\}$

Definition 1.7 (Truth Definition of **DEL**)**.** The *truth definition* of **DEL** is that of **EL**, enriched by the following clause:
$\mathfrak{M}, s \Vdash [\mathcal{E}, e]\xi$ iff
if $\mathfrak{M}, s \Vdash Pre_e$ then $\mathfrak{M} \times \mathcal{E}, (s, e) \Vdash \xi$.

With this, we have *reduction axioms* of **DEL**, which enable van Benthem (2008) to get the completeness theorem of **DEL** in an easy way. As an instance, we state the reduction axiom for knowledge modalities here. That is:

$$[\mathcal{E}, e]K_i\xi \leftrightarrow (Pre_e \rightarrow \bigwedge_{f \in E \textit{ and } e \sim_i f} K_i[\mathcal{E}, f]\xi) \tag{1}$$

It should be emphasized that, we have not specified the precondition language of **DEL** here. In some papers, e.g. Baltag et al. (1998) and van Benthem et al. (2008), it is specified to be **EL**−language[2]. In some other papers, e.g. Baltag and Moss (2004) and Renne (2008), the precondition language of **DEL** is specified to be **DEL** itself.

Note that if we want to have dynamic modalities in precondition language, we must define the syntax carefully, and treat *Pre* as a defined symbol, as in Baltag and Moss (2004). Otherwise, the attempt of employing the full **DEL** to be the precondition language may result self-referential semantics, hence is not suitable. What follows is an easy instance for this:

Example 1. For any pointed event model $\mathcal{E} = (E, \{\sim_i\}_{i \in I}, \{Pre_e\}_{e \in E}, e)$. If we take the full **DEL** to be the precondition language, then a self-referential paradox may rise.

For instance, if $Pre_e = [\mathcal{E}, e]\bot$, then we have:
$\mathfrak{M}, s \Vdash Pre_e$ iff
$\mathfrak{M}, s \Vdash [\mathcal{E}, e]\bot$ iff
$\mathfrak{M}, s \Vdash Pre_e \Rightarrow \mathfrak{M} \times \mathcal{E}, (s, e) \Vdash \bot$ iff
$\mathfrak{M}, s \Vdash Pre_e \Rightarrow \bot$ iff
$\mathfrak{M}, s \nVdash Pre_e$
That is, Pre_e is satisfied by a model iff it is not satisfied by the same model.

The observation above shows that, to allow the precondition language to contain dynamic update modalities, we should avoid self-reference at least.

In this paper, we will discuss a class of methods to avoid self-reference in general, and also a by-product of the class of methods.

2 Possible Methods of Avoiding Self-Reference

In this section, two different ideas of avoiding self-reference are considered. Due to the author's interests, we will focus on the second idea in later sections. However, the first idea is also enlighten.

[2] In (Baltag et al. 1999, Section 2.1), it is stated that "*PRE* is a map $PRE : K \rightarrow \mathcal{L}$", while the "logical language with epistemic actions" being denoted by $\mathcal{L}([\alpha])$. In van Benthem et al. (2008), we have "$\mathbf{pre} : S \rightarrow \mathcal{L}_{\mathbf{EL}}$ is the pre-condition function" directly. Note that the strings *PRE*, *Pre*, *pre* may have different meanings in different papers.

2.1 A Semantical Idea

An nature way of avoiding self-reference is to employ an order on events[3]. That is, each pointed event model has the form of $\mathcal{E} = (E, \{\sim_i\}_{i\in I}, \{Pre_e\}_{e\in E}, \leqslant, e)$, where $\leqslant$ is a linear ordering on E, and for each event e, Pre_e does not contain f, provided $e \leqslant f$. Obviously, in a fixed pointed event model, self-referential problems are avoided successfully.

However, this is not sufficient for avoiding referential loop between two or more pointed event models.

Example 2. $e \in E$, $f \in F$ and $g \in G$, together with $Pre_e = [\mathcal{F}, f]\bot$, $Pre_f = [\mathcal{G}, g]\bot$ and $Pre_g = [\mathcal{E}, e]\bot$, results a self-referential paradox.

To remedy this, we may assume that the **DEL** language has an ordered set of pointed event model symbols, which enables us to set a similar clause for pointed event models. That is, for each pointed event model $(\mathcal{E}, e)$, Pre_e can refer to $(\mathcal{F}, f)$, only if $(\mathcal{E}, e) > (\mathcal{F}, f)$.

If we let the order to be empty, then any events can refer to no events. The result is, we only let **EL**−language to be our precondition language.

2.2 A Syntactical Idea in General

Since most of our classical logics do not have self-referential problems, it might be helpful to consider some of them. In this subsection, a general observation is given.

We take basic epistemic modal logic **EL**, cf. Definition 1.3, for instance. To interpret $K\eta$, what we need to do is to care about a *model condition*, i.e., uncertainty relation, together with the interpretation of a subformula. For other kind of formulas, situations are similar. That is, all we need to observe are model conditions and interpretations of subformulas.

In other logics, such as first-order logic, temporal logic and **PDL** Blackburn et al. (2001), we encounter similar situations when interpreting formulas. In Janssen (1997), this is called "the principle of compositionality", i.e., *The meaning of a compound expression is a function of the meanings of its parts*.

With employing $\mathcal{L}$ to denote the set of syntactical strings, $\mathcal{M}$ to denote the set of models, we could express these situations as following:

$$\Lambda(\eta) := \Pi(\Lambda(\Theta_1(\eta)), \Lambda(\Theta_2(\eta)), \cdots) \tag{2}$$

with $\Lambda\colon \mathcal{L} \to \wp(\mathcal{M})$, $\Theta\colon \mathcal{L} \to \mathcal{L}$, which select a proper substring of the inputted string as the output, $\Pi\colon \wp(\mathcal{M})^{n(or\ \omega)} \to \wp(\mathcal{M})$ being a "model operation", which may contain operations corresponding to accessibility relations, Boolean connectives, product updates, and so on. Hence Λ is a *syntax* $\to$ *semantic*−operation, while Θ's being *syntax*−operations, and Π being a *semantic*−operation. In the illustration of $K\eta$ above, $n = 1$ with the only Θ exporting η if the inputted string has the form of $K\eta$, while Π checking whether all states(in fact, pointed epistemic models[4]) accessible from the current one are in $\Lambda(\Theta(K\eta))$.

Since all of classical logics have "regularly" syntaxes, that is, no string could be a proper substring of itself, when interpreting a formula, the operations

[3]This is suggested by Prof. Johan van Benthem during a discussion.

[4]Suppose that sRt in $\mathfrak{M}$. Then when we interpret $K\eta$ at s, we refer to whether or not $\mathfrak{M}, t \Vdash \eta$. Thus, we consider not only states, but also pointed epistemic models.

above will never refer to the interpretation of the formula itself[5]. If we can accomplish this on **DEL**(with dynamic modalities in precondition languages), we may also avoid self-referential problems. As we will see in the rest of this section, what we need to do is a syntactical repair.

In **DEL** van Benthem (2008), the precondition of an event e is denoted by Pre_e. Recall that in the truth definition of **DEL**(cf. Definition 1.7), we have:

$$\mathfrak{M}, s \Vdash [\mathcal{E}, e]\xi \;\; iff \;\; \mathfrak{M}, s \Vdash Pre_e \Rightarrow \mathfrak{M} \times \mathcal{E}, (s, e) \Vdash \xi$$

It should be emphasized that when interpreting $[\mathcal{E}, e]\xi$, we refer to the interpretation of Pre_e, which is of course not a proper substring of $[\mathcal{E}, e]\xi$. In the notation above, the output of a Θ operation is always a proper substring of the inputted string. Hence, we can not express this in the way we used above. Instead, we use:

$$\Lambda(\eta) := \Pi(\Lambda(\Omega(\Theta_1(\eta))), \Lambda(\Theta_2(\eta))) \tag{3}$$

where Θ_1 yields e when $[\mathcal{E}, e]\xi$ is inputted, Θ_2 yields ξ when $[\mathcal{E}, e]\xi$ is inputted, Π takes the two sets of pointed models(which are also inputs), while caring about the "product update" relation between pointed models. The only new operation, i.e., $\Omega\colon \mathcal{L} \to \mathcal{L}$, is an syntactical operation which yields Pre_e[6] when e is inputted.

From (2) to (3), the only alteration is the appearance of Ω operation. This indicates that **DEL** is in some sense special, provided we employ *Pre* as an initial symbol.

If we set no restrictions on Ω, then Ω could be any operation from $\mathcal{L}$ to $\mathcal{L}$. In Example 1, Ω yields $[\mathcal{E}, e]\bot$ when e is inputted, which enables Π to touch the interpretation of $[\mathcal{E}, e]$ when we are interpreting $[\mathcal{E}, e]$. The observation above shows how self-reference is involved in Example 1. It is a more complex, but also similar case in Example 2.

When proper restrictions are attached to Ω, we can transform the special situation (3) to the standard situation (2), and hence, avoid self-referential problems. $\Theta_1 : \mathcal{L} \to \mathcal{L}$, yields a proper substring of the inputted string. Hence if $\Omega : \mathcal{L} \to \mathcal{L}$ yields a substring(proper or improper) of the inputted string, then we could combine them to one operation, say, $\Theta_1^{\bowtie} : \mathcal{L} \to \mathcal{L}$, which yields a proper substring of the inputted string. In this way, we could express the situation by:

$$\Lambda(\eta) := \Pi(\Lambda(\Theta_1^{\bowtie}(\eta)), \Lambda(\Theta_2(\eta))) \tag{4}$$

which shares the form of (2). By the observation about (2) in previous paragraphs, we know that self-referential problems have been avoided. To do this, what we have done is a syntactical repair, i.e., to add a restriction on Ω, require it to produce a substring of the inputted string as output. In the terminology of **DEL**, that is:

Put preconditions explicitly in events.

[5]Here might be a point to be made clear. Assume that we want to interpret $\neg\phi$, we can select Θ as we do in propositional logic. But this time, instead of complement, we choose identity for Π. Then the resulting truth-definition clause would be: $u \Vdash \neg\phi$ iff $u \Vdash \phi$. Is that a paradox? *No.* It is of course a strange interpretation of $\neg$, but still acceptable. Recall that in Example 1, what we have is $u \Vdash \eta$ iff $u \nVdash \eta$!

[6]"Pre_e" is also a string.

Since preconditions are formulas, when we define an event[7], we have to involve at least one formula, i.e., the precondition of the event. This could be expressed by:

$$Event \ := \ \Gamma(\kappa) \tag{5}$$

where κ is the precondition of the event, Γ is a "non-decreasing" operation, which sets the inputted string as a substring(proper or improper) of the output. Note that Γ packs a precondition formula to be an event, while in (3), Ω opening the package and fetching the precondition formula out. Obviously, Γ and Ω are inversions of each other.

In summary, our syntactical idea is:

Define events by (5), where Γ is a "non-decreasing" operation.

We already have an instance of applying such a restriction. That is, Public Announcement Logic(**PAL**), which is treated as a special case of **DEL** by van Benthem (2008), Renne (2008), and also some other papers. All events in **PAL** share the form of $!\phi$, while the precondition of $!\phi$ being ϕ. In the expression of (5), that is, Γ yields $!\varepsilon$, if ε is inputted.

The key clause of **PAL**'s truth definition is:

$$\mathfrak{M}, s \Vdash [!\phi]\xi \ \ iff \ \ \mathfrak{M}, s \Vdash \phi \Rightarrow \mathfrak{M}|\phi, s \Vdash \xi$$

We could express this in the form of (3), where Θ_1 yields $!\phi$ when $[!\phi]\xi$ is inputted, Θ_2 yields ξ when $[!\phi]\xi$ is inputted, Ω is the syntactical operation which yields ϕ when $!\phi$ is inputted, Π takes the two sets of pointed models(which are also inputs), while caring about the "hard information update" relation between pointed models. Obviously, we could take $\Theta_1^{\bowtie}$ to be the combination $\Theta_1 \circ \Omega$. Since $\Theta_1 \circ \Omega$ is also an operation which yields a proper substring of the inputted string, we can transfer the form of (3) to the form of (4), i.e., (2). Since the situation of **PAL** meets the form of (2), **PAL** does not have self-referential problems.

Remark 1. *In Baltag et al. (1998) and van Benthem et al. (2008), we have an unstandard restriction on the operation Ω, i.e., the range of the operation is limited to the set of strings of **EL**-language. This restriction does not follow the general idea discussed above, but also works. Since event symbol e is not in **EL**, the string $\Omega(e)$ contains no e. Hence $\Lambda(\Omega(\Theta_1([\mathcal{E}, e]\xi)))$ does not refer to the interpretation of $[\mathcal{E}, e]$. In this way, we avoid self-referential problems in **DEL**, while paying a cost. That is, in the terminology of **DEL**, we take **EL**, which does not contain dynamic update modalities, as our precondition language. As we have showed in Example 1, this restriction is dropped when we employ dynamic into precondition language. Hence we will only focus on the general idea discussed above.*

From the observation above, it is clear that what we could do is to repair the syntax of **DEL** from van Benthem (2008), define events by (5), and make sure that the operation Γ is a "non-decreasing" operation.

Obviously, there are lots of capable Γ's, corresponding to lots of capable repairs. For instance, we could choose the "minimal" Γ, i.e., identity operation. We will consider this special case in Section 3 as an illustration, while observing another case in Section 4.2, for other motivations.

[7]Note that in the language of **DEL**, events are initial symbols, not defined symbols.

3 The "Minimal" Method: Formulaic Events

3.1 A Preliminary Discussion

As we have stated above, in this section, we will take the minimal "non-decreasing" Γ operation: *Identity*.

We have expressed the situation of **DEL** by (3), while Ω has Γ, which yields an event when a formula is inputted, as its inversion. Thus, in the truth definition of **DEL**, the Γ operation is in charge of "packing precondition formulas in events". Since we have taken the identity operation as Γ, then

$$\textit{Each event is just its precondition formula.}$$

Before introducing the resulting system, it might be helpful to make one point clear now. That is:

Remark 2. *Once events are defined to be their precondition formulas, they are determined by their preconditions. It might be interesting to consider whether events are determined by their preconditions, physically or philosophically. However, instead of caring about these discussions, what we will do is to present an illustration of our "minimal" syntactical method.*

3.2 The System feDEL

We define the language of **feDEL**[8] and pointed formulaic event models simultaneously at first. Since events are defined to be formulas, domains of event models are then sets of formulas, while uncertainty relations between events being relations between formulas. Thus, we will not only mention formulas when we define pointed formulaic event models, but also mention pointed formulaic event models when we define the language.

Definition 3.1 (Language of **feDEL** and Pointed formulaic event model)**.** Suppose that we have a finite number(e.g., $m \in \mathbb{N}$) of agents, which are denoted by $1, 2, \cdots, m$, then the language of **feDEL** is defined by the following inductive rules:

$$\phi ::= \bot | p | \neg\phi | \phi \vee \phi \vee \cdots | K_i\phi | [\mathfrak{E}]\phi$$

with 'p' standing for any propositional letter, and '$\mathfrak{E}$' standing for any *pointed formulaic event model*[9] which is defined as:

$$\mathfrak{E} ::= \phi, \underbrace{(\phi, \cdots)}_{n(or\ \omega)}, \underbrace{(\underbrace{(\underbrace{\phi, \cdots}_{\mu \leqslant n(or\ \in\omega^+)}), \cdots}_{m}), \cdots}_{n(or\ \omega)}$$

while each ϕ in the pointed formulaic event model occuring in the second part of it, i.e.: in $\underbrace{(\phi, \cdots)}_{n(or\ \omega)}$.

[8] **fe** stands for "formulaic event" here.

[9] The pointed formulaic event model is still a syntactical object, though we call it as a "model".

Note that we employ countably infinite disjunctions, and we allow pointed formulaic event models to have countably infinite many formulaic events.

From logical constant till knowledge modality, the above definition is usual, though we have noted that infinite disjunctions are involved. The reason why infinite disjunctions are allowed is that we want to manage infinite "event models" (cf. the reduction axioms for **DEL**, i.e.: (1)). The only clause which needs more explanation might be the one for dynamic modality.

The idea is simple[10]: *Put a pointed-graph into the language*. To make it clear, it might be helpful to write each component explicitly. Suppose that we have m agents and the following pointed formulaic event model:

$$\eta_j, (\underbrace{\eta_1,\cdots,\eta_j,\cdots}_{n(or\ \omega)}), \underbrace{(\underbrace{(\underbrace{\eta_{1.1.1},\eta_{1.1.2},\cdots}_{\mu_{1.1}\leqslant n(or\ \in\omega^+)}),(\underbrace{\eta_{1.2.1},\cdots}_{\mu_{1.2}\leqslant n(or\ \in\omega^+)}),\cdots}_{m}),(\underbrace{(\underbrace{\eta_{2.1.1},\cdots}_{\mu_{2.1}\leqslant n(or\ \in\omega^+)}),\cdots}_{m}),\cdots}_{n(or\ \omega)} \tag{6}$$

Then (the precondition of) the current formulaic event is η_j, with $\eta_1,\cdots,\eta_j,\cdots$ being the n (or $\omega - many$) formulaic events in the same pointed formulaic event model. For any x s.t. $0 < x \leqslant n(or\ 0 < x \in \omega)$, any y s.t. $0 < y \leqslant m$ and any z s.t. $0 < z \leqslant n(or\ 0 < z \in \omega)$, $\eta_{x.y.z}$ is the z-th formulaic event which is accessible for agent y from η_x. $\mu_{x.y}$ is employed to denote the number of formulaic events which are accessible for agent y from η_x[11].

We present an easy example here:

Example 3. In the following pointed formulaic event model:

$$\eta_1, (\eta_1,\eta_2,\eta_3), ((\eta_3),()), ((),(\eta_1,\eta_3)), ((),())$$

The first part, i.e. η_1, presents that the current formulaic event is η_1. The second part, i.e. (η_1,η_2,η_3), presents that these three are the only three formulaic events in the formulaic event model. The third part, i.e. $((\eta_3),()), ((),(\eta_1,\eta_3)), ((),())$ presents the accessibility relation. There are three pieces in this part, corresponding to the three formulaic events respectively. For instance, the second piece, i.e. $((),(\eta_1,\eta_3))$, indicates that from η_2, there are no accessible formulaic events for agent 1, with η_1, η_3 being the two formulaic events accessible for agent 2.

Diagrammatically, this can be shown by:

$$\begin{tikzcd} & \eta_1 \arrow[dr, "1"] & \\ \eta_2 \arrow[ur, "2"] \arrow[rr, "2"] & & \eta_3 \end{tikzcd}$$

[10]This idea is inspired by Professor Johan van Benthem in one of his lectures in Tsinghua Univ. Beijing, Oct. 2008.

[11]We explain the reason why we put $\mu_{x.y} \leqslant n(or \in \omega^+)$ in (6). If the pointed formulaic event model is finite, then it has a finite number, say n, of formulaic events. Hence for any formulaic event η_x, any agent y, there are at most n accessible formulaic events. In this case, $\mu_{x.y} \leqslant n$. If the pointed formulaic event model is infinite, then by Definition 3.1, it has $\omega-$many formulaic events. Hence, for each formulaic event and each agent, there are at most ω accessible formulaic events. In this case, $\mu_{x.y} \in \omega^+$, while $\omega^+ = \omega \cup \{\omega\}$.

In the notation of (6), we have: $\eta_{1.1.1} = \eta_3$, $\eta_{2.2.1} = \eta_1$, $\eta_{2.2.2} = \eta_3$, and $\mu_{1.1} = 1$, $\mu_{2.2} = 2$, $\mu_{1.2} = \mu_{2.1} = \mu_{3.1} = \mu_{3.2} = 0$.

By the example above, it is clear that what we have employed is a formal description of pointed graphs. Since the general notation is not convenient, in the following text, we will use:

$\overline{\eta}$ to denote $\{\underbrace{\eta_1, \cdots, \eta_j, \cdots}_{n(or\ \omega)}\}$, the domain of the pointed formulaic event model,

$\overrightarrow{\eta}$ to denote $(\underbrace{\eta_1, \cdots, \eta_j, \cdots}_{n(or\ \omega)})$, the list of all formulaic events in that model,

$\overline{\eta_{x.i}}$ to denote $\{\underbrace{\eta_{x.i.1}, \eta_{x.i.2}, \cdots}_{\mu_{x.i} \leqslant n(or\ \in \omega^+)}\}$, the set of formulaic events which are accessible by agent i from η_x, and

$\widetilde{\eta}$ to denote $\underbrace{(\underbrace{(\underbrace{\eta_{1.1.1}, \cdots}_{\mu_{1.1} \leqslant n(or\ \in \omega^+)}), \cdots}_{m}), \cdots}_{n(or\ \omega)}$, the accessibility relations of that model.

Hence a general pointed formulaic event model would be denoted as:

$$\eta, \overrightarrow{\eta}, \widetilde{\eta} \tag{7}$$

while " η_y is accessible from η_x by agent i " being denoted by

$$\eta_y \in \overline{\eta_{x.i}} \tag{8}$$

Like in **DEL**, we also start the discussion of semantic of **feDEL** with the definition of pointed product model.

Definition 3.2 (Pointed Product Model)**.** For any pointed epistemic model $(\mathfrak{M}, s) = (W, \{R_i | i \in I\}, V, s)$ and any pointed formulaic event model $\mathfrak{E} = \eta, \overrightarrow{\eta}, \widetilde{\eta}$, if $\mathfrak{M}, s \Vdash \eta$, then the *pointed product model* $(\mathfrak{M} \circ \mathfrak{E}, s \circ \eta)$ is $(W \circ \mathfrak{E}, \{(R \circ \mathfrak{E})_i | i \in I\}, V \circ \mathfrak{E}, s \circ \eta)$, where:

$W \circ \mathfrak{E} = \{\, u \circ \tau \,|\, u \in W \textit{ and } \tau \in \overline{\eta} \textit{ and } \mathfrak{M}, u \Vdash \tau \,\}$

$(R \circ \mathfrak{E})_i = \{\, (u \circ \eta_x, v \circ \eta_y) \in (W \circ \mathfrak{E}) \times (W \circ \mathfrak{E}) \,|\, uR_iv \textit{ and } \eta_y \in \overline{\eta_{x.i}} \,\}$, for each $i \in I$

$V \circ \mathfrak{E} = \{\, u \circ \tau \in W \circ \mathfrak{E} \,|\, u \in V \,\}$

The way we define pointed product model here is very similar with the way we used in **DEL**[12] (cf. Definition 1.6). Firstly, we calculate the cross product of an pointed epistemic model and a pointed (formulaic) event model. Secondly, we eliminate all the states $(u \circ \tau)$, if $\mathfrak{M}, u$ fails to satisfy (the precondition of) τ.

[12]We have altered some notations, e.g., using $s \circ \eta$ instead of (s, η). This new notation may help us to reduce the number of parentheses. In Definition 3.2, we have the notation of $\mathfrak{M} \circ \mathfrak{E}, s \circ \eta$. If we write the pointed formulaic event model explicitly, it is: $\mathfrak{M} \circ \eta, \overrightarrow{\eta}, \widetilde{\eta}, s \circ \eta$. Note that by (7), $\widetilde{\eta}$ has the form of $(\cdots), \cdots, (\cdots)$. In this case, it may be difficult to distinguish the original notation, i.e. (s, η) from $\widetilde{\eta}$.

Thirdly, we construct accessibility relations by pairs $(u \circ \eta_x, v \circ \eta_y)$, if (u, v) and (η_x, η_y) are in corresponding accessibility relations of $\mathfrak{M}$ and $\mathfrak{E}$[13], respectively. Lastly, the valuation is transferred as we did in **DEL**.

With the definition of pointed product model, we can state the truth definition of **feDEL**.

Definition 3.3 (Truth Definition of **feDEL**)**.** With employing corresponding clauses from epistemic logic van Benthem (2008) and the propositional fragment of $\mathcal{L}_{\omega_1\omega}$ Ebbinghaus et al. (1994), we only need to state the clause for dynamic modality: for any pointed epistemic model $\mathfrak{M}, s$ and any pointed formulaic event model $\mathfrak{E}$(with η as its current formulaic event):

$$\mathfrak{M}, s \Vdash [\mathfrak{E}]\xi \;\; iff \;\; \mathfrak{M}, s \Vdash \eta \Rightarrow \mathfrak{M}\circ\mathfrak{E}, s\circ\eta \Vdash \xi$$

It is not surprising that the truth definition of **feDEL** is similar with that of **DEL**. But this time, we have **feDEL** as the "precondition language" of itself.

The following reduction axioms, may also be acquaintances.

Theorem 1 (Reduction Axioms of **feDEL**)**.** *The following wffs are valid in* ***feDEL****:*[14]

(1) $[\eta, \overrightarrow{\eta}, \widetilde{\eta}]q \leftrightarrow (\eta \to q)$

(2) $[\eta, \overrightarrow{\eta}, \widetilde{\eta}]\neg\xi \leftrightarrow (\eta \to \neg[\eta, \overrightarrow{\eta}, \widetilde{\eta}]\xi)$

(3) $[\eta, \overrightarrow{\eta}, \widetilde{\eta}]\bigwedge\vec{\xi} \leftrightarrow \bigwedge\overrightarrow{[\eta, \overrightarrow{\eta}, \widetilde{\eta}]\xi}$

(4) $[\eta_x, \overrightarrow{\eta}, \widetilde{\eta}]K_i\xi \leftrightarrow (\eta_x \to \bigwedge_{\eta_y \in \overline{\eta_{x,i}}}(K_i[\eta_y, \overrightarrow{\eta}, \widetilde{\eta}]\xi))$

Proof. These validities could be shown by an analogue of the verification we used in **DEL**. The only difference is: what we have as "preconditions" here are not **EL**-formulas, but **feDEL**-formulas. Nevertheless, this difference does not matter since in the reduction axioms above, we treat "preconditions" as Boolean components without caring about structures inside them.

Here come the details. For any pointed epistemic model $(\mathfrak{M}, s) = (W, \{R_i\}_{i\in I}, V, s)$ and any pointed formulaic event model $\mathfrak{E} = \eta, \overrightarrow{\eta}, \widetilde{\eta}$, we have:

(1) $\mathfrak{M}, s \Vdash [\eta, \overrightarrow{\eta}, \widetilde{\eta}]q$ iff
$\mathfrak{M}, s \Vdash \eta \Rightarrow \mathfrak{M}\circ\mathfrak{E}, s\circ\eta \Vdash q$ iff
$\mathfrak{M}, s \Vdash \eta \Rightarrow s\circ\eta \in (V\circ\mathfrak{E})(q)$ iff
$\mathfrak{M}, s \Vdash \eta \Rightarrow s \in V(q)$ iff
$\mathfrak{M}, s \Vdash \eta \Rightarrow \mathfrak{M}, s \Vdash q$ iff
$\mathfrak{M}, s \Vdash \eta \to q$

(2) $\mathfrak{M}, s \Vdash [\eta, \overrightarrow{\eta}, \widetilde{\eta}]\neg\xi$ iff
$\mathfrak{M}, s \Vdash \eta \Rightarrow \mathfrak{M}\circ\mathfrak{E}, s\circ\eta \Vdash \neg\xi$ iff
$\mathfrak{M}, s \Vdash \eta \Rightarrow \{\mathfrak{M}, s \Vdash \eta \;\&\&\; \mathfrak{M}\circ\mathfrak{E}, s\circ\eta \Vdash \neg\xi\}$ iff
$\mathfrak{M}, s \Vdash \eta \Rightarrow \mathfrak{M}, s \Vdash \neg[\eta, \overrightarrow{\eta}, \widetilde{\eta}]\xi$ iff
$\mathfrak{M}, s \Vdash \eta \to \neg[\eta, \overrightarrow{\eta}, \widetilde{\eta}]\xi$

(3) $\mathfrak{M}, s \Vdash [\eta, \overrightarrow{\eta}, \widetilde{\eta}]\bigwedge\vec{\xi}$ iff
$\mathfrak{M}, s \Vdash \eta \Rightarrow \mathfrak{M}\circ\mathfrak{E}, s\circ\eta \Vdash \bigwedge\vec{\xi}$ iff
$\mathfrak{M}, s \Vdash \eta \Rightarrow \&\&\,(\overrightarrow{\mathfrak{M}\circ\mathfrak{E}, s\circ\eta \Vdash \xi})$ iff
$\&\&\,(\overrightarrow{\mathfrak{M}, s \Vdash \eta \Rightarrow \mathfrak{M}\circ\mathfrak{E}, s\circ\eta \Vdash \xi})$ iff

[13]The accessibility relations in formulaic event models are given in special notations, cf. (8).

[14]With employing $\neg$ in our language, we could define one of $\wedge$(including infinite case) and $\vee$(including infinite case) on the other. Hence it is sufficient to give the reduction axiom for $\wedge$. We denote finite conjunction $\xi_1 \wedge \cdots \wedge \xi_n$ or infinite conjunction $\xi_1 \wedge \xi_2 \wedge \cdots$ by $\bigwedge\vec{\xi}$.

$$
\begin{aligned}
&\&\& \,\overrightarrow{(\mathfrak{M}, s \Vdash [\eta, \overrightarrow{\eta}, \widetilde{\eta}]\xi)} \quad \text{iff} \\
&\mathfrak{M}, s \Vdash \bigwedge \overrightarrow{[\eta, \overrightarrow{\eta}, \widetilde{\eta}]\xi} \\
(4)\ &\mathfrak{M}, s \Vdash [\eta_x, \overrightarrow{\eta}, \widetilde{\eta}]K_i\xi \quad \text{iff} \\
&\mathfrak{M}, s \Vdash \eta_x \Rightarrow \mathfrak{M}\circ\mathfrak{E}, s\circ\eta_x \Vdash K_i\xi \quad \text{iff} \\
&\mathfrak{M}, s \Vdash \eta_x \Rightarrow (for\ any\ t\circ\eta_y\ s.t.\ (s\circ\eta_x, t\circ\eta_y) \in (R\circ\mathfrak{E})_i) \\
&\quad \mathfrak{M}\circ\mathfrak{E}, t\circ\eta_y \Vdash \xi \quad \text{iff} \\
&\mathfrak{M}, s \Vdash \eta_x \Rightarrow (for\ any\ \eta_y \in \overline{\eta_{x.i}})\,(for\ any\ t\ s.t.\ sR_it) \\
&\quad \{\mathfrak{M}, t \Vdash \eta_y \Rightarrow \mathfrak{M}\circ\mathfrak{E}, t\circ\eta_y \Vdash \xi)\} \quad \text{iff} \\
&\mathfrak{M}, s \Vdash \eta_x \Rightarrow (for\ any\ \eta_y \in \overline{\eta_{x.i}})\,(for\ any\ t\ s.t.\ sR_it) \\
&\quad \mathfrak{M}, t \Vdash [\eta_y, \overrightarrow{\eta}, \widetilde{\eta}]\xi \quad \text{iff} \\
&\mathfrak{M}, s \Vdash \eta_x \Rightarrow (for\ any\ \eta_y \in \overline{\eta_{x.i}})\,\mathfrak{M}, s \Vdash K_i[\eta_y, \overrightarrow{\eta}, \widetilde{\eta}]\xi \quad \text{iff} \\
&\mathfrak{M}, s \Vdash \eta_x \Rightarrow \mathfrak{M}, s \Vdash \textstyle\bigwedge_{\eta_y \in \overline{\eta_{x.i}}} K_i[\eta_y, \overrightarrow{\eta}, \widetilde{\eta}]\xi \quad \text{iff} \\
&\mathfrak{M}, s \Vdash \eta_x \rightarrow \textstyle\bigwedge_{\eta_y \in \overline{\eta_{x.i}}} K_i[\eta_y, \overrightarrow{\eta}, \widetilde{\eta}]\xi
\end{aligned}
\qquad \square
$$

Note that we do not need the reduction axiom for iteration case, since we can deal with these wffs from the innermost modality to the outermost one[15].

With these reduction axioms, we have similar completeness result as in **DEL** van Benthem (2008):

Corollary 1 (Completeness of **feDEL**)**.** *feDEL is axiomatized by EL(with infinitely disjunction)-axioms together with the reduction axioms stated in Theorem 1.*

3.3 Event model correspondence of feDEL and DEL

A pair of interesting questions might be: *Could all pointed event models w.r.t. **DEL** be stated in **feDEL**?* and *Could all pointed formulaic event models w.r.t. **feDEL** be stated in **DEL**?*

In **feDEL**, events are determined by their preconditions(cf. Remark 2). Thus if a **DEL** event model has two events with the same precondition, then the event model seems to be out of the range of **feDEL**. Also, we have formulaic events which have dynamic modalities in their precondition formulas, which seem difficult to be expressed by **DEL** event models. Nevertheless, by employing some tricks, we can get positive results.

Definition 3.4 (Event model correspondence between **feDEL** and **DEL**)**.** For any (**DEL**) pointed event model $\mathcal{E}, e$ and any (**feDEL**) pointed formulaic event model $\eta, \overrightarrow{\eta}, \widetilde{\eta}$:

$\mathcal{E}, e$ and $\eta, \overrightarrow{\eta}, \widetilde{\eta}$ are said to be *correspond* if

$$\mathfrak{M}, s \Vdash_{\mathbf{DEL}} [\mathcal{E}, e]\xi \ \ iff \ \ \mathfrak{M}, s \Vdash_{\mathbf{feDEL}} [\eta, \overrightarrow{\eta}, \widetilde{\eta}]\xi$$

for any pointed epistemic model $\mathfrak{M}, s$ and any **EL**−formula ξ.

A similar notion have appeared in van Eijck and Ruan (2004), with the name of "Same update effect". Since our discussion relates two class of event models(i.e., **DEL** event models and **feDEL** formulaic event models), and hence,

[15] It should be emphasized that: what we have employed are pointed formulaic event models which may have countably infinite many formulaic events. Still, we insisted on applying the wff-building operations for only finite many times(infinitely disjunctions are built by applying the corresponding operation just ones). Thus, we do not have infinitely iteration of dynamic modalities in **feDEL**.

cares about two kinds of updates together, we take the name "correspondence" instead.

Definition 3.5 (Formula correspondence between **feDEL** and **DEL**)**.** For any **DEL**−formula ϕ and any **feDEL**−formula ψ: ϕ and ψ are said to be *correspond* if

$$\mathfrak{M}, s \Vdash_{\mathbf{DEL}} \phi \ \ iff \ \ \mathfrak{M}, s \Vdash_{\mathbf{feDEL}} \psi$$

for any pointed epistemic model $\mathfrak{M}, s$.

Theorem 2 (Existence of corresponding event model)**.** *(1) For any (**DEL**) pointed event model, there is a (**feDEL**) pointed formulaic event model corresponding to it.*

*(2) For any (**feDEL**) pointed formulaic event model, there is a (**DEL**) pointed event model corresponding to it.*

Proof. (1) Suppose that a pointed event model $(\mathcal{E}, e) = (E, \{\sim_i\}_{i\in I}, \{Pre_e\}_{e\in E}, e)$ is given. Now we construct the corresponding pointed formulaic event model.

Firstly, we define $\overrightarrow{\eta}$. Since E is at most countable(cf. Definition 1.4), we can enumerate its elements, say, $e_1, e_2, \cdots$. Let:

$$\eta_j := Pre_{e_j} \underbrace{\vee\bot \cdots \vee \bot}_{z_j \ times}$$

where

$$z_j := \mu z.\ Pre_{e_j} \underbrace{\vee\bot \cdots \vee \bot}_{z \ times} \notin \{\, \eta_1, \cdots, \eta_{j-1} \,\}$$

Since Pre_{e_j} is an **EL**−formula, η_j is also an **EL**−formula. Hence we have:

$$\Vdash_{\mathbf{DEL}} Pre_{e_j} \leftrightarrow \eta_j$$

and

$$\Vdash_{\mathbf{feDEL}} Pre_{e_j} \leftrightarrow \eta_j \tag{9}$$

In this way, we transfer events with same preconditions to different, but equivalent formulas.

$\overrightarrow{\eta}$ is defined to be $(\eta_1, \eta_2, \cdots)$, while $\overline{\eta}$ being the collection of formulas in $\overrightarrow{\eta}$.

Secondly, we take the current formulaic event η to be Pre_e.

Thirdly, let $\overline{\eta_{x.i}}$ to be $\{\, \eta_y \,|\, e_x \sim_i e_y \,\}$. Since we have numbered all e_j's and all η_j's, we can specify $\eta_{x.i.1}, \eta_{x.i.2}, \cdots$ from $\overline{\eta_{x.i}}$. Then we can let $\widetilde{\eta}$ to be

$$((\eta_{1.1.1}, \eta_{1.1.2}, \cdots), (\eta_{1.2.1}, \cdots), \cdots), ((\eta_{2.1.1} \cdots), \cdots), \cdots$$

This finishes our construction of the pointed formulaic event model, i.e., $\eta, \overrightarrow{\eta}, \widetilde{\eta}$.

Now it is sufficient to verify the correspondence relation.

We employ an induction on the formula ξ, and use reduction axioms of **DEL** and **feDEL**.

$\triangleright$ $\mathfrak{M}, s \Vdash_{\mathbf{DEL}} [\mathcal{E}, e]q$ iff (reduction axioms of **DEL**)
$\mathfrak{M}, s \Vdash_{\mathbf{DEL}} Pre_e \rightarrow q$ iff (Pre_e is an **EL**−formula)
$\mathfrak{M}, s \Vdash_{\mathbf{feDEL}} Pre_e \rightarrow q$ iff (by (9))
$\mathfrak{M}, s \Vdash_{\mathbf{feDEL}} \eta \rightarrow q$ iff (reduction axioms of **feDEL**)
$\mathfrak{M}, s \Vdash_{\mathbf{feDEL}} [\eta, \overrightarrow{\eta}, \widetilde{\eta}]q$

▷ $\mathfrak{M}, s \Vdash_{\mathbf{DEL}} [\mathcal{E}, e]\neg\xi$ iff (reduction axioms of **DEL**)
$\mathfrak{M}, s \Vdash_{\mathbf{DEL}} Pre_e \to \neg[\mathcal{E}, e]\xi$
iff (i.h., Pre_e is an **EL**−formula, together with (9))
$\mathfrak{M}, s \Vdash_{\mathbf{feDEL}} \eta \to \neg[\eta, \overrightarrow{\eta}, \widetilde{\eta}]\xi$ iff (reduction axioms of **feDEL**)
$\mathfrak{M}, s \Vdash_{\mathbf{feDEL}} [\eta, \overrightarrow{\eta}, \widetilde{\eta}]\neg\xi$

▷ $\mathfrak{M}, s \Vdash_{\mathbf{DEL}} [\mathcal{E}, e] \bigwedge \overrightarrow{\xi}$ iff (reduction axioms of **DEL**)
$\mathfrak{M}, s \Vdash_{\mathbf{DEL}} \bigwedge \overrightarrow{[\mathcal{E}, e]\xi}$ iff (i.h.(many times))
$\mathfrak{M}, s \Vdash_{\mathbf{feDEL}} \bigwedge \overrightarrow{[\eta, \overrightarrow{\eta}, \widetilde{\eta}]\xi}$ iff (reduction axioms of **feDEL**)
$\mathfrak{M}, s \Vdash_{\mathbf{feDEL}} [\eta, \overrightarrow{\eta}, \widetilde{\eta}] \bigwedge \overrightarrow{\xi}$

▷ $\mathfrak{M}, s \Vdash_{\mathbf{DEL}} [\mathcal{E}, e_x]K_i\xi$ iff (reduction axioms of **DEL**)
$\mathfrak{M}, s \Vdash_{\mathbf{DEL}} Pre_{e_x} \to \bigwedge_{e_x \sim_i e_y} K_i[\mathcal{E}, e_y]\xi$
iff (i.h.[16], Pre_{e_x} is an **EL**−formula, together with (9))
$\mathfrak{M}, s \Vdash_{\mathbf{feDEL}} \eta_x \to \bigwedge_{\eta_y \in \widetilde{\eta_{x.i}}} K_i[\eta_y, \overrightarrow{\eta}, \widetilde{\eta}]\xi$
iff (reduction axioms of **feDEL**)
$\mathfrak{M}, s \Vdash_{\mathbf{feDEL}} [\eta_x, \overrightarrow{\eta}, \widetilde{\eta}]K_i\xi$

(2) The construction of corresponding **DEL** pointed event model and the verification of correspondence are almost just the inverse procedure of what we presented in (1). For formulaic event η_j with dynamic modalities, we can take the **feDEL**−equivalent **EL**−formula to be Pre_{e_j}. The reduction axioms of **feDEL** guarantee that there is such a formula. □

Note that in Definition 3.4, ξ is required to be an **EL**−formula. By reduction axioms of **DEL** and **feDEL**, for formulas with dynamic modalities, we can apply Theorem 2 on the inner most modality. Thus, this requirement is sufficient for us to get the following corollary:

Corollary 2 (Existence of corresponding formula)**.** *(1) For any **DEL**−formula ϕ, there is an **feDEL**−formula $f(\phi)$ corresponding to it.*
*(2) For any **feDEL**−formula ψ, there is a **DEL**−formula $g(\psi)$ corresponding to it.*

Proof. For notational convenience, **DEL** dynamic modalities are denoted by $\Sigma_1, \Sigma_2, \cdots$, while **feDEL** dynamic modalities are denoted by $\Delta_1, \Delta_2, \cdots$. Besides, for any **DEL**−formula α, by the reduction axioms of **DEL**, we can make the **DEL**−equivalent **EL**−formula of α, which is denoted by $\llbracket \alpha \rrbracket$. Similarly, the **feDEL**−equivalent **EL**−formula of a **feDEL**−formula β is denoted by $\|\beta\|$.

We only proof (1) here, the proof of (2) is quite similar.

Suppose that ϕ has dynamic modalities, say, $\Sigma_1, \Sigma_2, \cdots$. By Theorem 2, we can construct corresponding formulaic event models, say, $\Delta_1, \Delta_2, \cdots$.

Then $f(\phi) := \phi(\Sigma_1 \mapsto \Delta_1, \Sigma_2 \mapsto \Delta_2, \cdots)$. That is, we substitute all Σ's by corresponding Δ's.

Now we proof two claims together by an induction, i.e.:

$$\mathfrak{M}, s \Vdash_{\mathbf{DEL}} \phi \ iff \ \mathfrak{M}, s \Vdash_{\mathbf{feDEL}} f(\phi) \tag{10}$$

which is our target, and

$$\mathfrak{M}, s \Vdash_{\mathbf{DEL}} \llbracket \phi \rrbracket \ iff \ \mathfrak{M}, s \Vdash_{\mathbf{DEL}} \|f(\phi)\| \tag{11}$$

[16]By this induction, we are showing that our construction stated above can generate corresponding pointed formulaic event models, no matter which pointed event model is given in hand. Thus, we can also apply i.h. here, though it is w.r.t. $(\mathcal{E}, e_y)$, instead of $(\mathcal{E}, e_x)$.

as a by-product.

$\triangleright$ $\mathfrak{M}, s \Vdash_{\mathbf{DEL}} q$ iff $\mathfrak{M}, s \Vdash_{\mathbf{feDEL}} q$ iff $\mathfrak{M}, s \Vdash_{\mathbf{feDEL}} f(q)$.
$\mathfrak{M}, s \Vdash_{\mathbf{DEL}} \lceil\!\lceil q \rceil\!\rceil$ iff $\mathfrak{M}, s \Vdash_{\mathbf{DEL}} q$
iff $\mathfrak{M}, s \Vdash_{\mathbf{DEL}} \lfloor\!\lfloor q \rfloor\!\rfloor$ iff $\mathfrak{M}, s \Vdash_{\mathbf{DEL}} \lfloor\!\lfloor f(q) \rfloor\!\rfloor$.

$\triangleright$ $\mathfrak{M}, s \Vdash_{\mathbf{DEL}} \lceil\!\lceil \neg\phi \rceil\!\rceil$ iff $\mathfrak{M}, s \Vdash_{\mathbf{DEL}} \neg\lceil\!\lceil \phi \rceil\!\rceil$ iff (i.h. (11))
$\mathfrak{M}, s \Vdash_{\mathbf{DEL}} \neg\lfloor\!\lfloor f(\phi) \rfloor\!\rfloor$ iff $\mathfrak{M}, s \Vdash_{\mathbf{DEL}} \lfloor\!\lfloor \neg f(\phi) \rfloor\!\rfloor$
iff $\mathfrak{M}, s \Vdash_{\mathbf{DEL}} \lfloor\!\lfloor f(\neg\phi) \rfloor\!\rfloor$.
$\mathfrak{M}, s \Vdash_{\mathbf{DEL}} \neg\phi$ iff $\mathfrak{M}, s \Vdash_{\mathbf{DEL}} \lceil\!\lceil \neg\phi \rceil\!\rceil$
iff (since we have proved (11) for this case above)
$\mathfrak{M}, s \Vdash_{\mathbf{DEL}} \lfloor\!\lfloor f(\neg\phi) \rfloor\!\rfloor$ iff $\mathfrak{M}, s \Vdash_{\mathbf{feDEL}} \lfloor\!\lfloor f(\neg\phi) \rfloor\!\rfloor$
iff $\mathfrak{M}, s \Vdash_{\mathbf{feDEL}} f(\neg\phi)$.

$\triangleright$ The inductive steps for $\bigwedge$ and K_i are similar.
We also prove (11) firstly, and then prove (10) as a consequence.

$\triangleright$ $\mathfrak{M}, s \Vdash_{\mathbf{DEL}} \Sigma_i\phi$ iff $\mathfrak{M}, s \Vdash_{\mathbf{DEL}} \Sigma_i\lceil\!\lceil \phi \rceil\!\rceil$ iff (i.h. (11))
$\mathfrak{M}, s \Vdash_{\mathbf{DEL}} \Sigma_i\lfloor\!\lfloor f(\phi) \rfloor\!\rfloor$ iff (Theorem 2,
together with the fact that $\lfloor\!\lfloor f(\phi) \rfloor\!\rfloor$ is an **EL**−formula)
$\mathfrak{M}, s \Vdash_{\mathbf{feDEL}} \Delta_i\lfloor\!\lfloor f(\phi) \rfloor\!\rfloor$ iff $\mathfrak{M}, s \Vdash_{\mathbf{feDEL}} \Delta_i f(\phi)$
iff $\mathfrak{M}, s \Vdash_{\mathbf{feDEL}} f(\Sigma_i\phi)$.
$\mathfrak{M}, s \Vdash_{\mathbf{DEL}} \lceil\!\lceil \Sigma_i\phi \rceil\!\rceil$ iff $\mathfrak{M}, s \Vdash_{\mathbf{DEL}} \Sigma_i\phi$
iff (since we have proved (10) for this case above)
$\mathfrak{M}, s \Vdash_{\mathbf{feDEL}} f(\Sigma_i\phi)$ iff $\mathfrak{M}, s \Vdash_{\mathbf{feDEL}} \lfloor\!\lfloor f(\Sigma_i\phi) \rfloor\!\rfloor$
iff $\mathfrak{M}, s \Vdash_{\mathbf{DEL}} \lfloor\!\lfloor f(\Sigma_i\phi) \rfloor\!\rfloor$ □

Remark 3. *The results of Corollary 2 are trivial. For instance, we could let $f(\phi)$ to be $\lceil\!\lceil \phi \rceil\!\rceil$. However, the method of the proof is an illustration of generating corresponding formulas with "similar profiles" by employing Theorem 2.*

In the rest of this section, we consider the relationship between **feDEL** and the "special case" of **DEL**, namely, **PAL**. It was stated in van Benthem (2008) that the event model for a public announcement $!\eta$ has only one event with precondition η, and reflexive accessibility relations for all agents. By Theorem 2, we could state this pointed event model by the following pointed formulaic event model in **feDEL** :

$$\eta, (\eta), \underbrace{((\eta), \cdots, (\eta))}_{m}$$

where m is the number of agents.

4 Local Product Models

In Section 2.2, we state our syntactical repairing method in general. Recall that the key-point is to *put preconditions explicitly in events*. This repair improves the degree of formulas, and enables us to explore new notions. We will take "*local product model*" in this section as an example.

4.1 Knowledge-dependent Uncertainty Relations Between Events

In (Baltag and Moss 2004, Section 3.1), it is stated that the "actions"[17] there are "simple", since the appearances of actions to agents are uniform, that is, no matter which state an agent is in, he has the same uncertainty about events[18].

By the idea of Section 2.2, formulas are explicitly placed in events. Since satisfiabilities of formulas are sensitive to epistemic states, a nature question would be:

$$Can\ we\ define\ knowledge-dependent\ uncertainty\ relations\ between\ events?$$

Suppose that we have two events e_1 and e_2, with $\chi_{1.1}, \chi_{1.2}, \cdots$ and $\chi_{2.1}, \chi_{2.2}, \cdots$ as series of formulas respectively given in e_1 and e_2[19]. Then we could let[20]:

$$e \sim_i f \ \ iff \ \ \mathfrak{M}, s \Vdash \bigwedge_k \overrightarrow{\Diamond_i(\chi_{1.k} \leftrightarrow \chi_{2.k})} \tag{12}$$

by which, an agent may have different uncertainties about events at different epistemic states. It is clear that in (12), each event should have a same number of formulas. Besides, if we employ (12) in an event model, then it is no longer an independent model. Fortunately, we can omit uncertainty relations in the definition of event model, and put our remedy at the definition of product model. Since the resulting product models are sensitive to epistemic states, which indicates that the events are no longer "simple" in the sense of Baltag and Moss (2004), we will call them *local product models*.

In the next subsection, we construct a system to illustrate how local product models behavior.

Remark 4. *We have defined* ***feDEL****, which is our minimal method, in Section 3. Since formulas have been put into events, we can introduce local product model, instead of Definition 3.2. We may call the resulting system* ***L-feDEL***[21]*. But in that case, the idea of (12), i.e., an agent can tell two events iff he knows the preconditions of these two events are different, is not plausible. Suppose that we have an event "it rains" and an identity event "nothing happens". I may not be able to tell the precondition of "it rains", e.g., some atmospheric condition, but I can tell that it rains or not.*

4.2 The System L-pfeDEL

In this section, we play a method different from **feDEL**. In this method, we do not require events to be determined by their preconditions, and makes the semantic more plausible.

[17]The notion "action" in Baltag and Moss (2004) corresponds to "events" in this paper.

[18]It is the same case in van Benthem (2008), cf. Definition 1.4 in this paper.

[19]In Section 3, we have **feDEL** in which there is only one formula, i.e.: the precondition formula, in each event. But we can put more formulas in, provided we still meet the requirement of our syntactical idea.

[20]We use $\Diamond_i$ as an abbreviation of $\neg K_i \neg$.

[21]**L** stands for "local" here.

In the following definition of **L-pfeDEL**[22], each event is defined to be a pair of formulas, i.e.[23]:

$$event := (precondition, postcondition)$$

Definition 4.1 (Language of **L-pfeDEL**)**.** Suppose that we have an at most countable set of agents, which are denoted by $1, 2, \cdots$, then the language of **L-pfeDEL** is defined by the following inductive rules:

$$\phi ::= \bot|p|\neg\phi|\phi \vee \phi \vee \cdots |K_i\phi|[\mathfrak{E}]\phi$$

with 'p' standing for any propositional letter, and '$\mathfrak{E}$' standing for any *pointed pairwise formulaic event model* which is defined as:

$$\mathfrak{E} ::= \phi - \phi, (\underbrace{\phi - \phi, \cdots}_{n(or\omega)})$$

Note that in the language of **L-pfeDEL**, we do not include the uncertainty relations between pairwise formulaic events. The reason is we desire knowledge-dependent uncertainty relations between events, to be generated in Definition 4.2.

As in the case of **feDEL**, we have the following abbreviations:

$\overrightarrow{\varepsilon - \delta}$ denotes $(\underbrace{\varepsilon_1 - \delta_1, \cdots}_{n(or\omega)})$,

$\overline{\varepsilon - \delta}$ denotes $\{\underbrace{\varepsilon_1 - \delta_1, \cdots}_{n(or\omega)}\}$.

Definition 4.2 (Pointed Local Product Model)**.** For each pointed epistemic model $(\mathfrak{M}, s) = (W, \{R_i | i \in I\}, V, s)$, and each pointed pairwise formulaic event model $\mathfrak{E} = \varepsilon - \delta, \overrightarrow{\varepsilon - \delta}$, the *pointed local product model* $(\mathfrak{M} \circ_s \mathfrak{E}, s \circ \varepsilon - \delta)$ is:

$(W \circ \mathfrak{E}, \{(R \circ_s \mathfrak{E})_i \,|\, i \in I\}, V \circ \mathfrak{E}, s \circ \varepsilon - \delta)$ where:

$W \circ \mathfrak{E} = \{t \circ \varepsilon - \delta \,|\, t \in W \textit{ and } \varepsilon - \delta \in \overline{\varepsilon - \delta} \textit{ and } \mathfrak{M}, t \Vdash \varepsilon\}$

$(R \circ_s \mathfrak{E})_i = \{(t_1 \circ \varepsilon_1 - \delta_1, t_2 \circ \varepsilon_2 - \delta_2) \in (W \circ \mathfrak{E}) \times (W \circ \mathfrak{E}) \,|\, t_1 R_i t_2 \textit{ and } \mathfrak{M}, s \Vdash \Diamond_i(\varepsilon_1 \leftrightarrow \varepsilon_2) \wedge \Diamond_i(\delta_1 \leftrightarrow \delta_2)\}$

$V \circ \mathfrak{E} = \{t \circ \varepsilon - \delta \in W \circ \mathfrak{E} \,|\, t \in V\}$

The notion defined above is "local" since it is dependent to knowledge, i.e., dependent to the epistemic state where it occurs. The only clause which is sensitive to states is the one for uncertainty relations, which says that, an agent can tell two events iff either he knows the preconditions of these two events are different, or he knows the postconditions of these two events are different.

Now we can give the truth definition of **L-pfeDEL** as a routine.

[22] As **fe** in **feDEL**, **pfe** here means "pairwise-formulaic event", while **L** means "local".

[23] Postconditions are employed here to make the semantic of our instance more natural. This employ is not necessary. We can have local product models in any system which has formulas explicitly in events, cf. Remark 4. In one word, we treat postcondition as a tool, without caring about the meaning of it. Besides, van Benthem et al. (2006) has discussions about postcondition itself in detail.

Definition 4.3 (Truth Definition of **L-pfeDEL**). With employing corresponding clauses from epistemic logic van Benthem (2008) and propositional fragment of $\mathcal{L}_{\omega_1\omega}$ Ebbinghaus et al. (1994), we only need to state the clause for dynamic modality: for any pointed epistemic model $\mathfrak{M}, s$ and any pointed pairwise formulaic event model $\mathfrak{E}$ (with $\varepsilon - \delta$ as its current pairwise formulaic event):

$$\mathfrak{M}, s \Vdash [\mathfrak{E}]\xi \ \ iff \ \ \mathfrak{M}, s \Vdash \varepsilon \Rightarrow \mathfrak{M} \circ_s \mathfrak{E}, s \circ \varepsilon - \delta \Vdash \xi$$

The sensitivities of uncertainty relations between events to epistemic states bring something new into the field of **DEL**. However, these sensitivities also brings a technical difficulty, i.e., the reduction axioms. By an usual way, we may get a list of valid "reduction axioms". But in this time, the one for knowledge modality would have the form of

$$[\varepsilon - \delta, \overrightarrow{\varepsilon - \delta}]K_i\xi \leftrightarrow (\varepsilon \to \bigwedge_{\mathfrak{M}, s \Vdash \Diamond_i(\varepsilon \leftrightarrow \varepsilon') \wedge \Diamond_i(\delta \leftrightarrow \delta')} \nu)$$

for some formula ν, where $\Vdash$, $\mathfrak{M}$ are not syntactical symbols. This fact makes the above expression no longer a wff[24], unless we could find some way to determine that satisfiability syntactically.

There is a positive result that, we could still state **PAL** as a special "case" of **L-pfeDEL**, where the pointed pairwise formulaic event model corresponding to $!P$ is:

$$P - \top, (P - \top)$$

In fact, with our definition, any pairwise formulaic events are reflexive, provided our pointed epistemic model is reflexive, since $\Diamond_i(\varepsilon \leftrightarrow \varepsilon) \wedge \Diamond_i(\delta \leftrightarrow \delta)$ is globally true in reflexive models.

5 Conclusion

In this paper, formulaic events in the filed of **DEL** were discussed. At first, we stated a class of syntactical methods of avoiding self-referential problems in general. In a word, it was: *put preconditions explicitly in events*. Then the "minimal" method, i.e. **feDEL**, was founded and showed to be a suitable rewriting of **DEL**. After that, we went further to introduce local product models, while founding **L-pfeDEL**, which was also an instance of our syntactical methods, as a tiny step of exploring non-"simple"[25] events(or actions).

The work could be carried on in different ways. For instance, semantical methods (cf. Section 2.1) also worth exploring. During an e-mail discussion, Tomohiro Hoshi pointed out that while the semantical idea constructing an order in the set of pointed event models, the syntactical idea also constructs an order in the same set. Hence, it seems possible to join these two ideas. Also, it might be a challenging work to find out suitable reduction axioms for **L-pfeDEL**. Besides, we could check other methods in our syntactical-repair class.

[24] In **feDEL**, uncertainty relations are given syntactically. Similar for **DEL**, though in an implicit version.

[25] We use the word "simple" in the sense of (Baltag and Moss 2004, Section 3.1).

Acknowledgements This paper is finished after learning a course given by Professor Johan van Benthem in Tsinghua Univ. Beijing, 2008. The author is greatly indebted to Professor Johan van Benthem who has led me to the world of **DEL**, read earlier drafts of this paper, and given helpful suggestions. I am grateful to Doctor Alexandru Baltag, Doctor Hans van Ditmarsch, Tomohiro Hoshi, and Doctor Fenrong Liu(who is my supervisor), who have read earlier drafts and given suggestions to me. I also want to thank my classmate Minghui Ma, for all the time he shared with me, studying, arguing, or just talking during the course.

References

A. Baltag and L. Moss. Logics for epistemic programs. *Synthese*, 139:165–224, 2004.

A. Baltag, L. Moss, and S. Solecki. The logic of public announcements, common knowledge, and private suspicions. In I. Gilboa, editor, *Proceedings of the 7th Conference on Theoretical Aspects of Rationality and Knowledge (TARK VII)*, pages 43–56, Evanston, IL, U.S., 1998.

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

H. D. Ebbinghaus, J. Flum, and W. Thomas. *Mathematical Logic*. Springer-Verlag, New York, second edition, 1994.

D. Gabbay. *Labelled Deductive Systems*, volume 1. Clarendon press, Oxford, 1996.

T. Hoshi. Public announcement logic with protocal constraints. *Manuscript*, September 2008.

T. Janssen. Compositionality. In J. van Benthem and A. ter Meulen, editors, *Handbook of Logic and Linguistics*. Elsevier Science Publishers, Amsterdam, 1997.

B. Renne. *Dynamic Epistemic Logic with Justification*. PhD thesis, The City University of New York, New York, United States of America, 2008.

J. van Benthem. *Logical Dynamics of Information And Interaction*. Manuscript, 2008.

J. van Benthem, J. van Eijck, and B. Kooi. Logics of communication and change. *Information and Computation*, 244(11):1620–1662, 2006.

J. van Benthem, J. Gerbrandy, T. Hoshi, and E. Pacuit. Merging frameworks for interaction. *Manuscript*, August 2008.

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

J. van Eijck and J. Ruan. Action emulation. *Manuscript*, December 2004.

Ceteris paribus modalities and the future contingents problem

Carlo Proietti

Institut d' Histoire et de Philosophie des Sciences et des Techniques, Université de Paris 1 and Dipartimento di Filosofia, Università di Pisa

proietticarlo@hotmail.com

Abstract

This paper presents two systems of temporal logic, Λ_{CPT} and $\Lambda_{CPT@}$, with *ceteris paribus* modalities. The principal aim is to show how this approach can be useful to give an ockhamist solution to the *future contingents problem* along the same lines of A. Prior in Prior (1967). The interest of this work lies also in the fact that $\Lambda_{CPT@}$ represents an alternative modal account of supervaluationist and post-semantics approaches to temporal reasoning.

1 Introduction

The notion of *ceteris paribus* plays a fundamental role in the definition of many key concepts in different areas. For example, it is often suggested that economic and psychologic laws, but also scientific laws *tout court*, hold only *ceteris paribus*, i.e. if we consider only cases where a set of standard conditions is satisfied. In van Benthem et al. (2009) a fundamental distinction is made among two important senses in which we can use the notion of *ceteris paribus*. In the first sense, something holds *ceteris paribus* if it holds whenever a fixed set of *normal* conditions is satisfied. In the second sense, something holds *ceteris paribus* if it holds "everything else being equal", i.e. in all cases that are equivalent to the actual situation w.r.t. a given set of conditions.[1] The modal logic elaborated in van Benthem et al. (2009) for dealing with *ceteris paribus* preferences[2] represents a very useful logical tool for modeling other notions where *ceteris paribus*, in the sense of "everything else being equal", plays a central role.

In temporal reasoning we can find many notions of this second kind. In Proietti and Sandu (2008), we investigated possible solutions to the temporal version of *Fitch's paradox*, i.e. the argument against the *principle of discovery*, by

[1] See van Benthem et al. (2009) pp.98-101.

[2] This work re-elaborates and extends the seminal approach of G. H. von Wright in von Wright (1963).

employing an epistemic operator K defined as a *ceteris paribus* one. The satisfaction conditions for this operator were inspired and closely related to those of *ceteris paribus* preference operators. The central idea in Proietti and Sandu (2008) can be applied in modeling other important modal notions, e.g. modalities involved in the *future contingents problem* or in *counterfactual conditionals*.

In this paper, I will focus on the first topic and show how *ceteris paribus* temporal logic can represent an important alternative approach, w.r.t. mainstream ones, for solving this problem and the issues related to temporal reasoning with an open future. What is important is that, unlike many other philosophical solutions, we can provide an axiomatization for our *ceteris paribus-temporal* approach.

Nevertheless, I will also show that the language of simple propositional temporal logic is not expressive enough to deal with some issues presented by the future contingents problem and I will then recur to a more powerful modal language and semantics with *satisfaction operators* in order to settle this problem.

2 Discriminating among "similar" worlds and histories

The central idea of a modal logic for *ceteris paribus* preference is to introduce new modalities $[\Gamma]$ and $\langle\Gamma\rangle$, where Γ is some set of "basic formulas" of the language. The accessibility relation by which these modalities are defined is *modal equivalence*, between two points, on the set Γ. In other words, this condition tells that, when evaluating a formula $[\Gamma]\phi$ or $\langle\Gamma\rangle\phi$ at a point w, we should restrict our attention only to points being equivalent with w on a particular set of conditions dictated by the formulas in Γ.

The idea of considering sets of conditions discriminating with different sharpness among alternative states lies implicitly also in many other areas of reasoning where modalities may be applied. Consider for example the case of *counterfactual conditionals* such as "If elephants had wings they couldn't fly anyway". In the usual *Lewis-Stalnaker semantics* (*LSS*), the truth of such a conditional with a false antecedent is evaluated by considering the "most similar worlds" where the antecedent holds. In *LSS*, *similarity* is a previously given metric relation between worlds. Considering *equivalence* between points w.r.t. a set of formulas represents an alternative way to account for *similarity* between worlds. Moreover, two different sets Γ and Δ define similarity under two different respects that we may want to take into account. Thus, the kind of modalities introduced in van Benthem et al. (2009) has a potential advantage, over *LSS*, of considering not only one fixed similarity relation, but a large (potentially infinite) set.

In the next sections I will show how the very same intuition lies hidden behind a well known solution, due to Ockham, of the problem of future contingents.

2.1 The *future contingents* problem

The discussion about the *future contingents* problem dates back to Aristotle and Diodorus Chronus and has been a widely debated topic during the whole history of western philosophy. This was due to the fact that this problem is formulated as an argument whose conclusion consists in the negation of fortuitous events and, even more important, of free-will, as one can gather from the most famous presentation of it by Aristotle:

> In the case of that which is or which has taken place, propositions, whether positive or negative, must be true or false [. . .]
> When the subject, however, is individual, and that which is predicated of it relates to the future, the case is altered. For if all propositions whether positive or negative are either true or false, then any given predicate must either belong to the subject or not [. . .]
> Now if this be so, nothing is or takes place fortuitously, either in the present or in the future, and there are no real alternatives; everything takes place of *necessity* and is fixed. (see Aristotle (1941))

As the reader should have noticed, the absence of fortuitous events is a consequence, according to Aristotle, of the fact that statements concerning future events become, if true, *necessary*. Modal notions such as *necessity* and *possibility* were employed more explicitly by Diodorus in his famous *Master Argument*.[3] Medieval logicians called *necessity per accidens* the specific modal notion occurring in this context. Indeed, we are face to a particular *temporal* kind of necessity: the one that can be ascribed to past events since they are considered as *immutable*.

Nowadays, the ought-to-be diodorean argument is often presented step by step in a way that makes explicit a specific assumption on necessity *per accidens*, that is the *Principle of the Necessity of the Past* (PNP), as well as other fundamental assumptions in the argument, i.e. the *Principle of Necessity Transfer* (PNT) and an assumption on *Temporal Dependence* (TD). The argument runs as follows.

(1) If ϕ is the case at time T, then is necessary after T that it was the case that ϕ. (PNP)

(2) If it is necessary that ϕ and $\phi \Rightarrow \psi$,[4] then it is necessary that ψ. (PNT)

(3) It is true at t_1 that Peter will do X at t_3. (premise)

(4) It is necessary at t_2 (with $t_1 < t_2 < t_3$) that it was the case (at t_1) that Peter will do X at t_3. (by (1) and (3))

(5) It was the case (at t_1) that Peter will do X at t_3. $\Rightarrow$ Peter will do X at t_3 (TD)

(6) It is necessary at t_2 that Peter will do X at t_3. (by (4),(5) and (PNT))

[3] This last was actually an argument against free-will and indeterminism. The premises of this argument, according to the report of Epictetus, were:

(P1) Every true proposition concerning the past is *necessary*.

(P1) The *impossible* does not follow from the *possible*.

and the conclusion was

(C) *Possible* is only what is or will be the case.

[4] Here $\Rightarrow$ stands for the *strict conditional*.

(7) Peter can't avoid at t_2 to do X at t_3 (by (6) and the implication from necessity to unavoidability)

(8) Peter is not free at t_2 not to do X at t_3 (by (7) and the implication from unavoidability to absence of freedom)

Mainstream logico-philosophical solutions of this problem deny that all future-tensed statements about a future moment in time might have a definite truth value before this moment. In other words, they do not consider premises like (3), or others of the same form, as admissible.[5] However, there are many reasons of dissatisfaction with those approaches (two of them are mentioned in footnote 5) and, by this reason, a modal solution in a bivalent framework is also desirable.

2.2 The Ockham-Prior solution

William of Ockham brought forward the principal alternative solution[6], that consists in bringing into question the general applicability of (PNP) to premises like (3). Indeed, according to Ockham, the problem lies in the step from (3) to (4), which is made possible by the overly general formulation of the Principle of Necessity of the Past given with (1).

The crucial consideration made by Ockham is that, even if at a certain time, say t_2, it is true that ϕ = "it was the case at t_1 that Peter will do X at t_3", with $t_1 < t_2 < t_3$, ϕ should by no means be also *necessary per accidens* at t_2, and this because it concerns a future (w.r.t. t_2) point in time t_3. Thus, at t_2, ϕ is not purely about the past or, following the terminology introduced by M. Adams Adams (1967), does not concern a *hard past fact*[7]. Ockham's solution was firstly investigated with formal methods by Prior (see Prior (1967) chap. 7) in a system of temporal logic with an additional operator $\Box$ representing necessity *per accidens*.

Given a set $\Phi = \{p, q, r, \dots\}$ of propositional letters, we define the language $\mathcal{L}$ recursively as follows:

$$\mathcal{L} = \Phi \mid \bot \mid \neg\phi \mid \phi \wedge \psi \mid F\phi \mid P\phi \mid \Box\phi$$

We define also G as $\neg F \neg$, H as $\neg P \neg$ and $\Diamond$ (possibility *per accidens*) as $\neg\Box\neg$.

Since the past and present were considered by Ockham as *immutable* or *fixed*, whereas the future is *open*, it was natural for Prior to interpret this language in

[5]This line of thought gathers at least two different logical approaches, such as many-valued logics on the one hand and gappy semantics, including J. MacFarlane's *postsemantics* (see MacFarlane (2003) and MacFarlane (2007)), on the other hand. However, this very natural move of rejecting truth *ante rem* comes with many disadvantages. For example, a well-known problem with many-valued logics is that one should assign counterintuitive truth-values to some future-tensed statements, e.g. some future-tensed tautologies are not evaluated with truth-value 1. On the other side, a major inconvenient of gappy semantics (like supervaluationist approaches and their refinements), lies in the fact that no clear notion is available of what should be a *logical calculus* or a semantic notion of *logical consequence* for them. See Varzi (2007) for an exhaustive discussion on this point.

[6]There is actually a third possible line of answering the argument, that would consist in blocking the steps from (6) to (7), and then (8), i.e. the jump from necessity to the absence of freedom. This move is based on a conceptual distinction among logical determinism (based on necessity) and causal determinism (based on the impossibility of acting otherwise). An example of this answer to the diodorean argument was, in my view, given by Cicero in *De Fato*.

[7]Statements like the one in the example are instead called *soft* past facts.

tree-like frames. Frames of this kind are couples $\mathcal{F} = (\mathcal{T}, <)$, where $\mathcal{T}$ is a set with elements $m_1, m_2, \ldots$, intuitively *moments*, and $<$ is a strict ordering relation on $\mathcal{T}$ (i.e. transitive, asymmetric and irreflexive). $\mathcal{F}$ has a *tree-like* form when it fulfills the following condition[8]:

$$\forall m_1 \forall m_2 \forall m_3 ((m_1 < m_3 \wedge m_2 < m_3) \rightarrow (m_1 < m_2 \vee m_2 < m_1 \vee m_1 = m_2))$$

Let us define a *history* as a *maximal chain* in $\mathcal{T}$ for the relation $<$. The set of the histories $h_1, h_2, \ldots$ in $\mathcal{T}$ will be $\mathcal{H}(\mathcal{T})$. Given a moment m, H_m stands for the set of the histories containing m.

A model $\mathcal{M}$ based on a frame $\mathcal{F}$ is a couple $\mathcal{M} = (\mathcal{F}, V)$ where V is a function $V : \Phi \longrightarrow \mathcal{P}(\mathcal{T} \times \mathcal{H}(\mathcal{T}))$ satisfying the following *dependence condition*:

$$(m, h) \in V(p) \Rightarrow \forall h' \in H_m (m, h') \in V(p)$$

This condition means that every atomic sentence p, which we can suppose to represent an *immediate* or *purely present* fact, should be evaluated only relative to a moment m and not to a particular history. Nevertheless, satisfaction in the general case is defined with respect to a couple (m, h) of a moment and a history in a recursive way:

$$\begin{array}{rcl}
\mathcal{M}, (m,h) \models p & \text{iff} & (m,h) \in V(p) \\
\mathcal{M}, (m,h) \models \neg\phi & \text{iff} & \mathcal{M}, (m,h) \not\models \phi \\
\mathcal{M}, (m,h) \models \phi \wedge \psi & \text{iff} & \mathcal{M}, (m,h) \models \phi \text{ and } \mathcal{M}, (m,h) \models \psi \\
\mathcal{M}, (m,h) \models \phi \vee \psi & \text{iff} & \mathcal{M}, (m,h) \models \phi \text{ or } \mathcal{M}, (m,h) \models \psi \\
\mathcal{M}, (m,h) \models P\phi & \text{iff} & \text{for some } m' < m,\ \mathcal{M}, (m',h) \models \phi \\
\mathcal{M}, (m,h) \models F\phi & \text{iff} & \text{for some } m' > m,\ \mathcal{M}, (m',h) \models \phi \\
\mathcal{M}, (m,h) \models \Box\phi & \text{iff} & \text{for all } h',\ (h' \in H_m \Rightarrow \mathcal{M}, (m,h') \models \phi)
\end{array}$$

The recursive clause for $\Box$ is easy to understand. If we look to a specific m as a node where many future courses of events (histories) branch, then we can say that a particular ϕ is *necessary per accidens* (or *determined*), at a particular moment m, only if ϕ is true at m relative to every possible future course of events, i.e. it is not something that depends on what is the future history.

In this language, the general version of (PNP) occurring in the diodorean argument can be easily rendered as the schema:

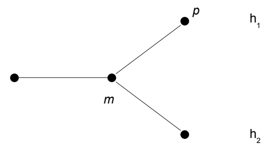
(PNP) $P\phi \rightarrow \Box P\phi$

Here, formulas representing *hard* past facts are typically those of the form $P\phi$, where ϕ does not contains any future-tense operator like F or G, whereas formulas such as $PF\phi$ or $PG\phi$ may sometimes represent *soft* past facts[9] In Prior's semantics it is a straightforward task to falsify (PNP) in the case of these last formulas, as it was requested by Ockham. This is showed in the model of Figure 1(a), where $(m, h_1) \models PFp$ but $(m, h_1) \not\models \Box PFp$.

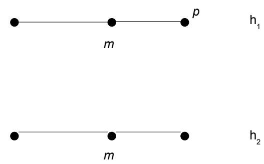
[8] Another natural condition, but which doesn't play any role in Prior's construction, is that every two points in $\mathcal{F}$ should have a *common root*:

$$\forall m_1 \forall m_2 \exists m_0 (m_0 \leq m_1 \vee m_0 \leq m_2)$$

[9] An exact distinction between formulas representing hard past and soft past facts cannot be traced in this context, due to the limited expressivity of the language.



(a) a tree-like model



(b) linear displaying

Figure 1: Ockhamist models

Despite all that, a controversial point about Prior's interpretation of Ockham is the absence, in tree-like models, of the *actual* future: many future histories are possible at a moment m, but no one of them is really *the* future history.[10] Many criticisms were addressed against this view, claiming that the *actual* future plays indeed an important role in Ockham's system of thought (at least because God knows it!). Observations of this kind, as well as other motivations, independent from Ockham's exegesis, motivated the search of so-called *actualist* solutions to the problem of future contingents[11]. These solutions are often based on tree-like models[12].

According to my view, in order to understand the essence of *necessity per accidens* and similar modalities, we should move from a slightly different perspective, making a kind of *gestalt switch*, that I will illustrate in the next section.

[10]Indeed, according to Prior, an evaluation w.r.t. a particular history is only *prima facie*.
[11]D. Lewis formulated many strong arguments for the indispensability of referring to the actual future in (Lewis (1986) pp. 203-209). Other arguments are given in Freddoso (1983).
[12]See for example Barcellan and Zanardo (1999), or Roy (2004) for an overview.

2.3 Multi-linear frames

It is important, for the points I want to stress, to notice that the tree-like structure of models is not essential for the ockhamist solution[13]. Indeed, an equivalent alternative way to represent the same tree in Figure 1(a) is to display it in a multi-linear frame as it is done in Figure 1(b).

Models of this kind are based on frames $\mathcal{F} = (\mathcal{T}, <)$, where $<$ is a *strict partial order* (SPO) with *no branching on the right* and *no branching on the left* (but it is not a *total* one). The set Φ is evaluated, as before, on subsets of $\mathcal{T} \times \mathcal{H}(T)$ and the satisfaction clauses for boolean and temporal operators are the same. The main difference is that we can let fall the *dependence condition*, and have different histories containing the same moment m but diverging at m or even at moments prior to it.

If we let fall the *dependence condition*, then the operator $\Box$ should be defined in an alternative way. Indeed, we need to consider as the same all histories sharing the same initial segment up to a certain m. This can be done via the following notion of equivalence between histories.

Definition 2.1 (Equivalence up to m)**.** Two histories h and h' are equivalent up to m, $h \sim_m h'$, if and only if for all $m' \leq m$ and for all $p \in \Phi$ we have $(m', h) \in V(p)$ iff $(m', h') \in V(p)$

We can then give the following satisfaction clause for $\Box$.

$$\mathcal{M}, (m, h) \models \Box\phi \quad \text{iff} \quad \text{for all } h', (h' \sim_m h \Rightarrow \mathcal{M}, (m, h') \models \phi)$$

and this has the same effect of falsifying (PNP) (see Figure 1(b)).

Multi-linear displaying is not just an alternative equivalent way to represent a priorean tree. Indeed, looking things from the multi-linear perspective, we can notice that $\Box$ (resp. $\Diamond$) is only one among many necessity (resp. possibility) operators that can be defined via different notions of equivalence between histories. We often make use of present-tensed notions of possibility which do not exclude histories diverging in the past. We can see it in the following example:

- Maybe someone entered in this caveau last night, but everything has been left untouched

where the "maybe" expresses a present alternative with a different past history, i.e. where someone entered the caveau (supposed that actually no one did).[14] We can define modal operators like this one just by taking a different equivalence relation: in this case one that does not discriminate between histories in which someone entered the caveau and histories in which no one did[15].

[13]This point was made also by R. Thomason and by A. J. Freddoso (Freddoso (1983)), who firstly proposed a linear solution of the kind that I am going to illustrate.

[14]This particular sentence could perhaps be paraphrased via an equivalent formula in Prior's language, but this is not so important: examples of this kind can be multiplied and it is unlikely to be the case that using the necessity *per accidens* operator we can capture all of them.

[15]More close to our topic, the debate on *future contingents* among medieval logicians is another example of the utility of discriminating among histories in many different ways. Contrary to Ockham, for whom the present is fixed, like the *hard past*, Duns Scotus' system admitted many alternative possible presents. We can interpret this difference via two alternative equivalence relations between histories: Ockham's equivalence relation should include the present moment in time, while Scotus' one should not.

The purpose of the next section is to introduce a logic for dealing with all these modalities at once. Achieving this general goal will also provide us with a logical framework for an ockhamist solution of the *future contingents* problem. This is even more important due to the fact that, as I mentioned before (see p. 4 footnote 5), a major trouble of mainstream *open-future* approaches to this problem, is that they lack an underlying calculus. Indeed, also in his tree-like ockhamist semantics in (Prior (1967) chap. 7), Prior gives only some of the valid schemes governing the necessity *per accidens* operator.

3 *Ceteris paribus* temporal language and logic.

Let us build a multi-modal language $\mathcal{L}_{\mathcal{CPT}}$, by replacing $\Box$ and $\Diamond$ with new modalities $[\Gamma]$ (a necessity operator) and $\langle\Gamma\rangle$ (a possibility operator). The new language is defined as follows:

$$\mathcal{L}_{\mathcal{CPT}} ::= \Phi \mid \neg\phi \mid \phi \wedge \psi \mid F\phi \mid P\phi \mid \langle\Gamma\rangle\phi$$

where Γ is any (possibly infinite) set of temporal formulas, let us call this fragment $\mathcal{L}_{\mathcal{T}}$, i.e. formulas built only by using boolean constructions and temporal operators such as P and F (or their duals H and G). We also define, in the usual way, $[\Gamma]$ as $\neg\langle\Gamma\rangle\neg$.

In order to have a more general approach to temporal semantics, we can abstract from moments and histories and consider, more generally, the class C_{CPT} of frames of kind $\mathcal{F} = (W, R_F, R_P)$, where W is any set of points, R_F and R_P are *strict partial orders* and R_P is the converse relation of R_F[16]. Satisfaction is defined in the usual way for boolean and temporal constructions.

As we hinted before, the satisfaction clauses for $[\Gamma]$ and $\langle\Gamma\rangle$ will be based on modal equivalence $\equiv_\Gamma$ between two points.

Definition 3.1 ($\equiv_\Gamma$). $w \equiv_\Gamma v$ if and only if for all $\phi \in \Gamma$

$$w \models \phi \text{ iff } v \models \phi$$

Definition 3.2 (Models). A model $\mathcal{M}$ is a tuple $(W, R_F, R_P, \{\equiv_\Gamma\}_{\Gamma \in \mathcal{L}_T}, V)$, where (W, R_F, R_P) belongs to C_{CPT}, V is a valuation and $\equiv_\Gamma$ is as in 3.1.

Satisfaction is defined in a natural way for the new necessity operators

$$\mathcal{M}, w \models [\Gamma]\phi \quad \text{iff} \quad \text{for all } v, (w \equiv_\Gamma v \Rightarrow \mathcal{M}, v \models \phi)$$

and the corresponding possibility operator $\langle\Gamma\rangle$ will have the following clause.

$$\mathcal{M}, w \models \langle\Gamma\rangle\phi \quad \text{iff} \quad \text{for some } v, (w \equiv_\Gamma v \text{ and } \mathcal{M}, v \models \phi)$$

A sound and complete axiomatization is given in the Appendix.

Now the question is how to apply the ockhamist multi-linear solution in this framework. In order to do that, we should find a Γ such that $\equiv_\Gamma$ mimics the notion of equivalence up to a certain moment. Let us consider the following fragment O (for Ockham) of $\mathcal{L}_{CPT}$

$$O = \Phi \mid \bot \mid \neg\phi \mid \phi \wedge \psi \mid P\phi$$

[16]We will maintain the symbol < as a shorthand for R_F (and > for R_P) in cases where this doesn't create any problem.

Indeed, $\equiv_O$ is a relation of equivalence *up to the present*, since it takes into account only present and past-tensed formulas, which is analogous to the relations of type $\sim_m$ introduced in the preceding section. It is then very natural, on this basis, to interpret the operator $[O]$ as a necessity *per accidens* operator. Nevertheless, this is exact only under some conditions. Indeed, by a closer examination, we can observe that the relation $\sim_m$ induces a partial isomorphism between the points of two different histories. But modal equivalence cannot guarantee isomorphism: at best it can impose, under some specific conditions, bisimilarity. Let us introduce some definitions and results in order to explain better this limitation and its consequences.

Definition 3.3 (R_P-bisimulation)**.** Consider two models $\mathcal{M} = (W, R_F, R_P, \{\equiv_\Gamma\}_{\Gamma\in\mathcal{L}_T}, V)$ and $\mathcal{M}' = (W', R'_F, R'_P, \{\equiv_\Gamma\}_{\Gamma\in\mathcal{L}_T}, V')$. A non-empty relation $Z \subseteq W \times W'$ is a R_P-bisimulation between $\mathcal{M}$ et $\mathcal{M}'$ iff the following conditions are satisfied :

(i) If wZw' then $w \in V(p)$ iff $w' \in V(p)$.

(ii) If wZw' and wR_Pv then for some v' in $\mathcal{M}'$, vZv' and $w'R'_pv'$ (*forth condition*).

(iii) If wZw' and $w'R'_pv'$ then for some v in $\mathcal{M}$, vZv' and wR_Pv (*back condition*).

Two points w and w' are R_P-bisimilar, we write $w \approx_P w'$, if there is an R_P-bisimulation Z such that wZw'. It is not difficult to show, by standard techniques, that if $w \approx_P w'$, then $w \equiv_O w'$. This can be proved by the same proceeding employed in Blackburn et al. (2001) sect. 2.2, theorem 2.20. The converse holds under the condition of the *Hennessy-Milner theorem* (Blackburn et al. (2001) p. 69). We say that a model $\mathcal{M}$ is *image-finite* w.r.t. to a relation R if and only if for each state w the set $\{v \mid wRv\}$ is finite. We can then prove the following result:

Theorem 1 (Corollary of *Hennessy-Milner theorem*)**.** *Let $\mathcal{M}$ and $\mathcal{M}'$ be two models that are image-finite w.r.t. R_P. Then, for every $w \in W$ and $w' \in W'$, $w \approx_P w'$ iff $w \equiv_O w'$.*

Proof. The difficult direction, from right to left, can be proved by showing that $\equiv_O$ is indeed an R_P-bisimulation. The proceeding is the same employed in Blackburn et al. (2001) (p. 69, thm. 2.24), but restricted to formulas in O. □

A function f from $\mathcal{M}$ to $\mathcal{M}'$ satisfying (i)-(iii) is a special case of a R_P-bisimulation, which should be called, using the standard terminology, a R_P-*bounded morphism*. It is easy to observe that the relation $\sim_m$ obtains between two histories h and h' if and only if there is a function f_m between their points such that $(m, h') = f_m(m, h)$ and f_m is a R_P-*bounded morphism* which is also a *bijection* between the predecessors of (m, h) and (m, h'). Actually, we can employ the Hennessy-Milner result to give a sufficient condition for $\equiv_O$ to force an R_P-bounded morphism fulfilling such condition. In the following result $R_P(w) = \{v \in W \mid wR_Pv\}$.

Theorem 2 (Conditions for similarity)**.** *Let $\mathcal{M}$ and $\mathcal{M}'$ be two models based on C_{CPT}-frames. If R_P is conversely well-founded, then, for every $w \in W$ and $w' \in W'$, $w \equiv_O w'$ if and only if there is a function f such that $w' = f(w)$, it is a R_P-bounded morphism and it is a bijection among $R_P(w)$ and $R_P(w')$.*

Proof. The direction from right to left is immediate. For the opposite direction, suppose that $w \equiv_O w'$. We want to show that $\equiv_O$ determines the function f fulfilling the above conditions. We start by defining $f(w) = w'$ and with the remark that, given the conditions on frames and the relation R_P, $R_P(w)$ and $R_P(w')$ are indeed two finite sets strictly and totally ordered by R_P. We can show that $R_P(w)$ and $R_P(w')$ have the same cardinality. Suppose that w has exactly n R_P-predecessors. In the class of frames we are considering, this fact is captured by the formula $\psi := \bigwedge_{0 \leq j \leq n} P^j \top \wedge P^n H \bot$. We have that $w \models \psi$ and, since $\psi \in O$ and $w \equiv_O w'$, then $w' \models \psi$ and thus w' has n predecessors. Moreover, both w and w' should have an immediate predecessor, say v and v', and it is easy to show that $v \equiv_O v'$. Suppose it was not the case; then there would be a $\psi_1 \in O$ such that $v \models \psi_1$ and $v' \not\models \psi_1$. Moreover, for $\phi := \psi_1 \wedge \bigwedge_{0 \leq j \leq n-1} P^j \top \wedge P^{n-1} H \bot$, we have that $v \models \phi$. No one among v' and its predecessors could satisfy this formula. Thus, $w \models P\phi$ and $w' \not\models P\phi$. But this contradicts $w \equiv_O w'$, since $P\phi \in O$. Thus, $v \equiv_O v'$ and we can take $f(v) = v'$. We can iterate this process by considering succively all the couples of immediate predecessors. It is not difficult to show that the correspondence thus established fulfills the required conditions. $\square$

This result has a precise meaning for our topic: if time is *discrete* and has a beginning, then modal equivalence guarantees the desired conditions; otherwise there might be counterexamples: one of them is illustrated in figure 2. This is a problem if we suppose time as being a *dense* order and/or with no beginning.

$$
\begin{array}{ccccccccccccccc}
\cdots & & -5 & & -4 & & -3 & & -2 & & -1 & & 0 & & \cdots \\
\cdots & \longleftarrow & \bullet p & \longleftarrow & \bullet p & \longleftarrow & \bullet & \longleftarrow & \bullet & \longleftarrow & \bullet p & \longleftarrow & \bullet p & \longleftarrow \cdots & h_1 \\
\cdots & \longleftarrow & \bullet & \longleftarrow & \bullet p & \longleftarrow & \bullet & \longleftarrow & \bullet p & \longleftarrow & \bullet & \longleftarrow & \bullet p & \longleftarrow \cdots & h_2
\end{array}
$$

Figure 2: Two R_P-bisimilar non isomorphic histories

It seems then that we cannot fully reproduce, via our language, the notion of *equivalence up to m* which is required to define *necessity per accidens*. Isomorphism is indeed a very strong condition which presumably cannot be reproduced by our basic linguistic means[17].

Nevertheless, we can do better than that by using a much expressive language and semantics, as we did in Proietti and Sandu (2008). Before doing that, we will illustrate another important motivation for proceeding this way. A reason for enriching the expressive power of our language is provided by an important puzzle in the discussion about future contingents: the so-called argument *de praesenti ad praeteritum*, which deserves an accurate presentation and discussion.

[17]Indeed, it is known that even in the richer language of first order logic elementary equivalence is not enough to guarantee isomorphism

4 The argument *de praesenti ad praeteritum*

The basic intuition according to which a sentence like "there will be a sea-battle tomorrow", uttered now by Jake, is neither true nor false seems to give an overwhelming argument against bivalence. According to J. MacFarlane (MacFarlane (2003)), this so-called *indeterminacy intuition* is counterbalanced by another kind of consideration, amounting to a *determinacy intuition*, that he introduces by the following example.

> But now what about someone who is assessing Jake's utterance from some point in the future ? Sally is hanging onto the mast, deafened by the roar of the cannon. She turns to Jake and says 'your assertion yesterday turned out to be true'. Sally's reasoning seems to be unimpeachable :
>
> Jake asserted yesterday that there would be a sea-battle today.
> There is a sea-battle today.
> So Jake's assertion was true.
>
> When we take this retrospective view, we are driven to assign a determinate truth-value to Jake's utterance. This is the *determinacy intuition*. (MacFarlane (2003) p. 322)

Arguments of this form are also called *de praesenti ad praeteritum*, i.e. reasonings from present to past. MacFarlane's principal effort is to reconcile these two opposed intuitions in a special semantic framework with truth-value gaps, that he calls *post-semantics*. His most important move consists in discriminating not only among different *contexts of utterance*, but also among different *contexts of assessment*: a particular sentence, like the one we are considering, can be nor true nor false, if assessed when Jake is uttering it. But it can also be true, if assessed in the context described in the quotation above, or either false, if assessed the day after when no sea-battle happened.

One of the motivations for this move is to account for determinacy by avoiding at the same time actualist solutions about truth. The main question is now if truth is always to be considered as a property of utterances in a given context. In the more recent MacFarlane (2007), MacFarlane admits that this is not always the case, indeed:

> ... in ordinary speech, truth and falsity are almost invariably predicated of propositions, as in the following:
>
> **(30)** What he said is false
>
> **(31)** Nothing George asserted in his talk is true
>
> **(32)** I know you believe he's dishonest, but that's false
>
> **(33)** It's true that it has been a hot summer
>
> **(34)** That was a false claim
>
> Aside from a few, relatively isolated examples, like
>
> **(35)** A truer sentence was never spoken

> people do not apply the predicate "true" and "false" to sentences or utterances, except in areas of philosophical incursion. It surprised me a bit when I realized this, because it is very common in *philosophical prose* to predicate "true" and "false" of sentences and utterances. These uses, however, must be understood as technical. (MacFarlane (2007) p. 16)

Propositions, in the sense illustrated by MacFarlane, have a *fixed truth-value* that is not context dependent. In our case, context independence means simply *atemporality*. In standard propositional temporal languages and semantics the word 'proposition' has not this meaning: a propositional letter p is "sometimes true, sometimes false". Nevertheless, a typical mean for context disambiguation consists in using some rigidifying operators, like satisfaction operators in hybrid logics. This is the main motivation for introducing, as it is done in the next section, a language with rigidifying operators and a model-theoretic interpretation for it. This step will be useful also for answering the other problems of expressivity hinted before.

5 The language $\mathcal{L}_{CPT@}$ and its semantics

The extended language $\mathcal{L}_{CPT@}$ is based, in addition to a set of propositional variables Φ, on another set $\mathcal{T}^*$, disjoint with respect to Φ. We will call it the set of *temporal indexes*. In addition we have, for every index m^*, a new *satisfaction operator* $@_{m^*}$. The language is built as follows:

$$\mathcal{L}_{\mathcal{CPT}@} ::= m^* \mid p \mid \bot \mid \neg\phi \mid \phi \wedge \psi \mid P\phi \mid F\phi \mid @_{m^*}\phi \mid \langle\Gamma\rangle\phi$$

Here Γ is, as before, any (possibly infinite) set of temporal formulas, including the constructions with temporal indexes and satisfaction operators.

Our language is always interpreted on strictly and partially ordered models of type $\mathcal{M} = (W, R_F, R_P, \{\equiv_\Gamma\}_{\Gamma\in\mathcal{L}_T}, V)$, but with a further specification concerning V. In standard hybrid logics, where indexes behave like names, we should restrict the interpretation of every index i to a single point: V(i) is a singleton. Here, every index should instead denote a single point in every chain, i.e. in every history.

Consider the relations R_F, R_P and their reflexive and transitive closures R_F^* and R_P^*. The relation $R_F^* \cup R_P^*$, is an equivalence relation which determines a partition of W in many equivalence classes $C_1, C_2, \ldots$. For every point w, we will denote by C_w the equivalence class of w. The valuation V should assign a single point in every class C_w to every index m^*. This point will be indicated as

$d_{C_w}(m)$, i.e. the denotation of m in C_w. We can then give the following definition:

$$\begin{array}{rll}
M, w \models p & \text{iff} & w \in V(p) \\
M, w \models m & \text{iff} & w \in V(m) \\
M, w \models \neg\phi & \text{iff} & M, w \not\models \phi \\
M, w \models \phi \wedge \psi & \text{iff} & M, w \models \phi \text{ and } M, w \models \psi \\
M, w \models F\phi & \text{iff} & \exists t'(tR_F t' \text{ and } M, t' \models \phi) \\
M, w \models P\phi & \text{iff} & \exists t'(tR_P t' \text{ and } M, t' \models \phi) \\
M, w \models @_m\phi & \text{iff} & M, d_{C_w}(m) \models \phi \\
M, w \models [\Gamma]\phi & \text{iff} & \text{for all } v, (w \equiv_\Gamma v \Rightarrow \mathcal{M}, v \models \phi)
\end{array}$$

Axioms of hybrid logic are still sound for this revised semantics (see Appendix). We suspect that completeness should be provable by adding these axioms to those of *ceteris paribus* logics, but this result is still to be proved.

What is the benefit of expanding the language in this way? First of all, we can fulfill MacFarlane's requirement of referring to *atemporal contents* (or *propositions*) in a very simple way: given any formula ϕ, $@_m\phi$ is the temporal disambiguation, referring to a particular moment, of ϕ. Indeed, atemporality of propositions is expressed by the following validity, in our models, of the following schema:

$$@_m\phi \leftrightarrow H@_m\phi \wedge @_m\phi \wedge G@_m\phi$$

Nevertheless, it is no difficult to verify that we can avoid the modal collapse $@_m\phi \leftrightarrow [\Gamma]@_m\phi$ in the same way we avoided the collapse $P\phi \leftrightarrow \Box P\phi$.

Indeed, suppose that we have a strict ordering $<_T$ on the set of temporal indexes. We can then define new necessity *per accidens* operators. Given any temporal index m, let us define

$$O_m = \{\pm @_{m^*} p \mid p \in \Phi \text{ and } m^* \leq_T m\} \cup \{\pm m' \mid m' \in \mathcal{T}^*\}$$

We can then define an equivalence relation $\equiv_{O_m}$ in the same way we did in section 6. Intuitively, two points are O_m-equivalent if they are named by the same temporal index, say m, and they satisfy the same atomic propositions in points named by indexes lower or equal than m.

In cases where we can attribute a name to every single point this condition is equivalent with isomorphism: for example in the case of models $\mathcal{T} \times \mathcal{H}(\mathcal{T})$, where every moment is named.

Let us now consider MacFarlane's account of the sea-battle example via the model in Figure 3, where p stands for "there is a sea battle", w is the moment in time, named by m, when Jake made the assertion and w' is the moment, named by m', when Sally considers that what Jake's asserted yesterday not only is but also was true. Sally's reasoning here is unimpeachable because of the logical equivalence of $@_{m'}p$ with $P@_{m'}p$ or also with $@_m@_{m'}p$. Nevertheless, there is still the possibility of accounting for MacFarlane's multiple contextuality by differentiating among simple truth and determination (or necessity *per accidens*): this last represents the notion of *super-truth* that a post-semanticist takes to be the genuine one. The difference among simple truth and necessity of a propositional content $@_{m'}p$ at two different points in time is represented by the different truth value, at points in the model of Figure 3 of (1) and (2). Indeed (1)

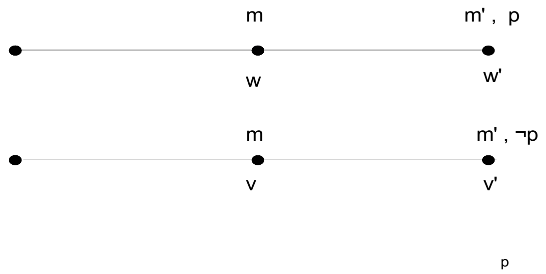


Figure 3:

is satisfied at w, where the content of Jake's assertion is simply true, whereas (2) is satisfied at w', meaning that the same propositional content is now necessary or super-true. (4) is instead satisfied at v', named by m, i.e the situation where $@_{m'}p$ is determinately not true.

$$w \models @_{m'}p \wedge \neg[O_m]@_{m'}p \tag{1}$$

$$w' \models @_{m'}p \wedge [O_m]@_{m'}p \tag{2}$$

$$v \models \neg @_{m'}p \wedge \langle O_m \rangle @_{m'}p \tag{3}$$

$$v'' \models \neg @_{m'}p \wedge \neg \langle O_m \rangle @_{m'}p \tag{4}$$

The operators $[O_m]$ can thus serve, in this context, to reproduce the difference among truth and super-truth and account for the determinacy and the indeterminacy intuitions.

As we said before, the relation $\equiv_{O_m}$ forces the semantic condition required by $\sim_m$ in models where every point is named and in every history the points are named in the same order. This last condition can be fulfilled, in hybrid logics, by axioms like $m \rightarrow Pm'$ or $m \rightarrow Fm'$.

6 Conclusions

The aim of this paper was to show how the *ceteris paribus* modal approach can be useful also in temporal puzzles like the *future contingents* problem. Not only it can account for the ockhamist solution, but it can also represent a basis for a modal translation of supervaluationist and postsemantical approaches to temporal reasoning. The main philosophical perplexity against the semantics I presented and, more generally, against the actualist semantics is that truth of future statements is, since bivalent, determined *ante rem*. This seems to be, after all, yet a bad point for free-will. But it is important for me to point out again that the truth predicate may be employed in different senses in natural reasoning and that identifying the "genuine" notion of truth with that of super-truth or determination could be problematic in many contexts: one of them is the argument *de praesenti ad praeteritum*, but many other convincing arguments in this sense were given in Lewis (1986) (see pp. 192-209). For this reason I find important to distinguish among simple truth and other notions and I think that modal logic can still be useful in this sense.

Appendix: Λ_{CPT}

Ceteris paribus temporal logic Λ_{CPT} is built over the following schemata:

(P)	All tautologies of propositional calculus
(K_F)	$G(p \to q) \to (Gp \to Gq)$
(K_P)	$H(p \to q) \to (Hp \to Hq)$
(Conv)	$(p \to HFp) \land (p \to GPp)$
(4_F)	$Gp \to GGp$
(4_P)	$Hp \to HHp$
(3_F)	$Fp \land Fq \to F(p \land Fq) \lor F(p \land q) \lor F(Fp \land q)$
(3_P)	$Pp \land Pq \to P(p \land Pq) \lor P(p \land q) \lor P(Pp \land q)$
$(K_{[\Gamma]})$	$[\Gamma](p \to q) \to ([\Gamma]p \to [\Gamma]q)$
$(T_{[\Gamma]})$	$[\Gamma]p \to p$
$(4_{[\Gamma]})$	$[\Gamma]p \to [\Gamma][\Gamma]p$
$(5_{[\Gamma]})$	$\langle\Gamma\rangle p \to [\Gamma]\langle\Gamma\rangle p$
(A1)	$\langle\Gamma\rangle\phi \to \phi$
(A2)	$\langle\Gamma\rangle\neg\phi \to \neg\phi$
(A3)	$\langle\Gamma'\rangle\phi \to \langle\Gamma\rangle\phi$

In axioms (A1)-(A3) $\phi \in \Gamma$ and $\Gamma \subseteq \Gamma'$. Inference rules are the following:

(MP)	If $\vdash \phi \to \psi$ and $\vdash \phi$ then $\vdash \psi$
(G-gen)	If $\vdash \phi$ then $\vdash G\phi$
(H-gen)	If $\vdash \phi$ then $\vdash H\phi$
($[\Gamma]$-gen)	If $\vdash \phi$ then $\vdash [\Gamma]\phi$
(Subst)	If $\vdash \phi$ then $\vdash \phi^\sigma$

It is possible to prove, by the same techniques used in van Benthem et al. (2009), the following completeness result.

Theorem 3 (Completeness for Λ_{CPT})**.** *For any set of formulas Δ and any formula ϕ*

$$\Delta \models_{C_{CPT}} \phi \text{ if and only if } \Delta \vdash_{\Lambda_{CPT}} \phi$$

Proof. (P)-(3_P) are usual axioms for temporal logics with transitive non branching time. Axioms $(K_{[\Gamma]})$-$(5_{[\Gamma]})$ are the S5 axioms for $[\Gamma]$; they are sound since $\equiv_\Gamma$ is an equivalence relation. Soundness of axioms (A1)-(A3) is easy to prove (see also van Benthem et al. (2009) pp. 104-105). The completeness direction is a more difficult task that works, as for preference logic, by *bulldozing* the canonical model, as I did in Proietti (2008). □

Appendix: $\Lambda_{CPT@}$

It is not a difficult task to verify, by usual methods, that axioms and rules employed in hybrid logics (see Blackburn et al. (2001) sect 7.3 or Blackburn and ten Cate (2006)) are still sound in our semantics. Another advantage of hybrid logics, which is preserved in this framework, is the possibility of defining frame properties, like asymmetry or irreflexivity, that are not definable with a simple modal propositional language. This means also that we need not to transform the canonical model via bulldozing in order to prove completeness. We have a set of K-axioms:

$$
\begin{array}{rl}
(P) & \text{All propositional tautologies} \\
(K_F) & G(p \to q) \to (Gp \to Gq) \\
(K_P) & H(p \to q) \to (Hp \to Hq) \\
(K_{@_m}) & @_m(p \to q) \to (@_m p \to @_m q) \\
(\text{Conv}) & @_m Fm' \leftrightarrow @_{m'} Pm
\end{array}
$$

A set of axioms specific for operators $@_m$

$$
\begin{array}{rl}
(\text{Selfdual}) & @_m p \leftrightarrow \neg @_m \neg p \\
(\text{Ref}) & @_m m \\
(\text{Agree}) & @_m @_{m'} p \leftrightarrow @_{m'} p \\
(\text{Intro}) & m \to (p \leftrightarrow @_m p) \\
(\text{Back}_F) & F @_m p \to @_m p \\
(\text{Back}_P) & P @_m p \to @_m p
\end{array}
$$

a set of temporal axioms

$$
\begin{array}{rl}
(\text{irreflexivity of } R_F) & m \to \neg Fm \\
(\text{irreflexivity of } R_P) & m \to \neg Pm
\end{array}
$$

$$
\begin{array}{rl}
(\text{transitivity of } R_F) & FFm \to Fm \\
(\text{transitivity of } R_P) & PPm \to Pm \\
(\text{Non branching}) & @_m Fm' \vee @_m m' \vee @_{m'} Fm
\end{array}
$$

and the axioms for *ceteris paribus* modalities, with the same restrictions of Λ_{CPT} for axioms (A1)-(A3).

$$
\begin{array}{rl}
(K_{[\Gamma]}) & [\Gamma](p \to q) \to ([\Gamma]p \to [\Gamma]q) \\
(T_{[\Gamma]}) & [\Gamma]p \to p \\
(4_{[\Gamma]}) & [\Gamma]p \to [\Gamma][\Gamma]p \\
(5_{[\Gamma]}) & \langle\Gamma\rangle p \to [\Gamma]\langle\Gamma\rangle p \\
(A1) & \langle\Gamma\rangle\phi \to \phi \\
(A2) & \langle\Gamma\rangle\neg\phi \to \neg\phi \\
(A3) & \langle\Gamma'\rangle\phi \to \langle\Gamma\rangle\phi
\end{array}
$$

With the following inference rules:

(MP)	If $\vdash \phi \to \psi$ and $\vdash \phi$ then $\vdash \psi$
($@_m$-gen)	If $\vdash \phi$ then $\vdash @_m\psi$
(G-gen)	If $\vdash \phi$ then $\vdash G\psi$
(H-gen)	If $\vdash \phi$ then $\vdash H\psi$
(Name)	If $\vdash @_m\phi$ and m hasn't any occurrence in ϕ, then $\vdash \psi$
(BG_F)	If $\vdash @_m Fm' \to @_{m'}\phi$ and $m' \neq m$ hasn't any occurrence in ϕ, then $\vdash @_m G\phi$
(BG_P)	If $\vdash @_m Pm' \to @_{m'}\phi$ and $m' \neq m$ hasn't any occurrence in ϕ, then $\vdash @_m H\phi$
(Subst)	If $\vdash \phi$ then $\vdash \phi^\sigma$

References

M. Adams. Is the existence of god a "hard" fact? *The Philosophical Review*, 76: 492–503, 1967.

Aristotle. De Interpretatione (On Interpretation). In R. McKeon, editor, *The Basic Works of Aristotle*. Random House, New York, 1941.

B. Barcellan and A. Zanardo. Actual futures in peircean branching-time logic, 1999. URL `citeseer.ist.psu.edu/326930.html`.

P. Blackburn and B. ten Cate. Pure extensions, proof rules, and hybrid axiomatics. *Studia Logica*, 84:277–322, 2006.

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

A. J. Freddoso. Accidental necessity and logical determinism. *The Journal of Philosophy*, 80:257–278, 1983.

D. Lewis. *On the Plurality of Worlds*. Blackwell, Oxford, 1986.

J. MacFarlane. Future contingents and relative truth. *The Philosophical Quarterly*, 53:321–336, 2003.

J. MacFarlane. Truth in the Garden of the forking Paths. In M. Kolbel, M. Garcia-Carpintero, editor, *Relative Truth*. Oxford University Press, Oxford, 2007.

A. N. Prior. *Past, Present and Future*. Oxford University Press, Oxford, 1967.

C. Proietti. Le problème des futurs contingents et le paradoxe de Fitch. Une étude commune de deux problèmes en logique modale. PhD Thesis, 2008.

C. Proietti and G. Sandu. Fitch's Paradox and *Ceteris Paribus* Modalities. forthcoming on *Synthese (Special issue): Knowability and beyond*, 2008.

O. Roy. L'ockhamisme de prior et l'ockhamisme actualiste en logique temporelle. URL `http://www.philos.rug.nl/~olivier/docs/acfas2004.pdf`. 2004.

J. van Benthem, P. Girard, and O. Roy. Everything else being equal. A modal logic approach to *ceteris paribus* preferences. *Synthese*, 38:83–125, 2009.

A. Varzi. Supervaluationism and its logics. *Mind*, 116:633–676, 2007.

G. H. von Wright. *The Logic of Preference*. University of Edinburgh Press, Edinburgh, 1963.