Logic and Interactive RAtionality

Yearbook 2009

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

Cover design by Nina Gierasimczuk.

Foreword

Following a one-year tradition, this yearbook contains a selection of papers that have been presented and discussed during the year 2009 at the seminar formerly known as *"Logics for Dynamics of Information and Preferences"* and now called *"Logic and Interactive Rationality (LIRA)"*. This seminar, a regular event at the Institute for Logic, Language and Computation (ILLC) of the University of Amsterdam for more than three years, has grown during 2009 to a national event, with thematic sessions organized in cooperation with the Department of Theoretical Philosophy and the Multi-Agent Systems Group of the University of Groningen.

Just as the 2008 edition, this yearbook reflects not only the work of the regular participants of the seminar (PhD students and researchers based in the Netherlands), but also the work of colleagues from around the world.

We gratefully thank the authors for their contributions and their collaboration through the editing process. In particular, we thank Johan van Benthem for initiating, stimulating and supporting this yearbook, and Nina Gierasimczuk for yet another beautiful cover.

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

Contents

Preface

Johan van Benthem

Much can be said, and has been said, about the year of 2009 in terms of financial and ecological crises. But this book shows the past year from a milder side: it has been remarkably good for logic and dynamics. The contributions in this book document talks at ILLC's long-standing Amsterdam Dynamics seminar, which by now is run as a nationwide Dutch affair with colleages from Groningen, Utrecht, and beyond, participating in thematic afternoons. And all that, if you check the list of editors of this volume, under benevolent international management.

If you read on, you will find lively and trail-blazing papers on many major themes in the current logical study of dynamic agency: questions, learning, awareness, preference, intentions, testimony and trust, communication networks, or strategies and powers in games. Many of these propose new models that bring out major aspects of rational agency, and many also propose new logics and explore their theory. There are also papers linking up with core topics in philosophical and mathematical logic, such as relevance and fixed-points, since logic of agency does not replace classical themes: it needs them. After a good dose of this mixture, you will be at the heart of things.

The authors of this book represent an active international community, and many of these papers have been presented in many countries. One such highlight was the second *LORI* Conference in Chongqing last October, where logical dynamics merged with mayong and river cruises. Chances are that you may cross the authors' path at some summer school or workshop near you. But until that day, the pages of this book should, and will, convey the spirit.

Johan van Benthem
February 2010

Continuous fragment of the μ-calculus

Gaëlle Fontaine

ILLC, University of Amsterdam
Gaelle.Fontaine@uva.nl

Abstract

In this paper we investigate the Scott continuous fragment of the modal μ-calculus. We discuss its relation with constructivity, where we call a formula constructive if its least fixpoint is always reached in at most ω steps. Our main result is a syntactic characterization of this continuous fragment. We also show that it is decidable whether a formula is continuous.

1 Introduction

This paper is a study into the fragment of the modal μ-calculus that we call continuous. Roughly, given a proposition letter p, a formula φ is said to be continuous in p if it monotone in p and if in order to establish the truth of φ at a point, we only need finitely many points at which p is true. The continuous fragment of the μ-calculus is defined as the fragment of the μ-calculus in which $\mu x.\varphi$ is allowed only if φ is continuous in x.

We prove the following two results. First, Theorem 2 gives a natural syntactic characterization of the continuous formulas. Informally, continuity corresponds to the formulas built using the operators $\vee$, $\wedge$, $\Diamond$ and μ. Second, we show in Theorem 3 that it is decidable whether a formula is continuous in p.

We believe that this continuous fragment is of interest for a number of reasons. A first motivation concerns the relation between continuity and another property, constructivity. The constructive formulas are the formulas whose fixpoint is reached in at most ω steps. Locally, this means that a state satisfies a least fixpoint formula if it satisfies one of its *finite* approximations. It is folklore that if a formula is continuous, then it is constructive. The other implication does not strictly hold. However, interesting questions concerning the link between constructivity and continuity remain. In any case, given our Theorem 2, continuity can be considered as the most natural candidate to approximate constructivity syntactically.

Next this fragment can be seen as a natural generalization of PDL in the following way. We define the completely additive formulas as the formulas built using the operators $\vee$, $\Diamond$ and μ. That is, the syntax is the same as for the continuous formulas, except that the conjunction is not allowed. Then it was observed by Yde Venema (personal communication) that PDL coincides with the fragment of the μ-calculus in which $\mu x.\varphi$ is allowed only if φ is completely additive. In this perspective, the continuous fragment appears as a natural extension of PDL.

Another reason for looking at this fragment (which also explains the name) is the link with Scott continuity. A formula is continuous in p iff it is continuous with respect to p in the Scott topology on the powerset algebra (with all other variables fixed). Scott continuity is of key importance in many areas of theoretical computer sciences where ordered structures play a role, such as domain theory (see, e.g., Abramsky and Jung (1994)). For many purposes, it is sufficient to check that a construction is Scott continuous in order to show that it is computationally feasible.

Finally our results fit in a model-theoretic tradition of so-called preservation results (see, e.g., Chang and Keisler (1973)). Giovanna D'Agostino and Marco Hollenberg have proved some results of this kind in the case of the μ-calculus (see, e.g., D'Agostino and Hollenberg (2000) and Hollenberg (1998)). Their proofs basically consist in identifying automata corresponding to the desired fragment and in showing that these automata give the announced characterization. The proof of our main result is similar as we also first start by translating our problem in terms of automata. We also mention that a version of our syntactic characterization in the case of first order logic has been obtained by Johan van Benthem in van Benthem (1996).

The paper is organized as follows. First we recall the syntax of the μ-calculus and some basic properties that will be used later on. Next we define the continuous fragment and we show how it is linked to Scott continuity, constructivity and PDL. Finally we prove our main result (Theorem 2) which is a syntactic characterization of the fragment and we show that it is decidable whether a formula is continuous (Theorem 3). We end the paper with questions for further research.

2 Preliminaries

We introduce the language and the Kripke semantic for the μ-calculus.

Definition 2.1. Let *Prop* be a finite set of proposition letters and let *Var* be a countable set of variables. The *formulas of the μ-calculus* are given by

$$\varphi \ ::= \ \top \mid p \mid x \mid \varphi \vee \varphi \mid \neg\varphi \mid \Diamond\varphi \mid \mu x.\varphi,$$

where p ranges over the set *Prop* and x ranges over the set *Var*. In $\mu x.\varphi$, we require that every occurrence of x is under an even number of negations in φ. The notion of *closed μ-formula* or *μ-sentence* is defined in the natural way.

As usual, we let $\varphi \wedge \psi$, $\Box\varphi$ and $\nu x.\varphi$ be abbreviations for $\neg(\neg\varphi \vee \neg\psi)$, $\neg\Diamond\neg\varphi$ and $\neg\mu x.\neg\varphi[\neg x/x]$. For a set of formulas Φ, we denote by $\bigvee \Phi$ the disjunction of formulas in Φ. Similarly, $\bigwedge \Phi$ denotes the conjunction of formulas in Φ.

Finally, we extend the syntax of the μ-calculus by allowing a new construct of the form $\nabla\Phi$, where Φ is a finite set of formulas. We will consider such a formula to be an abbreviation of $\bigwedge\{\Diamond\varphi : \varphi \in \Phi\} \wedge \Box \bigvee \Phi$. Remark that in Janin and Walukiewicz (1995), David Janin and Igor Walukiewicz use the notation $a \to \Phi$ for $\nabla_a\Phi$.

For reasons of a smooth presentation, we restrict to the unimodal fragment. All the results can be easily extended to the setting where we have more than one basic modality.

Definition 2.2. A *Kripke frame* is a pair (M, R), where M is a set and R a binary relation on M. A *Kripke model* $\mathcal{M}$ is a triple (M, R, V) where (M, R) is a Kripke frame and $V : Prop \to \mathcal{P}(M)$ a valuation.

If sRt, we say that t is a *successor* of s and we write $R(s)$ to denote the set $\{t \in M : sRt\}$. A *path* is a (finite or infinite) sequence $s_0, s_1, \ldots$ such that $s_i R s_{i+1}$ (for all $i \in \mathbb{N}$).

Definition 2.3. Given a μ-formula φ, a model $\mathcal{M} = (M, R, V)$ and an assignment $\tau : Var \to \mathcal{P}(M)$, we define a subset $[\![\varphi]\!]_{\mathcal{M},\tau}$ of M that is interpreted as the set of points at which φ is true. This subset is defined by induction in the usual way. We only recall that

$$[\![\mu x.\varphi]\!]_{\mathcal{M},\tau} = \bigcap\{U \subseteq M : [\![\varphi]\!]_{\mathcal{M},\tau[x:=U]} \subseteq U\},$$

where $\tau[x := U]$ is the assignment τ' such that $\tau'(x) = U$ and $\tau'(y) = \tau(y)$ for all $y \neq x$.

Observe that the set $[\![\mu x.\varphi]\!]_{\mathcal{M},\tau}$ is the least fixpoint of the map $\varphi_x : \mathcal{P}(M) \to \mathcal{P}(M)$ defined by $\varphi_x(U) := [\![\varphi]\!]_{\mathcal{M},\tau[x:=U]}$, for all $U \subseteq M$. Similarly, for a proposition letter p, we can define the map $\varphi_p : \mathcal{P}(M) \to \mathcal{P}(M)$ by $\varphi_p(U) := [\![\varphi]\!]_{\mathcal{M}[p:=U],\tau}$, where $\mathcal{M}[p := U]$ is the model (M, R, V') with $V'(p) = U$ and $V'(p') = V(p')$, for all $p' \neq p$.

If $s \in [\![\varphi]\!]_{\mathcal{M},\tau}$, we write $\mathcal{M}, s \Vdash_\tau \varphi$ and we say that φ *true* at $s \in \mathcal{M}$ under the assignment τ. If φ is a sentence, we simply write $\mathcal{M}, s \Vdash \varphi$.

A formula φ is *monotone* in a proposition letter p if for all models $\mathcal{M} = (M, R, V)$, all assignments τ and all sets $U, U' \subseteq M$ satisfying $U \subseteq U'$, we have $\varphi_p(U) \subseteq \varphi_p(U')$. The notion of monotonicity in a variable x is defined in an analogous way.

Finally we use the notation $\varphi \models \psi$ if for all models $\mathcal{M}$ and all points $s \in \mathcal{M}$, we have $\mathcal{M}, s \Vdash \varphi$ implies $\mathcal{M}, s \Vdash \psi$.

Now in order to determine if a sentence is true at a point, we can also look at the evaluation game.

Definition 2.4. Let φ be a sentence such that $\neg\psi$ is a subformula only if ψ is a proposition letter or the formula $\top$. We also assume that each variable is bound by at most one fixpoint. Thus, for every variable x occurring in φ, there is a unique formula ψ_x such that $\mu x.\psi_x$ or $\nu x.\psi_x$ is a subformula of φ.

We fix a model $\mathcal{M} = (M, R, V)$ and a point $s \in M$. We define the *evaluation game* for the sentence φ in the model $\mathcal{M}$ with starting position (s, φ) as the following game. The game is played between two players, the Duplicator and the Spoiler. The starting position is (s, φ). The rules, determining the admissible moves, together with the player who is supposed to make a move, are given in the table above.

If the match is finite, the player who gets stuck looses. Now suppose the match is infinite. Let Inf be the set of variables x such that positions of the form (t, x) are reached infinitely often. Let x_0 be a variable in Inf such that for all variables y in Inf, ψ_y is a subformula of ψ_{x_0}. If x_0 is bound by a μ-operator, then the Spoiler wins the match. Otherwise the Duplicator wins.

Position	Player	Admissible moves
$(t, \varphi_1 \vee \varphi_2)$	Duplicator	$\{(t, \varphi_1), (t, \varphi_2)\}$
$(t, \varphi_1 \wedge \varphi_2)$	Spoiler	$\{(t, \varphi_1), (t, \varphi_2)\}$
$(t, \Diamond\psi)$	Duplicator	$\{(u, \psi) : tRu\}$
$(t, \Box\psi)$	Spoiler	$\{(u, \psi) : tRu\}$
(t, x)	-	$\{(t, \psi_x)\}$
$(t, \mu x.\psi)$	-	$\{(t, \psi)\}$
$(t, \nu x.\psi)$	-	$\{(t, \psi)\}$
$(t, \neg\top)$ or $[(t, p)$ with $t \notin V(p)]$ or $[(t, \neg p)$ with $t \in V(p)]$	Duplicator	$\emptyset$
$(t, \neg p)$ with $t \in V(p)$	Duplicator	$\emptyset$
$(t, \top)$ or $[(t, p)$ with $t \in V(p)]$ or $[(t, \neg p)$ with $t \notin V(p)]$	Spoiler	$\emptyset$
$(t, \neg p)$ with $t \notin V(p)$	Duplicator	$\emptyset$

Without further notice, we will use the fact that the Duplicator has a winning strategy in the evaluation game in $\mathcal{M}$ with starting position (s, φ) iff $\mathcal{M}, s \Vdash \varphi$.

Proposition 1. *Let φ be a sentence such that no two distinct fixpoints bind the same variable and such that $\neg\psi$ is a subformula only if ψ is a proposition letter or the formula $\top$. Let $\mathcal{M} = (M, R, V)$ be a model and let $s \in M$. Then, the Duplicator has a winning strategy in the evaluation game in $\mathcal{M}$ with starting position (s, φ) iff $\mathcal{M}, s \Vdash \varphi$.*

When deciding whether a sentence is true at a point s, it only depends on the points accessible (in possibly many steps) from s. These points together with the relation and the valuation inherited from the original model form the submodel generated by s. We will use this notion later on and we briefly recall the definition.

Definition 2.5. Let $\mathcal{M} = (M, R, V)$ be a model. A subset N of M is *downward closed* if for all s and t, sRt and $t \in N$ imply that $s \in N$. N is *upward closed* if for all s and t, sRt and $s \in N$ imply that $t \in N$.

A model $\mathcal{N} = (N, S, U)$ is a *generated submodel* of $\mathcal{M}$ if $N \subseteq M$, N is upward closed, $S = R \cap (N \times N)$ and $U(p) = V(p) \cap N$, for all $p \in Prop$. If N' is a subset of M, we say that $\mathcal{N} = (N, S, U)$ is the *submodel generated* by N' if $\mathcal{N}$ is a generated submodel and if N is the smallest upward closed set containing N'.

In our proof, it will be often more convenient to work with a certain kind of Kripke models. That is, we will suppose that the models we are dealing with are trees such that each point (except the root) has infinitely many bisimilar siblings. We make this definition precise and we give the results needed to justify this assumption.

Definition 2.6. A point s is a *root* of a model $\mathcal{M} = (M, R, V)$ if for every t distinct from s, there is a path from s to t. $\mathcal{M}$ is a *tree* if it has a root, every point distinct from the root has a unique predecessor and R is acyclic (that is, there is no non-empty path starting at a point t and ending in t).

A model $\mathcal{M} = (M, R, V)$ is *ω-expanded* if it is a tree such that for all $s \in M$ and all successors t of s, there are infinitely many distinct successors of s that are bisimilar to t.

Proposition 2. *Let $\mathcal{M} = (M, R, V)$ be a model and let $s \in M$. There exists a tree $\mathcal{M}' = (M', R', V')$ that is ω-expanded such that s and the root s' of $\mathcal{M}'$ are bisimilar. In particular, for all μ-sentences φ, $\mathcal{M}, s \Vdash \varphi$ iff $\mathcal{M}', s' \Vdash \varphi$.*

Another way to look at formulas of the μ-calculus is to consider automata. In Janin and Walukiewicz (1995), David Janin and Igor Walukiewicz define a notion of automaton that operates on Kripke models and that corresponds exactly to the μ-calculus.

Definition 2.7. A *μ-automaton* $\mathbb{A}$ over a finite alphabet Σ is a tuple (Q, q_0, δ, Ω) such that Q is a finite set of states, $q_0 \in Q$ is the initial state, $\delta : Q \times \Sigma \to \mathcal{P}\mathcal{P}(Q)$ is the transition map and $\Omega : Q \to \mathbb{N}$ is the parity function.

Given a frame $\mathcal{M} = (M, R, V)$ with a labeling $L : M \to \Sigma$ and a point $s \in M$, an $\mathbb{A}$-game in $\mathcal{M}$ with starting position (s, q_0) is played between two players, the Duplicator and the Spoiler. The game is as follows: If we are in position (t, q) (where $t \in M$ and $q \in Q$), the Duplicator has to make a move. The Duplicator chooses a marking $m : Q \to \mathcal{P}\{u : tRu\}$ and then a description D in $\delta(q, L(t))$. If $u \in m(q)$, we say that u is *marked* with q.

The marking and the description have to satisfy the two following properties. First, if $q' \in D$, there exists a successor u of t that is marked with q'. Second, if u is a successor of t, there exists $q' \in D$ such that u is marked with q'. After the Duplicator has chosen a marking m, the Spoiler plays a position (u, q') such that $t \in m(q')$.

Either player wins the game if the other player cannot make a move. An infinite match $(s, q_0), (s_1, q_1), \dots$ is won by the Duplicator if the smallest element of $\{\Omega(q) : q \text{ appears infinitely often in } q_0, q_1, \dots\}$ is even.

We say that $(\mathcal{M}, s)$ is *accepted* by $\mathbb{A}$ if the Duplicator has a winning strategy in the $\mathbb{A}$-game in $\mathcal{M}$ with starting position (s, q_0).

Remark that a model (M, R, V) can be seen as a frame (M, R) with a labeling $L : M \to \mathcal{P}(Prop)$ defined by $L(t) = \{p \in Prop : t \in V(p)\}$, for all $t \in M$.

Theorem 1. *Janin and Walukiewicz (1995) For every μ-automaton $\mathbb{A}$ (over the alphabet $\mathcal{P}(Prop)$), there is a sentence φ such that for all models $\mathcal{M}$ and all points $s \in \mathcal{M}$, $\mathbb{A}$ accepts $(\mathcal{M}, s)$ iff $\mathcal{M}, s \Vdash \varphi$. Conversely, for every sentence φ, there is a μ-automaton $\mathbb{A}$ such that for all models $\mathcal{M}$ and all points $s \in \mathcal{M}$, $\mathbb{A}$ accepts $(\mathcal{M}, s)$ iff $\mathcal{M}, s \Vdash \varphi$.*

3 Continuity

We define the notion of continuity for a formula and we show the connection with the Scott continuity. We also mention that these formulas are constructive and that there is a natural connection with PDL.

Definition 3.1. Fix a proposition letter p. A sentence φ is *continuous in p* if for all models $\mathcal{M} = (M, R, V)$ and all $s \in M$, we have

$$\mathcal{M}, s \Vdash \varphi \text{ iff } \exists\, F \subseteq V(p) \text{ s.t. } F \text{ is finite and } \mathcal{M}[p := F], s \Vdash \varphi.$$

The notion of *continuity in x* (where x is a variable) is defined similarly.

That is, a formula φ is continuous in p iff it is monotone in p and whenever φ is true at a point in a model, we only need finitely many points where p is true in order to establish the truth of φ.

Continuity and Scott continuity

It does not seem very natural that a formula satisfying such a property should be called continuous. In fact, it is equivalent to require that the formula is Scott continuous with respect to p in the powerset algebra (with all other proposition letters fixed). In the next paragraph, we recall the definition of the Scott topology and we briefly show that the notion of Scott continuity and our last definition coincide.

Definition 3.2. Let $\mathcal{M} = (M, R, V)$ be a model. A family $\mathcal{F}$ of subsets of M is *directed* if for all $U_1, U_2 \in \mathcal{F}$, there exists $U \in \mathcal{F}$ such that $U \supseteq U_1 \cup U_2$.

A *Scott open set* in the powerset algebra $\mathcal{P}(M)$ is a family O of subsets of M that is closed under upset (that is, if $U \in O$ and $U' \supseteq U$, then $U' \in O$) and such that for all directed family $\mathcal{F}$ satisfying $\bigcup \mathcal{F} \in O$, the intersection $\mathcal{F} \cap O$ is non-empty.

As usual, a map $f : \mathcal{P}(M) \to \mathcal{P}(M)$ is *Scott continuous* if for all Scott open sets O, the set $f^{-1}[O] = \{f^{-1}(U) : U \in O\}$ is Scott open.

Fix a proposition letter p. A sentence φ is *Scott continuous in p* if for all models $\mathcal{M} = (M, R, V)$, the map $\varphi_p : \mathcal{P}(M) \to \mathcal{P}(M)$ is Scott continuous.

Remark that the Scott topology can be defined in an arbitrary partial order (see, e.g., Gierz et al. (1980)). It is a fairly standard result that a map f is Scott continuous iff it preserves directed joins. That is, for all directed family $\mathcal{F}$, we have $f(\bigcup \mathcal{F}) = \bigcup f[\mathcal{F}]$ (where $f[\mathcal{F}] = \{f(U) : U \in \mathcal{F}\}$). Now we check that our notion of continuity defined in a Kripke semantic framework is equivalent to the standard definition of Scott continuity.

Proposition 3. *A sentence is continuous in p iff it is Scott continuous in p.*

Proof. For the direction from left to right, let φ be a continuous sentence in p. Fix a model $\mathcal{M} = (M, R, V)$. We show that the map $\varphi_p : \mathcal{P}(M) \to \mathcal{P}(M)$ preserves directed joins.

Let $\mathcal{F}$ be a directed family. It follows from the monotonicity of φ that the set $\bigcup \varphi_p[\mathcal{F}]$ is a subset of $\varphi_p(\bigcup \mathcal{F})$. Thus, it remains to show that $\varphi_p(\bigcup \mathcal{F}) \subseteq \bigcup \varphi_p[\mathcal{F}]$. Take s in $\varphi_p(\bigcup \mathcal{F})$. That is, the formula φ is true at s in the model $\mathcal{M}[p := \bigcup \mathcal{F}]$. As φ is continuous in p, there is a finite subset F of $\bigcup \mathcal{F}$ such that φ is true at s in $\mathcal{M}[p := F]$. Now, since F is a finite subset of $\bigcup \mathcal{F}$ and since $\mathcal{F}$ is directed, there exists a set U in $\mathcal{F}$ such that F is a subset of U. Moreover, as φ is monotone, $\mathcal{M}[p := F], s \Vdash \varphi$ implies $\mathcal{M}[p := U], s \Vdash \varphi$. Therefore, s belongs to $\varphi_p(U)$ and in particular, s belongs to $\bigcup \varphi_p[\mathcal{F}]$. This finishes to show that $\varphi_p(\bigcup \mathcal{F}) \subseteq \bigcup \varphi_p[\mathcal{F}]$.

For the direction from right to left, let φ be a Scott continuous sentence in p. First we show that φ is monotone in p. Let $\mathcal{M} = (M, R, V)$ be a model. We check that $\varphi_p(U) \subseteq \varphi_p(U')$, in case $U \subseteq U'$. Suppose $U \subseteq U'$ and let $\mathcal{F}$ be the set $\{U, U'\}$. The family $\mathcal{F}$ is clearly directed and satisfies $\bigcup \mathcal{F} = U'$. Using the fact that φ_p preserves directed joins, we get that $\varphi_p(U') = \varphi_p(\bigcup \mathcal{F}) = \bigcup \varphi_p[\mathcal{F}]$. By definition of $\mathcal{F}$, we have $\bigcup \varphi_p[\mathcal{F}] = \varphi_p(U) \cup \varphi_p(U')$. Putting everything together, we obtain that $\varphi_p(U') = \varphi_p(U) \cup \varphi_p(U')$. Thus, $\varphi_p(U) \subseteq \varphi_p(U')$.

To show that φ is continuous in p, it remains to show that if $\mathcal{M}, s \Vdash \varphi$, then there exists a finite subset F of $V(p)$ such that $\mathcal{M}[p := F], s \Vdash \varphi$. Suppose that the formula φ is true at s in $\mathcal{M}$. That is, s belongs to $\varphi_s(V(p))$. Now let $\mathcal{F}$ be the family $\{F \subseteq V(p) : F \text{ finite}\}$. It is not hard to see that $\mathcal{F}$ is a directed family satisfying $\bigcup\mathcal{F} = V(p)$. Since φ_p preserves directed joins, we obtain $\varphi_p(V(p)) = \varphi_p(\bigcup\mathcal{F}) = \bigcup\varphi_p[\mathcal{F}]$. From $s \in \varphi_p(V(p))$, it then follows that $s \in \bigcup\varphi_p[\mathcal{F}]$. Therefore, there exists $F \in \mathcal{F}$ such that $s \in \varphi_p(F)$. That is, F is a finite subset of $V(p)$ such that $\mathcal{M}[p := F], s \Vdash \varphi$. □

Continuity and constructivity

A formula is constructive if its fixpoint is reached in at most ω steps. Formally, we have the following definition.

Definition 3.3. Fix a proposition letter p. A sentence φ is *constructive* in p if for all models $\mathcal{M} = (M, R, V)$, the least fixpoint of the map $\varphi_p : \mathcal{P}(M) \to \mathcal{P}(M)$ is equal to $\bigcup\{\varphi_p^i(\emptyset) : i \in \mathbb{N}\}$ (where φ_p^i is defined by induction by $\varphi_p^0 = \varphi_p$ and $\varphi_p^{i+1} = \varphi_p \circ \varphi_p^i$).

Locally, this means that given a formula φ constructive in p and a point s in a model at which $\mu p.\varphi$ is true, there is some natural number n such that s belongs to the finite approximation $\varphi_p^n(\emptyset)$. We observe that a continuous formula is constructive.

Proposition 4. *A sentence φ continuous in p is constructive in p.*

Proof. Let φ be a sentence continuous in p and let $\mathcal{M} = (M, R, V)$ be a model. We show that the least fixpoint of φ_p is $\bigcup\{\varphi_p^i(\emptyset) : i \in \mathbb{N}\}$.

Let $\mathcal{F}$ be the family $\{\varphi_p^i(\emptyset) : i \in \mathbb{N}\}$. It is enough to check that $\varphi_p(\bigcup\mathcal{F}) = \bigcup\mathcal{F}$. First remark that $\mathcal{F}$ is directed. Therefore, $\varphi_p(\bigcup\mathcal{F}) = \bigcup\varphi_p[\mathcal{F}]$. It is also easy to prove that $\bigcup\varphi_p[\mathcal{F}] = \bigcup\mathcal{F}$. Putting everything together, we obtain that $\varphi_p(\bigcup\mathcal{F}) = \bigcup\mathcal{F}$ and this finishes the proof. □

Remark that a constructive sentence might not be continuous.

Example 1. Let φ be the formula $\Box p \wedge \Box\Box\bot$. Basically, φ is true at a point s in a model if the depth of s is less or equal to 2 (that is, there are no t and t' satisfying $sRtRt'$) and all successors of s satisfy p. It is not hard to see that φ is not continuous in p. However, we have that for all models $\mathcal{M} = (M, R, V)$, $\varphi_p^2(\emptyset) = \varphi_p^3(\emptyset)$. In particular, φ is constructive in p.

Example 2. Let ψ be the formula $\nu x.p \wedge \Diamond x$. The formula ψ is true at a point s if there is a infinite path starting from s and at each point of this path, p is true. This sentence is not continuous in p. However, it is constructive, since for all models $\mathcal{M} = (M, R, V)$, we have $\psi_p(\emptyset) = \emptyset$.

Observe that in the previous examples we have $\mu p.\varphi \equiv \mu p.\Box\Box\bot$ and $\mu p.\psi \equiv \mu p.\bot$. Thus, there is a continuous sentence (namely $\Box\Box\bot$) that is equivalent to φ, modulo the least fixpoint operation. Similarly, there is a continuous sentence (the formula $\bot$) that is equivalent to ψ, modulo the least fixpoint operation. This suggests the following question (Yde Venema): given a constructive formula φ, can we find a continuous formula ψ satisfying $\mu p.\varphi \equiv \mu p.\psi$? The answer is still unknown and this could be a first step for further study of the relation between continuity and constructivity.

Decidability of constructivity is also an interesting question. We would like to mention that in Otto (1999), Martin Otto proved that it is decidable in

EXPTIME whether a basic modal formula $\varphi(p)$ is bounded. We recall that a basic modal formula $\varphi(p)$ is bounded if there is a natural number n such that for all models $\mathcal{M}$, we have $\varphi_p^n(\emptyset) = \varphi_p^{n+1}(\emptyset)$.

Continuity and PDL

We finish this section by few words about the connection between the continuous fragment and PDL. We start by defining the completely additive formulas.

Definition 3.4. Let P be a subset of *Prop* and let X be a subset of Var. The set of *completely additive formulas* with respect to $P \cup X$ is defined by induction in the following way:

$$\varphi \ ::= \ \top \mid p \mid x \mid \psi \mid \varphi \vee \varphi \mid \Diamond\varphi \mid \mu y.\chi,$$

where p is in P, x is in X, ψ is a formula of the μ-calculus such that the proposition letters of ψ and the variables of ψ do not belong to $P \cup X$ and χ is completely additive with respect to $P \cup X \cup \{y\}$.

We define the *completely additive fragment* as the fragment of the μ-calculus in which $\mu x.\varphi$ is allowed only if φ is completely additive with respect to x. As mentioned in the introduction, it was observed by Yde Venema that this fragment coincides with test-free PDL.

Similarly, we define the *continuous fragment* as the fragment of the μ-calculus in which $\mu x.\varphi$ is allowed only if φ is continuous in x. It is routine to check that any completely additive formula with respect to p is continuous in p (and the proof is similar to the proof of Lemma 1 below). In particular, the completely additive fragment is included in the continuous fragment. That is, PDL is a subset of the continuous fragment. We remark that this inclusion is strict. An example is the formula $\varphi = \mu x.(\Diamond(p \wedge x) \wedge \Diamond(q \wedge x))$. This formula belongs to the continuous fragment but is not equivalent to a formula in PDL. Roughly, the sentence φ is true at a point s if there is a finite binary tree-like submodel starting from s, such that each non-terminal node of the tree has a child at which p is true and a child at which q is true. This example was given by Johan van Benthem in van Benthem (2006).

4 Syntactic characterization of the continuous fragment

In this section, we give a characterization of the continuous fragment of the μ-calculus. The main result states that the sentences which are continuous in p are exactly the sentences such that p and the variables are only in the scope of the operators $\vee$, $\wedge$, $\Diamond$ and μ. These formulas are formally defined as the set $CF(p)$.

Definition 4.1. Let P be a subset of *Prop* and let X be a subset of Var. The set of formulas $CF(P \cup X)$ is defined by induction in the following way:

$$\varphi \ ::= \ \top \mid p \mid x \mid \psi \mid \varphi \vee \varphi \mid \varphi \wedge \varphi \mid \Diamond\varphi \mid \mu y.\chi,$$

where p is in P, x is in X, ψ is a formula of the μ-calculus such that the proposition letters of ψ and the variables of ψ do not belong to $P \cup X$ and χ belongs to $CF(P \cup X \cup \{y\})$. We abbreviate $CF(\{p\})$ to $CF(p)$.

As a first property, we mention that the formulas in $CF(P \cup X)$ are closed under composition.

Proposition 5. *If φ is in $CF(P \cup X \cup \{p\})$ and ψ is in $CF(P \cup X)$, then $\varphi[\psi/p]$ belongs to $CF(P \cup X)$.*

Proof. By induction on φ. □

Next we observe that the sentences in $CF(p)$ are continuous.

Lemma 1. *A sentence φ in $CF(p)$ is continuous in p.*

Proof. We prove by induction on φ that for all sets $P \subseteq Prop$ and $X \subseteq Var$, $\varphi \in CF(P \cup X)$ implies that φ is continuous in p and in x, for all $p \in P$ and all $x \in X$. We focus on the inductive step $\varphi = \mu y.\chi$, where χ is in $CF(P \cup X \cup \{y\})$. We also restrict ourselves to show that φ is continuous in p, for a proposition letter p in P.

Fix a proposition letter $p \in P$. First we introduce the following notation. For a model $\mathcal{M} = (M, R, V)$, an assignment τ and a subset U of W, we let $\chi_y^U : \mathcal{P}(M) \to \mathcal{P}(M)$ be the map defined by $\chi_y^U(W) = [\![\chi]\!]_{\mathcal{M}[p:=U],\tau[y:=W]}$, for all $W \subseteq M$. We also denote by $f(U)$ the least fixpoint of χ_y^U.

Now we show that φ is monotone in p. That is, for all models $\mathcal{M} = (M, R, V)$, all assignments τ and all subsets U, U' of M such that $U \subseteq U'$, we have $\mathcal{M}[p := U], s \Vdash_\tau \mu y.\chi$ implies $\mathcal{M}[p := U'], s \Vdash_\tau \mu y.\chi$. Fix a model $\mathcal{M} = (M, R, V)$, an assignment τ and sets $U, U' \subseteq M$ satisfying $U \subseteq U'$. Suppose $\mathcal{M}[p := U_0], s \Vdash_\tau \mu y.\chi$. That is, s belongs to the least fixpoint $f(U)$of the map χ_y^U. Since χ is monotone in p, we have that for all $W \subseteq M$, $\chi_y^U(W) \subseteq \chi_y^{U'}(W)$. It follows that the least fixpoint $f(U)$ of the map χ_y^U is a subset of the least fixpoint of the map $\chi_y^{U'}$. Putting this together with $s \in f(U)$, we get that s belongs to the least fixpoint of $\chi_y^{U'}$. That is, $\mathcal{M}[p := U'], s \Vdash_\tau \mu y.\chi$ and this finishes the proof that φ is monotone in p.

Next suppose that $\mathcal{M}, s \Vdash_\tau \mu y.\chi$, for a point s in a model $\mathcal{M} = (M, R, V)$. That is, s belongs to the least fixpoint L of the map χ_y. Now let $\mathcal{F}$ be the set of finite subsets of $V(p)$. We need to find a set $F \in \mathcal{F}$ satisfying $\mathcal{M}[p := F], s \Vdash_\tau \mu y.\chi$. Or equivalently, we have to show that there exists $F \in \mathcal{F}$ such that s belongs to the least fixpoint $f(F)$ of the map χ_y^F.

Let $\mathcal{G}$ be the set $\{f(F) : F \in \mathcal{F}\}$. It is routine to show that $\mathcal{G}$ is a directed family. Since L is the least fixpoint of χ_y, we have that for all $U \subseteq M$, $\chi_y(U) \subseteq U$ implies $L \subseteq U$. So if we can prove that $\chi_y(\bigcup \mathcal{G}) \subseteq \bigcup \mathcal{G}$, we will obtain $L \subseteq \bigcup \mathcal{G}$. Putting this together with $s \in L$, it will follow that $s \in \bigcup\{f(F) : F \in \mathcal{F}\}$. Therefore, in order to show that $s \in f(F)$ for some $F \in \mathcal{F}$, it is sufficient to prove that $\chi_y(\bigcup \mathcal{G}) \subseteq \bigcup \mathcal{G}$.

Assume $t \in \chi_y(\bigcup \mathcal{G})$. Since $\mathcal{G}$ is a directed family and χ is Scott continuous in y, we have $\chi_y(\bigcup \mathcal{G}) = \bigcup \chi_y(\mathcal{G})$. Thus, there exists $F_0 \in \mathcal{F}$ such that $t \in \chi_y(f(F_0))$. Now since χ is continuous in p, there exists a finite set $F_1 \subseteq V(p)$ such that $t \in \chi_y^{F_1}(f(F_0))$. Let F be the set $F_0 \cup F_1$. Since χ is monotone in p, $t \in \chi_y^{F_1}(f(F_0))$ implies $t \in \chi_y^F(f(F_0))$. It also follows from the monotonicity in p that for all $U \subseteq M$, $\chi_y^{F_0}(U) \subseteq \chi_y^F(U)$. Therefore, the least fixpoint $f(F_0)$ of $\chi_y^{F_0}$ is a subset of the least fixpoint $f(F)$ of χ_y^F. Using the fact that χ is monotone in y and the inclusion $f(F_0) \subseteq f(F)$, we obtain $\chi_y^F(f(F_0)) \subseteq \chi_y^F(f(F))$. Putting this together with $t \in \chi_y^F(f(F_0))$, we get $t \in \chi_y^F(f(F))$. Moreover, since $f(F)$ is a fixpoint of χ_y^F, we have $\chi_y^F(f(F)) = f(F)$. Hence, t belongs to $f(F)$. In particular, t belongs to $\bigcup \mathcal{G}$ and this finishes the proof. □

We also prove the converse: the sentences in $CF(p)$ are enough to characterize the continuous fragment of the μ-calculus. The proof is inspired by the one given by Marco Hollenberg in Hollenberg (1998), where he shows that a sentence is distributive in p over unions iff it is equivalent to $\langle\pi\rangle p$, for some p-free μ-program π.

Theorem 2. *A sentence φ is continuous in p iff it is equivalent to a sentence in $CF(p)$.*

Proof. By Lemma 1, we only need to prove the implication from left to right. Let φ be a sentence continuous in p. We need to find a formula χ in $CF(p)$ that is equivalent to φ.

The proof consists in constructing a finite set $\Pi \subseteq CF(p)$ such that

$$\varphi \equiv \bigvee\{\psi : \psi \in \Pi \text{ and } \psi \models \varphi\}. \tag{1}$$

Indeed, if there is such a set Π, we can define χ as the formula $\bigvee\{\psi : \psi \in \Pi \text{ and } \psi \models \varphi\}$. Clearly, χ belongs to $CF(p)$ and is equivalent to ψ.

We define Π as the set of sentences in $CF(p)$, which correspond to μ-automata with at most k states, where k is a natural number that we will define later and which depends on φ. In order to define k, we introduce the following notation. First, let $\mathbb{A} = (Q, q_0, \delta, \Omega)$ be a μ-automaton corresponding to φ. For $q \in Q$, let φ_q denote the sentence corresponding to the automaton we get from $\mathbb{A}$ by changing the initial state from q_0 to q.

Next we denote by Sort0 be the set of sentences of the form

$$\bigwedge\{p' : p' \in Prop\backslash\{p\}, p' \in \sigma\} \wedge \bigwedge\{\neg p' : p' \in Prop\backslash\{p\}, p' \notin \sigma\},$$

where σ is a subset of $Prop\backslash\{p\}$. For a point s in a model, there is a unique formula in Sort0 true at s. This formula gives us exactly the set of proposition letters in $Prop\backslash\{p\}$ which are true at s. Sort1 is the set of all sentences of the form

$$\bigwedge\{\varphi_q[\bot/p] : q \in S\} \wedge \bigwedge\{\neg\varphi_q[\bot/p] : q \notin S\},$$

where S is a subset of Q. Finally Sort2 contains all sentences of the form $\chi \wedge \nabla\Psi$, where $\chi \in$ Sort0 and Ψ is a subset of Sort1. As for the formulas in Sort0, it is easy to see that given a model $\mathcal{M}$ and a point s in $\mathcal{M}$, there is exactly one sentence in Sort1 and one sentence in Sort2 which are true at s. Remark finally that Sort0, Sort1 and Sort2 are sets of sentences which do not contain p.

Now we can define Π as the set of sentences in $CF(p)$, which correspond to μ-automata with at most $|\text{Sort2}| \cdot 2^{|Q|+1}$ states. Since there are only finitely many such automata modulo equivalence, Π is finite (up to equivalence). It is also immediate that Π is a subset of $CF(p)$. Thus it remains to show that equivalence (1) holds.

From right to left, equivalence (1) is obvious. For the direction from left to right, suppose that $\mathcal{M} = (M, R, V)$ is a model such that $\mathcal{M}, s \Vdash \varphi$, for some point s. We need to find a sentence $\psi \in \Pi$ satisfying $\psi \models \varphi$ and such that $\mathcal{M}, s \Vdash \psi$. Equivalently, we can construct an automaton $\mathbb{A}'$ corresponding to a formula

$\psi \in \Pi$ such that $\psi \models \varphi$ and $\mathcal{M}, s \Vdash \psi$. That is, we can construct an automaton $\mathbb{A}'$ with at most $|\text{Sort2}| \cdot 2^{|Q|+1}$ states, corresponding to a sentence in $CF(p)$, such that $\mathbb{A}'$ accepts $(\mathcal{M}, s)$ and satisfying $\mathbb{A}' \Vdash \mathbb{A}$ (that is, for all models $\mathcal{M}'$ and all $s' \in \mathcal{M}'$, if $\mathbb{A}'$ accepts $(\mathcal{M}', s')$, then $\mathbb{A}$ accepts $(\mathcal{M}', s')$).

By Proposition 2, we may assume that $\mathcal{M}$ is a tree with root s and that $\mathcal{M}$ is ω-expanded. Since φ is continuous, there is a finite subset F of $V(p)$ such that $\mathcal{M}[p := F], s \Vdash \varphi$. Let T be the minimal downward closed set that contains F. Using T, we define the automaton $\mathbb{A}'$. Roughly, the idea is to define the set of states of $\mathbb{A}'$ as the set T together with an extra point $a_\top$. However, we need to make sure that the set of states of $\mathbb{A}'$ contains at most $|\text{Sort2}| \cdot 2^{|Q|+1}$ elements. There is of course no guarantee that $T \cup \{a_\top\}$ satisfies this condition. The solution is the following. We define for every point in T its representation, which encodes the information we might need about the point. Then we can identify the points having the same representation in order to "reduce" the cardinality of T.

Before defining the automaton $\mathbb{A}'$, we introduce some notation. Given a point t in $\mathcal{M}[p := F]$, there is a unique sentence in Sort2 that is true at t. We denote it by $s2(t)$. Next if t belongs to F, we define the color $col(t)$ of t as 1 and otherwise, the color of t is 0. We let $Q(t)$ be the set $\{q \in Q : \mathcal{M}[p := F], t \Vdash \varphi_q\}$. Finally, we define the representation map $r : M \to (\text{Sort2} \times Q \times \{0,1\}) \cup \{a_\top\}$ by

$$r(t) = \begin{cases} (s2(t), Q(t), col(t)) & \text{if } t \in T, \\ a_\top & \text{otherwise.} \end{cases}$$

The automaton $\mathbb{A}' = (Q', q'_0, \delta', \Omega')$ is a μ-automaton over the alphabet $\text{Sort2} \times \{0,1\}$. Its set of states Q' is given by

$$Q' = \{r(t) : t \in T\} \cup \{a_\top\},$$

and its initial state q'_0 is $r(s)$. Next for all $(\sigma, i) \in \text{Sort2} \times \{0,1\}$, the set $\delta'(q', (\sigma, i))$ is defined by

$$\delta'(q', (\sigma, i)) = \begin{cases} \Big\{ r[R(u)] : u \in T \text{ and } r(u) = r(t) \Big\} & \\ \qquad\qquad\qquad \text{if } q' = r_t, \sigma = s2(t) \text{ and } i = col(t), \\ \{\{a_\top\}, \emptyset\} & \text{if } q' = a_\top, \\ \emptyset & \text{otherwise.} \end{cases}$$

Intuitively, when the automaton is in the state $q' = r(t)$ and it reads the label $(s2(t), col(t))$, the Duplicator has to pick a successor u of t that is in T and this induces a description in $\delta'(r(t), (s2(t), col(t)))$. As soon as the automaton reaches the state $a_\top$, either the match is finite or the automaton stays in the state $a_\top$. In all other cases, the Duplicator gets stuck.

Finally, the map Ω' is such that $\Omega'(a_\top) = 0$ and $\Omega'(q') = 1$, for all $q' \neq a_\top$. In other words, the only way the Duplicator can win an infinite match is to reach the state $a_\top$ and to stay there.

Remark that a model $\mathcal{M}' = (M', R', V')$ can be seen as a frame (M', R') with a labeling $L' : M' \to \text{Sort2} \times \{0,1\}$ defined by $L'(t') = (s2(t'), 1)$ if p is true at t' and $L'(t') = (s2(t'), 0)$ otherwise. Thus, the automaton $\mathbb{A}'$ can operate on models.

In order to extract the formula ψ from this automaton, it will be convenient to think of the alphabet of $\mathbb{A}'$ not being the set $\text{Sort2} \times \{0,1\}$ but the set $\mathcal{P}(\text{Sort2} \cup \{p\})$.

The idea is to see a pair $(\sigma, i) \in \mathrm{Sort2} \times \{0,1\}$ as the set $\{\sigma\}$ if $i = 0$ or as the set $\{\sigma, p\}$ if $i = 1$. More precisely, if $\rho \subseteq \mathrm{Sort2} \cup \{p\}$, the transition map would associate to the pair (q', ρ) the set $\delta'(q', (\sigma, 0))$ if $\rho = \{\sigma\}$ for some $\sigma \in \mathrm{Sort2}$, the set $\delta'(q', (\sigma, 1))$ if $\rho = \{\sigma, p\}$ for some $\sigma \in \mathrm{Sort2}$ and the empty set otherwise.

Now if we think to the formulas of Sort2 as proposition letters, it follows from Theorem 1 that $\mathbb{A}'$ is equivalent to a sentence ψ whose proposition letters belong to $\mathrm{Sort2} \cup \{p\}$. Such a formula ψ is also a sentence with proposition letters in *Prop*, in an obvious way. To finish the proof, we need to show that ψ is equivalent to a sentence which is in Π, ψ is true at s and $\psi \models \varphi$.

Claim 1. *ψ is equivalent to a sentence in Π.*

Proof. The intuition is the following. In order to win an $\mathbb{A}'$-match, the Duplicator has to reach the state $a_\top$ and then, the match is basically over. It seems natural that such a property can be expressed using only least fixpoints (and no greatest fixpoint).

Next we also need to make sure that in a formula corresponding to $\mathbb{A}'$, neither p nor any variable is in the scope of the operator $\Box$. This is guaranteed by the presence of the state $a_\top$ in any non-empty description that the Duplicator might pick. Very informally, each description corresponds to a subformula (of the sentence corresponding to the automaton) which starts with the operator ∇. Using the fact that $a_\top$ belongs to any of these descriptions (except the empty one) and corresponds to the sentence $\top$, we can show that the ∇ operator can be replaced by the modal operator $\Diamond$.

Formally the proof is the following. First observe that $\mathbb{A}'$ has at most $|\mathrm{Sort2}| \cdot 2^{|Q|+1}$ states. Thus in order to show that ψ is equivalent to a formula in Π, it is sufficient to show that ψ is equivalent to a sentence in $CF(p)$.

For $q' \in Q'$ and $S' \subseteq Q'$, we define the translation $tr(S', q')$ of q' with respect to S'. The translation $tr(S', q')$ is a formula in the language whose set of proposition letters is *Prop* and whose set of variables is $Var \cup Q'$. For those q' that are equal to $r(t) = (s2(t), Q(t), col(t))$ and $S' \subseteq Q'$, we have

$$\begin{aligned} tr(S', q') := {} & s2(t) \wedge col(t).p \wedge \\ & \bigvee \Big\{ \bigwedge \big\{ \Diamond \mu q''.tr(S' \backslash \{q''\}, s') : q'' \in r[R(u)] \text{ and } q'' \in S' \big\} \\ & \quad \wedge \bigwedge \big\{ \Diamond q'' : q'' \in r[R(u)] \text{ and } q'' \notin S' \big\} : u \in T \text{ and } r(u) = q' \Big\} \end{aligned}$$

where $col(t).p$ is p if $col(t) = 1$ and $\top$ if $col(t) = 0$. By convention, $\bigwedge \emptyset = \top$. For all $S' \subseteq Q'$, we define $tr(S', a_\top)$ by $\top$.

It is routine to show that $tr(S', q')$ is a well-defined sentence with proposition letters in $Prop \cup (Q' \backslash S')$ and that it belongs to $CF(\{p\} \cup (Q' \backslash S'))$. The proofs are by induction on the cardinality of S'. In particular, $tr(Q', q'_0)$ belongs to $CF(p)$. Therefore, in order to prove the claim, it is enough to show that ψ is equivalent to $tr(Q', q'_0)$.

Before proving that ψ is equivalent to a formula in $CF(p)$, we show that the translations satisfy the following property. For all $q' \in Q'$,

$$tr(Q', q') \text{ is equivalent to } \mu q'.tr(Q' \backslash \{q'\}, q'). \tag{2}$$

We give a sketch of the proof of (2). First, we observe that for all $S' \subseteq Q'$ and all $q' \notin S'$, we have

$$tr(S', q') = tr(S' \backslash \{q'\}, q')[\mu q'.tr(S' \backslash \{q'\}, q')/q']. \tag{3}$$

We skip the proof which is a standard proof by induction on the cardinality of S'. In order to prove (2), fix a state $q' \in Q'$. Now let χ be the formula $tr(Q' \backslash \{q'\}, q')$. It follows from the equality (3) that $tr(Q', q')$ is equivalent to $tr(Q' \backslash \{q'\}, q')[\mu q'.tr(Q' \backslash \{q'\}, q')/q']$. That is, $tr(Q', q')$ is equivalent to $\chi[\mu q'.\chi/q']$. Next, by definition of the fixpoint operator, we know that $\chi[\mu q'.\chi/q']$ is equivalent to $\mu q'.\chi$. Putting everything together, we obtain that $tr(Q', q')$ is equivalent to $\mu q'.\chi$. That is, $tr(Q', q')$ is equivalent to $\mu q'.tr(Q' \backslash \{q'\}, q')$ and this finishes the proof of (2).

Now we prove that ψ is equivalent to $tr(Q', q'_0)$. For let $\mathcal{M} = (M', R', V')$ be a model and let s' be a point in $\mathcal{M}'$. We need to show that

$$\mathcal{M}', s' \Vdash \psi \leftrightarrow tr(Q', q'_0). \tag{4}$$

By Proposition 2, we may assume that $\mathcal{M}'$ is a tree with root s', that is ω-expanded.

For the direction from left to right of (4), suppose that $\mathcal{M}', s' \Vdash \psi$. We know that the Duplicator has a winning strategy g in the $\mathbb{A}'$-game in $\mathcal{M}'$ with starting position (s', q'_0). Given a winning strategy f, we say that a point t' is marked with a state $q' \in Q'$ if there is a f-conform $\mathbb{A}'$-match (in $\mathcal{M}'$ with starting position (s', q'_0)) during which the point t' is marked with q'. Now we define a winning strategy g' (for the Duplicator in the $\mathbb{A}'$-game in $\mathcal{M}'$ with starting position $(s'q'_0)$) such that for all $t' \in \mathcal{M}'$, t' is marked with exactly one state $q'(t')$ and the set of successors marked with a state $q' \neq a_\top$ is finite.

We construct g' by induction on the "distance" to the root s'. It is immediate that s' is only marked with q'_0. Now assume that t' is only marked with q'. Thus there is an $\mathbb{A}'$-match where the Duplicator plays according to his winning strategy and during which the position (t', q') is reached. Then the Duplicator chooses a description D and a marking m. First we can modify the marking such that for each state $q'' \neq a_\top$, exactly one point is marked with q''. Next suppose that a successor u' of t' is marked with q'_1 and q'_2. Since $\mathcal{M}'$ is ω-expanded, we can pick a successor v' of t' that is bisimilar to u' and only marked with $a_\top$. We can then modify the marking such that u' is marked with q'_1 and v' is only marked with q'_2. It is not hard to see that this is still a winning strategy for the Duplicator and that, according to this strategy, every successor of t' is marked with a unique state and only finitely many of these successors are marked with a state $q'' \neq a_\top$. This completes the definition of g'.

Now in order to show the left to right implication of (4), we have to prove that $\mathcal{M}', s' \Vdash tr(Q', q'_0)$. The idea is to show that if t' is marked with q', then $\mathcal{M}', t' \Vdash tr(Q', q')$. In particular, this would imply that $\mathcal{M}', s' \Vdash tr(Q', q'_0)$, since s' is marked with q'_0. Thus, it is sufficient to prove that if t' is marked with q', then $\mathcal{M}', t' \Vdash tr(Q', q')$.

The proof is by induction on the distance $d_\top(t')$ that we define in the following way. For all $t' \in \mathcal{M}'$,

$$d_\top(t') := \begin{cases} 0 & \text{if } q'(t') = a_\top, \\ max\{d_\top(v') : t'R'v'\} + 1 & \text{otherwise.} \end{cases}$$

Remark that since the set of points marked with a state $q' \neq a_\top$ is finite, we have that $d_\top(t')$ is a natural number for all $t' \in \mathcal{M}'$.

For the basic case, we check that if $d_\top(t') = 0$ and t' is marked with q', then $\mathcal{M}', t' \Vdash tr(Q', q')$. This is immediate since $q' = a_\top$ and $tr(Q', a_\top) = \top$.

For the induction step, we fix a point t' marked with a state $r(t)$ ($t \in T$) and we assume that for all v' such that $d_\top(v') < d_\top(t')$, we have $\mathcal{M}', v' \Vdash tr(Q', q')$ if v' is marked with q'. In particular, for all successors v' of t', we have $\mathcal{M}', v' \Vdash tr(Q', q')$ if v' is marked with q'. We need to show that $\mathcal{M}', t' \Vdash tr(Q', r(t))$. Since t' is marked with $r(t)$ (and hence the Duplicator cannot get stuck in position $(t', r(t))$), we have $\delta'(r(t), L'(t')) \neq \emptyset$. Thus, it follows from the definition of δ' that $\mathcal{M}', t' \Vdash s2(t) \wedge col(t).p$. It remains to find $u \in T$ such that $r(u) = r(t)$ and such that $\bigwedge\{\Diamond\mu s'.tr\,(Q'\backslash\{s'\}, s') : s' \in r[R(u)]\}$ is true at t'.

Since t' is marked with $r(t)$, we know that if the Duplicator plays according to his winning strategy in an $\mathbb{A}'$-match, then the position $(t', r(t))$ can be reached. In the next move of the match, the Duplicator chooses a description in $\delta'(r(t), L'(t'))$. Let $u \in T$ be such that the Duplicator chooses the description $r[R(u)]$. It follows from the definition of δ' that $r(u) = r(t)$.

Next we prove that $\bigwedge\{\Diamond\mu q''.tr\,(Q'\backslash\{q''\}, q'') : q'' \in r[R(u)]\}$ is true at t'. For let q'' be a state in $r[R(u)]$. We have to show that $\mathcal{M}', t' \Vdash \Diamond\mu q''.tr\,(Q'\backslash\{q''\}, q'')$. Recall that by (2), we have that $\mu q''.tr\,(Q'\backslash\{q''\}, q'')$ is equivalent to $tr(Q', q'')$. Thus it is enough to check that $\mathcal{M}', t' \Vdash \Diamond tr(Q', q'')$. Since $r[R(u)]$ is the description chosen by the Duplicator, there exists a successor v' of t' that is marked with q''. By induction hypothesis, we know that $\mathcal{M}', v' \Vdash tr(Q', q'')$. It immediately follows that $\mathcal{M}', t' \Vdash \Diamond tr(Q', q'')$ and this finishes the proof from the left to right implication of (4).

For the converse direction of (4), assume that $\mathcal{M}', s' \Vdash tr(Q', q'_0)$. Thus, the Duplicator has a winning strategy h in the evaluation game with starting position $(s', tr(Q', q'_0))$. We say that a point t' is h-marked with a state q' if there is an h-conform evaluation game during which the Duplicator plays the position $(t', tr(S', q'))$ for some $S' \subseteq Q'$. Since $\mathcal{M}'$ is ω-expanded, we may assume that the strategy h is such that each point is h-marked with at most one state. We do not give details, as this is similar to the transformation from the strategy g to the strategy g'.

We define by induction a winning strategy for the Duplicator in the $\mathbb{A}'$-game in $\mathcal{M}'$ with starting position (s', q'_0). The idea is to ensure that if a point t' is marked with $q' \neq a_\top$ (in an $\mathbb{A}'$-match conform to our strategy), then t' is h-marked with q' in the evaluation game. The starting position of the $\mathbb{A}'$-game is (s', q'_0). It is immediate that s' is h-marked with q'_0, since any evaluation match starts with the position $(s', tr(Q', q'_0))$.

Suppose that we have defined the strategy until the $\mathbb{A}'$-match reaches the position (t', q'), where $q' \neq a_\top$ and t' is h-marked with q'. Thus there exists $t \in T$ such that $q' = r(t)$. Since t' is h-marked with q', there is $S' \subseteq Q'$ and an h-conform evaluation match during which we reach the position $(t', tr(S', q'))$. In particular, this position is winning for the Duplicator. By definition of the translation tr, we have that $\mathcal{M}', t' \Vdash s2(t) \wedge col(t)$. Thus, $\delta'(q', L'(t'))$ is $\{r[R(u)] : u \in T \text{ and } r(u) = q'\}$ and the Duplicator has to choose a description in this set. Since the position $(t', tr(S', q'))$ is played by the Duplicator in an h-conform evaluation match, we know that there exists $u \in T$ such that $r(u) = q'$ and the next position of this

h-conform evaluation match is

$$(t', \bigwedge \left\{\Diamond \mu q''.tr\,(S'\backslash\{q''\}, s') : q'' \in r[R(u)] \text{ and } q'' \in S'\right\} \wedge \\ \bigwedge \left\{\Diamond q'' : q'' \in r[R(u)] \text{ and } q'' \notin S'\right\}). \quad (5)$$

Now we define the strategy such that the Duplicator chooses the description $r[R(u)]$. Next we need to define a marking that is legal with respect to this description. Fix a state q'' in the chosen description. That is, q'' belongs to $r[R(u)]$. It follows from (5) that there is an h-conform evaluation match during which either the position $(t', \Diamond \mu q''.tr(S'\backslash\{q''\}, q''))$ or the position $(t', \Diamond q'')$ is reached. Thus, there exists a successor v' of t' such that the next position of this h-conform match is $(v', \mu q''.tr(S'\backslash\{q''\}, q''))$ or (v', q''). This implies that one of the next positions of this h-conform match is $(v', tr(S'', q''))$ for some $S'' \subseteq Q'$. Therefore, we can define our strategy in this $\mathbb{A}'$-match such that the point v' is marked with the state q''. Next observe that since T is finite and $\mathcal{M}$ is ω-expanded, there are successors of u which do not belong to T. In particular, $a_\top$ belongs to $r[R(u)]$. Therefore we can also mark with $a_\top$ all the successors of t'. It is easy to check that our marking is a legal marking.

It remains to show that this strategy is indeed winning. Since the formula $tr(Q', q'_0)$ contains only least fixpoints, the Duplicator can only wins finite matches in the evaluation game (with starting position $(s', tr(Q', q'_0))$). Thus, on every path starting from s', there are only finitely many points h-marked with a state in Q'. This implies that in any $\mathbb{A}'$-match conform to our strategy, we eventually reached the position $(t', a_\top)$ and the Duplicator wins the game. □

Claim 2. *$\mathcal{M}, s \Vdash \psi$.*

Proof. We need to provide a winning strategy for the Duplicator in the $\mathbb{A}'$-game in the model $\mathcal{M}$ with starting position (s, q_0). The strategy is defined by induction and we ensure that whenever a position (t, q') is played, then $q' = a_\top$ or $q' = r(t)$.

This certainly holds for the initial position (s, q_0). Now assume that the Duplicator has to respond to a position (t, q'). Assume first that $q' = a_\top$. If t has at least one successor, the Duplicator chooses the description $\{a_\top\}$ and he marks all the successors of t with $a_\top$. If t has no successor, the Duplicator picks the description $\emptyset$ and the match is over.

Now if $t \in T$ and $q' = r(t)$, the Duplicator chooses the description $r[R(t)]$. A successor v of t that belongs to T is marked with $r(v)$. All the successors of t are marked with $a_\top$. It is routine to show that such a strategy is well-defined and winning. □

Claim 3. *$\psi \models \varphi$.*

Proof. Suppose $\mathcal{M}' = (M', R', V')$ is a model such that $\mathcal{M}', s' \Vdash \psi$, for some point s'. That is, the Duplicator has a winning strategy in the $\mathbb{A}'$-game in $\mathcal{M}'$ with starting position (s', q'_0). We have to show that $\mathcal{M}', s' \Vdash \varphi$.

As before, we may assume that $\mathcal{M}'$ is a tree with root s' and that $\mathcal{M}'$ is ω-expanded. Recall from Claim 1 that we say that a point t' is marked with a state q' if there is an $\mathbb{A}'$-match during which the Duplicator plays according to his winning strategy and the point t' is marked with q'. As in Claim 1, each point t' of $\mathcal{M}'$ may assumed to be marked with a unique state $q'(t')$ of $\mathbb{A}'$ and given a point t', we may suppose that the set of successors marked with a state $q' \neq a_\top$ is finite.

Let T' be the set of points marked with a state $q' \neq a_\top$. Let F' be the set of $t' \in T'$ such that $q'(t')$ is of the form $(s2(t), Q(t), 1)$ for some $t \in T$. For v' in $\mathcal{M}'$, we let $col(v')$ be 1 if $v' \in F'$ and $col(v') = 0$ otherwise. We also define $L'(V')$ as the pair $(s2(v'), col(v'))$ (where $s2(v')$ is the only sentence in Sort2 which is true at v' in $\mathcal{M}'[p := F']$).

It is routine to check that the strategy for the Duplicator in the model $\mathcal{M}'$ (with starting position (s', q'_0)) remains a winning strategy in $\mathcal{M}'[p := F']$ (with starting position (s', q'_0)). Moreover since the Duplicator cannot get stuck in position $(t', q'(t'))$, we have $\delta'(q'(t'), L'(t')) \neq \emptyset$. Hence, if $q'(t') = r(t)$ for some $t \in T$, it follows from the definition of δ' that $L'(t') = (s2(t), col(t))$. In particular, if $q'(t') = (s2(t), Q(t), 1)$ for some $t \in T$, then $col(t') = 1$. In other words, F' is a subset of $V'(p)$. Since φ is monotone, it is then enough to prove $\mathcal{M}'[p := F'], s' \Vdash \varphi$ in order to show that $\mathcal{M}', s' \Vdash \varphi$.

To prove this, we need to find a winning strategy for the Duplicator in the $\mathbb{A}$-game in $\mathcal{M}'[p := F']$ with starting position (s', q_0). The idea is to make sure that if $t' \in T'$ and $q'(t') = r(t)$, then positions of the form (t', q) are played only if $q \in Q(t)$. This holds for the initial position (s', q_0), as $q'(s') = r(s)$ and $q_0 \in Q(s)$ (since $\varphi_{q_0} = \varphi$ and φ is true at s in $\mathcal{M}[p := F]$). We will see that following our strategy, as soon as a position (t', q) is reached with $t' \notin T'$, then the Duplicator can win.

Suppose that the Duplicator has to respond to a position (t', q) with $t' \in T'$, $q'(t') = r(t)$ and $q \in Q(t)$. By definition of the map q', there is some match of the $\mathbb{A}'$-game (in the model $\mathcal{M}'[p := F']$ with starting position (s', q'_0)) which is conform to the Duplicator's strategy and during which the position $(t', r(t))$ is reached. Following his winning strategy, the Duplicator has then to choose a description in $\delta(r(t), L'(t'))$. Let $u \in T$ be a point such that the Duplicator chooses the description $r[R(u)]$. Remark that by definition of δ', we have $r(t) = r(u)$. Putting this together with $q \in Q(t)$, we obtain that q belongs to $Q(u)$. Since $q \in Q(u)$, we have $\mathcal{M}[p := F], u \Vdash \varphi_q$. Therefore, the Duplicator has a winning strategy for the position (u, q) (in the $\mathbb{A}$-game in the model $\mathcal{M}[p := F]$).

Let m be a marking chosen according to this strategy. Suppose m is legal with respect to $D \in \delta(q, L(u))$, where $L(u) = \{p' \in Prop : \mathcal{M}[p := F], u \Vdash p'\}$. Since the position $(t', r(t))$ is winning for the Duplicator, he cannot get stuck. Hence, $\delta'(r(u), L'(t')) \neq \emptyset$. It follows from the definition of δ' that t' and u satisfy the same sentence in Sort2. In particular, they satisfy the same propositions letters of $Prop \backslash \{p\}$. Moreover, by definition of F', we also have $t' \in F'$ iff $u \in F$. Putting everything together, we obtain that the sets $L(t') = \{p' \in Prop : \mathcal{M}'[p := F'], t' \Vdash p'\}$ and $L(u)$ are the same. Thus, $\delta(q, L(t')) = \delta(q, L(u))$ and the description D is also available in $\delta(q, L(t'))$.

We define the marking $\bar{m}$ in response to (t', q) in the following way:

$$\begin{aligned} \bar{m}(\bar{q}) &= \{v' \notin T' : t'R'v' \text{ and } \mathcal{M}'[p := F'], v' \Vdash \varphi_{\bar{q}}\} \cup \\ &\quad\ \{v' \in T' : t'R'v', q'(v') = r(v) \text{ and } \bar{q} \in Q(v)\}, \end{aligned}$$

for all $\bar{q} \in Q$. We show that $\bar{m}$ is a legal marking with respect to D.

For suppose that $\bar{q} \in D$. We need to find a successor v'_0 of t' such that $v'_0 \in \bar{m}(\bar{q})$. Since m is a legal marking with respect to D, there exists a successor

v_0 of u such that $v_0 \in m(\bar{q})$. There are two cases.

First, suppose that v_0 belongs to T. Recall that following his winning strategy, the Duplicator chooses the description $r[R(u)]$ at position $(t', r(u))$. So there is a successor v_0' of t' that is marked with $r(v_0)$. Since v_0 belongs to $m(\bar{q})$, we have $\mathcal{M}[p := F], v_0 \Vdash \varphi_{\bar{q}}$. That is, $\bar{q} \in Q(v_0)$. Gathering everything together, we have that $v_0' \in T'$ is a successor of t' such that $q'(v_0') = r(v_0)$ and $\bar{q} \in Q(v_0)$. Thus, $v_0' \in \bar{m}(\bar{q})$ and we are done.

Next suppose that v_0 does not belong to T. Thus in $\mathcal{M}[p := F]$, there is no point from v_0 on where p holds. In particular, $\mathcal{M}[p := F], v_0 \Vdash \varphi_{\bar{q}}$ implies $\mathcal{M}[p := F], v_0 \Vdash \varphi_{\bar{q}}[\bot/p]$. Since $\mathcal{M}[p := F], v_0 \Vdash \varphi_{\bar{q}}[\bot/p]$, we have $\mathcal{M}[p := F], u \Vdash \Diamond\varphi_{\bar{q}}[\bot/p]$. As u and t' satisfy the same sentences in Sort2, it follows that $\mathcal{M}'[p := F'], t' \Vdash \Diamond\varphi_{\bar{q}}[\bot/p]$. Thus there is a successor v' of t' such that $\mathcal{M}'[p := F'], v' \Vdash \varphi_{\bar{q}}[\bot/p]$. Since $\varphi_{\bar{q}}[\bot/p]$ does not contain any p, this also means that $\mathcal{M}', v' \Vdash \varphi_{\bar{q}}[\bot/p]$.

Next observe that by construction of T' and by definition of the parity condition Ω', T' is finite. As $\mathcal{M}'$ is ω-expanded, we can choose a successor v_0' of t' that is bisimilar in $\mathcal{M}'$ to v' and such that $v_0' \notin T'$. Putting this together with $\mathcal{M}', v' \Vdash \varphi_{\bar{q}}[\bot/p]$, we obtain $\mathcal{M}', v_0' \Vdash \varphi_{\bar{q}}[\bot/p]$. Using again the fact that $\varphi_{\bar{q}}[\bot/p]$ does not contain any p, this implies $\mathcal{M}'[p := F'], v_0' \Vdash \varphi_{\bar{q}}[\bot/p]$. Now remark that by definition of δ', T' is downward closed. In particular, since $v_0' \notin T'$, no point in the model generated by v_0' belongs to T'. It follows that in the submodel of $\mathcal{M}'[p := F']$ generated by v_0', p holds nowhere. Therefore, $\mathcal{M}'[p := F'], v_0' \Vdash \varphi_{\bar{q}}[\bot/p]$ implies $\mathcal{M}'[p := F'], v_0' \Vdash \varphi_{\bar{q}}$. Hence, v_0' belongs to $\bar{m}(\bar{q})$.

To prove that $\bar{m}$ is a legal marking with respect to D, it remains to show that for all successors v' of t', there is a state $\bar{q} \in D$ such that $v' \in \bar{m}(\bar{q})$. Let v' be a successor of t'. There are again two cases.

First, suppose that v' belongs to T'. Thus, v' is marked with a state $r(v)$, for some successor v of u. Since m is a legal marking with respect to D, there is a state $\bar{q} \in D$ such that $v \in m(\bar{q})$. That is, $\bar{q} \in Q(v)$. By definition of $\bar{m}$, this means that $v' \in \bar{m}(\bar{q})$.

Finally assume that v' does not belong to T'. Let $s1(v')$ be the unique sentence in Sort1 that is true at v' in $\mathcal{M}'[p := F']$. Hence, $\Diamond s1(v')$ is true at t' in $\mathcal{M}'[p := F']$. Since u and t' satisfy the same sentence in Sort2, we get that $\mathcal{M}[p := F], u \Vdash \Diamond s1(v')$. Thus there is a successor v of u such that $\mathcal{M}[p := F], v \Vdash s1(v')$. As $s1(v')$ does not contain any p, we also have that $\mathcal{M}, v \Vdash s1(v')$. Since T is finite and $\mathcal{M}$ is ω-expanded, we can choose a successor v_0 of u that does not belong to T and that is bisimilar in $\mathcal{M}$ to v. In particular, $\mathcal{M}, v \Vdash s1(v')$ implies $\mathcal{M}, v_0 \Vdash s1(v')$. Using again the fact that $s1(v')$ does not contain any p, we get $\mathcal{M}[p := F], v_0 \Vdash s1(v')$.

Now since m is a legal marking with respect to D, there exists $\bar{q} \in D$ such that $v_0 \in m(\bar{q})$. That is, $\mathcal{M}[p := F], v_0 \Vdash \varphi_{\bar{q}}$. As before, we can show that in the submodel of $\mathcal{M}[p := F]$ generated by v_0, p holds nowhere. Therefore, $\mathcal{M}[p := F], v_0 \Vdash \varphi_{\bar{q}}$ implies $\mathcal{M}[p := F], v_0 \Vdash \varphi_{\bar{q}}[\bot/p]$. As $\mathcal{M}[p := F], v_0 \Vdash s1(v')$, v' and v_0 satisfy the same sentence in Sort1. Thus, from $\mathcal{M}[p := F], v_0 \Vdash \varphi_{\bar{q}}[\bot/p]$, we get $\mathcal{M}'[p := F'], v' \Vdash \varphi_{\bar{q}}[\bot/p]$. Moreover, since $v' \notin T'$, we know that in the submodel of $\mathcal{M}'[p := F']$ generated by v', p does not hold. Thus $\mathcal{M}'[p := F'], v' \Vdash \varphi_{\bar{q}}$ and $v' \in \bar{m}(\bar{q})$.

The Spoiler may respond to $\bar{m}$ in two ways. First, he may pick a position $(v', \bar{q})$ with $v' \notin T'$ and $\mathcal{M}'[p := F'], v' \Vdash \varphi_{\bar{q}}$. Then the Duplicator has a winning strategy from this point on. We continue with this strategy.

Next, the Spoiler may pick a position $(v', \bar{q})$ with $v' \in T'$, $q'(v') = r(v)$ and $\bar{q} \in Q(v)$. Then we continue with the strategy we have described here. Recall now that by definition of the parity condition Ω', T' is finite. Therefore, in any match played according to our strategy, the Spoiler will end up picking a position $(v', \bar{q})$ with $v' \notin T'$ and $\mathcal{M}'[p := F'], v' \Vdash \varphi_{\bar{q}}$. The Duplicator can then win and this completes the proof. □

As a corollary of this last proof, we obtain that it is decidable whether a formula is continuous in p.

Theorem 3. *It is decidable whether a formula is continuous in p.*

Proof. Fix a proposition letter p. Let Π be the set of sentences in $CF(p)$ which correspond to μ-automata with at most $|\text{Sort2}| \cdot 2^{|Q|+1}$ states. Now there are only finitely many such automata (modulo equivalence). There is also an effective translation from μ-automata to μ-sentences. Finally it is easy to verify whether a formula is in $CF(p)$. Therefore, we can compute Π.

It follows from the proof of Theorem 2 that a sentence φ is continuous in p iff $\varphi \equiv \{\psi : \psi \in \Pi \text{ and } \psi \models \varphi\}$. That is, φ is continuous in p iff there exists a subset Ψ of Π such that $\varphi \equiv \bigvee \Psi$. Therefore, in order to decide if φ is continuous in p, we can compute all the subsets Ψ of Π and check whether φ is equivalent to $\bigvee \Psi$. Since the μ-calculus if finitely axiomatizable and has the finite model property, it is decidable whether φ is equivalent to a disjunct $\bigvee \Psi$ and this completes the proof. □

□

Looking at the decision procedure presented in the proof of Theorem 3, we can see that the complexity is at most 4EXPTIME. That is, it involves four interlocked checking procedures, each of them being of complexity at most EXPTIME. This result is not very satisfying and we are looking for a better algorithm.

Finally, we mention that a similar syntactic characterization can be obtained in the case of basic modal logic. More precisely, a basic modal formula is continuous in p iff it belongs to the modal fragment $CF_m(p)$ of $CF(p)$. We give a formal definition of $CF_m(p)$ and a sketch of the proof in the appendix.

Definition 4.2. Let P be a subset of *Prop*. $CF_m(P)$ is defined by induction in the following way:

$$\varphi \ ::= \ \top \mid p \mid \psi \mid \varphi \vee \varphi \mid \varphi \wedge \varphi \mid \Diamond\varphi,$$

where p belongs to P and no proposition letters of ψ is in P.

Corollary 1. *A basic modal formula is continuous in p iff it belongs to $CF_m(p)$.*

Proof. We focus on the direction from left to right. Let φ be a basic modal formula continuous in p. It follows from Theorem 2 that φ is equivalent to a μ-sentence ψ in $CF(p)$. Remark that if we look carefully at the proof of Theorem 2, we can see that ψ is guarded (that is, each variable x in ψ is in the scope of a modal operator). We may also assume that each variable x in ψ is bound at most once in the formula. Thus for a variable x in ψ, there exists a unique subformula of the form $\mu x.\alpha_x$.

Recall that the dependency order $\leq$ on the bound variables of ψ is the least partial order such that if x occurs free in $\mu y.\alpha_y$, then $x \leq y$. Now we define the formula ψ^i ($i \in \mathbb{N}$) by induction on i. We let ψ^0 be the formula obtained by deleting all the μ-operators in ψ and we let ψ^{i+1} be the formula $\psi^i[\alpha_{x_n}/x_n]\dots[\alpha_{x_1}/x_1]$, where the sequence $x_1, \dots, x_n$ is a linear ordering of all bound variables of ψ such that if $x_i \leq x_j$, then $i \leq j$.

Let n be the modal depth of φ. Consider the formula ψ^n. Now let χ be the basic modal formula obtained by replacing all the variables by $\top$. It is clear that χ is a basic modal formula in $CF_m(p)$. We skip the proof but using the fact that ψ is guarded and equivalent to a basic modal formula of depth n, we can show that χ is equivalent to ψ. Thus we found a basic modal formula χ in $CF_m(p)$ that is equivalent to φ. □

5 Conclusion and further work

We defined the continuous fragment of the μ-calculus and showed how it relates to Scott continuity. We also started to investigate the relation between continuity and constructivity. Finally, we gave a syntactic characterization of the continuous formulas and we proved that it decidable whether a formula is continuous.

This work can be continued in various directions. To start with, it would be interesting to clarify the link between continuity and constructivity. In particular, we could try to answer the following question: given a constructive formula φ, can we find a continuous formula ψ satisfying $\mu p.\varphi \equiv \mu p.\psi$?

Next we observe that in the proof of Theorem 2, the construction of the automaton $\mathbb{A}'$ depends on the model $\mathcal{M}$ and the point s at which φ is true. Is it possible to construct an automaton $\mathbb{A}'$ by directly transforming the automaton $\mathbb{A}$ that is equivalent to φ? Such a construction might help us to find a better lower complexity bound for the decision procedure (for the membership of a formula in the continuous fragment).

We believe that it might be interesting to generalize our approach. As mentioned earlier, similar results to our characterization have been obtained by Giovanna D'Agostino and Marco Hollenberg in D'Agostino and Hollenberg (2000). Is there any general pattern that can be found in all these proofs?

We could also extend this syntactic characterization to other settings. For example, we can try to get a similar result if we restrict our attention to the class of finitely branching models.

Finally, we would like to mention that in Ikegami and van Benthem (2008), Daisuke Ikegami and Johan van Benthem proved that the μ-calculus is closed under taking product update. Using their method together with our syntactic characterization, it is possible to show that the set of continuous formulas is closed under taking product update.

Acknowledgements I am grateful to Johan van Benthem for proposing this work, interesting comments and suggestions for the introduction. Many thanks also to my supervisor Yde Venema for his valuable help and support. Finally special thanks to Balder ten Cate for comments and inspiring emails.

References

S. Abramsky and A. Jung. Domain Theory. In A. S., G. D. M., and M. T.S.M., editors, *Handbook for Logic in Computer Science*. Clarendon Press, Oxford, 1994.

C. Chang and H. Keisler. *Model Theory*. North Holland, 1973.

G. D'Agostino and M. Hollenberg. Logical Questions Concerning The mu-Calculus: Interpolation, Lyndon and Los-Tarski. *Journal of Symbolic Logic*, pages 310–332, 2000.

G. Gierz, K. Hofman, K. Keimel, J. Lawson, M. M., and D. Scott. *A Compendium of Continuous Lattices*. Springer-Verlag, 1980.

M. Hollenberg. *Logic and Bisimulation*. PhD thesis, Utrecht University, 1998. Zeno Institute of Philosophy.

D. Ikegami and J. van Benthem. Modal Fixed-Point Logic and Changing Models, 2008. ILLC Preprint PP-2008-19.

D. Janin and I. Walukiewicz. Automata for the modal μ-calculus and related results. In *Proceedings MFCS'95*, pages 552–562. Springer, 1995.

M. Otto. Eliminating Recursion in the mu-Calculus. In *Proceedings of STACS'99*. Springer, 1999.

J. van Benthem. Modal Frame Correspondences and Fixed-Points. *Studia Logica*, pages 133–155, 2006.

J. van Benthem. *Exploring Logical Dynamics*. Studies in Logic, Language and Information. CSLI Publications and Cambridge University Press, 1996.

On the Minimality of Definite Tell-tale Sets in Finite Identification of Languages

Nina Gierasimczuk and Dick de Jongh

Institute for Logic, Language, and Computation, University of Amsterdam
n.gierasimczuk@uva.nl, d.h.j.dejongh@uva.nl

Abstract

This work is concerned with notions of minimality of definite tell-tale sets used in finite identification of languages. We base our considerations on the observation that a class of languages is finitely identifiable from positive data if every element of this class includes a definite tell-tale set, i.e. a finite subset that separates the language from other possibilities (Mukouchi 1992). In this paper we discuss two notions of minimality of definite tell-tales in finite identification. We show that the problem of finding a minimal definite tell-tale for one language from a finitely identifiable finite class of finite sets is PTIME computable, while the problem of finding a minimal-size definite tell-tale set is NP-complete. In the last section, we restrict out attention to preset learners, finite learners who use in their identification task one fixed DFTT for each language. We define the non-effective fastest learning function that performs finite identification on the basis of minimal definite finite tell-tale sets. In connection to the fastest learner we show that the problem of finding the set of all minimal definite tell-tale sets for a finite language from a finite class is NP-hard. We conclude with results about recursion-theoretic complexity, showing that finding the minimal size DFTTs is not always possible in a recursive manner.

1 Introduction

To finitely identify a language means to be able to recognize it with certainty after receiving some (specific) finite sample of this language. Such a finite sample that suffices for finite identification is called *definite finite tell-tale* (DFTT, for short, see: (Mukouchi 1992)). One can see such a DFTT as the collection of the most characteristic elements of the set, but it has also a different connotation that is based on the *eliminative power* of its elements. We can think of the information that is carried by a particular sample of the language in a negative way, as showing which of the hypotheses are inconsistent with the information that has arrived, and thereby eliminating them. A set is finitely identifiable if it has comprised in a finite subset the power of eliminating all languages under consideration different from itself. These finite subsets are the definite finite tell-tales.

The characterization of finite identifiability (Mukouchi 1992) says that if a class of languages is finitely identifiable, then the identification can be done on the basis of corresponding DFTTs, i.e. finite subsets of the original languages that contain a sample that is essential for finite identifiability. A number of issues emerges when analyzing computational properties of definite finite tell-tales. Since finite tell-tales are by no means unique it is obviously useful to obtain small tell-tales. In this context we are particularly interested in the possibility of distinguishing two notions of minimality for DFTTs. A *minimal DFTT* is a DFTT that cannot be further reduced without losing its eliminative power with respect to a class of languages. A *minimal-size DFTT* of a set L, is a DFTT that is one of those which are smallest among all possible DFTTs of L.

In order to investigate the computational complexity of finding such minimal DFTTs, we will have to restrict ourselves to finite classes of languages. This may seem to be a very heavy restriction, but it creates the possibility of grasping important aspects of the complexity of finite identification. It also allows discussing the complexity of finite identifiability from the perspective of a teacher. In particular, it allows estimating how complex it is to find an optimal sample of a language that allows finite identifiability with respect to a certain class, and expose it to the pupil. We also investigate the analogous problems concerning recursion-theoretic complexity.

The plan of the paper is as follows. We introduce some basic notions, and provide the definition and characterization of finite identifiability. Then we present the refined notions of minimal DFTT, and minimal-size DFTT. We show that the problem of finding a minimal-size DFTT is NP-complete (using the Minimum Cover Problem (Garey and Johnson 1979)), while the finding any minimal DFTT is PTIME computable. Therefore, it can be argued that it is harder for a teacher to provide a minimal-size optimal sample, than just any minimal information. With this context in mind we restrict the attention to preset learners, learning functions which use for the identification of each language one fixed preset DFTT. In the end we define the non-effective fastest learning function that performs finite identification on the basis of minimal DFTTs. Its computational complexity in the finite case turns out to be NP-hard. Similar results are obtained in the analogous recursion-theoretic context. As so often the best is the enemy of the better.

2 Notation and Basic Definitions

Let U be an infinite recursive set, we call any $L \subseteq U$ a language. A class of languages $\mathcal{L} = \{L_1, L_2, \ldots\}$ is an indexed family of recursive languages if there is a computable function $f : \mathbb{N} \times U \to \{0, 1\}$, such that

$$f(i, w) = \begin{cases} 1 & \text{if } w \in L_i, \\ 0 & \text{if } w \notin L_i. \end{cases}$$

In the rest of this paper we will consider indexed families of recursive languages.

Later on, in order to consider computational complexity, we will devote our attention to finite classes of sets. Then we take the class $\mathcal{L}$ to be $\{L_1, L_2, \ldots, L_n\}$. We use $\mathcal{I}_{\mathcal{L}} = \{i \mid L_i \in \mathcal{L}\}$ to denote the set of indices of sets in $\mathcal{L}$.

Definition 2.1 (Positive presentation (text))**.** By a positive presentation ε of L we mean an infinite sequence of elements from L enumerating all and only the elements from L (allowing repetitions).

Definition 2.2 (Notation)**.** We will use the following notation:

- ε_n is the n-th element of ε;
- $\varepsilon|n$ is the sequence $(\varepsilon_0, \varepsilon_1, \ldots, \varepsilon_{n-1})$;
- $set(\varepsilon)$ is the set of elements that occur in ε;
- φ is a learning function — a recursive map from finite data sequences to indexes of hypotheses, $\varphi : U^* \to \mathbb{N} \cup \{\uparrow\}$. The function is allowed to refrain from giving a natural number answer, in that case the output is marked by $\uparrow$. In particular, as we will see below, in finite identification the function is defined to give a natural number as an answer from $I_{\mathcal{L}}$ only once.

Finite identifiability of a class of languages from positive data is defined by the following chain of conditions.

Definition 2.3 (Finite identification, FIN)**.** A learning function φ:

1. finitely identifies $L_i \in \mathcal{L}$ on ε iff, when inductively given ε, at some point φ outputs i, and stops;
2. finitely identifies $L_i \in \mathcal{L}$ iff it finitely identifies L_i on every ε for L_i;
3. finitely identifies $\mathcal{L}$ iff it finitely identifies every $L_i \in \mathcal{L}$;
4. a class $\mathcal{L}$ is finitely identifiable iff there is a learning function φ that finitely identifies $\mathcal{L}$.

In the last section we will relax the condition of recursivity of φ to discuss some case of non-effective finite identifiability.

3 Definite tell-tale sets and finite identification

A necessary and sufficient condition for finite identifiability has already been formulated in the literature (Mukouchi 1992). It involves a modified, stronger notion of finite tell-tale (Angluin 1980), namely the *definite finite tell-tale*.

Definition 3.1 (Mukouchi 1992)**.** A set S_i is a definite finite tell-tale for $L_i \in \mathcal{L}$ if

1. $S_i \subseteq L_i$,
2. S_i is finite, and
3. for any index j, if $S_i \subseteq L_j$ then $L_i = L_j$.

Finite identifiability can be then characterized in the following way.

Theorem 1 (Mukouchi 1992). *A class $\mathcal{L}$ is finitely identifiable from positive data iff there is an effective procedure that on input i produces all elements of a definite finite tell-tale of L_i.*

In other words, each set in a finitely identifiable class contains a finite subset that distinguishes it from all other sets in the class.

4 Eliminative Power and Finite Identifiability

Identifiability in the limit (Gold 1967) of a class of languages guarantees the existence of a reliable strategy that allows for convergence to a correct hypothesis for every language from the class. The exact moment at which a correct hypothesis has been reached is not known and in general can be uncomputable. Things are different in finite identifiability. Here, the learning function is allowed to answer only once. Hence, the conjecture is based on certainty. In other words, the learner must know that the answer she gives is true, because there is no place for a change of mind later.

Knowing that one hypothesis is true means to be able to exclude all other possibilities. In this section we define the notion of *eliminative power* of a piece of information, which reflects the informative strength of data with respect to a certain class of sets.

Definition 4.1. Let us take $\mathcal{L}$ — an indexed class of recursive languages, and $x \in \bigcup \mathcal{L}$. The eliminative power of x with respect to $\mathcal{L}$ is determined by a function $El_{\mathcal{L}} : \bigcup \mathcal{L} \rightarrow \wp(\mathbb{N})$, such that:

$$El_{\mathcal{L}}(x) = \{i \mid x \notin L_i \;\&\; L_i \in \mathcal{L}\}.$$

Additionally, we will use $El_{\mathcal{L}}(X)$ for $\bigcup_{x \in X} El_{\mathcal{L}}(x)$.

In other words, function $El_{\mathcal{L}}$ takes x and outputs the set of indexes of all the sets in $\mathcal{L}$ that are inconsistent with x, and therefore in the light of the information x they can be "eliminated".

We can now characterize finite identifiability in terms of the eliminative power.

The following is easy to observe.

Proposition 1. *A set S_i is a definite tell-tale set of $L_i \in \mathcal{L}$ iff*

1. $S_i \subseteq L_i$,
2. S_i *is finite, and*
3. $El_{\mathcal{L}}(S_i) = \mathcal{I}_{\mathcal{L}} - \{i\}$.

Moreover, from Theorem 1 we know that finite identifiability of an indexed family of recursive languages requires that every set in a class has a DFTT. Formally:

Theorem 2. *A class $\mathcal{L}$ is finitely identifiable from positive data iff there is an effective procedure that for any i supplies a finite set $S_i \subseteq L_i$, such that*

$$El_{\mathcal{L}}(S_i) = \mathcal{I}_{\mathcal{L}} - \{i\}.$$

4.1 The Computational Complexity of Finite Identifiability Check

As has already been mentioned in the introduction, we aim at analyzing the computational complexity of finding DFTTs. In order to do that we restrict to finite classes of finite sets. One may ask about the purpose of further reduction of sets that are already finite. In fact, if a finite class of finite sets is finitely identifiable, then each element of the class is already its own DFTT. However, finite sets can be much larger than their DFTTs. For example, we can take a class of the following form:

$$\mathcal{L} = \{L_i = \{2i, 2^i \text{ first odd integers}\} \mid i = 1, \ldots, n\}.$$

In the case of $\mathcal{L}$ a reduction to the minimal information that suffices for finite identification makes a significant difference in the length of the process of learning.

Theorem 3. *Checking whether a finite class of finite sets is finitely identifiable is polynomial with respect to the size of the class, i.e. the number of sets in the class and the maximal cardinality of sets in the class.*

Proof. The procedure consists of computing $El_{\mathcal{L}}(x)$ and checking whether for each $L_i \in \mathcal{L}$, $El(L_i) = \mathcal{I}_{\mathcal{L}} - \{i\}$.

Let us first focus on computing $El_{\mathcal{L}}(x)$ for $\bigcup \mathcal{L}$. We take $\mathcal{L}$ and assume that $|\mathcal{L}| = m$, and that the largest set in $\mathcal{L}$ has n elements.

In the first steps we have to obtain $\bigcup \mathcal{L}$. After this,we list for each element of $\bigcup \mathcal{L}$ the indices of the sets to which the element does not belong. In this step we have computed $El_{\mathcal{L}}(x)$ for each $x \in \bigcup \mathcal{L}$. All components of this procedure can clearly be performed in polynomial time with respect to m and n. It remains to check whether for all $L_i \in \mathcal{L}$, $\bigcup_{x \in L_i} El_{\mathcal{L}}(x) = \mathcal{I}_{\mathcal{L}} - \{i\}$. This involves essentially only the operation of sum. □

From this analysis we conclude that checking whether a finite class of finite sets is finitely identifiable is quite easy, polynomial task. Nevertheless, as we saw in the example in the beginning of this section, it can be time consuming if n is a large number.

5 Finding a Minimal Definite Finite Tell-tale

We are now ready to introduce one of the two nonequivalent notions of minimality of the DFTTs. We will call S_i a minimal DFTT of L_i in $\mathcal{L}$ if and only if all the elements of the sets in S_i are essential for finite identification of L_i in $\mathcal{L}$, i.e. taking an element out of the set S_i will decrease its eliminative power with respect to $\mathcal{L}$, and hence it will no longer be a DFTT. We will show that a language can have many minimal DFTTs of different cardinalities. This will give us cause to introduce another notion of minimality — minimal-size DFTT.

Definition 5.1. Let us take a finitely identifiable indexed family of recursive languages $\mathcal{L}$, and $L_i \in \mathcal{L}$. A minimal DFTT of L_i in $\mathcal{L}$ is an $S_i \subseteq L_i$, such that

1. S_i is a DFTT for L_i in $\mathcal{L}$, and

2. $\forall X \subset S_i \; El_{\mathcal{L}}(X) \neq \mathcal{I}_{\mathcal{L}} - \{i\}$.

Theorem 4. *Let $\mathcal{L}$ be a finitely identifiable finite class of finite sets. Finding a minimal DFTT of $L_i \in \mathcal{L}$ can be done in polynomial time.*

Proof. Assume that the class $\mathcal{L}$:

1. is finitely identifiable;
2. is finite;
3. consists only of finite sets.

From the assumptions 1 and 3, we know that for each $L_i \in \mathcal{L}$ a DFTT exists, in fact L_i is its own DFTT.

The following procedure yields a minimal DFTT for each $L_i \in \mathcal{L}$.

We want to find a set $X \subseteq L_i$ such that

$$El(X) = \mathcal{I}_{\mathcal{L}} - \{i\}, \text{ but } \forall Y \subset X \; El(Y) \neq \mathcal{I}_{\mathcal{L}} - \{i\}.$$

First we set $X := L_i$.

We look for $x \in X$ such that $El(X - \{x\}) = \mathcal{I}_{\mathcal{L}} - \{i\}$. If there is no such element, X is the desired DFTT. If there is such x exists, we set $X := X - \{x\}$, and repeat the procedure.

Let $|L_i| = n$, where $|\cdot|$ stands for cardinality. The number of comparisons needed for finding a minimal DFTT of L_i in $\mathcal{L}$ is bounded by n^2. □

5.1 Example

Let us consider the class

$$\mathcal{L} = \{L_1 = \{1,3,4\}, L_2 = \{2,4,5\}, L_3 = \{1,3,5\}, L_4 = \{4,6\}\}.$$

The procedure of finding minimal DFTTs for sets in $\mathcal{L}$ is as follows.

1. We construct a list of the elements from $\bigcup \mathcal{L}$.
2. With each element x from $\bigcup \mathcal{L}$ we associate $El_{\mathcal{L}}(x) = \{i \,|\, x \notin L_i\}$, i.e. the set of indices of sets to which x does not belong (names of sets that are inconsistent with x). Table 1 shows the result of the two first steps for $\mathcal{L}$.
3. The next step is to find minimal DFTTs for every set in the class $\mathcal{L}$. As an example, let us take the first set $L_1 = \{1,2,3\}$. We order elements of L_1, and take the first element of the ordering. Let it be 1. We compute $El_{\mathcal{L}}(L_1 - \{1\})$, it turns out to be $\{2,3,4\}$. We therefore accept the set $\{3,4\}$ as a smaller DFTT for L_1. Then we try to further reduce the obtained DFTT, by checking the next element in the ordering, let it be 3. $El_{\mathcal{L}}(\{3,4\} - \{3\}) = \{4\} \neq \{2,3,4\}$, so 3 cannot be subtracted without loss of eliminative power. We perform the same check for the last singleton $\{4\}$. It turns out that $\{3,4\}$ cannot further be reduced. We give $\{3,4\}$ as a minimal DFTT of L_1.[1]
4. We perform the same procedure for all the sets in $\mathcal{L}$. As a result we get minimal DFTTs for each $L_i \in \mathcal{L}$ presented in table 2.

[1]Checking only singletons is enough because the eliminative power of sets is defined as the sum of the eliminative power of its elements.

x	$\mathbf{El}_{\mathcal{L}}(\mathbf{x})$
1	$\{2,4\}$
2	$\{1,3,4\}$
3	$\{2,4\}$
4	$\{3\}$
5	$\{1,4\}$
6	$\{1,2,3\}$

Table 1: Eliminative power of the elements in $\bigcup \mathcal{L}$ with respect to $\mathcal{L}$

set	a minimal DFTT
$\{1,3,4\}$	$\{3,4\}$
$\{2,4,5\}$	$\{4,5\}$
$\{1,3,5\}$	$\{3,5\}$
$\{4,6\}$	$\{6\}$

Table 2: DFTTs of $\mathcal{L}$

6 Finding a Minimal-Size Definite Finite Tell-tale

We can use the notion of eliminative power to construct a procedure for finding minimal-size DFTTs of a finitely identifiable class $\mathcal{L}$.

Let us again take $\mathcal{L}$ — a class of finite sets. We assume that $|\mathcal{L}| = m$.

To find a DFTT of minimal size for the set $L_i \in \mathcal{L}$, one has to perform a search through all the subsets of L_i starting from singletons, looking for the first X_i, such that $El(X_i) = \mathcal{I}_{\mathcal{L}} - \{i\}$.

DFTTs of minimal size need not be unique. Which one is encountered first depends on the manner of performing the search. Below we describe the example discussed before.

6.1 Example

Let us consider again the class from Example 5.1, namely

$$\mathcal{L} = \{L_1 = \{1,3,4\}, L_2 = \{2,4,5\}, L_3 = \{1,3,5\}, L_4 = \{4,6\}\}.$$

1. We construct a list of the elements from $\bigcup \mathcal{L}$.

2. With each element x from $\bigcup \mathcal{L}$ we associate $El_{\mathcal{L}}(x) = \{i | x \notin L_i\}$, i.e. the set of hypotheses for sets to which x does not belong (names of sets that are inconsistent with x). Table 1 presents the result of the two first steps for $\mathcal{L}$.

3. The next step is to find minimal-size DFTTs for every set in the class $\mathcal{L}$. As an example, let us take the first set $L_1 = \{1,3,4\}$. We are looking for $X \subseteq L_1$ of minimal size, such that $El_{\mathcal{L}}(X) = \mathcal{I}_{\mathcal{L}} - \{L_1\}$.

(a) We look for $\{x\}$ such that $x \in L_1$ and $El_{\mathcal{L}}(\{x\}) = \{2, 3, 4\}$. There is no such singleton.

(b) We look for $\{x, y\}$ such that $x, y \in L_1$ and $El_{\mathcal{L}}(\{x\}) = \{2, 3, 4\}$. There are two such sets: $\{1, 4\}$ and $\{3, 4\}$.

4. We perform the same procedure for all $L_i \in \mathcal{L}$. As a result we get minimal-size DFTTs for each of $\mathcal{L}$, the result is presented in table 3.

set	minimal-size DFTTs
$\{1, 3, 4\}$	$\{1, 4\}$ or $\{3, 4\}$
$\{2, 4, 5\}$	$\{2\}$
$\{1, 3, 5\}$	$\{1, 5\}$ or $\{3, 5\}$
$\{4, 6\}$	$\{6\}$

Table 3: Minimal-size DFTTs of $\mathcal{L}$

Let us now compare the two resulting reductions of sets from $\mathcal{L}$ (see Table 3). The case of L_2 shows that the two procedures give different outcomes. The actual difference in this case is not huge, but in bigger sets it can be significant.

set	a minimal DFTT	minimal-size DFTTs
$\{1, 3, 4\}$	$\{3, 4\}$	$\{1, 4\}$ or $\{3, 4\}$
$\{2, 4, 5\}$	$\{4, 5\}$	$\{2\}$
$\{1, 3, 5\}$	$\{3, 5\}$	$\{1, 5\}$ or $\{3, 5\}$
$\{4, 6\}$	$\{6\}$	$\{6\}$

Table 4: A comparison of minimal and minimal-size DFTTs of $\mathcal{L}$

6.2 Running time

Let us now analyze and discuss the running time of such a procedure.

First we need to compute $El_{\mathcal{L}}(x)$ for $\bigcup \mathcal{L}$. From the Theorem 3 we know that it can be done in polynomial time.

Now, let us approximate the number of steps needed to find a minimal-size DFTT of a chosen set $L_i \in \mathcal{L}$. We again assume that $|\mathcal{L}| = m$, and L_i has n elements.

In the procedure described above we performed a search through, in the worst case, all combinations from 1 to $|L_i|$, to find the right set $X \subseteq L_i$, such that $El_{\mathcal{L}}(X)$ satisfies the condition of eliminative all hypothesis but h_i. So, for each set L_i, the number of comparisons that have to be performed is:

$$n + \frac{n!}{2!(n-2)!} + \frac{n!}{3!(n-3)!} + \ldots + 1 = 2^{n-1}$$

6.3 Computational Complexity

It is costly to find minimal-size DFTTs. As we have seen above, our procedure leads to a complete search through the large space of all subsets of a language. We call this computational problem the MINIMAL-SIZE DFTT Problem, and define it formally below. The problem is checking whether $L_i \in \mathcal{L}$ has a DFTT of size k or smaller.

Definition 6.1 (MINIMAL-SIZE DFTT Problem).

Instance A finite class of finite sets $\mathcal{L}$, a set $L_i \in \mathcal{L}$, and a positive integer $k \leq |L_i|$.

Question Is there a minimal DFTT $X_i \subseteq L_i$ of size $\leq k$?

We are going to show that the MINIMAL-SIZE DFTT Problem is NP-complete by pointing out that it is equivalent to the MINIMUM COVER Problem, which is know to be NP-complete (Karp 1972). Let us recall it below.

Definition 6.2 (MINIMAL COVER Problem).

Instance: Collection P of subsets of a finite set F, positive integer $k \leq |P|$.

Question: Does P contain a cover for X of size k or less, i.e. a subset $P' \subseteq P$ with $|P'| \leq k$ such that every element of X belongs to at least one member of P'?

Theorem 5. *The* MINIMAL-SIZE DFTT *Problem is NP-complete.*

Proof. First let us observe that by Theorem 2, MINIMAL-SIZE DFTT Problem is equivalent to the following Problem:

Definition 6.3 (MINIMAL-SIZE DFTT′ Problem).

Instance: Collection $\{El(x)|x \in L_i\}$, positive integer $k \leq |L_i|$.

Question: Does $\{El(x)|x \in L_i\}$ contain a cover for $\mathcal{I}_{\mathcal{L}} - \{i\}$ of size k or less, i.e. a subset $Y_i \subseteq \{El(x)|x \in L_i\}$ with $|Y_i| \leq k$ such that every element of $\{El(x)|x \in L_i\}$ belongs to at least one member of Y_i?

It is easy to observe that MINIMAL-SIZE DFTT′ Problem is a notational variant of MINIMUM COVER Problem, i.e. we take $F = \mathcal{I}_{\mathcal{L}}$, $P = \{El(x)|x \in L_i\}$ (and therefore $|P| = |L_i|$), and $X = \mathcal{L}_{\mathcal{L}} - \{i\}$. Therefore MINIMAL-SIZE DFTT′ Problem is NP-complete. Since MINIMAL-SIZE DFTT′ Problem is equivalent to MINIMAL-SIZE DFTT Problem, we conclude that MINIMAL-SIZE DFTT Problem is also NP-complete. □

The MINIMAL-SIZE DFTT Problem may have to be solved by an optimal ("good") teacher, who is expected to give only relevant information, and guarantee fast learning. Of course such a teacher may have insight in the construction of $\mathcal{L}$ and then may have knowledge of minimal-size DFTTs in a different manner.

7 Preset Learning

In this section we will discuss the notion of *preset learner*, i.e. a learner that exclusively makes use of DFTTs for each language of the class. This concept is based on the intuition that this is the most simple-minded way of going about identifying a language finitely. Moreover it is very easy to teach to a learner. The notion is most important in the case of finite classes of languages because in the infinite case there seems to be not much other recourse for finite learners anyway. Again the complexity of finding minimal size DFTTs is the main theme of this section.

We begin with a general discussion without restricting $\mathcal{L}$ to being finite. Let us take a finitely identifiable class $\mathcal{L}$, and $L_i \in \mathcal{L}$ and consider the collection S^i of all minimal DFTTs of $L_i \in \mathcal{L}$.

Proposition 2. *ε is an environment for $L_i \in \mathcal{L}$ iff for some $n \in \mathbb{N}$, $set(\varepsilon|n)$ is a superset of some S^i_j in S^i.*

Proof. ($\Rightarrow$) Since $\mathcal{L}$ is finitely identifiable, at some stage $\varepsilon|n$ in ε for L_i a DFTT S'_i for L_i has to occur. Therefore there has to exist a minimal $S^i_j \subseteq S'_i \subseteq set(\varepsilon|n)$.

($\Leftarrow$) trivial. □

This means that for every environment of $L_i \in \mathcal{L}$ there is a finite point at which the elements of some minimal definite tell-tale have been enumerated and the learner can safely and economically conclude L_i. Using Proposition 2 we can then define a finite learner who decides on the right language at the first moment that this is possible. In order to define the fastest preset learner we have to consider a learner that can make use of any minimal DFTT. Therefore, for every set the learner has access to the set of all minimal DFFTs. So, strictly speaking it is not a preset learner.

Definition 7.1. The fastest preset learner φ_{fast} is defined in the following way.

$$\varphi_{fast}(\varepsilon|n) = \begin{cases} i & \text{if } \exists S^i_j \in S^i (S^i_j \subseteq set(\varepsilon|n) \ \&\ S^i_j \not\subseteq set(\varepsilon|n-1)), \\ \uparrow & \text{otherwise.} \end{cases}$$

The function φ_{fast} will identify the collection $\mathcal{L}$ because an environment for a language L_i will give at a proper time a (minimal) DFTT for that language. On the other hand, if $\varphi_{fast}(\varepsilon|n) = i$, then $set(\varepsilon|k)$ for $k < n$ does not have enough eliminative power to exclude all other hypotheses in $\mathcal{L}$, and hence no other successful learners could conjecture L_i either.

Clearly φ_{fast} is not defined in an effective manner and therefore does not adhere to Definition 2. If we are in the case of a finite set of finite languages the definition is effective. In the next subsection we discuss its complexity.

7.1 Computational complexity

Let us again go back to the case of finite class of finite sets. To compute the set S^i of all minimal DFTTs of $L_i \in \mathcal{L}$ we need to perform the procedure for finding a minimal DFTT for all possible orderings of elements in L_i. Therefore the simple procedure described earlier has to be performed $n!$ times. This indicates that finding the set of all minimal DFTTs is quite costly. In fact we show that the problem is NP-hard.

Proposition 3. *Finding S^i of $L_i \in \mathcal{L}$ is NP-hard.*

Proof. It is easy to observe that once we enumerate S^i of L_i, we can find a minimal-size DFTT of L_i in polynomial time, by simply picking one of the smallest sets in S^i. This means that the MINIMAL-SIZE DFTT Problem for L_i can be polynomially reduced to the problem of finding S^i for L_i. □

The fastest learner does not have to work this way. Since in the finite case languages are their own DFTTs, the Learner can use these to arrive at a conclusion as quickly as possible. Fast finite identification on the basis of DFTTs of finite classes of finite sets has then the same complexity as regular finite identification.

7.2 Recursion-theoretic Complexity

In this section we return to the case of infinite classes of languages. We will present results concerning minimal DFTTs and minimal-size DFTTs. We show that for recursive (preset) learning functions there are definite restrictions with regard to the minimality of the DFTTs they can use.

In the following we will use the manner of speech where we will say for example that e is a Turing machine if we mean that e is an integer that codes a Turing machine, M_e, and that $f(a) = \{b, c\}$ if $f(a)$ codes the finite set containing just b and c.

Let us now recall the Kleene's T-Predicate.

Definition 7.2 (T-predicate)**.** $T(e, x, y)$ holds iff e is a Turing machine and y is a computation of e with input x.

With the use of the T-predicate the Halting Problem can be formulated in the following way:

$$M_e(e) \downarrow \iff \exists y T(e, e, y).$$

Theorem 6. *There exists a preset learner φ with an effective function f that gives a DFTT $f(i) = X_i$ for each L_i, but for which no recursive function g exists such that for each i, $g(i)$ is the set of all minimal DFTTs for L_i.*

Proof. Let us consider the following class of finite sets $\mathcal{L} = \{L_i | i \in \mathcal{L}\}$, such that:

$$L_i = \{2i, 2(\mu y T(i, i, y)) + 1\}.$$

We can easily observe that $|L_i| = 2 \iff M_i(i) \downarrow$. Note that it is decidable for an arbitrary natural number whether it is a member of L_i. If n is even it should be $2i$. If n is odd $\frac{n-1}{2}$ should code a computation of M_i on i. Minimal DFTTs of L_i are $\{2i\}$, and, in case $M_i(i) \downarrow$, $\{2(\mu y T(i, i, y)) + 1\}$. It is clear that the function f such that

$$f(i) = \{\{2i\}, \{2(\mu y T(i, i, y)) + 1\}\}$$

cannot be given by a total recursive function, because its existence would solve Halting Problem. □

We can improve on this theorem by showing that a minimal-size DFTT can not always be given effectively.

Theorem 7. *There exists preset learner φ with an effective function f that gives a DFTT X_i for each L_i, but for which no recursive function g exists such that for each i, $g(i)$ is a minimal size DFTT for L_i.*

Proof. Let $j : \mathbb{N}^2 \to \mathbb{N}$ be a recursive pairing function (bijection) with inverses j_1 and j_2. We now consider the class $\mathcal{L} = \{L_i | i \in \mathcal{L}\}$ such that:

$$L_i = \{3j_1(i), 3j_2(i) + 1, 3(\mu y T(i, i, y)) + 2\}.$$

Note that a minimal-size DFTT of L_i is $\{3j_1(i), 3j_2(i)+1\}$ or $\{3(\mu y T(i, i, y))+2\}$ if $M_i(i) \downarrow$. Clearly the minimal size DFTTs of L_i cannot be given by a total recursive function g, because the existence of g would solve Halting Problem. □

The next question that comes to mind is whether it is always possible to effectively obtain a minimal DFTT. Our further work on the subject involves an attempt to solve this problem.

8 Conclusions and future work

We used a characterization of finite identification of languages from positive data to discuss the complexity of finding an optimal teaching strategy in finite identification. We introduced two notions: minimal DFTT and minimal-size DFTT. By viewing the informativeness of examples as their power to eliminate certain conjectures, we have checked the computational complexity of 'finite teachability' from minimal DFTT and minimal-size DFTT. In the former case the problem turns out to be PTIME computable, while the latter falls into the NP-complete class of problems. This suggests that it is easy to teach in a way that avoids irrelevant information but it is potentially difficult to teach in the most effective way. In the last section we restrict our attention to preset learners and we use our results on minimality to define a non-effective fastest preset learner who has access to all minimal DFTTs in a class and decides on a language as soon as possible. The essential part of this method is to find a set of all minimal DFTTs for a language. We show that this problem is NP-hard. We show some related results in the analogous recursion-theoretic context.

Some links between finite identification and dynamic epistemic logic (see e.g. van Ditmarsch, van der Hoek, and Kooi 2007, for an introduction) have already been established (Gierasimczuk 2009, Dégremont and Gierasimczuk 2009). For dynamic epistemic logic the restriction to finite sets of finite languages is very natural, so the analysis of its complexity that we present in this paper can be applied to strengthen this connection.

In the future we want to investigate more complex languages that consist of strings. We will be able to analyze another notion of minimality of relevant information, based not on the amount of data but on its simplicity.

References

D. Angluin. Inductive inference of formal languages from positive data. *Information and Control*, 45(2):117–135, 1980.

C. Dégremont and N. Gierasimczuk. Can doxastic agents learn? On the temporal structure of learning. In *Proc. 2nd LORI*, volume 5834 of *LNAI*, pages 90–104. Springer, 2009.

M. R. Garey and D. S. Johnson. *Computers and Intractability*. W. H. Freeman and Co., San Francisco, 1979.

N. Gierasimczuk. Bridging learning theory and dynamic epistemic logic. *Synthese*, 169(2):371–384, 2009.

E. Gold. Language identification in the limit. *Information and Control*, 10: 447–474, 1967.

R. M. Karp. Reducibility among combinatorial problems. In R. E. Miller and J. W. Thatcher, editors, *Complexity of Computer Computations*, pages 85–103. Plenum Press, 1972.

Y. Mukouchi. Characterization of finite identification. In *Proc. AII '92*, pages 260–267, London, UK, 1992. Springer-Verlag.

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

Twelve Angry Men: A Dynamic-Epistemic Study of Awareness, Implicit and Explicit Information

Davide Grossi and Fernando R. Velázquez-Quesada

Institute for Logic, Language and Computation, Universiteit van Amsterdam.
D.Grossi@uva.nl, F.R.VelazquezQuesada@uva.nl

Abstract

By moving from a suggestive example, the paper analyzes how information flows among agents involved in a deliberation. By interchanging information, agents become aware of details, draw the attention of the group to some issues, perform inferences and announce what they know. The proposed framework, based on the paradigm of dynamic epistemic logic, captures how knowledge results from step-wise multi-agent interaction.

1 Introduction

A jury faces the following task:

> "You've listened to the testimony [...] It's now your duty to sit down and try and separate the facts from the fancy. One man is dead. Another man's life is at stake. If there's a reasonable doubt [...] as to the guilt of the accused [...], then you must bring me a verdict of not guilty. If there's no reasonable doubt, then you must [...] find the accused guilty. However you decide, your verdict must be *unanimous*."

This is the setting of Sydney Lumet's 1957 movie "12 Angry Men" and represents a paradigmatic example of a collective decision-making scenario. These kind of scenarios, where a group of agents have to establish whether a given state-of-affairs holds or not (Kornhauser and Sager 1986), have been object of extensive study in the literature on judgment aggregation (List and Puppe 2009). However, while judgment aggregation focuses on the social-theoretic aspects of such decision-making processes, like properties of voting rules, possibility of reaching collective judgments with 'desirable' properties and others, this paper looks at the deliberation phase that typically precedes the very act of voting and, in particular, at its knowledge-related aspects.[1]

To best illustrate the problem at issue, we will consider the following example, taken from the mentioned movie.

Example 1 (12 Angry Men)**.** *The jury members are engaged in the deliberation that will lead to their unanimous verdict. An excerpt from their discussion is shown below.*

A: Now, why were you rubbing your nose like that?
H: If it's any of your business, I was rubbing it because it bothers me a little.
A: Your eyeglasses made those two deep impressions on the sides of your nose.
A: I hadn't noticed that before.
A: The woman who testified that she saw the killing had those same marks on the sides of her nose.
...
G: Hey, listen. Listen, he's right. I saw them too. I was the closest one to her. She had these things on the side of her nose.
...
D: What point are you makin'?
D: She had dyed hair, marks on her nose. What does that mean?
A: Could those marks be made by anything other than eyeglasses?
...
D: How do you know what she saw? How does he know all that? How do you know what kind of glasses she wore? Maybe they were sunglasses! Maybe she was far-sighted! What do you know about it?
C: I only know the woman's eyesight is in question now.
...
C: Don't you think the woman may have made a mistake?
B: Not guilty.

Agent *A* supports the idea that the defendant cannot be proven guilty beyond any reasonable doubt. Agent *H*'s action of rubbing his nose makes *A* aware of an issue that has not been considered before: marks on the nose. When he considers the issue, he remembers that the witness of the killing had such marks, and he announces it. Now everyone knows (in particular, *G* remembers) that the woman had marks on the side on her nose. Then, *A* draws an inference and announces what he has concluded: the marks are due to the use of glasses.

After the announcement, it is *C* who infers that the woman's eyesight is in question now. Finally, *B* makes the last reasoning step and announces to everybody that the defendant is not guilty beyond any reasonable doubt.

This example can be formally represented using the Dynamic Epistemic Logic (*DEL*) framework (van Ditmarsch et al. 2007), but by using its standard version we would get a single announcement that makes everyone knows that the defendant is not guilty beyond any reasonable doubt.[2] This one-shot solution, though simple and useful in some settings, abstracts from interesting subtleties involved in the example, where the agents' final knowledge is the result of several informational actions.

The present paper, an extension of Grossi and Velázquez-Quesada (2009), uses the principles of the *DEL* framework. Semantically, we define a model

[1]Recent literature on judgment aggregation has actually recognized the formal analysis of the deliberative phase preceding voting to be one of the key open research questions of the field (Dietrich and List 2007).

[2]A more adequate formalization can be achieved by using the notion of *belief* instead of *knowledge*, but we will stick to the latter to show how, even when restricted to it, our finer representation sheds some light on the small steps that leads to the final conclusion.

for representing the information a set of agents have at a certain stage, and we define operations over the model to represent actions that change this information; syntactically, we use a modal propositional language for describing the model and then we incorporate new modalities for describing the effects of the operations. Nevertheless, we assume a more fine-grained perspective on knowledge and on how it changes through different actions. Instead of using the strong Epistemic Logic notion of *knowledge* with all its omniscience properties, we extend the components of the standard *DEL* model and the syntax of the standard *DEL* language in order to work with the weaker notions of explicit and implicit knowledge as well as the notion of awareness. Then we define actions that modify the model-components in a fine-grained way, allowing us to represent a variety of informational actions.

Before presenting the definitions of the static and the dynamic parts of our framework (Section 3 and 4, respectively), let us settle down how we will understand the three notions of information in which our work is based: *awareness of*, *implicit information* and *explicit information*.

2 Awareness, implicit and explicit information

The first step towards a formal representation of the dynamics of information that takes place in our example is to analyze the involved notions of information.

We start with the strongest notion, that of *explicit information*. This information is directly available to the agent without any further effort. In our running example, all members of the Jury are explicitly informed (in this case, they *explicitly know*) that a killing has taken place, a boy is being accused of the killing, and a woman has testified affirming that she saw the killing.

There is also information that is not directly available; information that follows from what they explicitly known but should be "put in the light" by means of some reasoning step. In our example, at some stage, agent D is explicitly informed that the witness had marks on her nose. From that information it follows that she wears glasses, but D is just *implicitly* informed about it; he needs to perform an inference step to reach that conclusion.

But even if D does not have explicit information about the witness using glasses at that point, he considers it as a possibility, just like he considers possible for the accused to be innocent or guilty. Such possibilities are part of the current discussion; more syntactically, they are part of the agent's current language. On the other hand, before H rubs his nose, the possibility of the woman having or not marks in her nose is not considered by the agents: the agents are not *aware of* that possibility. Note that being not *aware of* a possibility does not imply that an agent have or does not have information about it. In our example, while H was completely uninformed about the possibility of the witness having marks on her nose, A knew that the witness had such marks, but he just disregarded that information.

Here is a more mathematical example of the three mentioned notions. Consider an agent trying to prove that if $p \to q$ and p are the case, then so is q. She is *explicitly* informed that $p \to q$ and p are the case, but she is informed about q just *implicitly*, since she needs to perform an inference step in order to make it explicit. While trying to prove q, the agent is not *aware of* r, s and other atomic proposition. Again, this does not say that the agent has or not information about *r*, *s* and so; it just says that it is not part of the information the agent entertains right now.

Relation between these notions. The relation we assume between *implicit* and *explicit* information is standard: explicit information is implicit information that has been "put in the light" by some reasoning mechanism. Therefore, explicit information is always part of the implicit one.

The relation between implicit information and information we are aware of can be seen from two different perspectives. We could assume that the agent's implicit information is everything that the agent can get to know, including what she would get if she became aware of every issue. Then, the information the agent is *awareness of* would be part of her *implicit* information. From our discussion before it can be seen that we will adopt another perspective: the information the agent is aware of actually defines her *language*, and no informational notion can go beyond it. In particular, the agent cannot have implicit information about something that is not part of her language. Therefore, implicit information is part of the information the agent is aware of.

Extending *DEL*. The main components of our running example are the different notions of information the agents have and the way these notions change through different informational actions. A good initial idea for analyzing the situation is to use the *DEL* setting in its basic form, *Public Announcement Logic (PAL)*, in order to describe the conversation that takes place in the Jury room, but we would get the undesirable outcome of a single announcement that makes all the agent know that the accused is not guilty beyond any reasonable doubt.

The reason for this is that the static part of *DEL*, classical *Epistemic Logic (EL)*, deals with a notion of information that, besides assuming a uniform language for all the agents, has the property of being closed under logical consequence.[3] Then, a simple act of announcing that "*. . . the woman who testified that she saw the killing had those same marks on the sides of her nose*" is enough to make all the agents know that there is a reasonable doubt about the guiltiness of the accused. Even the *awareness* extension introduced in Fagin and Halpern (1988) is not enough since, as we will explain later in Section 3.2, yields either a notion of *awareness* that does not define a language, or else a notion of explicit information that is closed under logical consequence. As we mentioned before, this closure, useful in some situations, hides important informational actions, like the act of becoming aware of some topic, or the act of applying an inference step. Being our main goal to describe these fine-grained notions and their correspondent informational actions, we have chosen to define an extension of the *DEL* framework that is adequate for our purposes.

Outline of the paper. In the following section we introduce our approach for representing the three notions that have been presented, together with some of the properties we get and a brief comparison with other approaches. Then we show how we can use it to describe the information the members of the Jury have before the discussion takes place. Accordingly, Section 4 presents the definitions of what we consider are the three informational actions that are relevant in our example, again presenting some of their properties and

[3]This is the case assuming the usual semantic model, Kripke models, and the traditional semantic definition for the modal operator. Other semantical approaches, like neighborhood models, would give us weaker systems, but would still suffer from closure under logical equivalence.

comparing it with some other works. These definitions allow us to describe how the agents' information change throughout the scene. Finally, Section 5 provides conclusions and pointers to future work.

3 The static framework

In this section we define the formal language that allow us to describe situations like the Jury's example, together with the semantic model and the formulas' semantic interpretation.

3.1 Basic language, models and interpretation

Definition 3.1 (Language $\mathcal{L}$)**.** Let $\mathtt{P}$ be a set of atomic propositions and let $\mathtt{A}$ be a set of agents. Formulas φ and rules ρ of the language $\mathcal{L}$ are given by

$$\varphi ::= p \mid {}^{[i]}p \mid \mathrm{A}_i\,\varphi \mid \mathrm{R}_i\,\rho \mid \neg\varphi \mid \varphi \vee \psi \mid \Box_i\varphi$$
$$\rho ::= (\{\varphi_1, \ldots, \varphi_{n_\rho}\}, \psi)$$

where $p \in \mathtt{P}$ and $i \in \mathtt{A}$. We denote by $\mathcal{L}_f$ the set of formulas of $\mathcal{L}$, and by $\mathcal{L}_r$ its set of rules. Other boolean connectives ($\wedge, \rightarrow, \leftrightarrow$) as well as the diamond modalities $\Diamond_i$ are defined as usual ($\Diamond_i\varphi := \neg\Box_i\neg\varphi$, for the last case).

The language $\mathcal{L}$ extends that of Epistemic Logic (*EL*) with three new basic components: ${}^{[i]}p$, $\mathrm{A}_i\,\varphi$ and $\mathrm{R}_i\,\rho$. Formulas of the form ${}^{[i]}p$ indicate that *agent i has proposition p available (at her disposal)* for expressing her information, and will be used to define the notion of *awareness of*. Formulas of the form $\mathrm{A}_i\,\varphi$ (*access* formulas) and $\mathrm{R}_i\,\rho$ (*rule-access* formulas) indicate that *agent i can access formula* φ and *rule* ρ, respectively. While the first will be used to express the information the agent can access, the second will be used to express the *processes* the agent can use. These processes, in our case syntactic rules, will allow the agent to extends her explicit information, and deserve a brief discussion.

The rules. Let us go back to our example for a moment, and consider how the three notions of information change. The subjects an agent considers, the *awareness of* notion, change as a consequence of someone else introducing a new issue, and the *implicit information* notion changes accordingly. On the other hand, changes in explicit information are not the result of external influences but the result of the agent's own reasoning steps. In order to perform such reasoning steps the agent needs certain extra information, just like knowing that the length of the legs of a right triangle is useless for getting the length of the hypothenuse unless we know the Pythagoras theorem or, in a simpler setting, just like knowing p and $p \rightarrow q$ is useless to infer q unless we know the *modus ponens* rule. In our "logical" setting, the more natural way of representing this extra information is by syntactic rules that allow the agent to infer consequences of what she explicitly known. The formal definition of the situations in which a rule can be applied and of its outcome are provided in Section 4.

When dealing with rules, the following definitions will be useful.

Definition 3.2 (Premises, conclusion and translation)**.** Let ρ be a rule in $\mathcal{L}_r$ of the form $(\{\varphi_1, \ldots, \varphi_{n_\rho}\}, \psi)$. We define

$$\mathsf{pm}(\rho) := \{\varphi_1, \ldots, \varphi_{n_\rho}\} \quad \text{the } \textit{set of premises} \text{ of } \rho$$
$$\mathsf{cn}(\rho) := \psi \quad \text{the } \textit{conclusion} \text{ of } \rho$$

Moreover, we define its *translation* $\mathrm{tr}(\rho) \in \mathcal{L}_f$ as an implication whose antecedent is the (finite) conjunction of the rule's premises and whose consequent is the rule's conclusion:

$$\mathrm{tr}(\rho) := \bigwedge \mathsf{pm}(\rho) \rightarrow \mathsf{cn}(\rho)$$

Availability of formulas. Formulas of the form ${}^{[i]}p$ allow us to express the availability of atomic propositions. The notion can be extended to express availability of formulas of the whole language in the following way.

Definition 3.3. Let i, j be agents in $\mathtt{A}$. Define

$$\begin{array}{rclrclrcl}
{}^{[i]}({}^{[j]}\varphi) & := & {}^{[i]}\varphi & {}^{[i]}(\neg\varphi) & := & {}^{[i]}\varphi & {}^{[i]}\rho & := & {}^{[i]}\,\mathrm{tr}(\rho) \\
{}^{[i]}(\mathrm{A}_j\,\varphi) & := & {}^{[i]}\varphi & {}^{[i]}(\varphi \vee \psi) & := & {}^{[i]}\varphi \wedge {}^{[i]}\psi & & & \\
{}^{[i]}(\mathrm{R}_j\,\rho) & := & {}^{[i]}\rho & {}^{[i]}(\Box_j\varphi) & := & {}^{[i]}\varphi & & &
\end{array}$$

Intuitively, formulas of the form ${}^{[i]}\varphi$ express that φ is available to agent i, and this happens whenever all the atoms in φ are available to i. For example, ${}^{[i]}(\neg p)$ is defined by ${}^{[i]}p$, that is, the formula $\neg p$ is available to agent i whenever p is available to her. On the other hand, ${}^{[i]}(p \vee q)$ is given by ${}^{[i]}p \wedge {}^{[i]}q$, that is, $p \vee q$ is available to agent i whenever *both* p and q are available to her.

Note how the definition of availability for agent i in the case of formulas involving other agent j (${}^{[j]}\varphi$, $\mathrm{A}_j\,\varphi$, $\mathrm{R}_j\,\rho$ and $\Box_j\varphi$) discard any reference to j. With this definition, we are assuming that all agents are "available" to each other, that is, all agents can talk about any other agent. Some other approaches, like van Ditmarsch and French (2009), consider also the possibility of agents that are not necessarily aware of all other agents. We will not pursue that possibility here, but we emphasize that this idea has interesting consequences, as we will mention once we provide our definitions for the *awareness of*, *implicit* and *explicit information* notions in Section 3.2.

Having defined the language $\mathcal{L}$, we now define the semantic model in which the formulas will be interpreted.

Definition 3.4 (Semantic model)**.** Let $\mathtt{P}$ be the set of atomic propositions and $\mathtt{A}$ the set of agents. A *semantic model* is a tuple $M = \langle W, R_i, V, \mathsf{PA}_i, \mathsf{AC}_i, \mathsf{R}_i\rangle$ where:

- $\langle W, R_i, V\rangle$ is a standard Kripke model with W the non-empty *set of worlds*, $R_i \subseteq W \times W$ an *accessibility relation* for each agent $i \in \mathtt{A}$ and $V : W \rightarrow \wp(\mathtt{P})$ the *atomic valuation*.
- $\mathsf{PA}_i : W \rightarrow \wp(\mathtt{P})$ is the *propositional availability function*, returning the set of atomic propositions agent $i \in \mathtt{A}$ has at disposal at each possible world.
- $\mathsf{AC}_i : W \rightarrow \wp(\mathcal{L}_f)$ is the *access set function*, returning the set of formulas agent $i \in \mathtt{A}$ can access at each possible world.
- $\mathsf{R}_i : W \rightarrow \wp(\mathcal{L}_r)$ is the *rule set function*, returning the set of rules agent $i \in \mathtt{A}$ can access at each possible world.

The pair (M, w) with M a semantic model and w a world in it is called a *pointed semantic model*. We denote by $\mathbf{M}$ the class of all semantic models.

Our semantic model extends Kripke models with three functions, PA_i, AC_i and R_i, that allow us to give semantic interpretation to the new formulas.

Definition 3.5 (Semantic interpretation)**.** Let $M = \langle W, R_i, V, \mathsf{PA}_i, \mathsf{AC}_i, \mathsf{R}_i \rangle$ be a semantic model, and take a world $w \in W$. The *satisfaction* relation $\models$ between formulas of $\mathcal{L}$ and the pointed semantic model (M, w) is given by

$(M,w) \models p$	iff	$p \in V(w)$	$(M,w) \models \mathrm{A}_i\,\varphi$	iff	$\varphi \in \mathsf{AC}_i(w)$
$(M,w) \models \neg\varphi$	iff	$(M,w) \not\models \varphi$	$(M,w) \models {}^{[i]}p$	iff	$p \in \mathsf{PA}_i(w)$
$(M,w) \models \varphi \vee \psi$	iff	$(M,w) \models \varphi$ or $(M,w) \models \psi$	$(M,w) \models \mathrm{R}_i\,\rho$	iff	$\rho \in \mathsf{R}_i(w)$
$(M,w) \models \Box_i\varphi$	iff	for all $u \in W$, R_iwu implies $(M,u) \models \varphi$			

As it becomes evident from Definitions 3.4 and 3.5, our logic is based on a sorted language where special atoms are introduced to represent signatures in the object language (${}^{[i]}\varphi$) and direct access to information ($\mathrm{A}_i\,\varphi$ and $\mathrm{R}_i\,\rho$). These special atoms, interpreted by the dedicated valuation functions PA_i, AC_i and R_i, respectively, are used to capture our finer notions of information. We will discuss this choice in Section 3.2, but first we show that the basic epistemic axiom system is sound and complete for our framework. More precisely,

Theorem 1 (Sound and complete axiom system for $\mathcal{L}$ w.r.t. **M**)**.** *The axiom system of Table 1 is sound and strongly complete for formulas of $\mathcal{L}$ w.r.t. models in* **M**.

Prop	$\vdash \varphi$	for φ a propositional tautology
K	$\vdash \Box_i(\varphi \to \psi) \to (\Box_i\varphi \to \Box_i\psi)$	for every agent i
Dual	$\vdash \Diamond_i\varphi \leftrightarrow \neg\Box_i\neg\varphi$	for every agent i
MP	If $\vdash \varphi \to \psi$ and $\vdash \varphi$, then $\vdash \psi$	
Nec	If $\vdash \varphi$, then $\vdash \Box_i\varphi$	for every agent i

Table 1: Axiom system for $\mathcal{L}$ w.r.t. **M**

Sketch of proof. Soundness is proved by showing that axioms are valid and rules preserve validity. For completeness, construct the standard modal canonical model M and, for each maximal consistent set of formulas w, define propositional availability, access set and rule set functions in the following way:

$$\mathsf{PA}_i(w) := \{p \in \mathrm{P} \mid {}^{[i]}p \in w\} \qquad \mathsf{AC}_i(w) := \{\varphi \in \mathcal{L}_f \mid \mathrm{A}_i\,\varphi \in w\}$$
$$\mathsf{R}_i(w) := \{\rho \in \mathcal{L}_r \mid \mathrm{R}_i\,\rho \in w\}$$

With these definitions, it is easy to show that the new formulas also satisfy the *Truth Lemma*, that is,

$$(M,w) \models {}^{[i]}p \ \text{ iff } \ {}^{[i]}p \in w \qquad (M,w) \models \mathrm{A}_i\,\varphi \ \text{ iff } \ \mathrm{A}_i\,\varphi \in w$$
$$(M,w) \models \mathrm{R}_i\,\rho \ \text{ iff } \ \mathrm{R}_i\,\rho \in w$$

This gives us completeness. □

Note how there are no axioms for formulas of the form ${}^{[i]}p$, $\mathrm{A}_i\,\varphi$ and $\mathrm{R}_i\,\rho$. As mentioned, such formulas are simply special *atoms* for the dedicated *valuation functions* PA_i, AC_i and R_i. Moreover, these functions do not have any special property and there is no restriction in the way they interact with each other (but see Section 3.2 for some interaction properties). Just like axiom systems for Epistemic Logic do not require axioms for atomic propositions, our system does not require axioms for these special atoms.

On the other hand, Definition 3.3 gives us validities expressing the behaviour of ${}^{[i]}\varphi$. For example, ${}^{[i]}({}^{[j]}\varphi) \leftrightarrow {}^{[i]}\varphi$, ${}^{[i]}(\mathrm{A}_j\,\varphi) \leftrightarrow {}^{[i]}\varphi$ and ${}^{[i]}(\mathrm{R}_j\,\rho) \leftrightarrow {}^{[i]}\,\mathrm{tr}(\rho)$ are all valid in **M**-models.

3.2 Defining the relevant notions

With the language, semantic model and semantic interpretation defined, it is now time to formalize the notions informally introduced in Section 2.

Definition 3.6. The notions of *awareness*, *implicit information* and *explicit information* are defined as follows.

Agent i is *aware of* formula φ	$\mathrm{Aw}_i(\varphi) := \Box_i\,{}^{[i]}\varphi$
Agent i is *aware of* rule ρ	$\mathrm{Aw}_i(\rho) := \Box_i\,{}^{[i]}\,\mathrm{tr}(\rho)$
Agent i is *implicitly informed* about formula φ	$\mathrm{Im}_i(\varphi) := \Box_i({}^{[i]}\varphi \wedge \varphi)$
Agent i is *implicitly informed* about rule ρ	$\mathrm{Im}_i(\rho) := \Box_i({}^{[i]}\,\mathrm{tr}(\rho) \wedge \mathrm{tr}(\rho))$
Agent i is *explicitly informed* about formula φ	$\mathrm{Ex}_i(\varphi) := \Box_i({}^{[i]}\varphi \wedge \varphi \wedge \mathrm{A}_i\,\varphi)$
Agent i is *explicitly informed* about rule ρ	$\mathrm{Ex}_i(\rho) := \Box_i({}^{[i]}\,\mathrm{tr}(\rho) \wedge \mathrm{tr}(\rho) \wedge \mathrm{R}_i\,\rho)$

The *awareness of* notion. The notion of *awareness of* for agent i is defined in terms of the worlds the agent can access and the formulas she has at her disposal in each one of them. We say that agent i is aware of φ if and only if she has at her disposal all the atoms in φ in all the worlds she considers possible. With this definition, the *awareness of* notion actually defines the language of the agent. First, if the agent is aware of a formula φ, then she is aware of all the atoms in the formula. But not only that: if the agent is aware of a set of atomic propositions, then she is aware of all formulas built from such atoms.

The first statement, Proposition 1 below, follows from this lemma.

Lemma 1. *Let (M, w) be a pointed semantic model and i be an agent. Let φ be a formula in $\mathcal{L}$, and denote by $\mathsf{atm}(\varphi)$ the set of atomic propositions occurring in φ.*

If i has φ at her disposal, that is, if $(M, w) \models {}^{[i]}\varphi$, then she has at her disposal all atoms in it, that is, $(M, w) \models {}^{[i]}p$ for every $p \in \mathsf{atm}(\varphi)$. In other words, the formula ${}^{[i]}\varphi \rightarrow {}^{[i]}p$ is valid for every $p \in \mathsf{atm}(\varphi)$.

Sketch of proof. The proof is a simple induction on φ. The base case is immediate, and each inductive case follows from its respective entry in Definition 3.3 and its inductive hypothesis. □

Proposition 1. *Let (M, w) be a pointed semantic model and i be an agent. Let φ be a formula in $\mathcal{L}$.*

If i is aware of φ, that is, if $(M, w) \models \mathrm{Aw}_i(\varphi)$, then she is aware of all its atoms, that is, $(M, w) \models \mathrm{Aw}_i(p)$ for every $p \in \mathsf{atm}(\varphi)$. In other words, the formula $\mathrm{Aw}_i(\varphi) \rightarrow \mathrm{Aw}_i(p)$ is valid for every $p \in \mathsf{atm}(\varphi)$.

Proof. Suppose $(M, w) \models \mathrm{Aw}_i(\varphi)$. Then, $(M, w) \models \Box_i\,{}^{[i]}\varphi$, that is, $(M, u) \models {}^{[i]}\varphi$ for every u such that $R_i w u$. Pick any such u; by Lemma 1, $(M, u) \models {}^{[i]}p$ for every $p \in \mathsf{atm}(\varphi)$. Hence, $(M, w) \models \Box_i\,{}^{[i]}p$, that is $(M, w) \models \mathrm{Aw}_i(p)$, for every $p \in \mathsf{atm}(\varphi)$. □

The second statement, Proposition 2 below, follows from this lemma.

Lemma 2. *Let (M, w) be a pointed semantic model and i be an agent. Let $\{p_1, \ldots, p_n\} \subseteq$ $\mathtt{P}$ be a subset of atomic propositions.*

If i has all atoms in $\{p_1, \ldots, p_n\}$ at her disposal, that is, if $(M, w) \models {}^{[i]}p_k$ for every $k \in \{1, \ldots, n\}$, then she has at her disposal any formula built from such atoms, that is, $(M, w) \models {}^{[i]}\varphi$ for any formula φ built from $\{p_1, \ldots, p_n\}$. In other words, the formula ${}^{[i]}p_k \rightarrow {}^{[i]}\varphi$ is valid for every φ built from $\{p_1, \ldots, p_n\}$.

Sketch of proof. The proof, by induction on φ, simply follows Definition 3.3. □

Proposition 2. *Let (M, w) be a pointed semantic model and i be an agent. Let $\{p_1, \ldots, p_n\} \subseteq \mathtt{P}$ be a subset of atomic propositions.*

If i is aware of all atoms in $\{p_1, \ldots, p_n\}$, that is, if $(M, w) \models \mathrm{Aw}_i(p_k)$ for every $k \in \{1, \ldots, n\}$, then she is aware of any formula built from such atoms, that is, $(M, w) \models \mathrm{Aw}_i(\varphi)$ for any formula φ built from $\{p_1, \ldots, p_n\}$. In other words, the formula $\mathrm{Aw}_i(p_k) \rightarrow \mathrm{Aw}_i(\varphi)$ is valid for every φ built from $\{p_1, \ldots, p_n\}$.

Proof. Suppose $(M, w) \models \mathrm{Aw}_i(p_k)$ for every $k \in \{1, \ldots, n\}$. Then, $(M, w) \models \Box_i {}^{[i]}p_k$, that is, $(M, u) \models {}^{[i]}p_k$ for every $k \in \{1, \ldots, n\}$ and every u such that $R_i wu$. Pick any such u; by Lemma 2, $(M, u) \models {}^{[i]}\varphi$ for any formula φ built from $\{p_1, \ldots, p_n\}$. Hence, $(M, w) \models \Box_i {}^{[i]}p_k$, that is $(M, w) \models \mathrm{Aw}_i(\varphi)$. □

As mentioned before, our *awareness of* notion assumes that every agents is aware of each other. We could drop this assumption and, following van Ditmarsch and French (2009), extend PA_i-sets to provide not only the atoms but also the agents agent *i* has at her disposal in each possible world. Formulas of the form ${}^{[i]}\varphi$ can be redefined accordingly: for example, ${}^{[i]}(\Box_j\varphi) := {}^{[i]}\varphi \wedge {}^{[i]}j$, with ${}^{[i]}j$ true at (M, w) iff $j \in \mathsf{PA}_i(w)$. This gives us a more fine-grained *awareness of* notion, and has interesting consequences.

First, we can represent agents that are not aware of *themselves* by simply not including *i* in the adequate PA_i-sets. Moreover, if an agent *i* is not aware of any agent ($\mathsf{PA}_i(w) \cap \mathtt{A} = \emptyset$ for every $w \in W$), then we have an agent whose explicit information can only be *propositional*. In particular, in a single-agent setting, the agent's explicit information will be completely non-introspective *unless she becomes aware of herself*. In the classical *EL* approach to knowledge (Kripke frames with equivalence accessibility relations), agents have full introspection, both positive and negative. In some works, like van Benthem and Velázquez-Quesada (2009) and our setting, that is not the case, but still the agent can reach introspection by performing the adequate inference. With the mentioned extension, introspection becomes a matter not only of the adequate inference, but also a privilege of *self-aware* agents.

The *implicit information* notion. This notion defines everything the agent can get to know without changing her *awareness of* notion and provided she has the tools (the rules) to perform the necessary inferences. It is defined as everything that is true in all the worlds the agent considers possible, modulo her current language. Here are some of the properties of this notion.

Proposition 3. *Let (M, w) be a pointed semantic model and let i be an agent.*

1. *The agent has implicit information of all the validities involving her current awareness: if φ is a validity built from $\{p_1, \ldots, p_n\} \subseteq \mathsf{P}$ and $(M, w) \models \mathrm{Aw}_i(p_k)$ for every $k \in \{1, \ldots, n\}$, then $(M, w) \models \mathrm{Im}_i(\varphi)$.*

2. *The agent's implicit information is closed under logical consequence: if $(M, w) \models \mathrm{Im}_i(\varphi \to \psi) \land \mathrm{Im}_i(\varphi)$, then $(M, w) \models \mathrm{Im}_i(\psi)$*

Proof. Take (M, w). For the first, Proposition 2 gives us $(M, w) \models \mathrm{Aw}_i(\varphi)$, that is, $(M, w) \models \Box_i\, {}^{[i]}\varphi$. But φ is a validity, so $(M, w) \models \Box_i \varphi$. Hence, $(M, w) \models \mathrm{Im}_i(\varphi)$. For the second, suppose $(M, w) \models \mathrm{Im}_i(\varphi \to \psi) \land \mathrm{Im}_i(\varphi)$. Then we have $(M, w) \models \Box_i(\varphi \to \psi) \land \Box_i \varphi$ and, by K axiom, $(M, w) \models \Box_i \psi$. But we also have $(M, w) \models \Box_i\, {}^{[i]}(\varphi \to \psi)$, hence $(M, w) \models \Box_i\, {}^{[i]}\psi$. Therefore, $(M, w) \models \mathrm{Im}_i(\psi)$. □

The *explicit information* notion This is the strongest of the three notions: for the agent to be explicitly informed about φ, she needs to be aware and to have implicit information about it. Since AC-sets do not have any special requirement, nothing needs to be explicitly known, and the notion does not have any closure property. This suits us well, since the agent's explicit information does not *need* to have any of these requirements. We can easily imagine an agent that is not explicitly informed about any validity, or another whose explicit information is not closed under logical consequence.

We emphasize that the *explicit information* notion does not *need* to have any special property, but that does not mean that it cannot. From our *dynamic* perspective, explicit information does not need built-in properties that guarantee the agent has certain amount of minimal information; what it needs is the appropriate set of actions that explains how the agent gets this information.

Their hierarchy. By simply unfolding their definitions, it follows that our three notions behave as stated in Section 2.

Proposition 4 (The hierarchy of the notions)**.** *In* **M***-models, we have Explicit information ⊆ Implicit information and Implicit information ⊆ Awareness of for both information about formulas and information about rules. In the formulas case, the stated properties make $\mathrm{Ex}_i(\varphi) \to \mathrm{Im}_i(\varphi)$ and $\mathrm{Im}_i(\varphi) \to \mathrm{Aw}_i(\varphi)$ valid.*

Interaction between the components. Though in principle there is no relation between the different components of the model, extra properties yield particular kinds of agents.

- If available atoms are preserved by the indistinguishability relation, that is, if $p \in \mathsf{PA}_i(w)$ implies $p \in \mathsf{PA}_i(u)$ for all worlds u such that $R_i wu$, then agent i's information satisfies what we call *weak introspection on available atoms*, a property characterized by the formula ${}^{[i]}p \to \Box_i\, {}^{[i]}p$.

- In a similar way, if accessible formulas are preserved by the indistinguishability relation, that is, if $\varphi \in \mathsf{AC}_i(w)$ implies $\varphi \in \mathsf{AC}_i(u)$ for all worlds u for which $R_i wu$, then agent i's information satisfies *weak introspection on accessible formulas*, characterized by $\mathrm{A}_i\, \varphi \to \Box_i \mathrm{A}_i\, \varphi$.

These two properties give us an agent whose available atoms and access to formulas are uniform through her accessibility relation. As observed in van Benthem and Velázquez-Quesada (2009), reflexive models that satisfy them also satisfy the following formulas:

- $\Box_i\, {}^{[i]}\varphi \leftrightarrow {}^{[i]}\varphi$,
- $\Box_i({}^{[i]}\varphi \wedge \varphi) \leftrightarrow ({}^{[i]}\varphi \wedge \Box_i\varphi)$,
- $\Box_i({}^{[i]}\varphi \wedge \varphi \wedge \mathrm{A}_i\,\varphi) \leftrightarrow ({}^{[i]}\varphi \wedge \Box_i\varphi \wedge \mathrm{A}_i\,\varphi)$.

This shows how, under the mentioned properties, our definitions for the three notions coincide with the definition of explicit information of Fagin and Halpern (1988), where access to formulas (and in our case availability of atoms) falls outside the scope of the modal operator. [4]

More interestingly, our basic semantic models do not impose restriction for formulas in access sets. In particular, they can contain formulas involving atomic propositions that are not in the corresponding propositional availability set, that is, $\mathrm{A}_i\,\varphi \wedge \neg({}^{[i]}\varphi)$ is satisfiable. If we ask for formulas in access sets to be built only from available atoms (if we ask for $\varphi \in \mathsf{AC}_i(w)$ to imply $\mathsf{atm}(\varphi) \subseteq \mathsf{PA}_i(w)$) then we can represent only what we call *strong unawareness*: if the agent is unaware of φ, then becoming aware of it does not give her any explicit information about φ, simply because φ (or any formula involving it) cannot be in her access set. This property is characterized by the formula $\mathrm{A}_i\,\varphi \rightarrow {}^{[i]}\varphi$.

On the other hand, our unrestricted setting allows us to additionally represent what we call *weak unawareness*: becoming aware of atoms in φ can give the agent explicit information about φ because φ can be already in her access set. This allow us to model a *remembering* notion: I am looking for the keys in the bedroom, and then when someone introduces the possibility for them to be in the kitchen, I remember that actually I left them next to the oven.

Other possibilities and approaches

Here we make a brief recap of some other possibilities for defining these notions both in our same setting as well as in different approaches.

Syntactic awareness vs. semantic awareness. The proposed formalization of the *awareness of* notion is based on the intuition that, at each state, each agent has only a particular subset of the language at disposal for phrasing her knowledge, so to say. This intuition is modeled via dedicated atoms ${}^{[i]}p$ and their inductive extension to any formula (Definition 3.3).

This is an eminently syntactic way to look at the availability of bits of language to agents and, thus, to look at awareness. An alternative model-theoretic approach can be obtained via a relation holding between states equivalent up to a signature $\mathtt{P}_i \subseteq \mathtt{P}$ (Grossi 2009).

Definition 3.7 (Equivalence up to a signature)**.** Let $\mathtt{P}$ be a countable set of atoms, W a set of states and V a valuation function. The states $w, w' \in W$ are *equivalent up to a signature* $\mathtt{P}_i \in \wp(\mathtt{P})$ ($\mathtt{P}_i$*-equivalent*) if and only if, for any $p \in \mathtt{P}_i$, we have $p \in V(w)$ iff $p \in V(w')$. When $w, w' \in W$ are $\mathtt{P}_i$-equivalent, we write $\sim_i$. [5]

Intuitively, the relation $\sim_i$ links states that are indistinguishable if the atoms in $\mathtt{P} - \mathtt{P}_i$ are neglected. Such relation can then be used as accessibility relation, thereby yielding an extension of $\mathsf{S5}$ in which the availability of certain formulas to agents gets a semantics in terms of a notion of indistinguishability yielded by the set of atoms considered. So, the fact that agent i can make use of φ in

[4] A more detailed comparison between the works is provided in Section 3.2.

[5] This notion of states *propositionally* equivalent up to a signature can be easily extended to a notion of states *modally* equivalence up to a signature, as done in van Ditmarsch and French (2009).

eliciting her information means that she can distinguish, thanks to the atoms she has at disposal, between φ-states and $\neg\varphi$-states which, in turn, boils down to the truth of $[\sim_i]\varphi \vee [\sim_i]\neg\varphi$, where $[\sim_i]$ gets the obvious interpretation.[6]

We have the following result relating the $^{[i]}\varphi$ and $[\sim_i]\varphi$ formulas, and thus the syntactic and the semantic approach.

Proposition 5 (Syntactic vs. semantic propositional availability)**.** *Let* $\mathtt{P}$ *be a countable set of atoms,* W *a set of states and* V *a valuation function. For each propositional availability function* PA_i *there exists a propositional equivalence relation* $\sim_i$ *modulo signature* $\mathtt{P}_i \subseteq \mathtt{P}$ *such that, for any state* w *and boolean formula* φ*:*

$$(W, \mathsf{PA}_i, V), w \models {}^{[i]}\varphi \quad \Longrightarrow \quad (W, \sim_i, V), w \models [\sim_i]\varphi \vee [\sim_i]\neg\varphi.$$

The implication is strict.

Proof. The proof can be obtained by construction. Each formula $^{[i]}\varphi$ is either true or false at a pointed model $((W, \mathsf{PA}_i, V), w)$ and, by Definition 3.3, that depends only on the truth values of its atoms in that model. Now, put $\mathtt{P}_i := \{p \mid (W, \mathsf{PA}_i, V), w \models {}^{[i]}p\}$. By Definition 3.7 we obtain the relation $\sim_i$. The desired implication is proved via a simple induction on the complexity of φ. That the implication is strict, follows directly from the fact that the truth of $[\sim_i]\varphi$-formulas is preserved under substitution of equivalents, while this does not hold for $^{[i]}\varphi$-formulas. □

In other words, the syntactic representation of the availability of atoms to agents is stronger than the model-theoretic one. In particular, the failure of the inverse implication in Proposition 5 constitutes the main reason for the assumption of a syntactic view in the present paper.

Other approaches to awareness. There are other approaches to the notion of awareness in the literature; here we briefly mention some of them and how they relate to our approach.

In the classic "*Belief, awareness, and limited reasoning*" (Fagin and Halpern 1988), the notion of awareness is modelled by assigning to each agent a set of formulas in each possible world. Such sets, in principle, lack of any particular property, but several possibilities are mentioned trough that paper. Two are the main differences with respect to our setting. First, our notion is given by a set of atomic propositions and *all the formulas that can be built from them* (the $^{[i]}\varphi$ formulas), while Fagin and Halpern's notion can include p and q without including $p \wedge q$. But even if their set of formulas is exactly those formulas built from some particular set of atoms, the definitions still differ since, in their case, the notion is defined relative to the set of formulas *of the evaluation point*: the agent is aware of φ at world w iff φ is in the correspondent set *of the world* w. This differs from our definition, where we look not at the atomic propositions the agent has at her disposal in the evaluation point, but at those she has *in every world she considers possible* (our $\mathrm{Aw}_i(\varphi)$).

Putting aside the notion of awareness of an agent, (van Ditmarsch and French 2009) presents an approach similar to ours: each possible world associates to each agent a set of atomic propositions, and the notion of awareness of is defined in terms of such set. Nevertheless, they follow Fagin and Halpern:

[6] We refer the interested reader to (Grossi 2009) for more details.

their notion indeed defines a *language*, like ours, but it does it relative to the evaluation point and not to all the worlds the agent considers possible.

Other definitions of explicit information. The formal definition of explicit information/knowledge has several variants in the literature, and in particular: the $\Box_i\varphi \wedge \mathrm{A}_i\,\varphi$ (implicit information plus their awareness) of Fagin and Halpern (1988), van Ditmarsch and French (2009), and the $\mathrm{A}_i\,\varphi$ of Duc (1997), Jago (2006), van Benthem (2008), Velázquez-Quesada (2009). A simple inspection to Definition 3.6 shows that we opted here for yet another definition, along the lines presented in Velázquez-Quesada (2009), in which all the ingredients of explicit information fall under the scope of the modal box. We will compare the our $^{[i]}\varphi$-less version $\Box_i(\varphi \wedge \mathrm{A}_i\,\varphi)$ with the two we have mentioned.

The main drawback of the works using $\mathrm{A}_i\,\varphi$ is that the agent's explicit information is limited to propositional formulas. The reason for this is that the definition itself does not guarantee intuitive properties, like explicit information being implicit information or, in the case of knowledge (equivalence relations), knowledge being true. To get them, they need to ask for formulas in AC_i-sets to be true, and this requirement is preserved through the needed model operations only when formulas in these sets are purely propositional.

With respect to $\Box_i\varphi \wedge \mathrm{A}_i\,\varphi$, the difference is that we put the A_i-part of the definition under the scope of the modal operator $\Box_i$. This choice guarantees that once information is interpreted as knowledge (i.e., when considering equivalence accessibility relations), we get an implicit form of both positive and negative introspection:

$$\mathrm{Ex}_i(\varphi) \rightarrow \mathrm{Im}_i(\mathrm{Ex}_i(\varphi)) \quad \text{and} \quad \neg\mathrm{Ex}_i(\varphi) \rightarrow \mathrm{Im}_i(\neg\mathrm{Ex}_i(\varphi)).$$

The proof can be obtained by simple modal principles. Intuitively, the formulas say that if i has explicit knowledge that φ, then she implicitly knows (that is, she should be in principle able to infer) that she has explicit knowledge that φ. Conversely, if she does not have explicit knowledge that φ then she implicitly knows that she does not have explicit knowledge. These intuitively appealing properties are not satisfied by the $\Box_i\varphi \wedge \mathrm{A}_i\,\varphi$ definition, even when considering equivalence accessibility relations.

As mentioned before, the definitions become equivalent when the accessibility relation is reflexive and the AC_i-sets are preserved through accessibility relations. Under such conditions, we get $\Box_i(\varphi \wedge \mathrm{A}_i\,\varphi) \leftrightarrow (\Box_i\varphi \wedge \mathrm{A}_i\,\varphi)$.

But the main problem is that an explicit information notion given simply by implicit information plus awareness is not adequate for our purposes. Using an *awareness of* definition following Fagin and Halpern (1988) does not define a language, and using a definition following van Ditmarsch and French (2009) gives us an explicit information notion closed under logical consequence. But while their definition has two components, φ and $\mathrm{A}_i\,\varphi$, ours has three: φ, $\mathrm{A}_i\,\varphi$ *and* $^{[i]}\varphi$. Let us see why.

With our definition, implicit information ($\Box_i\varphi$) plus awareness ($\Box_i\,^{[i]}\varphi$) is not enough to get explicit information: the agent should also somehow get to know that φ is true. While trying to prove a theorem, we may be aware of the relevant notions and the theorem can be true, but still we could fail to see it. Our AC_i-sets contain formulas that have been somehow "acknowledged" as true, being inference and observation (the last one covering communication) the most common actions that do it. This way we get a notion of awareness that indeed defines the language of an agent, but we also get a notion of explicit information that is not closed under logical consequence.

3.3 Working with *knowledge*

Our current definitions do not guarantee that the agent's information is *true*, simply because the real world does not need to be among the ones the agent considers possible. In order to work with true information, that is, with the notion of *knowledge*, we can simply work in models where the accessibility relations are reflexive, but following the standard *EL* approach we will also assume symmetry and transitivity.

Definition 3.8 (Class $\mathbf{M}_K$)**.** A semantic model $M = \langle W, R_i, V, \mathsf{PA}_i, \mathsf{AC}_i, \mathsf{R}_i \rangle$ is in the class $\mathbf{M}_K$ if and only if R_i is an *equivalence* relation for all agents i.

The following proposition follows from the reflexivity of R.

Proposition 6. *In $\mathbf{M}_K$-models, implicit and explicit information are* true *information for both formulas and rules. In other words,* $\mathrm{Im}_i(\varphi) \rightarrow \varphi$ *and* $\mathrm{Ex}_i(\varphi) \rightarrow \varphi$ *are valid in the case of formulas, and* $\mathrm{Im}_i(\rho) \rightarrow \mathrm{tr}(\rho)$ *and* $\mathrm{Ex}_i(\rho) \rightarrow \mathrm{tr}(\rho)$ *in the case of rules.*

When working with models in $\mathbf{M}_K$, we will use the term *knowledge* instead of the term *information*, that is, instead of talk about implicit and explicit information, we will talk about *implicit* and *explicit* knowledge. A sound and complete axiom system for validities of $\mathcal{L}$ in $\mathbf{M}_K$-models is given by the standard $\mathsf{S5}$ system, which extends the basic one with axioms *T* ($\Box_i\varphi \rightarrow \varphi$), *4* ($\Box_i\varphi \rightarrow \Box_i\Box_i\varphi$) and *5* ($\neg\Box_i\varphi \rightarrow \Box_i\neg\Box_i\varphi$) for every agent *i*.

3.4 The example

We are now in the position to start a formal analysis of Example 1. The table below indicates the information state of the relevant members of the jury at the beginning of the conversation. The relevant atomic propositions are `gls` (*the woman wears glasses*), `mkns` (*she has marks in the nose*), `esq` (*her eyesight is in question*) and `glt` (*the accused is guilty beyond any reasonable doubt*). The relevant rules, abbreviated as $\varphi \rightarrow \psi$ with φ the (conjunction of the) premise(s) and ψ the conclusion, are $\sigma_1 : \mathtt{mkns} \rightarrow \mathtt{gls}$, $\sigma_2 : \mathtt{gls} \rightarrow \mathtt{esq}$ and $\sigma_3 : \mathtt{esq} \rightarrow \neg\mathtt{glt}$.

A	$\Box_A(\mathrm{tr}(\sigma_1) \wedge \mathrm{R}_A\,\sigma_1)$ $\Box_A(\mathrm{tr}(\sigma_2) \wedge \mathrm{R}_A\,\sigma_2)$ $\Box_A(\mathrm{tr}(\sigma_3) \wedge \mathrm{R}_A\,\sigma_3)$	$\Box_A(\mathtt{mkns} \wedge \mathrm{A}_A\,\mathtt{mkns})$	$\mathrm{Aw}_A(\mathtt{glt})$ $\mathrm{Aw}_A(\mathtt{esq})$
B	$\Box_B(\mathrm{tr}(\sigma_1) \wedge \mathrm{R}_B\,\sigma_1)$ $\Box_B(\mathrm{tr}(\sigma_2) \wedge \mathrm{R}_B\,\sigma_2)$ $\Box_B(\mathrm{tr}(\sigma_3) \wedge \mathrm{R}_B\,\sigma_3)$		$\mathrm{Aw}_B(\mathtt{glt})$
C	$\Box_C(\mathrm{tr}(\sigma_1) \wedge \mathrm{R}_C\,\sigma_1)$ $\Box_C(\mathrm{tr}(\sigma_2) \wedge \mathrm{R}_C\,\sigma_2)$ $\Box_C(\mathrm{tr}(\sigma_3) \wedge \mathrm{R}_C\,\sigma_3)$	$\Box_C(\mathtt{mkns} \wedge \mathrm{A}_C\,\mathtt{mkns})$	$\mathrm{Aw}_C(\mathtt{glt})$
G	$\Box_G(\mathrm{tr}(\sigma_1) \wedge \mathrm{R}_G\,\sigma_1)$ $\Box_G(\mathrm{tr}(\sigma_2) \wedge \mathrm{R}_G\,\sigma_2)$ $\Box_G(\mathrm{tr}(\sigma_3) \wedge \mathrm{R}_G\,\sigma_3)$		$\mathrm{Aw}_G(\mathtt{glt})$

In words, all the agents know – in the standard epistemic sense – that if somebody has some signs on her nose that means she wears glasses, that if she wears glasses then we can question her eyesight, and that someone with questioned eyesight cannot be a credible eye-witness. Also, all the agents can in principle follow this line of reasoning because each one of them has access to these rules in all the worlds each one considers possible. However, only A and C have access to the bit of information which is needed to trigger the inference, namely, that the witness had those peculiar signs on her nose. This is, nonetheless, not enough since no agent is considering the atoms $\mathtt{mkns}$ and $\mathtt{gls}$ in their "working languages": they are not aware of these issues. The only bit of language they are considering concerns the defendant being guilty or not and, in A's case, the concern about the witness eyesight.

All in all, the key aspect here is that the bits of information that can possibly generate explicit knowledge are spread across the group. The effect of the deliberation is to share this bits through dedicated announcements, which is the topic of the next section.

4 Dynamics of information

Our framework allow us to describe the information of agents at some given point in time. It is time to provide the tools that allow us to describe how this information changes. Three are the informational actions relevant for our paper: become aware, inference and public announcement.

The first one, the *awareness* action, makes the agent aware of a given atomic proposition q; it is the processes through which the agent extends her current language, and it can be interpreted simply as the introduction of a topic in a conversation. The second one, the *inference* action, allows the agent extend the information she can access by the application of a rule. This is the process through which the agent "acknowledges" that certain formula is true, making explicit her implicit information. The third one, the *announcement* action, represents the agent's interaction with the external world: she announces to the others something that she explicitly knows.

For each one of the described actions, we define a model operation representing it.

Definition 4.1. Let $M = \langle W, R_i, V, \mathsf{PA}_i, \mathsf{AC}_i, \mathsf{R}_i \rangle$ be a semantic model.

- Take $q \in \mathtt{P}$ and $j \in \mathtt{A}$. The *awareness* operation yields the model $M_{q \rightsquigarrow j} = \langle W, R_i, V, \mathsf{PA}'_i, \mathsf{AC}_i, \mathsf{R}_i \rangle$, differing from M just in the propositional availability function of agent j, which is given by

$$\mathsf{PA}'_j(w) := \mathsf{PA}_j(w) \cup \{q\} \qquad \text{for every } w \in W$$

In words, the operation ${}^{q} \rightsquigarrow^{j}$ adds the atomic proposition q to the propositional availability set of the agent j in all worlds of the model.

- Take $\sigma \in \mathcal{L}_r$ and $j \in \mathtt{A}$. The *inference* operation yields the model $M_{j \hookrightarrow \sigma} = \langle W, R_i, V, \mathsf{PA}_i, \mathsf{AC}'_i, \mathsf{R}_i \rangle$, differing from M just in the access set function of the agent j, which is given by

$$\mathsf{AC}'_j(w) := \begin{cases} \mathsf{AC}_j(w) \cup \{\mathsf{cn}(\sigma)\} & \text{if } \sigma \in \mathsf{R}_j(w) \text{ and } \mathsf{pm}(\sigma) \subseteq \mathsf{AC}_j(w) \\ \mathsf{AC}_j(w) & \text{otherwise} \end{cases}$$

for every world $w \in W$. In words, the operation ${}^{j} \hookrightarrow^{\sigma}$ adds the conclusion of σ to the access set of an agent j at a world w iff her rule and access sets at w contain σ and its premises, respectively.

- Take $\chi \in \mathcal{L}_f$ and $j \in \mathtt{A}$; recall that $\mathsf{atm}(\chi)$ denotes the set of atomic propositions occurring in χ. The *announcement* operation yields the model $M_{j:\chi!} = \langle W', R'_i, V', \mathsf{PA}'_i, \mathsf{AC}'_i, \mathsf{R}'_i \rangle$, given by

  $$W' := \{w \in W \mid (M, w) \models \chi\}, \qquad R' := R \cap (W' \times W')$$

  and, for all $w \in W'$ and $i \in \mathtt{A}$,

  $$V'(w) := V(w) \qquad \mathsf{PA}'_i(w) := \mathsf{PA}_i(w) \cup \mathsf{atm}(\chi)$$
  $$\mathsf{AC}'_i(w) := \mathsf{AC}_i(w) \cup \{\chi\}$$

  In words, the operation ${}^{j} : \chi!$ removes worlds where χ does not hold, restricting the accessibility relation and the valuation to the new domain. It also extends propositional availability sets with the atomic propositions occurring in χ and extends access sets with χ itself, preserving rule sets as in the original model.

While the first two operations affect the model components of just one agent, the third one affects those of all agents. Indeed, while the awareness operation ${}^{q} \rightsquigarrow^{j}$ affects only agent j's PA-sets and the inference operation ${}^{j} \hookrightarrow^{\sigma}$ affects only agent j's AC-sets, the announcement affects the accessibility relation as well as the PA- and the AC-sets of *every* agent. But note that affecting just the model-components of a single agent, like our first two operations do, does not imply that other agent's information does not change, as we will discuss below.

It can be easily proved that the three operations preserve models in $\mathbf{M}_K$.

Proposition 7. *If M is a $\mathbf{M}_K$-model, so are $M_{q \rightsquigarrow j}$, $M_{j \hookrightarrow \sigma}$ and $M_{j:\chi!}$.*

In order to express the effect of this operations over the agent's knowledge, we extend the language $\mathcal{L}$ with three new modalities, $\langle {}^{q}\rightsquigarrow^{j} \rangle$, $\langle {}^{j}\hookrightarrow^{\sigma} \rangle$ and $\langle {}^{j}\!: \chi! \rangle$, representing each one of our operations (their "boxed" versions are defined as their correspondent dual, as usual). We call this language *extended* $\mathcal{L}$; the semantic interpretation of the new formulas is as follows.

Definition 4.2 (Semantic interpretation)**.** Let $M = \langle W, R_i, V, \mathsf{PA}_i, \mathsf{AC}_i, \mathsf{R}_i \rangle$ be a semantic model, and take a world $w \in W$. Define the following formulas

$$\mathsf{Pre}({}^{j}\hookrightarrow^{\sigma}) := \mathrm{Ex}_j(\sigma) \wedge \bigwedge_{\varphi \in \mathsf{pm}(\sigma)} \mathrm{Ex}_j(\varphi) \qquad \mathsf{Pre}({}^{j}\!: \chi!) := \mathrm{Ex}_j(\chi)$$

expressing the precondition for ${}^{j} \hookrightarrow^{\sigma}$ and ${}^{j} : \chi!$, respectively. Then,

$$\begin{array}{lll} (M, w) \models \langle {}^{q}\rightsquigarrow^{j} \rangle\, \varphi & \text{iff} & (M_{q \rightsquigarrow j}, w) \models \varphi \\ (M, w) \models \langle {}^{j}\hookrightarrow^{\sigma} \rangle\, \varphi & \text{iff} & (M, w) \models \mathsf{Pre}({}^{j}\hookrightarrow^{\sigma}) \text{ and } (M_{j \hookrightarrow \sigma}, w) \models \varphi \\ (M, w) \models \langle {}^{j}\!: \chi! \rangle\, \varphi & \text{iff} & (M, w) \models \mathsf{Pre}({}^{j}\!: \chi!) \text{ and } (M_{j:\chi!}, w) \models \varphi \end{array}$$

The semantic definitions rely on Proposition 7: the given operations preserve models in the relevant class, so we can evaluate formulas in them. Moreover, the precondition of each operation reflects its intuitive meaning: an agent can extend her language at any point, but for applying an inference with σ she needs to know explicitly the rule and all its premises. For announcing χ, the agent simply needs to *know* it explicitly.

To get a sound and complete axiom system for the extended language, we use a standard *DEL* technique. We extend our previous "static" system (Tables 1 plus axioms *T*, *4* and *5*) with *reduction axioms*: valid formulas indicating how to translate a formula with the new modalities to a provably equivalent one without them. Then, completeness follows from the completeness of the basic system (see van Benthem and Kooi (2004) for an extensive explanation of this technique).

Theorem 2 (Sound and complete axiom system for extended $\mathcal{L}$ w.r.t. $\mathbf{M}_K$)**.** *The axioms and rules of Table 1 plus axioms T, 4 and 5 plus axioms and rules of Table 2 (with $\top$ the always true formula) form a sound and strongly complete axiom system for formulas in extended $\mathcal{L}$ with respect to models in $\mathbf{M}_K$.*

$\vdash \langle {}^{q}{\rightsquigarrow}^{j} \rangle\, p \leftrightarrow p$	$\vdash \langle {}^{q}{\rightsquigarrow}^{j} \rangle\, {}^{[i]}p \leftrightarrow {}^{[i]}p$ for $i \neq j$
$\vdash \langle {}^{q}{\rightsquigarrow}^{j} \rangle\, \neg\varphi \leftrightarrow \neg\langle {}^{q}{\rightsquigarrow}^{j} \rangle\, \varphi$	$\vdash \langle {}^{q}{\rightsquigarrow}^{j} \rangle\, {}^{[j]}p \leftrightarrow {}^{[j]}p$ for $p \neq q$
$\vdash \langle {}^{q}{\rightsquigarrow}^{j} \rangle\, (\varphi \vee \psi) \leftrightarrow \big(\langle {}^{q}{\rightsquigarrow}^{j} \rangle\, \varphi \vee \langle {}^{q}{\rightsquigarrow}^{j} \rangle\, \psi\big)$	$\vdash \langle {}^{q}{\rightsquigarrow}^{j} \rangle\, {}^{[j]}q \leftrightarrow \top$
$\vdash \langle {}^{q}{\rightsquigarrow}^{j} \rangle\, \Box_i\varphi \leftrightarrow \Box_i\langle {}^{q}{\rightsquigarrow}^{j} \rangle\, \varphi$	$\vdash \langle {}^{q}{\rightsquigarrow}^{j} \rangle\, \mathrm{A}_i\, \varphi \leftrightarrow \mathrm{A}_i\, \varphi$
If $\vdash \varphi$, then $\vdash [{}^{q}{\rightsquigarrow}^{j}]\, \varphi$	$\vdash \langle {}^{q}{\rightsquigarrow}^{j} \rangle\, \mathrm{R}_i\, \rho \leftrightarrow \mathrm{R}_i\, \rho$
$\vdash \langle {}^{j}{\hookrightarrow}^{\sigma} \rangle\, p \leftrightarrow \mathsf{Pre}({}^{j}{\hookrightarrow}^{\sigma}) \wedge p$	$\vdash \langle {}^{j}{\hookrightarrow}^{\sigma} \rangle\, {}^{[i]}p \leftrightarrow \mathsf{Pre}({}^{j}{\hookrightarrow}^{\sigma}) \wedge {}^{[i]}p$
$\vdash \langle {}^{j}{\hookrightarrow}^{\sigma} \rangle\, \neg\varphi \leftrightarrow (\mathsf{Pre}({}^{j}{\hookrightarrow}^{\sigma}) \wedge \neg\langle {}^{j}{\hookrightarrow}^{\sigma} \rangle\, \varphi)$	$\vdash \langle {}^{j}{\hookrightarrow}^{\sigma} \rangle\, \mathrm{A}_i\, \varphi \leftrightarrow \mathsf{Pre}({}^{j}{\hookrightarrow}^{\sigma}) \wedge \mathrm{A}_i\, \varphi$ for $i \neq j$
$\vdash \langle {}^{j}{\hookrightarrow}^{\sigma} \rangle\, (\varphi \vee \psi) \leftrightarrow \big(\langle {}^{j}{\hookrightarrow}^{\sigma} \rangle\, \varphi \vee \langle {}^{j}{\hookrightarrow}^{\sigma} \rangle\, \psi\big)$	$\vdash \langle {}^{j}{\hookrightarrow}^{\sigma} \rangle\, \mathrm{A}_j\, \varphi \leftrightarrow \mathsf{Pre}({}^{j}{\hookrightarrow}^{\sigma}) \wedge \mathrm{A}_j\, \varphi$ for $\varphi \neq \mathsf{cn}(\sigma)$
$\vdash \langle {}^{j}{\hookrightarrow}^{\sigma} \rangle\, \Box_i\varphi \leftrightarrow (\mathsf{Pre}({}^{j}{\hookrightarrow}^{\sigma}) \wedge \Box_i[{}^{j}{\hookrightarrow}^{\sigma}]\, \varphi)$	$\vdash \langle {}^{j}{\hookrightarrow}^{\sigma} \rangle\, \mathrm{A}_j\, \mathsf{cn}(\sigma) \leftrightarrow \mathsf{Pre}({}^{j}{\hookrightarrow}^{\sigma})$
If $\vdash \varphi$, then $\vdash [{}^{j}{\hookrightarrow}^{\sigma}]\, \varphi$	$\vdash \langle {}^{j}{\hookrightarrow}^{\sigma} \rangle\, \mathrm{R}_i\, \rho \leftrightarrow \mathsf{Pre}({}^{j}{\hookrightarrow}^{\sigma}) \wedge \mathrm{R}_i\, \rho$
$\vdash \langle {}^{j}{:}\, \chi! \rangle\, p \leftrightarrow \mathsf{Pre}({}^{j}{:}\, \chi!) \wedge p$	$\vdash \langle {}^{j}{:}\, \chi! \rangle\, {}^{[i]}p \leftrightarrow \mathsf{Pre}({}^{j}{:}\, \chi!) \wedge {}^{[i]}p$ for $p \notin \mathsf{atm}(\chi)$
$\vdash \langle {}^{j}{:}\, \chi! \rangle\, \neg\varphi \leftrightarrow (\mathsf{Pre}({}^{j}{:}\, \chi!) \wedge \neg\langle {}^{j}{:}\, \chi! \rangle\, \varphi)$	$\vdash \langle {}^{j}{:}\, \chi! \rangle\, {}^{[i]}p \leftrightarrow \mathsf{Pre}({}^{j}{:}\, \chi!)$ for $p \in \mathsf{atm}(\chi)$
$\vdash \langle {}^{j}{:}\, \chi! \rangle\, (\varphi \vee \psi) \leftrightarrow \big(\langle {}^{j}{:}\, \chi! \rangle\, \varphi \vee \langle {}^{j}{:}\, \chi! \rangle\, \psi\big)$	$\vdash \langle {}^{j}{:}\, \chi! \rangle\, \mathrm{A}_i\, \varphi \leftrightarrow \mathsf{Pre}({}^{j}{:}\, \chi!) \wedge \mathrm{A}_i\, \varphi$ for $\varphi \neq \chi$
$\vdash \langle {}^{j}{:}\, \chi! \rangle\, \Box_i\varphi \leftrightarrow (\mathsf{Pre}({}^{j}{:}\, \chi!) \wedge \Box_i[{}^{j}{:}\, \chi!]\, \varphi)$	$\vdash \langle {}^{j}{:}\, \chi! \rangle\, \mathrm{A}_i\, \chi \leftrightarrow \mathsf{Pre}({}^{j}{:}\, \chi!)$
If $\vdash \varphi$, then $\vdash [{}^{j}{:}\, \chi!]\, \varphi$	$\vdash \langle {}^{j}{:}\, \chi! \rangle\, \mathrm{R}_i\, \varphi \leftrightarrow \mathsf{Pre}({}^{j}{:}\, \chi!) \wedge \mathrm{R}_i\, \varphi$

Table 2: Extra axioms for extended $\mathcal{L}$ w.r.t. $\mathbf{M}_K$

For the diamond modalities of the three operations, $\langle {}^{q}{\rightsquigarrow}^{j} \rangle$, $\langle {}^{j}{\hookrightarrow}^{\sigma} \rangle$ and $\langle {}^{j}{:}\, \chi! \rangle$, the reduction axioms in the case of atoms, $\neg$, $\vee$ and $\Box_i$ (left column of the table) are standard: the operations do not affect atomic propositions, distribute over $\vee$ and commute with $\neg$ and $\Box_i$ modulo their preconditions (in the last case, the diamond modality of the operation turns into a box).

The interesting cases are those expressing how propositional availability, access and rule sets are affected. For the ${}^{q}\rightsquigarrow^{j}$ operation, the axioms indicate that only q is added exactly to the propositional availability sets of agent j, leaving the rest of the components of the model as before. For the ${}^{j}\hookrightarrow^{\sigma}$ operation, the axioms indicate that only the access sets of agent j are modified, and the modification consist in adding the conclusion of the applied rule. Finally, axioms of the ${}^{j} : \chi!$ operation indicate that while rule sets are not affected, propositional availability sets *of every agent* are extended with the atoms of χ and access sets are extended with χ itself.

Though the *awareness* and the *inference* operations affect only the model components of the agent who performs them, they are in some sense *public*. In the *awareness* case, ${}^{q} \rightsquigarrow^{j}$ makes ${}^{[j]}q$ true in every world in the model and, in particular, makes it true in every world any agent i considers possible, that is, $\Box_i\, {}^{[j]}q$ becomes true everywhere. This does not say that every agent becomes aware of the fact that j is now aware of q, but it does say that they will be as soon as they become aware of p and, moreover, they will be explicitly informed about it as soon as they have access to ${}^{[i]}q$. The *inference* operation ${}^{j} \hookrightarrow^{\sigma}$ behaves in a similar way since it makes $\Box_i \mathrm{A}_j\, \mathsf{cn}(\sigma)$ true in every world of the model. A further refinement of these operations, reflecting better the private character of the corresponding actions, can be found in van Benthem and Velázquez-Quesada (2009).

Basic operations. We have introduced only those operations that have a direct interpretation in our setting. One can easily imagine situations like our running example in which becoming aware, applying inference and talking to people are the relevant actions that change the agents' information. Nevertheless, from a technical point of view, two of our operations can be decomposed into more basic ones.

Access sets AC are modified through rule-based *inference*, which add the conclusion of the rule whenever their premises and rule itself are present. But following van Benthem (2008) and van Benthem and Velázquez-Quesada (2009), we can define a more basic operation $+\chi^{j}_{\psi}$ that adds an arbitrary formula χ to the access set of agent j on those worlds satisfying ψ. The formal definition of this operation is straightforward, and so is the semantic definition of a formula representing it with no precondition needed. Then, we can define our inference operation in the following way:

$$\langle {}^{j}\hookrightarrow^{\sigma} \rangle\varphi := \mathsf{Pre}({}^{j}\hookrightarrow^{\sigma}) \wedge \langle +\mathsf{cn}(\sigma)^{j}_{\psi} \rangle\varphi$$

with $\psi := \mathrm{R}_j\, \sigma \wedge \mathrm{A}_j\, \mathsf{pm}(\sigma)$.

In a similar way, we can define a *restriction* operation $\chi!$ in the *Public Announcement Logic* style by simply restricting the model to those worlds satisfying χ. Then, our *announcement* operation ${}^{j} : \chi!$ becomes an appropriate precondition and a sequence of operations: a restriction with χ, then awareness operations (once for every atom in χ) and additions of χ for every agent. Assuming a finite set of them $i_1, \ldots, i_m$, we have

$$\langle {}^{j}\!: \chi! \rangle\varphi := \mathsf{Pre}({}^{j}\!: \chi!) \wedge \langle \chi! \rangle\Big(\langle {}^{q_1}\rightsquigarrow^{i_1} \rangle \cdots \langle {}^{q_n}\rightsquigarrow^{i_1} \rangle \langle +\chi^{i_1}_{\top} \rangle\Big) \cdots \Big(\langle {}^{q_1}\rightsquigarrow^{i_m} \rangle \cdots \langle {}^{q_n}\rightsquigarrow^{i_m} \rangle \langle +\chi^{i_m}_{\top} \rangle\Big)\varphi$$

with $q_1, \ldots, q_n$ the atomic propositions occurring in χ. Note that once the *restriction* operation $\chi!$ has taken place, the rest of the operations can be performed in any order, yielding exactly the same model. They can even be performed at the same time, suggesting the idea of *parallel* model operations that, though interesting, will not be pursued here.

4.1 Some properties of the operations

The operations behave as expected, witness the following proposition.

Proposition 8.

- *The formula* $[{}^{q}\!\rightsquigarrow^{j}]\,\mathrm{Aw}_j(q)$ *is valid: after* ${}^{q}\rightsquigarrow^{j}$ *the agent* j *is aware of* q*.*
- *The formula* $[{}^{j}\!\hookrightarrow^{\sigma}]\,\mathrm{Ex}_j(\mathsf{cn}(\sigma))$ *is valid: after* ${}^{j}\hookrightarrow^{\sigma}$ *the agent* j *is explicitly informed about* $\mathsf{cn}(\sigma)$*.*
- *For* χ propositional *and any agent* i, $[{}^{j}\!:\chi!]\,\mathrm{Ex}_i(\chi)$ *is valid: after* ${}^{j}:\chi!$ *any agent* i *is explicitly informed about* χ*.*

Proof. Pick any pointed semantic model (M, w). The first property is straightforward: the operation puts q in the PA_j-set of every world in the model, so in particular $\Box_j\,{}^{[j]}q$ is true at w.

For the second one, we cover the three ingredients for explicit information. After the inference operation, the agent is aware of $\mathsf{cn}(\sigma)$ because the precondition of the operation tells us that she was already aware of σ; this gives us $\Box_j\,{}^{[i]}\mathsf{cn}(\sigma)$. Moreover, after the operation, $\mathsf{cn}(\sigma)$ is in the AC_j-set of every world that already had σ and its premises, so in particular it is in every world R_j-accessible from w since the precondition of the operation requires that σ and its premises were already there; this gives us $\Box_j \mathrm{A}_j\,\mathsf{cn}(\sigma)$. Finally, observe that the ${}^{j}\hookrightarrow^{\sigma}$ operation only affects formulas containing $\mathrm{A}_j\,\mathsf{cn}(\sigma)$; hence, $\mathsf{cn}(\sigma)$ itself cannot be affected. Because of the precondition, we know that $\mathsf{cn}(\sigma)$ holds in every world R_j-accessible from w in M; then, it is still true at every world R_j-accessible from w *in* $M_{j\hookrightarrow\sigma}$; this gives us $\Box_j\mathsf{cn}(\sigma)$. Therefore, $\mathrm{Ex}_j(\mathsf{cn}(\sigma))$ holds at w in $M_{j\hookrightarrow\sigma}$.

The third case is also straightforward. The operation guarantees that, after it, $\Box_i({}^{[i]}\chi \wedge \mathrm{A}_i\,\chi)$ is true at w. Moreover, the new model contains only worlds in which χ was true, and, since propositional formulas depend just on the valuations, χ should still be true at each one of them. Hence, $\Box_i\chi$ is true at w and therefore we have $\mathrm{Ex}_i(\chi)$ true at w in $M_{j:\chi!}$. □

The property for announcements cannot be extended to arbitrary χ's because of the well-know Moore-type formulas of the form $p \wedge \neg\Box_i p$ that become false after being announced, and therefore cannot be explicitly known. It is interesting, though, to observe how our setting differs from the standard *PAL*. In the latter, after $p \wedge \neg\Box_i p$ is announced, every agent gets to know that p is true (and precisely because of that $\neg\Box_i p$ is not true anymore). But in our setting, announcing $p \wedge \neg\Box_i p$ or even $p \wedge \neg\mathrm{Ex}_i(p)$ does not guarantee that the agents will be informed about p explicitly. This is because, though $p \wedge \neg\mathrm{Ex}_i(p)$ is introduced to the AC_i-set of every world w, nothing assures us that p will be there. For agents whose set of rules allow them to 'break down' conjunctions, only a further inference step is needed to make p explicit information.

4.2 The example

Let us go back to the discussion room of Example 1. In Section 3.4, the static part of our framework allowed us to present a still image describing the agents' information before the discussion (Section 3.4). Here, the dynamic part allows us to "press play", and see a video describing how the agents interact and how their information evolves.

Stage 1. Agent D's action of scratching his nose makes A aware of both mkns and gls. Moreover, he becomes aware of the three relevant rules, since he was already questioning the eyesight of the woman (esq).

$$\langle {}^{\mathtt{mkns}}\!\rightsquigarrow^{A}\rangle\, \langle {}^{\mathtt{gls}}\!\rightsquigarrow^{A}\rangle \big(\mathrm{Aw}_A(\mathtt{mkns}) \wedge \mathrm{Aw}_A(\mathtt{gls}) \wedge \mathrm{Aw}_A(\mathtt{mkns} \to \mathtt{gls}) \wedge \mathrm{Aw}_A(\mathtt{gls} \to \mathtt{esq}) \wedge \mathrm{Aw}_A(\mathtt{esq} \to \neg\mathtt{glt})\big)$$

Stage 2. By becoming aware of mkns, A can introduce it into the discussion. Moreover, mkns becomes part of his explicit knowledge, and he announces it.

$$\langle {}^{A}\!: \mathrm{Aw}_A(\mathtt{mkns})!\rangle \big(\mathrm{Aw}_{JURY}(\mathtt{mkns}) \wedge \mathrm{Ex}_A(\mathtt{mkns}) \wedge \langle {}^{A}\!: \mathtt{mkns}!\rangle\, \mathrm{Ex}_{JURY}(\mathtt{mkns})\big)$$

Stage 3. In particular, the simple introduction of mkns to the discussion makes it part of G's explicit knowledge, since he was just unaware of it.

$$\Box_G(\mathtt{mkns} \wedge \mathrm{A}_G\, \mathtt{mkns}) \wedge \neg\mathrm{Aw}_G(\mathtt{mkns}) \wedge \langle {}^{A}\!: \mathrm{Aw}_A(\mathtt{mkns})!\rangle\, \mathrm{Ex}_G(\mathtt{mkns})$$

Stage 4. Now, A can apply the rule sgns $\rightarrow$ gls and, after doing it, he announces the conclusion gls.

$$\langle {}^{A}\!\hookrightarrow^{\mathtt{mkns}\to\mathtt{gls}}\rangle \big(\mathrm{Ex}_A(\mathtt{gls}) \wedge \langle {}^{A}\!: \mathtt{gls}!\rangle\, \mathrm{Ex}_{JURY}(\mathtt{gls})\big)$$

Stage 5. With gls in his explicit knowledge (from A's announcement), C can apply gls $\rightarrow$ esq, announcing esq after it.

$$\langle {}^{C}\!\hookrightarrow^{\mathtt{gls}\to\mathtt{esq}}\rangle \big(\mathrm{Ex}_C(\mathtt{esq}) \wedge \langle {}^{C}\!: \mathtt{esq}!\rangle\, \mathrm{Ex}_{JURY}(\mathtt{esq})\big)$$

Stage 6. Finally, B draws the last inference and announces the conclusion.

$$\langle {}^{B}\!\hookrightarrow^{\mathtt{esq}\to\neg\mathtt{glt}}\rangle \big(\mathrm{Ex}_B(\neg\mathtt{glt}) \wedge \langle {}^{B}\!: \neg\mathtt{glt}!\rangle\, \mathrm{Ex}_{JURY}(\neg\mathtt{glt})\big)$$

Stages 1-6 could be written in one formula, and given Proposition 8, it is not difficult to check that such formula is a logical consequence of the information stated on Section 3.4.

5 Conclusions and further work

We have defined a framework to represent not only different notions of agents' information (awareness of, implicit information and explicit information), but also the way they evolve through certain epistemic actions. The framework is expressive enough to deal with situations like our running example, an excerpt of Sydney Lumet's 1957 movie "12 Angry Men".

Among the questions that arise from the present work, we mention three that we consider interesting. (1) We have discussed individual notions of knowledge, but there is also the important notion of *common knowledge*. It will be interesting to look at implicit and explicit versions of the concept, as well as how it is affected by epistemic actions. (2) We have focused on the notion of knowledge, but there are several other notions, like *belief* and *safe belief* that are worthwhile to investigate from our fine-grained perspective. (3) We have provided a *dynamic logic* approach. A future research line consists in looking at correspondences between our proposal and work on dialogues in argumentation theory Prakken and Vreeswijk (2002).

Acknowledgments. Davide Grossi is supported by **Nederlandse Organisatie voor Wetenschappelijk Onderzoek** (VENI grant Nr. 639.021.816).

References

F. Dietrich and C. List. Strategy-proof judgment aggregation. *Economics and Philosophy*, 23:269–300, 2007.

H. N. Duc. Reasoning about rational, but not logically omniscient, agents. *Journal of Logic and Computation*, 7(5):633–648, 1997.

R. Fagin and J. Y. Halpern. Belief, awareness, and limited reasoning. *Artificial Intelligence*, 34(1):39–76, 1988. doi: 10.1016/0004-3702(87)90003-8.

D. Grossi. A note on brute vs. institutional facts: Modal logic of equivalence up to a signature. In *Normative Multi-Agent Systems*, 2009. URL `http://drops.dagstuhl.de/opus/volltexte/2009/1910`.

D. Grossi and F. R. Velázquez-Quesada. *Twelve Angry Men*: A study on the fine-grain of announcements. In He et al. (2009), pages 147–160. doi: 10.1007/978-3-642-04893-7_12.

X. He, J. F. Horty, and E. Pacuit, editors. *Logic, Rationality, and Interaction, Second International Workshop, LORI 2009, Chongqing, China, 2009. Proceedings*, volume 5834 of *Lecture Notes in Computer Science*, 2009. Springer. doi: 10.1007/978-3-642-04893-7.

M. Jago. Rule-based and resource-bounded: A new look at epistemic logic. In T. Ågotnes and N. Alechina, editors, *Proceedings of the Workshop on Logics for Resource-Bounded Agents, ESSLLI 2006*, pages 63–77, Malaga, Spain, 2006.

L. A. Kornhauser and L. G. Sager. Unpacking the court. *Yale Law Journal*, 96 (1):82–117, 1986.

C. List and C. Puppe. Judgment aggregation: A survey. In *Oxford Handbook of Rational and Social Choice*. Oxford University Press, 2009.

H. Prakken and G. Vreeswijk. Logics for defeasible argumentation. In D. Gabbay and F. Guenthner, editors, *Handbook of Philosophical Logic*, volume 4, pages 218–319. Kluwer Academic Publishers, Dordrecht, 2nd edition, 2002.

J. van Benthem. Merging observation and access in dynamic logic. *Journal of Logic Studies*, 1(1):1–17, 2008.

J. van Benthem and B. Kooi. Reduction axioms for epistemic actions. In R. Schmidt et. al., editors, *Advances in Modal Logic (Technical Report UMCS-04-09-01)*, pages 197–211. University of Manchester, 2004.

J. van Benthem and F. R. Velázquez-Quesada. Inference, promotion, and the dynamics of awareness. Institute for Logic, Language and Computation, 2009. URL `http://www.illc.uva.nl/Publications/ResearchReports/PP-2009-43.text.pdf`.

H. van Ditmarsch and T. French. Awareness and forgetting of facts and agents. In *Proceedings of the 2009 IEEE/WIC/ACM International Joint Conference on Web Intelligence and Intelligent Agent Technologies (WI-IAT 2009)*, Milan, 2009.

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

F. R. Velázquez-Quesada. Inference and update. *Synthese (Knowledge, Rationality and Action)*, 169(2):283–300, July 2009. doi: 10.1007/s11229-009-9556-2.

F. R. Velázquez-Quesada. Dynamic logics for explicit and implicit information. In He et al. (2009), pages 325–326. doi: 10.1007/978-3-642-04893-7_31.

Relevant Epistemic Logic

Ondrej Majer and Michal Peliš

Institute of Philosophy, Academy of Sciences of the Czech Republic
majer@site.cas.cz, pelis.michal@gmail.com

Abstract

We propose a system of epistemic logic much weaker than the standard modal frameworks, which is based on the relevant logic R, extended with a distinctive epistemic modality K. The intended interpretation is that $K\varphi$ holds (relative to a given information state s) if there is another information state (a source) available at s, confirming φ.

1 Introduction

The problem of representation of epistemic states and their changes has been discussed for a long time. The classical solution takes knowledge operator as a standard necessity-like modal operator and interprets the standard modal axioms (K, T, 4, 5) as epistemic properties (closure, truth, positive introspection, negative introspection). The most popular formalization (used also in computer science) is based on the epistemic version of S5, in which knowledge turns out to be an indistinguishability between epistemic states.

This approach has been extensively criticized (see, e.g., Fagin et al. (2003) and Duc (2001)) for being unrealistically strong. Agents it represents are 'too perfect'—they are, e.g., logically omniscient (they know all the logical truths) and fully introspective (they are explicitly aware of their both positive and negative knowledge). For these reasons such representations are sometimes called epistemic logics of *implicit* knowledge.

The introspection axioms can be omitted if we use systems weaker than S4, however, the omniscience appears in all systems of normal epistemic logics. One possibility to solve this was to use dynamic epistemic logic. In Duc (2001) we can find solutions based on modifications of standard Kripke semantics (awareness and impossible worlds) as well as solutions based on a combination of temporal and epistemic logic and complexity approaches (algorithmic knowledge). In our approach we shall avoid omniscience by using a weaker system than that of a normal modal logic, namely the framework of distributive relevant logic.

There have been some proposals combining epistemic and relevant frameworks (see Cheng (2000) and Wansing (2002)), but as they have a different aim then our approach, we are not going to discuss them here. From a purely technical point of view there are a number of ways to introduce modalities in the relevant framework—Greg Restall in Restall (2000) provides a nice general overview. We take a different approach—instead of adding modalities externally we use notions already contained in the relevant framework to build our knowledge operator. There are several reasons for that. The relevant framework is complex on its own and adding completely independent modalities on the top of it would make it even harder to deal with. Second reason is an interpretation of the relevant framework itself. It has been criticized for its seeming (in our opinion) non-intuitivness and a lack of a generally accepted clear philosophical interpretation. Providing an epistemic reading to some of the components of the relevant logic we would like to contribute to the collection of its interpretations. We also think that relevant framework very naturally represents our prototypical example of a rational agent—a scientist dealing with scientific data.

2 Agent

Imagine a scientist performing some experiments in a laboratory. Besides data from her own observations she has obviously access to other sources of data relevant for her research (articles, databases, etc.). Information available to her has two basic ingredients: experimental data ('facts')—inputs and outputs of experiments/observations—and 'laws'—generalizations extracted from the experimental data.

If we consider these two kinds of data from the point of view of a logical framework, we can, with some simplification, say that basic 'facts' are typically represented by atoms and their conjunctions and disjunctions, while basic 'laws' are represented by conditionals (and their combinations).

Abstracting from our current motivation, the question of an adequate interpretation of conditionals has a long history. It has been discussed since the beginning of 20th century (Hugh MacColl, Carl Irvin Lewis). The majority of solutions which have been produced agree, that a material implication does not reflect intuitions about conditionals and the way they are used in standard communication and that an adequate representation should require some connection between the antecedent and the consequent. This is in an accordance with our 'laboratory' motivation. The connection we are looking for is a regularity or a law-like connection between antecedent and consequent data and it is clear that the material implication is not an appropriate representation of this kind of connection as, among other things, it connects any two arbitrary formulas. For example, it holds for any α, β that

1. $\alpha \rightarrow (\beta \rightarrow \alpha)$
2. $(\alpha \wedge \neg\alpha) \rightarrow \beta$
3. $\alpha \rightarrow (\beta \vee \neg\beta)$

In our epistemic reading the material implication would make a 'law' from every two 'facts', which would obviously make the representation useless. It also admits not very useful 'laws' with the tautological consequent (as in 3).

The list of undesired properties does not end here. Our prototypical scientist is obviously faulty—due to an error her data may contain some contradicting pieces. But material implication obviously cannot deal with errors in the data. One such error corrupts all the remaining data (from a contradiction everything follows, see 2).

The tautologies 1–3 are just examples of the paradoxes of material implication. The fact, that these 'paradoxes' were completely solved only in the systems of relevant logics, the obvious choice for a conditional for our scientific agent is relevant implication.

3 Relational semantics

We criticized the modal tradition of accessing epistemic logic, however, like to keep some of its basic principles. We would like to have a notion closed to that of a possible world in the sense of a set of formulas representing an epistemic state of an agent. We also want to make agent's knowledge dependent not only on her current epistemic state, but also on some related or alternative epistemic states. These requirements naturally lead to some sort of a relational semantics. From what we discussed above it, should be clear that we cannot use a standard Kripke semantics with possible worlds and a binary accessibility relation, we need a more general relational structure.

Our agent is imperfect (in fact all the human agents are). One consequence of it we already mentioned: she can obtain contradicting data. She also might be unable to decide about truth/falsity of every formula (of a given language). Thus we have to weaken the notion of possible world in order to account for these cases. We shall call the new entity allowing to accommodate an inconsistent and/or incomplete data *situation*. We also replace the standard relation of epistemic accessibility with the relation of *independent confirmation*. This notion will be introduced in the section 4.

Our point of departure will be the distributive relevant logic R of Anderson and Belnap. Although the most natural way to introduce relevant logics is certainly proof theoretical (see, e.g., Paoli (2002)), we base our framework on the relational Routley-Meyer semantics, as developed by Mares (2004), Restall (1999), Paoli (2002), and others, on which we shall define epistemic modalities.

We give an informal exposition of structures in the relevant frame and definition of connectives (for formal definitions see the Appendix A).

3.1 Relevant frame

A relevant frame is a structure $\mathbf{F} = \langle S, L, C, \trianglelefteq, R \rangle$, where S is a non-empty set of situations (states), $L \subseteq S$ is a non-empty set of designated *logical situations*, $C \subseteq S^2$ is a *compatibility* relation, $\trianglelefteq\, \subseteq S^2$ is a relation of *involvement*, $R \subseteq S^3$ is an *relevance* relation.

A model $\mathbf{M}$ is a relevant frame with the relation $\Vdash$, where $s \Vdash \varphi$ has the same meaning as in Kripke frames—that s carries the information that the formula φ is true ($\varphi \in s$ if we consider states to be sets of formulas).

Situations Situations (sometimes also called information states) play in our framework the same role as possible worlds in Kripke frames. We assume, they consist of data immediately available to the agent. Like possible worlds, we can see situations as sets of formulas, but, unlike possible worlds, situations might be incomplete (neither φ nor $\neg\varphi$ is true in s) or inconsistent (both φ and $\neg\varphi$ are true in s).

Conjunction and disjunction Classical (weak) conjunction and disjunction correspond to the situation when the agent combines local data, i.e., data from her current situation. They behave in the same way as in the case of classical Kripke frames—their validity is given locally:

$$s \Vdash \psi \wedge \varphi \text{ iff } s \Vdash \psi \text{ and } s \Vdash \varphi$$

$$s \Vdash \psi \vee \varphi \text{ iff } s \Vdash \psi \text{ or } s \Vdash \varphi$$

Weak connectives are the only ones which are defined locally. The truth of negation and implication depends also on the data in situations, related to the actual ones, so they are modal by nature. It is possible to define strong conjunction and disjunction as well (see the Appendix A).

Implication Implication is a modal connective in the sense that it depends on a neighborhood of a current situation, which is given by the ternary *relevance* relation R. In fact it is analogous to the strong (necessary) implication in a standard Kripke frame. We know that an implication $(\varphi \rightarrow \psi)$ is necessarily true in a given world in a Kripke frame iff in all worlds, accessible from the given one, where the antecedent holds, the consequent holds as well. In other words, the implication $(\varphi \rightarrow \psi)$ holds necessarily if it holds through all the neighborhood of the given world. We can read the relevant implication in a very same way, except the neighborhood of a situation s is given by pairs of situations (y, z) such that (s, y, z) are related by R. We shall call y, z antecedent and consequent situations, respectively. We say that the implication $(\varphi \rightarrow \psi)$ holds at the situation s iff it is the case that for every antecedent situation y where φ (the antecedent of the implication) holds, ψ (the consequent of the implication) holds at the corresponding consequent situation z.

$$s \Vdash (\varphi \rightarrow \psi) \text{ iff } (\forall y, z \in S)(Rsyz \text{ implies } (y \Vdash \varphi \text{ implies } z \Vdash \psi))$$

The relation R reflects in our interpretation actual experimental setups. Antecedent situations correspond to some initial data (outcome of measurements or observations) of some experiment, while the related consequent situations correspond to the corresponding resulting data of the experiment. Implication then corresponds to some (simple) kind of a rule: if I observe in my current situation, that at every experiment (represented by a couple antecedent–consequent situation) each observation of φ is followed by an observation of ψ, then I accept 'ψ follows φ' as a rule.

Negation In Kripke models the *negation* of a formula φ is true at a world iff φ is not true there. As situations can be incomplete and/or inconsistent, this is not an option any more. Negation becomes a modal connective and its meaning depends on the worlds related to the given world by a binary modal relation C known as *compatibility*. Informally we can see the compatible situations as information sources our scientist wants to be consistent with. (Imagine data of different research groups working on related subject.) Relevant negation does not correspond straightforwardly to 'necessary false'. We do not require that the negated formula in question is false in the neighborhood of the given world, we just require no world in the neighborhood contains this formula unnegated.

The formula $\neg\varphi$ holds at $s \in S$ iff it is not 'possible' (in the standard modal sense with respect to the relation C) that φ; at no situation s', compatible with ('accessible from') the situation s, it is the case that φ (either s' is incomplete with respect to φ or $\neg\varphi$ holds there).

$$s \Vdash \neg\varphi \text{ iff } (\forall s' \in S)(sCs' \text{ implies } s' \nVdash \varphi)$$

Informally speaking, the agent can explicitly deny some information (a piece of data) only if no research group in her neighborhood claims it is true. This condition also has a normative side: she has to be skeptical in the sense that she denies everything not positively supported by any of her colleagues (in the situations related to her actual situation).

If we want to grant negative facts the same basic level as positive facts, we can read the clause for the definition of compatibility in the other direction: the agent can relate her actual situation just to the situations which do not contradict her negative facts.

Properties of the compatibility relation obviously determin the kinds of negation obtained. We shall not discuss them here, let us just note, that we assume compatibility is symmetric, but it is in general neither reflexive (inconsistent situations are not self-compatible) nor transitive. (For a formal definition see the Appendix A.)

Logical situations The framework we presented so far is very weak: there are just few tautologies valid in all situations and some of the important ones—those being usually considered as basic logical laws—are missing. For example the almost uniformly accepted identity axiom ($\alpha \to \alpha$) and the Modus Ponens rule fail to hold in every situation.

This is closely connected to the question how to define truth in a relevant frame (model). If we take a hint from Kripke frames, we should equate truth in a frame with truth in every situation. But this would gives us an extremely weak system with some very unpleasant properties (cf. Restall (1999)). Designers of relevant logics took a different route; instead of requiring truth in all situations, they identify the truth in a frame just with the truth in all logically 'well behaved' situations. These situations are called *logical*. In order to satisfy the 'good behavior' of a situation l it is enough to require that all the information in any antecedent situation related to l is contained in the corresponding consequent situation as well: for each $x, y \in S$, $Rlxy$ implies $|x| \subseteq |y|$, where $|s|$ is the set of all formulas, which are true in the situation s.

It is easy to see that situations constrained in this way validate both the identity axiom and (implicative) Modus Ponens.

Involvement Involvement is a relation resembling the persistence relation in intuitionistic logic—we can see it as a relation of information growth. However not every two situations which are in inclusion with respect to the validated formulas are in the involvement relation. We require that such an inclusion is observed or witnessed and only the logical situations can play the role of a witness.

$$x \trianglelefteq y \text{ iff } (\exists l \in L)(Rlxy)$$

This completes our exposition of relational semantics for relevant logics. We now move to epistemic modalities.

4 Knowledge

As we already mentioned, we are not going to introduce epistemic modalities as an external notion. The relevant framework with the motivation we presented already contains enough modal notions to define an epistemic operator we need. We therefore decided to use these notions rather than introduce new ones.

In the classical epistemic frame what an agent knows in a world w is defined as what is true in all epistemic alternatives of w, which are given by the corresponding accessibility relation. Our idea of the agent as a scientist processing some kind of data requires a different approach.

We assume our agent in her current situation s observes (has a direct approach to) some data, represented by formulas which are true at s. She is aware of the fact that these data might be unreliable (or even inconsistent). In order to accept some of the current data as knowledge the agent requires a confirmation from some 'independent' sources.

We require from a source of a current situation s to satisfy the following conditions:

1. A source shall be more elementary (it should not contain more data) than the current situation. (A source is below s in the $\trianglelefteq$-relation.)

2. The data from the source should not contradict the data in the current situation. (A source is compatible with s.)

3. A source shall be different from the current situation.

Definition 4.1 (Knowledge)**.** $s \Vdash K\varphi$ iff

$$(\exists x \in S)(sC^{\triangleleft}x \text{ and } x \Vdash \varphi),$$

where

$$sC^{\triangleleft}x \text{ iff } sCx,\ x \trianglelefteq s,\ \text{and } x \neq s.$$

In short, φ is known iff there is an source ('lower' compatible situation different from the actual one) validating φ.

We allowed our agent to deal with inconsistent data in order to get a more realistic picture. However, the agent should be able to separate inconsistent data. The modality we introduced provides us with just such an appropriate filter. Let us assume both φ and $\neg\varphi$ are in s (e.g., our agent might received such inconsistent information from two different sources). The agent considers both φ and $\neg\varphi$ as available data, but neither of them is confirmed information as according to the definition, no situation compatible to s can contain either φ or $\neg\varphi$.

4.1 Basic properties

It is to be expected that our system blocks all the undesirable properties of both material and strict implication. Moreover, we ruled out the validity of some of the properties of 'classical' epistemic logics that we have criticized, in particular, both positive and negative introspection, as well as some closure properties.

Let us have a relevant frame $\mathbf{F} = \langle S, L, C, \trianglelefteq, R \rangle$. Recall that the truth in the frame **F** corresponds to the truth in the logical situations of **F** (under any valuation). We will also consider a stronger notion—truth in all situations of **F** (under any valuation). This notion is interesting from the point of view of our motivation as our agent might happen to be in other actual situations than the logical ones.

Factivity Our approach makes the truth axiom **T** valid. For any situation (not only a logical one) $s \in S$, if φ is known at s ($s \Vdash K\varphi$), then there is a $\trianglelefteq$-lower compatible witness with φ true, which makes φ to be true at s as well. Thus, formula

$$K\alpha \to \alpha$$

is valid.

K-axiom In our interpretation the validity of the axiom **K** is not well motivated. **K** would in fact correspond to a 'distribution of confirmation': if an implication is confirmed, then the confirmation of the antecedent implies the confirmation of the consequent. Which does not need to be the case.

$$\not\models K(\alpha \to \beta) \to (K\alpha \to K\beta)$$

Introspection We defined knowledge as independently confirmed data. In this reading the axioms **4** and **5** rather than to introspection correspond to a 'second order confirmation' (if α is confirmed then the confirmation of α is confirmed as well, similarly for the negative introspection). It is easy to see that both axioms fail.

$$\not\models K\alpha \to KK\alpha$$

$$\not\models \neg K\alpha \to K\neg K\alpha$$

Necessity and negation There is a difference between $s \nVdash K\varphi$ and $s \Vdash \neg K\varphi$. The former simply says that φ is not confirmed at the current situation s, while the latter is stronger (at least in the case of selfcompatible situations), it says that φ is not confirmed in any situation compatible with s. From this point of view it is uncontroversial that both $K\varphi$ (confirmation in the current situation) and $\neg K\varphi$ (the lack of confirmation in the compatible situations) might be true in some situation s (in this case s is not compatible with itself). On the other hand no information can be confirmed in a current situation, if the corresponding negative data are available

$$\not\models K(\varphi) \wedge \neg\varphi$$

4.2 Closure properties

In the introduction we criticized too strong closure properties of the standard modal representations of knowledge. In fact the question how strong conditions shall be imposed on epistemic states to obtain an adequate representation is one of the crucial choices of the knowledge representation. It is also closely related to the problem of logical omniscience.

We can see the machinery of 'logical expansion' as having two basic ingredients. One is knowledge of all the tautologies of the logical system in question guaranteed by the necessity rule. The other is Modal Modus Ponens, which produces all consequences of any new piece of (non-logical) information.

Our system turns out to be extremely weak and avoids both of these closure properties and some more. It can be seen as anti-logical and pragmatic—in a sense that our agent believes (accepts) just what is (or was) observed. Even the data corresponding to logical laws have to be confirmed.

Necessitation rule The necessitation rule,

$$\frac{\varphi}{K\varphi}$$

common to all normal epistemic logics, guarantees among other things that all the tautologies of the logical system in question are known. In our framework this would mean that all the logical truths are confirmed. This is in general not the case. Let us assume that φ is valid formula (i.e., $l \Vdash \varphi$, for every logical situation l). The necessity rule would imply the validity of $K\varphi$. However, for $l \Vdash K\varphi$ an confirmation from a different source is required, so there must be a situation x such that $x \Vdash \varphi$ and $lC^{\triangleleft}x$, which, in general, does not need to be the case.

Modal Modus Ponens Closure of knowledge with respect to logical consequence, which is a part of logical omniscience (if an agent knows both φ and $\varphi \to \psi$, then she knows ψ as well) is forced by the validity of the modal Modus Ponens:

$$\frac{K\alpha \quad K(\alpha \to \beta)}{K\beta}$$

It is easy to see that it does not hold in our system. As we noted above, **K** is in fact a 'distribution of confirmation'. If both an implication and its antecedent are confirmed, there is no reason the consequent needs to be confirmed as well.

Let us note, that the weaker version of modal Modus Ponens holds

$$\frac{K\alpha \quad K(\alpha \to \beta)}{\beta}$$

however, it cannot be considered as any kind of omniscience. It just says that if both α and $(\alpha \to \beta)$ are confirmed, then β is a part of currently available data.

This rule holds not only in logical situations, but in all situations. If $K\alpha$ and $K(\alpha \to \beta)$ are true in an $s \in S$, then $s \Vdash \alpha$ and $s \Vdash \alpha \to \beta$ because of **T** axiom. It follows from the assumption $Rsss$ and the definition of implication, that $s \Vdash \beta$ as well.

Contradiction Contradiction in relevant logic is non-explosive: φ and $\neg\varphi$ might hold in a contradictory situation, but it does not entail an arbitrary formula ψ. (this would require an R-connection to situation where ψ holds).[1]

$$\not\models (\varphi \wedge \neg\varphi) \to \psi$$

As we noted above, a contradiction cannot be known (it is never confirmed).

$$\not\models K(\varphi \wedge \neg\varphi)$$

This has a trivial consequence, that knowledge of contradiction implies anything ($\models K(\varphi \wedge \neg\varphi) \to \psi$), so, in particular knowledge of contradiction implies knowledge of anything ($\models K(\varphi \wedge \neg\varphi) \to K(\psi)$). Nevertheless this does not lead to any kind of explosion as there is no such situation in which the antecedent is true. In standard models, $K(\varphi \wedge \neg\varphi)$ is never true either, but the reason is that $\varphi \wedge \neg\varphi$ is not true in any state (possible world). In our framework the situation is different: $\varphi \wedge \neg\varphi$ can be true in some situation (the agent obtained contradictory data), but $K(\varphi \wedge \neg\varphi)$ cannot.

Adjunction Modal adjunction also does not hold—if $K\alpha$ and $K\beta$ are true in s, then obviously $(\alpha \wedge \beta)$ is true there, but $K(\alpha \wedge \beta)$ need not be.[2] Our agent is really careful here. Even if each of α and β are confirmed separately, their conjunction is not accepted as knowledge, unless there is a single source confirming both of them (which in general does not need to be the case).

Modal disjunction rule In our system knowledge distributes with disjunction. It holds that

$$\frac{K(\alpha \vee \beta)}{K\alpha \vee K\beta}$$

Given a disjunction is confirmed in a current situation by a certain source, one of the disjuncts must be confirmed by it as well, because a disjunction is true at the source if at least one of the disjuncts is.

5 Conclusion

The original motivation for the project of relevant epistemic framework was an idea of a knowledge representation, which avoids the frequently criticized features of traditional modal representation, in particular, logical omniscience. We defined a new epistemic operator using standard parts of the relational semantics of distributive relevant logic—the relations of compatibility and involvement. The motivation we had in mind was an idea of a scientific agent-observer, whose knowledge is identified with the notion of independently confirmed data.

We obtained an epistemic operator which is extremely weak and has almost no closure properties. Our initial requirements were met, but we are aware of

[1] The explosion does not occur even in the case of the strong conjunction; $(\varphi \otimes \neg\varphi) \to \psi$ does not hold.

[2] The same negative result holds also for strong conjunction. If $K\alpha$ and $K\beta$ are true in s, then $(\alpha \otimes \beta)$ is true in s (because of the truth axiom **T** and, moreover, $Rsss$ for this case), but $K(\alpha \otimes \beta)$ need not be true in s.

the fact, that our solution will not satisfy everybody. First problem is the use of relevant framework, which, for many, seems to be too complex and lacking a clear interpretation. We find on a contrary the framework very elegant and hope our motivation suggested one more interpretation of the framework. Another problem is that the notion of knowledge represented by our operator is very weak—it certainly does not cover all the aspects of knowledge as it is generally understood and in particular it does not allow straightforwardly represent any reasoning processes. It was not our aim to define a generally applicable representation (and it does not seem there is one on the market) and the standard representations are rather 'overdetermined', so our attempt might be seen as showing the 'lower' end of the scale.

Our project is still work in progress. An earlier stage of this project was reflected in the article Majer and Peliš (2009). Another article dealing with the questions of axiomatization, completeness and properties of our system is in preparation.

Acknowledgements Work on this text was supported in part by grant no. P401/10/1504 and in part by grant no. IAA900090703 of the Grant Agency of the Academy of Sciences of the Czech Republic. We wish to thank to Marta Bílková, Libor Běhounek, Greg Restall and the audience of the LIRA seminar for valuable comments.

A Relevant logic R

There are more formal systems that can be called relevant logic. From the proof-theoretical viewpoint, all of them are considered to be substructural logics (see Restall (2000) and Paoli (2002)). Here we present the axiom system and (Routley-Meyer) semantics from Mares (2004) with some elements from Restall (1999).

A.1 Syntax

We use the language of classical propositional logic with signs for atomic formulas $\mathcal{P} = \{p, q, \ldots\}$, formulas being defined in the usual way:

$$\varphi ::= p \mid \neg\psi \mid \psi_1 \vee \psi_2 \mid \psi_1 \wedge \psi_2 \mid \psi_1 \rightarrow \psi_2$$

Axiom schemes

1. $A \rightarrow A$
2. $(A \rightarrow B) \rightarrow ((B \rightarrow C) \rightarrow (A \rightarrow C))$
3. $A \rightarrow ((A \rightarrow B) \rightarrow B)$
4. $(A \rightarrow (A \rightarrow B)) \rightarrow (A \rightarrow B)$
5. $(A \wedge B) \rightarrow A$
6. $(A \wedge B) \rightarrow B$

7. $A \to (A \vee B)$

8. $B \to (A \vee B)$

9. $((A \to B) \wedge (A \to C)) \to (A \to (B \wedge C))$

10. $(A \wedge (B \vee C)) \to ((A \wedge B) \vee (A \wedge C))$

11. $\neg\neg A \to A$

12. $(A \to \neg B) \to (B \to \neg A)$

Strong logical constants $\otimes$ (group conjunction, fusion) and $\oplus$ (group disjunction) are definable by implication and negation:

- $(A \oplus B) \equiv_{def} \neg(\neg A \to B)$
- $(A \otimes B) \equiv_{def} \neg(\neg A \oplus \neg B)$

Rules

Adjunction From A and B infer $A \wedge B$.

Modus Ponens From A and $A \to B$ infer B.

A.2 Routley-Meyer semantics

An R-frame is a quintuple $\mathbf{F} = \langle S, L, C, \trianglelefteq, R \rangle$, where S is a non-empty set of situations and $L \subseteq S$ is a non-empty set of logical situations. The relations $C \subseteq S^2$, $\trianglelefteq \subseteq S^2$, and $R \subseteq S^3$ were introduced in section 3, here we sum up their properties.

Properties of the relation R The basic property of R:

$$\text{if } Rxyz,\ x' \trianglelefteq x,\ y' \trianglelefteq y, \text{ and } z \trianglelefteq z', \text{ then } Rx'y'z'.$$

This means that the relation R is monotonic with respect to the involvement relation.

Moreover it is required that:

(r1) $Rxyz$ implies $Ryxz$

(r2) $R^2(xy)zw$ implies $R^2(xz)yw$, where R^2xyzw iff $(\exists s)(Rxys$ and $Rszw)$.

(r3) $Rxxx$

(r4) $Rxyz$ implies Rxz^*y^*

Properties of the relation C Compatibility between two states is inherited by the states involved in them ('less informative states'):

$$\text{If } xCy,\ x_1 \trianglelefteq x, \text{ and } y_1 \trianglelefteq y, \text{ then } x_1 C y_1.$$

Moreover, we require the following properties:

(c1) symmetricity xCy implies yCx

(c2) directedness $(\forall x)(\exists y)(xCy)$

(c3) convergence $(\forall x)(\exists y(xCy)$ implies $(\exists x^*)(xCx^*$ and $\forall z(xCz$ implies $z \trianglelefteq x^*)))$

(c4) $x \trianglelefteq y$ implies $y^* \trianglelefteq x^*$

(c5) $x^{**} \trianglelefteq x$

Model R-model **M** is a R-frame **F** with a valuation function $v : \mathcal{P} \longrightarrow 2^S$. The truth of a formula at a situation is defined in the following way:

- $s \Vdash p$ iff $s \in v(p)$
- $s \Vdash \neg\varphi$ iff $s^* \nVdash \varphi$
- $s \Vdash \psi \wedge \varphi$ iff $s \Vdash \psi$ and $s \Vdash \varphi$
- $s \Vdash \psi \vee \varphi$ iff $s \Vdash \psi$ or $s \Vdash \varphi$
- $s \Vdash (\varphi \rightarrow \psi)$ iff $(\forall y, z)(Rsyz$ implies $(y \Vdash \varphi$ implies $z \Vdash \psi))$

As we already said, the truth of a formula in a model, resp. in a frame, is defined as truth in all logical situations of this model/frame. As usual, R-tautologies are formulas true in all relevant frames. Whenever φ is a R-tautology, we write $\models \varphi$ and say that φ is a valid formula.

The condition (r1) validates the implicative version of *Modus Ponens* (axiom schema 3). It does not validate the conjunctive version $(A \wedge (A \rightarrow B)) \rightarrow B$, which requires (r3). (r2) corresponds to the *exchange rule* $(A \rightarrow (B \rightarrow C)) \rightarrow (B \rightarrow (A \rightarrow C))$, which is derivable from the axioms given above. (r4) validates *contraposition* (axiom schema 12). If we work without the Routley star, this can be rewritten as: $Rxyz$ implies $(\forall z'Cz)(\exists y'Cy)(Rxy'z')$.

Directedness and convergence conditions are necessary for the definition of the Routley star. From (c1) we obtain the validity of $(A \rightarrow \neg\neg A)$ and from the last condition (c5) we get the axiom schema 11.

References

J. Cheng. A strong relevant logic model of epistemic processes in scientific discovery. In E. Kawaguchi, H. Kangassalo, H. Jaakkola, and I. Hamid, editors, *Information Modelling and Knowledge Bases XI*, pages 136–159. IOS Press, 2000.

H. N. Duc. *Resource-Bounded Reasoning about Knowledge*. PhD thesis, Faculty of Mathematics and Informatics, University of Leipzig, 2001.

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

O. Majer and M. Peliš. Epistemic logic with relevant agents. In M. Peliš, editor, *The Logica Yearbook 2008*, pages 123–135. College Publications, 2009.

E. Mares and R. Meyer. The semantics of **R4**. *Journal of Philosophical Logic*, 22: 95–110, 1993.

E. D. Mares. *Relevant Logic*. Cambridge University Press, 2004.

F. Paoli. *Substructural Logics: A Primer*. Kluwer, 2002.

G. Restall. Four-valued semantics for relevant logics (and some of their rivals). *Journal of Philosophical Logic*, 24:139–160, 1995.

G. Restall. Information flow and relevant logics. In *Logic, Language and Computation: The 1994 Moraga Proceedings*, 1994.

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

G. Restall. Negation in relevant logics: How i stopped worrying and learned to love the routley star. In D. Gabbay and H. Wansing, editors, *What is Negation?*, volume 13 of *The Applied Logic Series*, pages 53–76. Kluwer, 1999.

G. Restall. Simplifid semantics for relevant logics (and some of their rivals). *Journal of Philosophical Logic*, 22:481–511, 1993.

H. Wansing. Diamonds are a philosopher's best friends. *Journal of Philosophical Logic*, 31:591–612, 2002.

Toward a Dynamic Logic of Questions

Johan van Benthem and Ştefan Minică

Institute for Logic Language and Computation
johan.vanbenthem@uva.nl, s.a.minica@uva.nl

Abstract

Questions are triggers for explicit events of 'issue management'. We give a complete logic in dynamic-epistemic style for events of raising, refining, and resolving an issue, all in the presence of information flow through observation or communication. We explore extensions of the framework to multi-agent scenarios and long-term temporal protocols. We also sketch a comparison with some alternative accounts.

Keywords: question, issue management, logical dynamics.

1 Introduction and motivation

Questions are different from statements, but they are just as important in driving reasoning, communication, and general processes of investigation. The first logical studies merging questions and propositions seem to have come from the Polish tradition: cf. Wisniewski (1995). A forceful modern defender of this dual perspective is Hintikka, who has long pointed out how any form of inquiry depends on an interplay of inference and answers to questions. Cf. Hintikka et al. (2002) and Hintikka (2007) on the resulting 'interrogative logic', and the epistemological views behind it. These logics are mainly about *general inquiry* and learning about the world. But there is also a related stream of work on the *questions* in natural language, as important speech acts with a systematic linguistic vocabulary. Key names are Groenendijk & Stokhof: cf. Groenendijk and Stokhof (1997), Groenendijk (1999), and the recent 'inquisitive semantics' of Groenendijk (2008) ties this in with a broader information-oriented 'dynamic semantics'. Logic of inquiry and logic of questions are related, but there are also differences in thrust: a dynamic logic of 'issue management' that fits our intuitions is not necessarily the same as a logic of speech acts that must make do with what natural language provides.

In this paper, we do not choose between these streams, but we propose a different technical approach. Our starting point is a simple observation. Questions are evidently important informational actions in human agency. Now the latter area is the birth place of *dynamic-epistemic logic* of explicit events that make information flow. But surprisingly, existing dynamic-epistemic systems do not give an explicit account of what questions do! In fact, central examples in the area have questions directing the information flow (say, by the Father in the puzzle of the Muddy Children) – but the usual representations in systems like *PAL* or *DEL* leave them out, and merely treat the answers, as events of public announcement. Can we make questions themselves first-class citizens in dynamic-epistemic logic, and get closer to the dynamics of inquiry? In Baltag (2001), Baltag has shown that we can. We will take this further, following a methodology that has already worked in other areas, and pursuing the same issues here: what are natural acts of inquiry, and how can dynamic logics bring out their structure via suitable recursion axioms? Moreover, by doing so, we at once get an account of non-factual questions, multi-agent aspects, temporal sequences, and other themes that have already been studied in a *DEL* setting.

2 A toy system of information and issues

The methodology of dynamic-epistemic logic starts with a static base logic describing states of some informational phenomenon, and identifies relevant informational state-changing events. Then, dynamic modalities are added to the base language, and their complete logic is determined on top of the given logic of the static models. To work in the same style, we first need a convenient static semantics to 'dynamify'. We take such a model from existing semantics of public questions, considering only one agent first, for simplicity. We will work in the style of epistemic logic and public announcement logic *PAL*, though our dynamic logic of questions will also have its differences.

2.1 Epistemic issue models

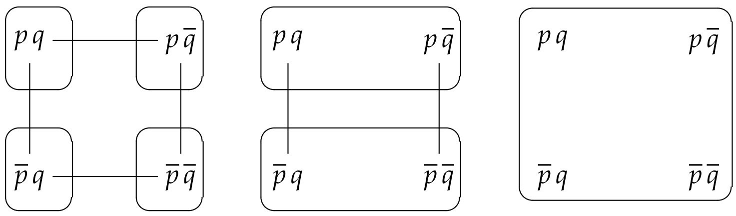
We work over standard epistemic models. In this setting, a simple framework for representing questions uses an equivalence relation over some relevant domain of alternatives, that we will call the 'issue relation'. This idea is found in many places, from linguistics (cf. Groenendijk and Stokhof (1997)) to learning theory (cf. Kelly (1996)): the current 'issue' is a partition of the set of options, with partition cells standing for the areas where we would like to be. This partition may be induced by a conversation whose current focus are the issues that have been put on the table, or a game where finding out about certain issues has become important to further play, a learning scenario for the language fed to us by our environment, or even a whole research program with an agenda determining what is currently under investigation. The 'alternatives' may range from simple finite settings like deals in a card game to complex infinite histories representing a total life experience. Formally, all this reduces to:

Definition 2.1 (Epistemic Issue Model)**.** *An* epistemic issue model *is a structure* $M = \langle W, \sim, \approx, V\rangle$ *where:*

- W *is a set of possible worlds or states (epistemic alternatives),*
- $\sim$ *is an equivalence relation on* W *(epistemic indistinguishability),*
- $\approx$ *is an equivalence relation on* W *(the abstract issue relation),*
- $V : \mathrm{P} \to \wp(W)$ *is a valuation for atomic propositions* $p \in \mathrm{P}$*.*

We could introduce models with more general relations, to account for, say, lack of epistemic introspection into the current issue, or belief instead of knowledge. While this is important eventually, equivalence relations will suffice for the points that we will make in this paper.

Figure 1: Examples of Epistemic Issue Models



In Figure 1 we illustrate intuitively the previous formal definition. Here and in subsequent diagrams, epistemic indistinguishability is represented by lines linking possible worlds, and the issue relation is represented by partition cells. We use the usual conventions and skip reflexive and transitive relations. We assume that the actual world is the top left one and in some case we also use double lines instead of partition cells to represent issue relations. With this understanding, Figure 1 depicts, from left to right, an epistemic issue model in which nothing is known and everything is an issue, a second one in which q is known in the actual world and the issue is to find out about p, and, finally, one in which everything is known in the actual world, and nothing is an issue.

2.2 Information and issues: language and semantics

To work with these structures, we need matching modalities in our language. Here we make a minimal choice of modal and epistemic logic for state spaces plus two modalities describing the issue structure. First, $K\varphi$ talks about knowledge or semantic information of an agent, its informal reading is 'φ is known', and its explanation is as usual: 'φ holds in all epistemically indistinguishable worlds'. To further describe our models, we add a universal modality $U\varphi$ saying that 'φ is true in all worlds'. Next, we use $Q\varphi$ to say that, locally in a given world, the current structure of the issue-relation has φ true: 'φ holds in all issue-equivalent worlds'. While convenient, this local notion does not express the global assertion that the current issue *is* φ, this will be defined later.

Finally, we find a need for a notion that mixes the epistemic and issue relations, talking (roughly) about what would be the case if the issue were resolved given what we already know. Technically, we add an intersection modality $R\varphi$ saying that "φ holds in all epistemically indistinguishable and issue equivalent worlds". While such modalities are frequent in many settings, they complicate axiomatization. We will assume the standard device of adding *nominals* naming single worlds (cf. Girard (2008), Liu (2008) for recent instances of this technique in the *DEL* setting).[1]

[1] As one illustration, working with nominals requires a modified valuation function in Definition 2.1, to a $V : \mathtt{P} \uplus \mathtt{N} \to \wp(W)$ mapping every proposition $p \in \mathtt{P}$ to a set of states $V(p) \subseteq W$, but every nominal $i \in \mathtt{N}$ to a singleton set $V(i)$ of a world $w \in W$.

Definition 2.2 (Static Language)**.** *The language $\mathcal{L}_{\mathbf{EL_Q}}(\mathtt{P},\mathtt{N})$ has disjoint countable sets* P *and* N *of propositions and nominals, respectively, with $p \in \mathtt{P}$, $i \in \mathtt{N}$. Its formulas are defined by the following inductive syntax rule:*

$$i \mid p \mid \bot \mid \neg\varphi \mid (\varphi \wedge \psi) \mid U\varphi \mid K\varphi \mid Q\varphi \mid R\varphi$$

When needed, dual existential modalities $\widehat{U}$, $\widehat{K}$, $\widehat{Q}$ and $\widehat{R}$ are defined as usual. Customary shortcuts to express disjunction and other boolean connectives are also used in their standard way. Formulas in this static language receive their meaning in the following way:

Definition 2.3 (Interpretation)**.** *Formulas are interpreted in models M at worlds w by the following recursive clauses:*

$$\begin{array}{lll}
M \models_w p & \textit{iff} & w \in V(p), \\
M \models_w i & \textit{iff} & w \in V(i), \\
M \models_w \neg\varphi & \textit{iff} & \textit{not } M \models_w \varphi, \\
M \models_w \varphi \wedge \psi & \textit{iff} & M \models_w \varphi \textit{ and } M \models_w \psi, \\
M \models_w U\varphi & \textit{iff} & \textit{for all } w \in W : M \models_w \varphi, \\
M \models_w K\varphi & \textit{iff} & \textit{for all } v \in W : w \sim v \textit{ implies } M \models_v \varphi, \\
M \models_w Q\varphi & \textit{iff} & \textit{for all } v \in W : w \approx v \textit{ implies } M \models_v \varphi, \\
M \models_w R\varphi & \textit{iff} & \textit{for all } v \in W : w\,(\sim \cap \approx)\,v \textit{ implies } M \models_v \varphi.
\end{array}$$

For instance, with this language we can express that the structure of the current issue settles fact φ with the following formula:

$$U(Q\varphi \vee Q\neg\varphi)^2$$

Here is how we say that an agent considers it possible that fact φ is not settled by the structure of the current issue:

$$\widehat{K}(\varphi \wedge \widehat{Q}\neg\varphi)$$

The next example says that an agent knows locally that a certain fact φ would be settled by the issue, while it is not settled globally:

$$KQ\varphi \wedge \neg U(Q\varphi \vee Q\neg\varphi)$$

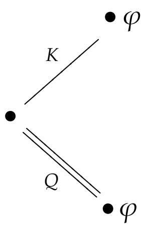
As for the third modality of 'resolution', it describes intuitively what agents would know if the current issue is resolved. Thus, we can say that in the current epistemic situation φ is neither known by the agent nor settled by the structure of the issue, but it is true upon resolution:

$$\neg Q\varphi \wedge \neg K\varphi \wedge R\varphi$$

A more complex example is when a fact is neither known nor settled in any world of the model, but it is true in all indistinguishable and issue-equivalent worlds, and it would be settled by a resolution action:

$$\neg\widehat{U}(K\varphi \vee Q\varphi) \wedge UR\varphi$$

[2] We use the term 'settling' in a technical sense, as saying that the issue answers (either explicitly or implicitly) the question whether φ holds. In natural language, there is also the notion of 'settling an issue', an event of finding out which partition cell we are in. This will be one of our later actions of 'issue management', that of *resolution*.



These examples show that our language can express quite complex notions about questions. Many such notions have been considered in the literature about questions and information flow, but often restricted to factual questions, and without the benefit of a uniform formal language.

We end with a technical point: the *intersection modality* $R\varphi$ cannot be defined in terms of K and Q. In particular, $\widehat{R}\varphi$ is not equivalent with $\widehat{K}\varphi \wedge \widehat{Q}\varphi$, witness the counterexample to the left. However, the use of 'nominals' i from hybrid logic helps us to completeness, by the valid converse:

$$\widehat{K}(i \wedge \varphi) \wedge \widehat{Q}(i \wedge \varphi) \rightarrow \widehat{R}\varphi$$

2.3 Static logic of information and issues

As for reasoning with our language, we have the valid implications $\widehat{K}\varphi \rightarrow \widehat{U}\varphi$, $\widehat{O}\varphi \rightarrow \widehat{U}\varphi$, $\widehat{Q}\varphi \rightarrow \widehat{U}\varphi$, $\widehat{R}\varphi \rightarrow \widehat{U}\varphi$. The following are not, in general, valid implications $R\varphi \rightarrow Q\varphi$, $R\varphi \rightarrow \neg Q\varphi$, $K\varphi \rightarrow Q\varphi$, $Q\varphi \rightarrow K\varphi$, $R\varphi \rightarrow K\varphi$.

More generally, we write $\models \varphi$ if the static formula φ is true in every model at every world. The static epistemic logic $\mathbf{EL_Q}$ of information and questions in our models is the set of all validities:

$$\mathbf{EL_Q} = \{\varphi \in \mathcal{L}_{\mathbf{EL_Q}} : \ \models \varphi\}$$

Definition 2.4 (Axiomatization)**.** *The proof system EL_Q contains the customary (epistemic) S5 axioms for K, Q and R:*

1. *$Kp \rightarrow p$ (Truth), $Kp \rightarrow KKp$, $\neg Kp \rightarrow K\neg Kp$ (Full Introspection)*
2. *$p \rightarrow Q\widehat{Q}p$, $p \rightarrow \widehat{Q}p$, $\widehat{Q}\widehat{Q}p \rightarrow \widehat{Q}p$ (equivalence relation for issue),*
3. *$p \rightarrow R\widehat{R}p$, $p \rightarrow \widehat{R}p$, $\widehat{R}\widehat{R}p \rightarrow \widehat{R}p$ (equivalence relation for resolution),*

together with the characteristic axiom for intersection:

4. *$\widehat{K}i \wedge \widehat{Q}i \leftrightarrow \widehat{R}i$.*

In addition, it contains a standard hybrid logic with a universal modality:

5. *$\Box(p \rightarrow q) \rightarrow (\Box p \rightarrow \Box q), \Box \in \{U, K, R, Q\}$ (Distribution)*
6. *$\neg\Box\neg p \leftrightarrow \Diamond p, \Diamond, \Box \in \{U, K, R, Q\}$ (Duality)*
7. *$p \rightarrow U\widehat{U}p$, $p \rightarrow \widehat{U}p$, $\widehat{U}\widehat{U}p \rightarrow \widehat{U}p$,*
8. *$\widehat{U}i, \Diamond p \rightarrow \widehat{U}p, \Diamond \in \{\widehat{K}, \widehat{R}, \widehat{Q}\}$*
9. *$\Diamond(i \wedge p) \rightarrow \Box(i \rightarrow p), \Box \in \{U, K, R, Q\}$ (Nominals)*
10. *From $\vdash_{PC} \varphi$ infer φ (Prop), From φ and $\varphi \rightarrow \psi$ infer ψ (MP)*
11. *From φ infer $\Box\varphi$, for $\Box \in \{U, K, R, Q\}$ (Necessitation)*

12. *From* φ *and* $\sigma_{\text{sort}}(\varphi) = \psi$ *infer* ψ*, where* σ_{sort} *is 'sorted'*[3]

13. *From* $i \to \varphi$ *infer* φ*, for* i *not occurring in* φ

14. *From* $\widehat{U}(i \wedge \Diamond j) \to \widehat{U}(j \wedge \varphi)$ *infer* $\widehat{U}(i \wedge \Box\varphi)$*, for* $\Diamond \in \{\widehat{K}, \widehat{R}, \widehat{Q}\}$*,* $i \neq j$*, and* j *not occurring in* φ*.*

We write $\vdash_{EL_Q} \varphi$ if φ is provable in the proof system EL_Q. These laws of reasoning derive many intuitive principles. For instance, here is the simple proof that agents have introspection about the current public issue:

$$U(Qp \vee Q\neg p) \quad \vdash_{EL_Q} \quad UU(Qp \vee Q\neg p) \quad \vdash_{EL_Q} \quad KU(Qp \vee Q\neg p)$$

Here are some simple derivable principles connecting our modalities:

$$U\varphi \to K\varphi, \qquad U\varphi \to Q\varphi, \qquad U\varphi \to R\varphi \qquad K\varphi \to R\varphi, \qquad Q\varphi \to R\varphi$$

Further technical details of proofs are irrelevant to our purposes here. We refer to ten Cate (2005) for hybrid modal proof systems and completeness theorems. This standard machinery leads to this expected result:

Theorem 1 (Completeness of $\mathbf{EL_Q}$). *For every formula* $\varphi \in \mathcal{L}_{\mathbf{EL_Q}}(\mathtt{P}, \mathtt{N})$*:*

$$\models \varphi \quad \textit{if and only if} \quad \vdash_{EL_Q} \varphi$$

3 Dynamic logic of issue management

In dynamic epistemic logic, the next step is now to identify basic events of information flow, and expand the logic accordingly. This situation is very analogous with logic of questions and events of 'issue management'.

3.1 Basic actions of issue management

To identify basic actions that change the issue relation in a given model, we first look at some pictures. As before, epistemic indistinguishability is represented by lines linking possible worlds, and the issue relation is represented by rectangular partition cells. For simplicity, we start with the initial issue as the universal relation, represented as a frame border.

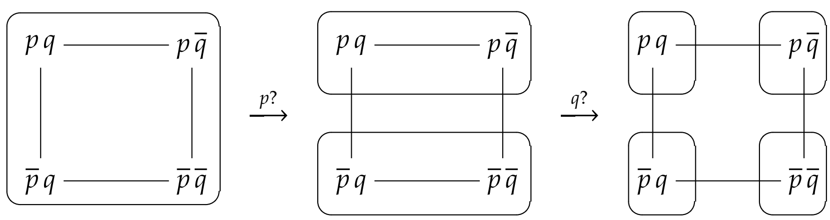


Figure 2: Effects of Asking Yes/No Questions.

[3]The technical notion 'sorted' and its uses are explained in ten Cate (2005).

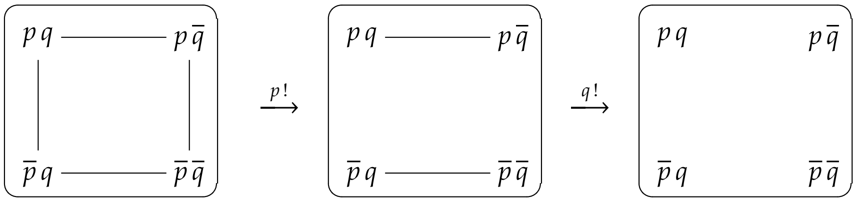


Figure 3: Almost Symmetrical Effects of 'Soft' Announcing.

In Figure 2, the first transition records the effect of asking a question: the issue relation is split into p and $\neg p$ cells. The second transition ilustrates the effect of asking a second question: the issue partition is further refined.

In Figure 3, the first transition is an announcement: the indistinguishability links between p and $\neg p$ worlds are removed. The second transition shows how a second announcement further refines the epistemic partition. Here and henceforth, we use a special sort of event that is congenial to this setting, viz. the *link-cutting announcements* $\varphi!$ of van Benthem & Liu Benthem and Liu (2007). These do not throw away worlds, but merely cut all links between φ- and $\neg\varphi$-worlds, keeping the whole model available for further reference.

In this way, there is a symmetry between a question and a soft announcement. One refines the issue, the other the information partition:

Definition 3.1 (Questions & Announcements)**.** *Let* $\overset{\varphi}{\equiv}_M = \{(w,v) \mid ||\varphi||^M_w = ||\varphi||^M_v\}$. *The execution of a* $\varphi?$ *action in a given model* M *results in a changed model* $M_{\varphi?} = \langle W_{\varphi?}, \sim_{\varphi?}, \approx_{\varphi?}, V_{\varphi?}\rangle$*, while a* $\varphi!$ *action results in* $M_{\varphi!} = \langle W_{\varphi!}, \sim_{\varphi!}, \approx_{\varphi!}, V_{\varphi!}\rangle$*:*

$$\begin{aligned} W_{\varphi?} &= W & \qquad W_{\varphi!} &= W \\ \sim_{\varphi?} &= \sim & \sim_{\varphi!} &= \sim \cap \overset{\varphi}{\equiv}_M \\ \approx_{\varphi?} &= \approx \cap \overset{\varphi}{\equiv}_M & \approx_{\varphi!} &= \approx \\ V_{\varphi?} &= V & V_{\varphi!} &= V \end{aligned}$$

The symmetry in this mechanism would be lost if we let $p!$ be an executable action only if it is *truthful*. For, the corresponding question $p?$ is executable in every world in a model, even those not satisfying p. The results that will follow can easily be stated for both kinds of announcement: truthful or not.

One attractive feature of this setting is that it suggests further natural operations on information and issues. In particular, Figure 4 contains two more management actions. In the first example two Yes/No questions $p?$ and $q?$ are asked, and then a global *resolving* action follows on the epistemic relation. In the second, two announcements $p!$ and $q!$ are made, and a *refinement* action follows on the issue relation, adjusting it to what agents already know. These operations are natural generalisations of asking and announcing:

Definition 3.2 (Resolution and Refinement)**.** *An execution of the 'resolve' action* $!$*, and of the 'refine' action* $?$ *in model* M *results in changed models* $M_! = \langle W_!, \sim_!, \approx_!, V_!\rangle$, $M_? = \langle W_?, \sim_?, \approx_?, V_?\rangle$*, respectively, with:*

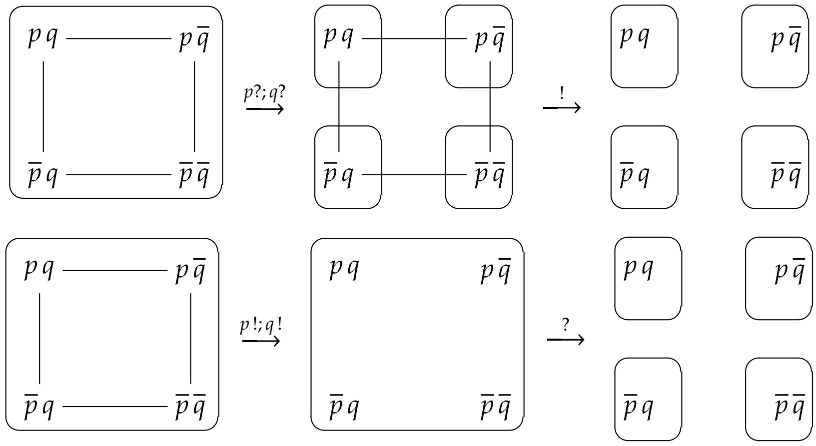


Figure 4: Resolving and Refining Actions.

$$\begin{array}{rcl} W_? & = & W \\ \sim_? & = & \sim \\ \approx_? & = & \approx \cap \sim \\ V_? & = & V \end{array} \qquad \begin{array}{rcl} W_! & = & W \\ \sim_! & = & \sim \cap \approx \\ \approx_! & = & \approx \\ V_! & = & V \end{array}$$

Again, the two actions are symmetric. As a way of understanding this, we could introduce a new agent whose role is that of an 'issue manager', dual to the epistemic information agent.

It is also useful to have one more issue management action # that *simultaneously* changes both equivalence relations. The effect of executing this in model M is a new model $M_\# = \langle W_\#, \sim_\#, \approx_\#, V_\# \rangle$ with:

$$W_\# = W, \quad \sim_\# = \approx_\# = \sim \cap \approx, \quad V_\# = V$$

Here is a summary of our repertoire of issue management actions:

$[\varphi!]$	'Soft' announcement	$[\varphi?]$	Question
$[!]$	Resolution	$[?]$	Refinement
$[\#]$	Simultaneous resolution	or	'parallel refinement'

3.2 Semantic properties of issue management

Our basic actions satisfy some intuitive principles. In particular, our three last ones form an algebra under composition, witness the following Table:

;	!	?	#
!	!	#	#
?	#	?	#
#	#	#	#

With more specific management actions of questions and announcements, the picture is more diverse. In particular, composing these operations is complex, and many prima facie laws do not hold, witness:

stefan:observation 1 (Composition)**.** *The following equations are not valid in* DEL_Q*:*

$$(11)\ \varphi!;! = !;\varphi! \qquad (12)\ \varphi!;? = ?;\varphi! \qquad (13)\ \varphi!;\# = \#;\varphi!$$
$$(14)\ \varphi?;! = !;\varphi! \qquad (15)\ \varphi?;? = ?;\varphi! \qquad (16)\ \varphi?;\# = \#;\varphi?$$
$$(17)\ \varphi?;\psi! = \psi!;\varphi?$$

Some of these observations crucially involve non-factual formulas. For instance, $\varphi?;\psi! = \psi!;\varphi?$ and $\varphi?;\psi? = \varphi?\cdot\psi?$ are both valid for factual φ and only fail for non-factual φ. We will include proofs and counterexamples in the full version of the paper.

Next, let us see how some known features of information management in *PAL* fare with our new issue management actions.

Repetition In *PAL*, repeating the same assertion $!\varphi$ has no new effects when its content φ is *factual*. But as the Muddy Children puzzle shows, repeating the same epistemic assertion can be informative, and lead to new effects, or in the above short-hand notation:

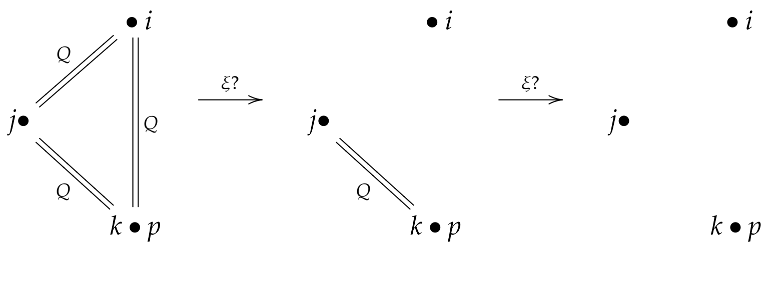
$$\varphi!\,;\varphi!\ \neq\ \varphi!$$

The reason is that when the model has changed, epistemic operators may change truth values. What about DEL_Q: is asking a question once the same as asking it twice? Again, for factual questions, this is clearly so, given the above semantics: the issue relation no longer changes in the second step. But when the question itself can refer to the issue relation, things are different:

stefan:observation 2 (Iteration)**.** *The equation* $\varphi?;\varphi? = \varphi?$ *is invalid in* DEL_Q.

Figure 5 has a counterexample for $\xi := (\widehat{Q}i \rightarrow (j \vee k)) \wedge ((\widehat{Q}j \wedge p) \rightarrow \widehat{Q}i)$. Both updates for this question, computed as above, change the model.

Figure 5: Effects of asking the same question twice.



Composition Next comes a difference with *PAL*. Public announcement satisfies a valid composition principle that gives the effects of two consecutive announcements with just a single one:

$$\varphi!;\psi! = (\varphi \wedge [\varphi]\psi)!$$

But as was already observed for preference change in Benthem and Liu (2007) and Benthem et al. (2008), this need not hold for more complex model changes.[4]

stefan:observation 3 (Proper Iteration)**.** *There is no question composition principle.*

Proof. If there were one single assertion having just the same effect as a sequence $\varphi?;\psi?$, then, starting with the issue configured as the universal relation on the domain of a model, such a sequence will always induce a two, not four, element partition; this refutation is also depicted in Figure 4.[5] □

Related to this are dynamic properties of ordering. While action order makes no difference with purely factual assertions or questions, it does when the content may be of an explicit epistemic or issue-related nature.

We have seen that information update and questions have many subtleties. It is time for a dynamic epistemic logic of issues that can reason about these.

3.3 Issue management: language and semantics

In order to talk explicitly about the above changes, dynamic modalities are added to the earlier static language of information and issues:

Definition 3.3 (Dynamic Language)**.** *Language* $\mathcal{L}_{\mathbf{DEL_Q}}(\mathtt{P},\mathtt{N})$ *is defined by adding the following clauses to Definition 2.2:* $[\varphi!]\psi \mid [\varphi?]\psi \mid [?]\varphi \mid [!]\varphi$

These are interpreted by adding the following clauses to Definition 2.3:

Definition 3.4 (Interpretation)**.** *Formulas are interpreted in M at w by the following clauses, where models $M_{\varphi?}$, $M_{\varphi!}$, $M_?$ and $M_!$ are as defined above:*

$$\begin{array}{lll} M \models_w [\varphi!]\psi & \textit{iff} & M_{\varphi!} \models_w \psi, \\ M \models_w [\varphi?]\psi & \textit{iff} & M_{\varphi?} \models_w \psi, \\ M \models_w [?]\varphi & \textit{iff} & M_? \models_w \varphi \\ M \models_w [!]\varphi & \textit{iff} & M_! \models_w \varphi \end{array}$$

Important insights about the relation between knowledge, questions and answers can now be expressed in our formal language. We can, for instance, say that a certain question $\psi?$ is entailed in an epistemic-issue model:

$$U(\widehat{Q}i \to [\psi?]\widehat{Q}i) \qquad \textit{for all } i$$

Intuitively this just says that the question does not change the issue structure. We can also express the fact that a sequence of questions entails $\psi?$ with:

$$U([\varphi_0?]\cdots[\varphi_n?]\widehat{Q}i \to [\varphi_0?]\cdots[\varphi_n?][\psi?]\widehat{Q}i)^6 \qquad \textit{for all } i$$

We can also express new notions of entailment, like, for instance, the notion of epistemic global entailment of an arbitrary announcement $\psi!$:

$$U(\widehat{R}i \to [\psi!]\widehat{R}i) \qquad \textit{for all } i$$

[4] The composition principle also fails in *PAL* with *protocols*, our topic in Section 6.

[5] This Fact is not a big obstacle. We could easily extend our language with multiple questions, that do not just change partitions on a single-formula basis.

[6] This generalizes previous definitions of entailment in which questions can only be factual.

A small modification of this in which we relax the previous requirement of abstract global entailment can capture local compliance of answers:

$$[\varphi_0?]\cdots[\varphi_n?](\psi \wedge \widehat{Ri}) \rightarrow [\varphi_0?]\cdots[\varphi_n?][\psi!]\widehat{Ri}^7 \qquad \textit{for all } i$$

Moreover, our language can express basic laws of interrogative reasoning. For instance, we can say that an agent knows in advance that the effect of a question followed by its resolution leads to knowledge of the relevant issue:

$$K[\varphi?][\,!\,]U(K\varphi \vee K\neg\varphi)$$

3.4 Dynamic logic of informational issues

We have seen that effects of asking questions are not always easy to keep straight, but also, that there is an interesting structure to management operations on models. Both purposes call for a complete dynamic epistemic logic of questions. Satisfaction and validity are defined as before. The dynamic epistemic logic of questioning is the set of all semantic validities:

$$\mathbf{DEL_Q} = \{\varphi \in \mathcal{L}_{\mathbf{DEL_Q}}(\mathtt{P}, \mathtt{N}) : \ \models \varphi\}$$

We introduce a proof system by adding the reduction axioms below to the earlier proof system EL_Q for the static fragment of the logic.

What follows is a long list of mostly operator commutation principles, interspersed with clauses where 'something happens'. This difference reflects the workings of our semantics of information and issue management:

Definition 3.5 (Reduction Axioms)**.** *The proof system DEL_Q extends the earlier static logic EL_Q by the following reduction axioms and inference rule:*

1. $[\varphi?]a \leftrightarrow a$ *(Asking & Atoms)*

2. $[\varphi?]\neg\psi \leftrightarrow \neg[\varphi?]\psi$ *(Asking & Negation)*

3. $[\varphi?](\psi \wedge \chi) \leftrightarrow [\varphi?]\psi \wedge [\varphi?]\chi$ *(Asking & Conjunction)*

4. $[\varphi?]U\psi \leftrightarrow U[\varphi?]\psi$ *(Asking & Universal Modality)*

5. $[\varphi?]K\psi \leftrightarrow K[\varphi?]\psi$ *(Asking & Knowledge)*

6. $[\varphi?]R\psi \leftrightarrow (\varphi \wedge R(\varphi \rightarrow [\varphi?]\psi)) \vee (\neg\varphi \wedge R(\neg\varphi \rightarrow [\varphi?]\psi))$ *(Asking & Resolution)*

7. $[\varphi?]Q\psi \leftrightarrow (\varphi \wedge Q(\varphi \rightarrow [\varphi?]\psi)) \vee (\neg\varphi \wedge Q(\neg\varphi \rightarrow [\varphi?]\psi))$ *(Asking & Partition)*

8. $[\,!\,]a \leftrightarrow a$ *(Resolving & Atoms)*

9. $[\,!\,]\neg\varphi \leftrightarrow \neg[\,!\,]\varphi$ *(Resolving & Negation)*

10. $[\,!\,](\psi \wedge \chi) \leftrightarrow [\,!\,]\psi \wedge [\,!\,]\chi$ *(Resolving & Conjunction)*

11. $[\,!\,]U\varphi \leftrightarrow U[\,!\,]\varphi$ *(Resolving & Universal Modality)*

[7] Again, this generalizes notions of compliance restricted to propositional formulas.

12. $[\,!\,]K\varphi \leftrightarrow R[\,!\,]\varphi$ *(Resolving & Knowledge)*
13. $[\,!\,]R\varphi \leftrightarrow R[\,!\,]\varphi$ *(Resolving & Resolution)*
14. $[\,!\,]Q\varphi \leftrightarrow Q[\,!\,]\varphi$ *(Resolving & Partition)*
15. $[\varphi!]a \leftrightarrow a$ *(Announcement & Atoms)*
16. $[\varphi!]\neg\psi \leftrightarrow \neg[\varphi!]\psi$ *(Announcement & Negation)*
17. $[\varphi!](\psi \wedge \chi) \leftrightarrow [\varphi!]\psi \wedge [\varphi!]\chi$ *(Announcement & Conjunction)*
18. $[\varphi!]U\psi \leftrightarrow U[\varphi!]\psi$ *(Announcement & Universal Modality)*
19. $[\varphi!]K\psi \leftrightarrow (\varphi \wedge K(\varphi \rightarrow [\varphi!]\psi)) \vee (\neg\varphi \wedge K(\neg\varphi \rightarrow [\varphi!]\psi))$ *(Answer & Knowledge)*[8]
20. $[\varphi!]R\psi \leftrightarrow (\varphi \wedge R(\varphi \rightarrow [\varphi!]\psi)) \vee (\neg\varphi \wedge R(\neg\varphi \rightarrow [\varphi!]\psi))$ *(Answer & Resolution)*
21. $[\varphi!]Q\psi \leftrightarrow Q[\varphi!]\psi$ *(Announcement & Partition)*
22. $[\,?\,]a \leftrightarrow a$ *(Refining & Atoms)*
23. $[\,?\,]\neg\varphi \leftrightarrow \neg[\,!\,]\varphi$ *(Refining & Negation)*
24. $[\,?\,](\psi \wedge \chi) \leftrightarrow [\,!\,]\psi \wedge [\,!\,]\chi$ *(Refining & Conjunction)*
25. $[\,?\,]U\varphi \leftrightarrow U[\,!\,]\varphi$ *(Refining & Universal Modality)*
26. $[\,?\,]K\varphi \leftrightarrow K[\,!\,]\varphi$ *(Refining & Knowledge)*
27. $[\,?\,]R\varphi \leftrightarrow R[\,!\,]\varphi$ *(Refining & Resolution)*
28. $[\,?\,]Q\varphi \leftrightarrow R[\,!\,]\varphi$ *(Refining & Partition)*
29. *From* φ *infer* $\Box\varphi$*, for* $\Box \in \{[\cdot ?], [\cdot !], [!], [?]\}$ *(Necessitation)*

We write $\vdash_{DEL_Q} \varphi$ if φ is provable in the proof system DEL_Q.

Theorem 2 (Soundness). *The reduction axioms in DEL_Q are sound.*

Proof. By standard modal arguments. We discuss two cases that go beyond mere commutation of operators. The first (*Asking & Partition*) explains how questions refine a partition:

$$[\varphi?]Q\psi \leftrightarrow (\varphi \wedge Q(\varphi \rightarrow [\varphi?]\psi)) \vee (\neg\varphi \wedge Q(\neg\varphi \rightarrow [\varphi?]\psi))$$

From left to right. Assume that $M \models_w [\varphi?]Q\psi$, then we also have $M_{\varphi?} \models_w Q\psi$. In case $M \models_w \varphi$, the new issue relation locally refined the old one to φ-worlds, and hence we get the left-hand disjunct on the right. The other case yields the right-hand disjunct. *From right to left.* Properly viewed, the preceding explanation already established an equivalence.

[8] If we assume truthfulness as a precondition of executing an announcement action this axiom (and other ones with a similar structure) does not need the right conjunct and will correspond to the standard DEL axioms for announcement.

Our second illustration (*Resolving & Knowledge*) shows how resolution changes knowledge making crucial use of our intersection modality:

$$[\,!\,]K\varphi \leftrightarrow R[\,!\,]\varphi$$

$M \models_w [\,!\,]K\varphi$ is equivalent to $M_! \models_w K\varphi$, which is equivalent to $\forall v \in W_! : w \sim_! v$ implies $M_! \models_v \varphi$. As $\sim_! = \sim \cap \approx$, the semantics of our dynamic modality tells us that $\forall v \in W : w\,(\sim \cap \approx)\,v$ implies $M \models_v [\,!\,]\varphi$, which is equivalent to $M \models_w R[\,!\,]\varphi$, as desired. □

Theorem 3 (Completeness of $\mathbf{DEL_Q}$)**.** *For every formula* $\varphi \in \mathcal{L}_{\mathbf{DEL_Q}}(\mathtt{P}, \mathtt{N})$*:*

$$\models \varphi \quad \textit{if and only if} \quad \vdash_{DEL_Q} \varphi.$$

Proof. This is a standard *DEL*-style translation argument. Working inside out, the reduction axioms translate dynamic formulas into corresponding static ones. In the end completeness for the static base logic is invoked. □

3.5 Discussion

So far we have given a logic of information and questions in standard *DEL* style. This calculus can derive many further principles, for instance:

The following formula is provable for all factual φ: $\varphi \to [\varphi?][!]K\varphi$.

Proof. [9]

1	$\varphi \to (\varphi \wedge R(\varphi \to \varphi)) \vee (\neg\varphi \wedge R(\neg\varphi \to \varphi))$	*PC*
2	$\varphi \to (\varphi \wedge R(\varphi \to [\varphi?]\varphi)) \vee (\neg\varphi \wedge R(\neg\varphi \to [\varphi?]\varphi))$	Lemma 1
3	$\varphi \to [\varphi?]R\varphi$	*Ak & R*
4	$\varphi \to [\varphi?]R[!]\varphi$	Lemma 1
5	$\varphi \to [\varphi?][!]K\varphi$	*Rs & K*

□

Note that even steps in the previous proof crucially depend on φ being factual, and they would fail otherwise. This subtlety of the system is illustrated in Figure 6. If we take a complex φ like, for instance, $Q\neg Kp$ we can see in this example that initially it is true in every world of the model but this is not the case anymore after a φ? question is asked and a resolution action [!] is performed. In the resulting model φ is false in at least one world.

Such aspects are not always easy to keep straight and our logic provides a way of keeping track of even more complicated cases.

But our analysis really shows its power (compared with other approaches to questions) when we consider the following two extensions: *multi-agent scenarios* and *protocols for investigation*. These two extensions will be the topics of the next two sections.

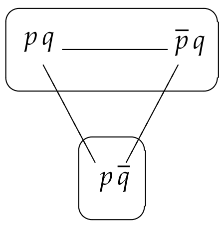


Figure 6: $\varphi = Q\neg Kp$

[9] Lemma 1: For factual φ and q ranging over management actions we have: $[q]\varphi \leftrightarrow \varphi$. The proof proceeds by induction using *Action & Atoms* axioms for the base case and *Action & Negation* or *Action & Conjunction* axioms, respectively, for the inductive step.

However, there is also a remaining desideratum right at the present level:

'Hidden validities'. Like with *PAL*, the current axiomatization leave unfinished business. While reduction axioms work on a formula-by-formula basis, they need not describe the general *schematic laws* of the system, such as the earlier composition law for consecutive assertions, that hold under arbitrary substitutions of formulas for proposition letters.[10] This deficit becomes even more urgent here. We saw that our model-changing operations of issue management had a nice algebraic structure. For instance, it is easy to see that resolving is idempotent and commutes with refinement:

$$!\,;\,! = ! \qquad \text{and} \qquad !\,;\,? = ?;\,!.$$

But $\mathbf{DEL_Q}$ does not state such facts explicitly, since, by working only from innermost occurrences of dynamic modalities, the completeness argument needed no recursion axioms with stacked modalities like [!] [!]. Yet this is obviously crucial information for a logic of issue management, and so, axiomatizing the schematic validities for operator stacking remains open.

4 Multi-agent question scenarios

Questions are typical multi-agent events, even though many logics of questions ignore this feature. In our setting, it should be easy to add more agents, and it is. We briefly explain the issues that arise, and solutions that are possible in DEL. Introducing a static multi-agent logic of information and issues is a routine step. Language, semantics, and logic are as before, now adding agent indices where needed. This language can make crucial distinctions like something being an issue for one agent but not (yet) for another, or some agent knowing what the issue is for another. Perhaps less routine would be adding group knowledge to our static base because an answer to a question often makes a fact common knowledge in the group {Questioner, Answerer}. Also, groups might have collective issues not owned by individual members, and this might be related to the earlier global actions of refinement and resolution. Even so, we leave extensions with common or distributed knowledge as an open problem.

Agent-specific preconditions So far, we only had impersonal public questions raising issues. Now consider real questions asked by one agent to another:

i asks j: "Is φ the case?"

As usual, even such simple Yes/No questions show what is going on. The preconditions for normal episodes of this kind are at least the following:

(a) Agent i does not know whether φ is the case,

(b) Agent i thinks it possible that j knows whether φ is the case,

[10] Note that the above reduction axioms for atoms typically do not have this substitution property – though many of our more complex reductions axioms do.

Of course, we are not saying that all questions have these preconditions. For instance, rhetorical or Socratic questions do not. Typology of questions is an empirical matter, beyond the scope of our logic. We will just assume that some preconditions are given for a question-answer episode, say pre($?\varphi, i, j$). Then we can proceed with our analysis. Which question actions? Do we really need new actions in this setting? One option would be to just split the event into two: publicly announcing the precondition, and then raising the issue in the earlier pure non-informative sense:

$$!pre(?\varphi, i, j); ?\varphi$$

This may have side-effects, however, since announcing the precondition may change the model and hence truth values of φ, making the effect of $?\varphi$ different from what it would have been before the announcement. One option is then to change both the epistemic relation and their issue relation of the model in parallel. Still, things works fine for factual φ, and we do not pursue this complication here.

Multi-agent scenarios with privacy The real power of dynamic epistemic logic only unfolds with informational actions that involve privacy of information and issues, and differences in observation. Many situations are like this. Say, someone asks a question to the teacher, but you do not hear which one. Private questions can be followed by public answers, and vice versa. Typical scenarios in this setting would be these:

(a) One of two agents raises an issue, but the other only sees that it is $?p$ or $?q$. What will happen?

(b) One agent raises an issue $?p$ privately, while the other does not notice. What will happen?

These are the things that product update was designed to solve. We will not do a DEL version of our logic in detail, but only show the idea. First, we need issue event models, where events can have epistemic links, but also an issue relation saying intuitively which events matter to which agent. Here is the essence of the update mechanism:

Definition 4.1 (Product update for questions)**.** Given an epistemic issue model M and issue event model E, the product model $M \times E$ is the standard DEL product with an issue relation $(s, e) \approx (t, f)$ iff $s \approx t$ and $e \approx f$.

We can easily see that this definition produces the desired effects by considering the simplest scenario one can imagine, depicted in Figure 7. We have a public question asked in a single-agent structure. We use two abstract epistemic events, or signals, to model the two possible answers to a Yes/No-question. The role of the issue relation in the action model is to highlight the signals that the agents considers important for the modelled question.

An interpretation point that should be considered here is our use of sets. A question involves two signals, neither of which is obviously the actual event, since nothing has happened yet. This is why we use designated sets in event models, and eventually, to keep things in harmony, in static epistemic issue models as well. One can perhaps circumvent this by reinterpretation, but the extension with sets has been proposed independently, and it works well.

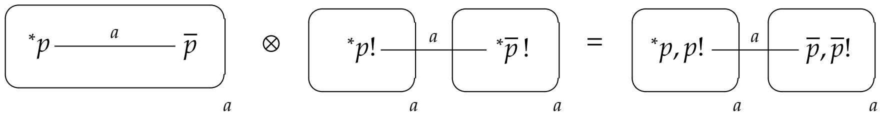


Figure 7: Product update for an impersonal (public or agent-independent) atomic factual question p? in a single-agent epistemic-issue structure.

The product-update effects are still the expected ones when we enrich the structures considered by allowing complex questions and many agent-interaction. Figure 8 ilustrates such a situation: the possible answers are the events that are distinguished in the action model, the issue relation gets refined and the epistemic uncertainty remains unchanged for both agents.

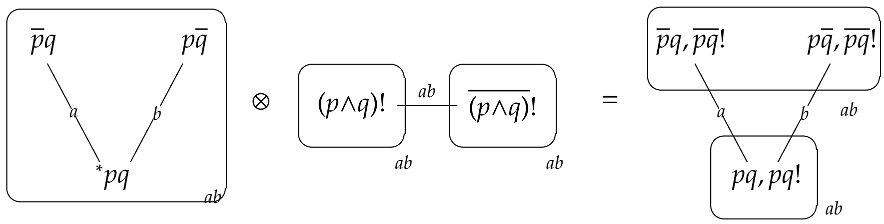


Figure 8: Product update for an agent-independent (impersonal, public) complex factual question $(p \wedge q)$? in a multi-agent environment.

Mixing preconditions and issue change. As long as we only consider factual preconditions for each epistemic event, questions have their expected effects: refining the issue relation by considering possible answers. If we consider the interactive aspect of asking and answering questions and also allow epistemic preconditions of questions then we have to account for the fact that asking a question can be an informative event in multi-agent interaction.

Such epistemic effects of raising agent-dependent questions go beyond refinement of the issue relation. Consider the example in Figure 9 where by asking a question a simultaneously gives b the answer to the very issue that he raises. In order to deal with such situations one has to consider the effects of preconditions for both possible informative answers and issue-raising question execution in each possible world.

The simplest way to deal with this complication is to place question preconditions in a *syntactic conjunction* with preconditions for each possible answer. If we do so, then executing a product update for a question can be informative. Consider our example where b ends up knowing more after the question $(p \wedge q)$ was asked by a, before any answer was given (see Figure 9)

This mechanism can also deal with private questions where other agents are misled about what has taken place. This is similar to product update for belief in Chapter 7, where equivalence relations give way to arbitrary directed relations (not necessarily reflexive) indicating what worlds an agent considers possible from the current one. Scenario (b) above will work with an event model with $!p$, $!\neg p$, *id* (the identity event that can happen anywhere), where agent 2 has a directed arrow pointing from the real event, say signal $!p$ in a world satisfying p, to *id*. We do not pursue these issues here, but refer to Baltag 2001, Minica 2010 for more sophisticated scenarios.

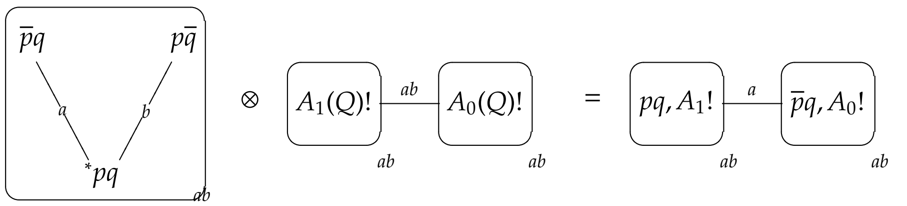


Figure 9: Product update for an agent-dependent (personalized, still public) complex factual question $(p \wedge q)$? with complex epistemic preconditions inside a multi-agent epistemic-issue environment (agent a asks question Q to agent b). Here $Q{=}(p \wedge q)?$, $\mathtt{pre}(Q){=}\neg(K_a(p \wedge q) \vee \neg K_a(p \wedge q)) \wedge \widehat{K_a}(K_b(p \wedge q) \vee K_b\neg(p \wedge q))$, $\mathtt{pre}(A_1(Q)) = p \wedge q \wedge \mathtt{pre}(Q)$, and $\mathtt{pre}(A_0(Q)) = \neg(p \wedge q) \wedge \mathtt{pre}(Q)$.

5 Temporal protocols for inquiry

We have now shown how our logic can do explicit issue management, while we can also deal with multi-agent scenarios that make questions come into their own as social acts. Still, this leaves out one further crucial theme: long-term temporal procedure in agency. Single information updates usually make sense only in an ongoing process of conversation, experiment, study, and so on. This is especially true for questions in their role of directing discourse or procedure. Not everything can be asked, due to social convention, limited measuring apparatus, and so on. Thus, it makes sense to incorporate procedure into our dynamic logics, toward a more realistic theory of inquiry. Like the earlier social multi-agent perspective, this essential temporal structure seems largely absent from existing logics of questions.

Here are two illustrations. First, there is a hierarchy of factual questions ("p?"), epistemic ("Did you know that p?") and procedural ones ("What would your brother say if I asked him p?") that determines how informational processes perform. In general, atomic Yes/No questions take more time to reach a propositional goal than complex ones. Next, an informational process may have procedural goals. Consider a standard Weighing Problem: "You have 9 pearls, 1 lighter than the others that are of equal weight. You can weigh 2 times with a standard balance. Find the lighter pearl." The process is a tree of histories of weighing acts with both solutions and failures. This structure is standard for procedurs of investigation, as found, e.g., in learning theory.

To deal with such structure, van Benthem, Gerbrandy, Hoshi & Pacuit 2009 adds temporal 'protocols' to PAL and DEL and axiomatizes their complete logic. A similar set of results is beyond the scope of this paper, but it is presented in van Benthem & Minica 2009. That paper defines temporal question protocols that constrain histories of investigation. These support an epistemic temporal logic of issues that can be axiomatized completely. It differs from our dynamic logic of questions in that there is no more reduction to a static base language. In the crucial recursive step for issue management, questions now obey this modified equivalence (stated with an existential action modality):

$$\langle ?\varphi\rangle Q\psi \leftrightarrow \langle ?\varphi\rangle\top \wedge ((\varphi \wedge Q(\varphi \rightarrow \langle ?\varphi\rangle\psi) \vee (\neg\varphi \wedge Q(\neg\varphi \rightarrow \langle ?\varphi\rangle\psi))$$

Here, the formula $\langle ?\varphi\rangle\top$ says that the question is allowed by the protocol. This logic of investigation does not reduce every assertion to a static one.

Of course, this is just a start: there are many further issues in the temporal logic of questions, that link up with *DEL* studies of belief (Degremont (2010)) and learning (Gierasimczuk (2010)).

6 Further directions

We have shown so far how various aspects of public inquiry are analyzed within the dynamic-epistemic methodology, making questions a natural companion to announcements and other informational events. While we have done this for knowledge, it would work equally well with dynamic logics of belief. In addition, we mention a few more general issues on our agenda:

Questioning games. An interesting further development links our dynamic analysis of questions to epistemic games for public announcements by Ågotnes and van Ditmarsch. In such games players have to find optimal announcement in order to reach their, possibly conflicting, epistemic goals. In Ågotnes and Ditmarsch (2009) new solution concepts are proposed for such games, in which the value of a question can receive a precise definition.[11] In strategic interactions an optimal question need not be the most informative one, and different preferences may arise in different epistemic scenarios.

Update, inference, and syntactic awareness dynamics. While *DEL* has been largely about observation-based semantic information, some recent proposals include more finely grained information produced by inference or introspection. The same sort of move makes sense in our current setting. For instance, yet another effect of asking a question is of making agents aware that something is an issue. This fits well in the syntactic approach to inferential and other fine-grained information in Benthem and Velazquez-Quesada (2009), with questions providing one reason for acts of 'awareness promotion'.

Multi-agent behaviour over time. A single question is hard to 'place' outside of some temporal scenario. For instance, questions as much as arguments drive argumentation, and serve as ways of either underpinning assertions, or calling them into doubt. Our study of *protocols* was one step in this direction, but we also need to make our dynamic logics of questions work in the analysis of conversation or *games*. This also makes sense in *learning* scenarios, where asking successive local questions is a natural addition to the usual input streams of answers (cf. Kelly (1996)).

Structured issues and agenda dynamics. But to us, the most striking limitation of our current approach is the lack of structure in our epistemic issue models. Both in conversation and in general inquiry, the *agenda* of relevant

[11] Another connection is game-theoretic 'value of questions' in signaling games (van Rooij (2005)).

issues is much more delicate than just some equivalence relation. For instance, there are less or more important issues. This reflects a general point on the informational side, where logics of ordered propositions have been used to model belief revision and preference ordering (cf. Liu (2008)).

7 Comparisons with other approaches

We have mentioned several other approaches to the logic of questions. There is the tradition of erotetic logic in the sense of Wisniewski (1995), or the later classic Belnap and Steel (1976).

More directly connected to our approach is the program of Hintikka for *interrogative logic* Hintikka et al. (2002). Questions are treated here as requests for new information, intertwined with deductive moves in 'interrogative tableaux'. There is a theory of answerhood, and an analysis of various types of questions in a predicate-logical setting. The framework has a number of nice theoretical results, including meta-theorems about the scope of questioning in finding the truth about some given situation. A merge might be of interest, bringing out Hintikka's concerns in an explicit dynamic epistemic setting.

But the closest comparison to our approach is inquisitive semantics (Groenendijk (2008), Ciardelli and Roelofsen (2009)) that gives propositions an 'interrogative meaning' in a universe of information states over propositional valuations, with sets of valuations expressing issues. At some level of abstraction, the ideas in this system are close to ours: information dynamics, questions change current partitions, etc. Indeed, inquisitive semantics poses immediate open problems for our dynamic logic. In particular, how can we generalize the analysis in this paper to arbitrary covers instead of partitions?

Comparing the two approaches must be left to another occasion. Here, we just note one key difference of methodology. Inquisitive semantics puts the dynamic information about questions in a new account of the *interrogative meaning* of sentences in a propositional language. This is not classical declarative meaning, and hence a deviant propositional logic emerges. By contrast, dynamic-epistemic logic gives an explicit account of questions and other actions of issue management, and once this is done, the base logic can stay classical. The distinction is similar to that between 'implicit' intuitionistic and 'explicit' epistemic approaches to knowledge (van Benthem 1993, 'Reflections on Epistemic Logic' Benthem (1993)). Connecting the two approaches to knowledge can be delicate (cf. van Benthem (2008)).

8 Conclusion

The dynamic calculi of questions in this paper show how dynamic-epistemic logic can incorporate a wide range of issue management beyond mere information handling. We have shown how these systems can be used to explore properties of issue management beyond what is found in other logics of questions, including complex epistemic assertions, many agents, explicit dynamics, and temporal protocols.[12]

[12]However, we have not arrived at a definite conclusion about the relation between our dynamic logics and alternatives. Perhaps all are needed to get the full picture of issue management.

Even so, we do our systems are only a first step - still removed from the complex syntactic structures of issues that give direction to rational agency. The insight itself that the latter are crucial comes from other traditions, as we have observed, but we hope to have shown that dynamic-epistemic logic has something of interest to contribute.

Acknowledgements We have profited from comments on seminar presentations and drafts of this paper from Alexandru Baltag, Cedric Degremont, Viktoria Denisova, Jan van Eijck, Hans van Ditmarsch, Jeroen Groenendijk, Solomon Feferman, Tomohiro Hoshi, Thomas Icard, Lena Kurzen, Floris Roelofsen, Katsuhiko Sano, Sonja Smets, George Smith, Cristina Unger, and Fernando Velázquez-Quesada.

References

T. Ågotnes and H. van Ditmarsch. But what will everyone say? - public announcement games. (to appear), 2009.

A. Baltag. Logics for insecure communication. In *J. van Benthem (ed.), Proceedings of TARK'01 (Eighth Conference on Rationality and Knowledge)*, pages 111–122. Morgan Kaufmann, 2001.

N. Belnap and T. Steel. *The logic of questions and answers*. Yale University Press, 1976.

J. van Benthem. Reflections on epistemic logic. *Logique et Analyse*, (34), 1993.

J. van Benthem. The information in intuitionistic logic. *Synthese*, 2(167), 2008.

J. van Benthem and F. Liu. Dynamic logic of preference upgrade. *Journal of Applied Non-Classical Logic*, (17:157-182), 2007.

J. van Benthem and F. Velázquez-Quesada. Inference, promotion and the dynamics of awareness. (working paper ILLC, to appear in KRA), 2009.

J. van Benthem, P. Girard, and O. Roy. Everything else being equal: A modal logic for *ceteris paribus* preferences. *Journal of Philosophical Logic*, (38:83-125), 2008.

I. Ciardelli and F. Roelofsen. Generalized inquisitive logic. 2009.

C. Degremont. The Temporal Mind. *Ph.D. Thesis*, (ILLC), 2010.

N. Gierasimczuk. to appear. *Ph.D. Thesis*, (ILLC), 2010.

P. Girard. Modal logic for belief and preference change. *Ph.D. Thesis*, (Stanford University), 2008.

J. Groenendijk. The logic of interrogation: Classical version. In T. Matthews and M. Strolovitch, editors, *SALT IX: Semantics and Linguistic Theory*. 1999.

J. Groenendijk. Inquisitive semantics: Two possibilities for disjunction. *Seventh International Tibilisi Symposium on Language, Logic, and Computation*, 2008.

J. Groenendijk and M. Stokhof. Questions. In J. van Benthem and A. ter Meulen, editors, *Handbook of the Logic and Language*. Elsevier, 1997.

J. Hintikka. *Socratic Epistemology: Explorations of Knowledge-seeking by Questioning*. Cambridge University Press, 2007.

J. Hintikka, I. Halonen, and A. Mutanen. Interrogative logic as a general theory of reasoning. In D. Gabbay, R. Johnson, H. Ohlbach, and J. Woods, editors, *Handbook of the Logic of Argument and Inference*. Elsevier, 2002.

K. Kelly. *The Logic of Reliable Inquiry*. Oxford University Press, 1996.

F. Liu. Changing for the better - preference dynamics and agent diversity. *Ph.D. Thesis*, (ILLC), 2008.

B. ten Cate. *Model Theory for Extended Modal Languages*. 2005.

R. van Rooij. Questions and relevance. *in: Questions and Answers: Theoretical and Applied Perspectives (Proceedings of 2nd CoLogNET-ElsNET Symposium, 96-107*, (ILLC), 2005.

A. Wisniewski. *The Posing of Questions*. Kluwer Academic Publishers, 1995.

Inquisitive Logic

Ivano Ciardelli and Floris Roelofsen

Department of Mathematics, University of Padova; Department of Linguistics, University of Massachusetts Amherst

iciardel@math.unipd.it, floris.roelofsen@gmail.com

Abstract

This paper investigates a generalized version of *inquisitive semantics* and the associated logic. The connection with intuitionistic logic and several intermediate logics is explored and a sound and complete axiomatization is thereby established.

1 Introduction

Traditionally, logic is concerned with argumentation. As a consequence, formal investigations of the semantics of natural language are usually focussed on the *descriptive* use of language, and the meaning of a sentence is identified with its *informative* content. Stalnaker (1978) gave this informative notion a dynamic and conversational twist by taking the meaning of a sentence to be its potential to update the common ground, where the common ground is viewed as the conversational participants' shared information. Technically, the common ground is taken to be a set of possible worlds, and a sentence provides information by eliminating some of these possible worlds.

Of course, this picture is limited in several ways. First, it only applies to sentences that are used exclusively to provide information. Even in a typical informative dialogue, utterances may serve different purposes as well. Second, the given picture does not take into account that updating the common ground is a *cooperative* process. One speech participant cannot simply change the common ground all by herself. All she can do is *propose* a certain change. Other speech participants may react to such a proposal in several ways. These reactions play a crucial role in the dynamics of conversation.

In order to overcome these limitations, *inquisitive semantics* starts with a different picture. It views propositions as proposals to update the common ground. Crucially, these proposals do not necessarily specify just one way of updating the common ground. They may suggest alternative ways of doing so. Formally, a proposition consists of one or more *possibilities*. Each possibility is a set of possible worlds and embodies a possible way to update the common ground. If a proposition consists of two or more possibilities, it is *inquisitive*:

it invites the other participants to provide information such that at least one of the proposed updates may be established. Inquisitive propositions raise an issue. They give direction to a dialogue. Thus, inquisitive semantics directly reflects that a primary use of language lies in the exchange of information in a cooperative dynamic process of raising and resolving issues.

Groenendijk (2009) and Mascarenhas (2009) first defined an inquisitive semantics for the language of propositional logic, focussing on the philosophical and linguistic motivation for the framework, and delineating some of its basic logical properties. The associated logic was axiomatized by Mascarenhas (2009), while a sound and complete sequent calculus was established independently by Sano (2009). Several linguistic applications of the framework are discussed by Balogh (2009).

In this paper, we consider a generalized version of the semantics proposed by Groenendijk (2009) and Mascarenhas (2009). This generalized semantics was first considered in lecture notes by Groenendijk (2008). Initially, it was thought to give the same results as the original semantics. Upon closer examination, however, Mascarenhas, Groenendijk, and Ciardelli observed that the two systems are different, and Ciardelli (2008) first argued that these differences speak in favor of the generalized semantics. Groenendijk and Roelofsen (2009) adopted the generalized semantics, and developed a formal pragmatic theory based on it.

The aim of the present paper is threefold. First, we will investigate and present some of the key features of the generalized semantics in a systematic way. Second, we will analyse the logic that the semantics gives rise to. In particular, we will explore the connection with intuitionistic logic and several well-known intermediate logics, which, among other things, will lead to a sound and complete axiomatization of inquisitive logic. Finally, we will argue that the generalized semantics is better-behaved than the original version of inquisitive semantics. In fact, we will define an entire hierarchy of parameterized versions of inquisitive semantics, and argue that only the generalized version, which can be seen as the limit case of the hierarchy, really behaves satisfactorily.

The paper is organized as follows. Section 2 introduces the generalized version of inquisitive semantics and presents some key features of the system. Section 3 investigates the associated logic, leading up to a sound and complete axiomatization. Section 4 shows that the *schematic fragment* of inquisitive logic (the logic itself is not closed under uniform substitution) coincides with the well-known Medvedev logic of finite problems. This is particularly interesting as it yields a sort of finitary pseudo-axiomatization of Medvedev logic (which is known not to be finitely axiomatizable). Finally, section 6 presents a translation of inquisitive logic into intuitionistic logic, showing that the former can be identified with the disjunctive-negative fragment of the latter.

2 Generalized inquisitive semantics

We assume a language $\mathcal{L}_{\mathcal{P}}$, whose expressions are built up from $\bot$ and a (finite or countably infinite) set of proposition letters $\mathcal{P}$, using binary connectives $\wedge$, $\vee$ and $\rightarrow$. We will also make use of three abbreviations: $\neg\varphi$ for $\varphi \rightarrow \bot$, $!\varphi$ for $\neg\neg\varphi$, and $?\varphi$ for $\varphi \vee \neg\varphi$. The first is standard, the second and the third will become clear shortly.

2.1 Indices, states, and support

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

Definition 2.1 (Indices)**.** A $\mathcal{P}$–*index* is a subset of $\mathcal{P}$. The set of all indices, $\wp(\mathcal{P})$, will be denoted by $\mathcal{I}_{\mathcal{P}}$. We will simply write $\mathcal{I}$ and talk of *indices* in case $\mathcal{P}$ is clear from the context.

Definition 2.2 (States)**.** A $\mathcal{P}$–*state* is a set of $\mathcal{P}$–indices. The set of all states, $\wp\wp(\mathcal{P})$, will be denoted by $\mathcal{S}_{\mathcal{P}}$. Again, reference to $\mathcal{P}$ will be dropped whenever possible.

The meaning of a sentence will be defined in terms of the notion of *support* (just as, in a classical setting, the meaning of a sentence is usually defined in terms of truth). Support is a relation between states and formulas. We write $s \models \varphi$ for 's supports φ'.

Definition 2.3 (Support)**.**

1. $s \models p$ iff $\forall w \in s : p \in w$
2. $s \models \bot$ iff $s = \emptyset$
3. $s \models \varphi \wedge \psi$ iff $s \models \varphi$ and $s \models \psi$
4. $s \models \varphi \vee \psi$ iff $s \models \varphi$ or $s \models \psi$
5. $s \models \varphi \to \psi$ iff $\forall t \subseteq s$: if $t \models \varphi$ then $t \models \psi$

It follows from the above definition that the empty state supports any formula φ. Thus, we may think of $\emptyset$ as the *inconsistent* state. The following two basic facts about support can be established by a straightforward induction on the complexity of φ:

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

Proposition 2 (Singleton states behave classically)**.** *For any index w and formula φ:*

$$\{w\} \models \varphi \iff w \models \varphi$$

where $w \models \varphi$ means: φ is classically true under the valuation w. In particular, $\{w\} \models \varphi$ or $\{w\} \models \neg\varphi$ for any formula φ.

It follows from Definition 2.3 that the support-conditions for $\neg\varphi$ and $!\varphi$ are as follows.

Proposition 3 (Support for negation)**.**

1. *$s \models \neg\varphi$ iff $\forall w \in s : w \models \neg\varphi$*
2. *$s \models !\varphi$ iff $\forall w \in s : w \models \varphi$*

Proof. Clearly, since ! abbreviates double negation, item 2 is a particular case of item 1. To prove item 1, first suppose $s \models \neg\varphi$. Then for any $w \in s$ we have $\{w\} \models \neg\varphi$ by persistence, and thus $w \models \neg\varphi$ by Proposition 2.
Conversely, if $s \not\models \neg\varphi$, then there must be $t \subseteq s$ with $t \models \varphi$ and $t \not\models \bot$. Since $t \not\models \bot$, $t \neq \emptyset$: thus, taken $w \in t$, by persistence and the classical behaviour of singleton states we have $w \models \varphi$. Since $w \in t \subseteq s$, it is not the case that $v \models \neg\varphi$ for all $v \in s$. □

The following construction will often be useful when dealing with cases where the set of propositional letters is infinite.

Definition 2.4. Let $\mathcal{P} \subseteq \mathcal{P}'$ be two sets of propositional letters. Then for any $\mathcal{P}'$–state s, the *restriction* of s to $\mathcal{P}$ is defined as $s\restriction_{\mathcal{P}} := \{w \cap \mathcal{P} \mid w \in s\}$.

The following fact, which can be established by a straightforward induction on the complexity of φ, says that whether or not a state s supports a formula φ only depends on the 'component' of s that is concerned with the letters in φ.

Proposition 4 (Restriction invariance)**.** *Let $\mathcal{P} \subseteq \mathcal{P}'$ be two sets of propositional letters. Then for any $\mathcal{P}'$–state s and any formula φ whose propositional letters are in $\mathcal{P}$:*

$$s \models \varphi \iff s\restriction_{\mathcal{P}} \models \varphi$$

2.2 Possibilities, propositions, and truth-sets

In terms of support, we define the *possibilities* for a sentence φ and the *meaning* of sentences in inquisitive semantics. We will follow the common practice of referring to the meaning of a sentence φ as the *proposition* expressed by φ. We also define the *truth-set* of φ, which embodies the *classical meaning* of φ.

Definition 2.5 (Truth sets, possibilities, propositions)**.** Let φ be a formula.

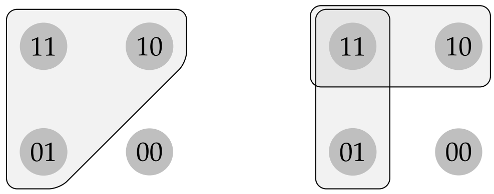
1. A *possibility* for φ is a maximal state supporting φ, that is, a state that supports φ and is not properly included in any other state supporting φ.
2. The *proposition* expressed by φ, denoted by $[\varphi]$, is the set of possibilities for φ.
3. The *truth set* of φ, denoted by $|\varphi|$, is the set of indices where φ is classically true.

Notice that $|\varphi|$ is a state, while $[\varphi]$ is a set of states. The classical meaning of φ is the set of all indices that make φ true. In inquisitive semantics, meaning is defined in terms of support rather than directly in terms of truth. It may be expected, then, that the proposition expressed by φ would be defined as the set of all states supporting φ. Rather, though, it is defined as the set of all *maximal* states supporting φ, that is, the set of all *possibilities* for φ. This is motivated by the fact that propositions are viewed as proposals, consisting of one or more alternative possibilities. If one state is included in another, we do not regard these two states as *alternatives*. This is why we are particularly interested in *maximal* states supporting a formula. Technically, however, the proposition expressed by φ still fully determines which states support φ and which states do not: the next result establishes that a state supports φ iff it is included in a possibility for φ.

Proposition 5 (Support and possibilities)**.** *For any state s and any formula φ:*

$$s \models \varphi \iff s \text{ is contained in a possibility for } \varphi$$

Proof. If $s \subseteq t$ and t is a possibility for φ, then by persistence $s \models \varphi$. For the converse, first consider the case in which the set $\mathcal{P}$ of propositional letters is finite. Then there are only finitely many states, and therefore if s supports φ,



(a) $|p \vee q|$ (b) $[p \vee q]$

Figure 1: The truth-set of $p \vee q$, and the proposition it expresses.

then obviously s must be contained in a *maximal* state supporting φ, i.e. in a possibility.
If $\mathcal{P}$ is infinite, given a $\mathcal{P}-$state $s \models \varphi$, consider its restriction $s\restriction_{\mathcal{P}_\varphi}$ to the (finite!) set $\mathcal{P}_\varphi$ of propositional letters occurring in φ. By Proposition 4, $s\restriction_{\mathcal{P}_\varphi} \models \varphi$, and thus $s\restriction_{\mathcal{P}_\varphi} \subseteq t$ for some $\mathcal{P}_\varphi-$state t which is a possibility for φ.
Now, consider $t^+ := \{w \in \mathcal{I}_\mathcal{P} \,|\, w \cap \mathcal{P}_\varphi \in t\}$. For any $w \in s$ we have $w \cap \mathcal{P}_\varphi \in (s\restriction\mathcal{P}_\varphi) \subseteq t$, so $w \in t^+$ by definition of t^+; this proves that $s \subseteq t^+$. Moreover, we claim that t^+ is a possibility for φ.
First, since $t^+\restriction_{\mathcal{P}_\varphi} = t$ and $t \models \varphi$, it follows from Proposition 4 that $t^+ \models \varphi$. Now, consider a state $u \supseteq t^+$ with $u \models \varphi$: then $u\restriction_{\mathcal{P}_\varphi} \supseteq t^+\restriction_{\mathcal{P}_\varphi} = t$ and moreover, again by Proposition 4, $u\restriction_{\mathcal{P}_\varphi} \models \varphi$; but then, by the maximality of t it must be that $u\restriction_{\mathcal{P}_\varphi} = t$. Now, for any $w \in u$, $w \cap \mathcal{P}_\varphi \in u\restriction_{\mathcal{P}_\varphi} = t$, so $w \in t^+$ by definition of t^+: hence, $u = t^+$. This proves that t^+ is indeed a possibility for φ. □

Example 1 (Disjunction)**.** *Inquisitive semantics crucially differs from classical semantics in its treatment of disjunction. This is illustrated by figures 1(a) and 1(b). These figures assume that $\mathcal{P} = \{p, q\}$; index* 11 *makes both p and q true, index* 10 *makes p true and q false, etcetera. Figure 1(a) depicts the truth set—that is, the classical meaning—of $p \vee q$: the set of all indices that make either p or q, or both, true. Figure 1(b) depicts the proposition associated with $p \vee q$ in inquisitive semantics. It consists of two possibilities. One possibility is made up of all indices that make p true, and the other of all indices that make q true. So in the inquisitive setting, $p \vee q$ proposes two alternative ways of enhancing the common ground, and invites a response that is directed at choosing between these two alternatives.*

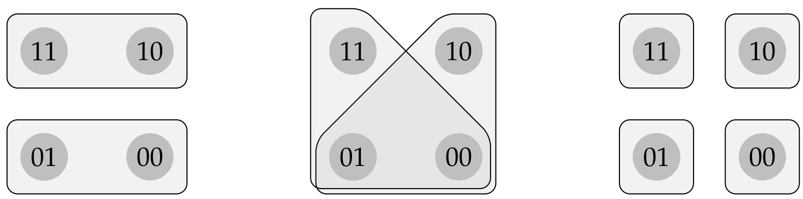
As an immediate consequence of Proposition 3, the possibilities for a (doubly) negated formula can be characterized as follows.

Proposition 6 (Negation)**.**

1. $[\neg\varphi] = \{|\neg\varphi|\}$
2. $[!\varphi] = \{|\varphi|\}$

2.3 Inquisitiveness and informativeness

Recall that propositions are viewed as proposals to change the common ground of a conversation. If $[\varphi]$ contains more than one possibility, then we say that φ is *inquisitive*. If the proposal expressed by φ is not rejected, then the indices that are not included in any of the possibilities for φ will be eliminated. If



(a) $?p$ (b) $p \to ?q$ (c) $?p \wedge ?q$

Figure 2: A polar question, a conditional question, and a conjoined question.

there are such indices—that is, if the possibilities for φ do not cover the entire space—then we say that φ is *informative*.

Definition 2.6 (Inquisitiveness and informativeness)**.**

- φ is *inquisitive* iff $[\varphi]$ contains at least two possibilities;
- φ is *informative* iff $[\varphi]$ proposes to eliminate certain indices: $\bigcup[\varphi] \neq \mathcal{I}$

Definition 2.7 (Questions and assertions)**.**

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

Definition 2.8 (Contradictions and tautologies)**.**

- φ is a *contradiction* iff it is only supported by the inconsistent state, i.e. iff $[\varphi] = \{\emptyset\}$
- φ is a *tautology* iff it is supported by all states, i.e. iff $[\varphi] = \{\mathcal{I}\}$

It is easy to see that a formula is a contradiction iff it is a classical contradiction. This does not hold for tautologies. Classically, a formula is tautological iff it is not informative. In the present framework, a formula is tautological iff it is neither informative nor inquisitive. Classical tautologies may well be inquisitive.

Example 2 (Questions)**.** *Figure 2 depicts the propositions expressed by the polar question $?p$, the conditional question $p \to ?q$, and the conjoined question $?p \wedge ?q$. Recall that $?p$ abbreviates $p \vee \neg p$. So $?p$ is an example of a classical tautology that is inquisitive: it invites a choice between two alternatives, p and $\neg p$. As such, it reflects the essential function of polar questions in natural language. For instance,* Is it raining? *invites a choice between two alternative possibilities, the possibility that it is raining and the possibility that it is not raining.*

Example 3 (Disjunction, continued)**.** *It is clear from Figure 1(b) that $p \vee q$ is* both *inquisitive* and *informative: $[p \vee q]$ consists of two possibilities, which, together, do not cover the set of all indices. This means that $p \vee q$ is* neither *a question* nor *an assertion.*

The following propositions give some sufficient syntactic conditions for a formula to be a question or an assertion, respectively. The straightforward proofs have been omitted.

Proposition 7. *For any two formulas φ, ψ:*

1. *$?\varphi$ and $?\psi$ are questions;*
2. *if φ and ψ are questions, then $\varphi \wedge \psi$ is a question;*
3. *if φ or ψ is a question, then $\varphi \vee \psi$ is a question;*
4. *if ψ is a question, then $\varphi \to \psi$ is a question.*

Proposition 8. *For any propositional letter p and formulas φ, ψ:*

1. *p is an assertion;*
2. *$\bot$ is an assertion;*
3. *if φ and ψ are assertions, then $\varphi \wedge \psi$ is an assertion;*
4. *if ψ is an assertion, then $\varphi \to \psi$ is an assertion.*

Note that items 2 and 4 of Proposition 8 imply that any negation is an assertion, which we already knew from Remark 6. Of course, $!\varphi$ is also always an assertion.

Using Proposition 8 inductively we obtain the following corollary showing that disjunction is the only source of inquisitiveness in our propositional language.[1]

Corollary 1. *Any disjunction-free formula is an assertion.*

In inquisitive semantics, the informative content of a formula φ is captured by the union $\bigcup[\varphi]$ of all the possibilities for φ. For φ proposes to eliminate all indices that are not in $\bigcup[\varphi]$. In a classical setting, the informative content of φ is captured by $|\varphi|$. Hence, the following result can be read as stating that inquisitive semantics agrees with classical semantics as far as informative content is concerned.

Proposition 9. *For any formula φ: $\bigcup[\varphi] = |\varphi|$.*

Proof. According to Proposition 2, if $w \in |\varphi|$, then $\{w\} \models \varphi$. But then, by Proposition 5, $\{w\}$ must be included in some $t \in [\varphi]$, whence $w \in \bigcup[\varphi]$. Conversely, any $w \in \bigcup[\varphi]$ belongs to a possibility for φ, so by persistence and the classical behaviour of singletons we must have that $w \in |\varphi|$. □

We end this subsection with a definition of equivalence between two formulas, several characterizations of questions and assertions, and a remark about the behaviour of the operators ? and !.

Definition 2.9 (Equivalence)**.**
Two formulas φ and ψ are *equivalent*, $\varphi \equiv \psi$, iff $[\varphi] = [\psi]$.

It follows immediately from Proposition 5 that $\varphi \equiv \psi$ just in case φ and ψ are supported by the same states.

[1] In the first-order case there will be a close similarity between disjunction and the existential quantifier, and the latter will be a source of inquisitiveness as well.

Proposition 10 (Characterization of questions)**.**
For any formula φ, the following are equivalent:

1. *φ is a question*
2. *φ is a classical tautology*
3. *$\neg\varphi$ is a contradiction*
4. *$\varphi \equiv ?\varphi$*

Proof. Equivalence (1 ⇔ 2) follows from the definition of questions and Proposition 9. (2 ⇔ 3) and (4 ⇒ 3) are immediate from the fact that a formula is a contradiction in the inquisitive setting just in case it is a classical contradiction. For (3 ⇒ 4), note that for any state s, $s \models ?\varphi$ iff $s \models \varphi$ or $s \models \neg\varphi$. This means that, if $\neg\varphi$ is a contradiction, $s \models ?\varphi$ iff $s \models \varphi$. In other words, $\varphi \equiv ?\varphi$. □

Note that an interrogative $?\varphi = \varphi \vee \neg\varphi$ is always a classical tautology, and therefore, by the equivalence (1 ⇔ 2), always a question. Furthermore, the equivalence (1 ⇔ 4) guarantees that $?\varphi \equiv ??\varphi$, which means that ? is idempotent.

Proposition 11 (Characterization of assertions)**.**
For any formula φ, the following are equivalent:

1. *φ is an assertion*
2. *if $s_j \models \varphi$ for all $j \in J$, then $\bigcup_{j\in J} s_j \models \varphi$*
3. *$|\varphi| \models \varphi$*
4. *$\varphi \equiv !\varphi$*
5. *$[\varphi] = \{|\varphi|\}$*

Proof.

(1 ⇒ 2) Suppose φ is an assertion and let t be the unique possibility for φ. If $s_j \models \varphi$ for all $j \in J$, then by Proposition 5 each s_j must be a subset of t, whence also $\bigcup_{j\in J} s_j \subseteq t$. Thus, by persistence, $\bigcup_{j\in J} s_j \models \varphi$.

(2 ⇒ 3) By Proposition 2, $\{w\} \models \varphi$ iff $w \in |\varphi|$. Then if φ satisfies condition (2), $|\varphi| = \bigcup_{w\in|\varphi|}\{w\} \models \varphi$.

(3 ⇒ 4) Suppose $|\varphi| \models \varphi$; by Proposition 5, $|\varphi|$ must be included in some possibility s for φ; but also, by Corollary 9, $s \subseteq |\varphi|$, whence $|\varphi| = s \in [\varphi]$. Moreover, since any possibility for φ must be included in $|\varphi|$ we conclude that $|\varphi|$ must be the *unique* possibility for φ. Thus, $[\varphi] = \{|\varphi|\}$.

(4 ⇔ 5) Since $[!\varphi] = \{|\varphi|\}$ (see Remark 6), obviously $\varphi \equiv !\varphi \iff [\varphi] = \{|\varphi|\}$.

(5 ⇒ 1) Immediate. □

Note that (1 $\Leftrightarrow$ 5) states that a formula is an assertion if and only if its meaning consists of its classical meaning. In this sense, assertions behave classically. Also note that (1 $\Leftrightarrow$ 4), together with the fact that $!\varphi$ is always an assertion, implies that $!\varphi \equiv !!\varphi$. That is, ! is idempotent.

The operators ! and ? work in a sense like projections on the 'planes' of assertions and questions, respectively. Moreover, the following proposition shows that the inquisitive meaning of a formula φ is completely determined by its 'purely informative component' $!\varphi$ and its 'purely inquisitive component' $?\varphi$.

Proposition 12 (Division in theme and rheme)**.** *For any formula φ, $\varphi \equiv !\varphi \wedge ?\varphi$.*

Proof. We must show that for any state s, $s \models \varphi$ iff $s \models !\varphi \wedge ?\varphi$. Suppose $s \models !\varphi \wedge ?\varphi$. Then, since $s \models ?\varphi$, s must support one of φ and $\neg\varphi$; but since $s \models \neg\neg\varphi$, s cannot support $\neg\varphi$. Thus, we have that $s \models \varphi$. The converse is immediate by the definitions of ! and ? and Proposition 3. □

2.4 Support, inquisitiveness, and informativeness

The basic notion in the semantics, as we have set it up here, is the notion of support. In terms of support, we defined possibilities and propositions, and in terms of possibilities we defined the notions of inquisitiveness and informativeness. We have tried to make clear how possibilities, propositions, inquisitiveness, and informativeness should be thought of intuitively, but we have not said much as to how the notion of support itself should be interpreted. It is important to emphasize that support should *not* be thought of as specifying conditions under which an agent with information state s can *truthfully utter* a sentence φ (this is a common interpretation of the notion of support in dynamic semantics, cf. Groenendijk et al. 1996). Rather, in the present setting support should be thought of as specifying conditions under which a sentence φ is *insignificant* or *redundant* in a state s, in the sense that, given the information available in s, φ is neither informative nor inquisitive. This intuition can be made precise by defining notions of inquisitiveness and informativeness *relative to a state*.

Definition 2.10 (Relative semantic notions)**.** Let φ be a formula, and s a state. Then:

- a *possibility* for φ in s is a maximal substate of s supporting φ;
- φ is *inquisitive* in s iff there are at least two possibilities for φ in s;
- φ is *informative* in s iff there is at least one index in s that is not included in any possibility for φ in s.

These notions allow us to formally establish the connection between support on the one hand, and inquisitiveness and informativeness on the other.

Proposition 13 (Support, inquisitiveness, and informativeness)**.**
A state s supports a formula φ iff φ is neither inquisitive in s nor informative in s.

Proof. Suppose that $s \models \varphi$. Then there is only one possibility for φ in s, namely s itself. So φ is not informative and not inquisitive in s. Conversely, if φ is not inquisitive in s, then there is only one possibility t for φ in s. If, moreover, φ is not informative in s, then t must be identical to s. By definition, $t \models \varphi$. So $s \models \varphi$ as well. □

3 Inquisitive logic

We are now ready to start investigating the logic that inquisitive semantics gives rise to. We begin by specifying the pertinent notions of entailment and validity.

Definition 3.1 (Entailment and validity)**.** A set of formulas Θ *entails* a formula φ in inquisitive semantics, $\Theta \models_{\mathsf{InqL}} \varphi$, if and only if any state that supports all formulas in Θ also supports φ. A formula φ is *valid* in inquisitive semantics, $\models_{\mathsf{InqL}} \varphi$, if and only if φ is supported by all states.

If no confusion arises, we will simply write $\models$ instead of $\models_{\mathsf{InqL}}$. We will also write $\psi_1, \dots, \psi_n \models \varphi$ instead of $\{\psi_1, \dots, \psi_n\} \models \varphi$. Note that, as expected, $\varphi \equiv \psi$ iff $\varphi \models \psi$ and $\psi \models \varphi$.

The intuitive interpretation of support in terms of insignificance or redundancy carries over to entailment: one can think of $\varphi \models \psi$ as saying that, whenever we are in a state where φ is redundant—i.e., neither informative nor inquisitive—ψ is as well. Or, in more dynamic terms, whenever we are in a state where the information provided by φ has been accommodated and the issue raised by φ has been resolved, ψ does not provide any new information and does not raise any new issue.

The following proposition states that if ψ is an assertion, inquisitive entailment boils down to classical entailment.

Proposition 14. *If ψ is an assertion, $\varphi \models \psi$ iff $|\varphi| \subseteq |\psi|$.*

Proof. Follows from Proposition 11 and the definition of entailment. □

We have already seen that the ! operator turns any formula into an assertion. We are now ready to give a more precise characterization: for any formula φ, $!\varphi$ is *the most informative assertion entailed by φ.*

Proposition 15. *For any formula φ and any assertion χ, $\varphi \models \chi \iff\ !\varphi \models \chi$.*

Proof. Fix a formula φ and an assertion χ. The right-to-left implication is obvious, since it is clear from Proposition 3 that $\varphi \models\ !\varphi$. For the converse direction, suppose $\varphi \models \chi$. Any possibility $s \in [\varphi]$ supports φ and therefore also χ, whence by Proposition 5 it must be included in a possibility for χ, which must be $|\chi|$ by Proposition 11 on assertions. But then also $|\varphi| = \bigcup[\varphi] \subseteq |\chi|$ whence $!\varphi \models \chi$ by Proposition 14. □

Most naturally, since a question does not provide any information, it cannot entail informative formulas.

Proposition 16. *If φ is a question and $\varphi \models \psi$, then ψ must be a question as well.*

Proof. If φ is a question, it must be supported by every singleton state. If moreover $\varphi \models \psi$, then ψ must also be supported by every singleton state. But then, since singletons behave like indices, ψ must be a classical tautology, that is, a question. □

Definition 3.2 (Logic)**.** Inquisitive logic, InqL, is the set of formulas that are valid in inquisitive semantics.

Proposition 17. *A formula φ is in InqL if and only if $\mathcal{I} \models \varphi$.*

Proof. The left-to-right direction is trivial. The right-to-left direction follows immediately from the fact that support is persistent. □

Proposition 18. *A formula is in InqL if and only if it is both a classical tautology and an assertion.*

Proof. If $\varphi \in$ InqL, it is supported by all states. In particular, it is supported by $\mathcal{I}$, which means that it is an assertion, and it is supported by all singleton states, which means, by Proposition 2, that it is a classical tautology. Conversely, if φ is an assertion, there is only one possibility for φ. If, moreover, φ is a classical tautology, this possibility must be $\mathcal{I}$. But then, by persistence, φ must be supported by all states. □

Thus, InqL coincides with classical logic as far as assertions are concerned: in particular, it agrees with classical logic on the whole disjunction-free fragment of the language.

Remark 1. *Although InqL is closed under the* modus ponens *rule, it is* not *closed under uniform substitution. For instance, $\neg\neg p \to p \in$ InqL for all proposition letters, but $\neg\neg(p \vee q) \to (p \vee q) \notin$ InqL. We will return to this feature of the logic in section 4.*

3.1 Disjunction property, deduction theorem, compactness, and decidability

We proceed by establishing a few basic properties of inquisitive logic and entailment.

Proposition 19 (Disjunction property)**.** *InqL has the* disjunction property*. That is, whenever a disjunction $\varphi \vee \psi$ is in InqL, at least one of φ and ψ is in InqL as well.*

Proof. If $\varphi \vee \psi \in$ InqL then, by Proposition 17, $\mathcal{I} \models \varphi \vee \psi$. This means, by definition of support, that $\mathcal{I} \models \varphi$ or $\mathcal{I} \models \psi$. But then, another application of Proposition 17 yields that φ or ψ must be in InqL. □

Proposition 20 (Deduction theorem)**.** *For any formulae $\theta_1, \ldots, \theta_n, \varphi$:*

$$\theta_1, \ldots, \theta_n \models \varphi \iff \models \theta_1 \wedge \cdots \wedge \theta_n \to \varphi$$

Proof. $\theta_1, \ldots, \theta_n \models \varphi$

$\iff$ for any $s \in \mathcal{S}$, if $s \models \theta_i$ for $1 \leq i \leq n$, then $s \models \varphi$

$\iff$ for any $s \in \mathcal{S}$, if $s \models \theta_1 \wedge \cdots \wedge \theta_n$, then $s \models \varphi$

$\iff \mathcal{I} \models \theta_1 \wedge \cdots \wedge \theta_n \to \varphi$

$\iff \theta_1 \wedge \cdots \wedge \theta_n \to \varphi \in$ InqL □

Theorem 2 (Compactness)**.** *For any set Θ and any formula φ, if $\Theta \models \varphi$ then there is a* finite *set $\Theta_0 \subseteq \Theta$ such that $\Theta_0 \models \varphi$.*

Proof. Since our set $\mathcal{P}$ of propositional letters is countable, so must be Θ, so we can write $\Theta = \{\theta_k \mid k \in \omega\}$. Now for any $k \in \omega$, let $\gamma_k = \theta_0 \wedge \cdots \wedge \theta_k$, and define $\Gamma = \{\gamma_k \mid k \in \omega\}$. It is clear that Γ and Θ are equivalent, in the sense that for any state s, $s \models \Gamma \iff s \models \Theta$, so we have $\Gamma \models \varphi$. Moreover, for $k \geq k'$ we have $\gamma_k \models \gamma_{k'}$.

If we can show that there is a formula $\gamma_k \in \Gamma$ such that $\gamma_k \models \varphi$, then this will mean that $\{\theta_0, \ldots, \theta_k\} \models \varphi$, and since $\{\theta_0, \ldots, \theta_k\}$ is a finite subset of Θ we will be done.

For any $k \in \omega$ let $\mathcal{P}_k$ be the set of propositional letters occurring in φ or in γ_k. By the definition of the formulas γ_k, it is clear that for $k \leq k'$ we have $\mathcal{P}_k \subseteq \mathcal{P}_{k'}$.

Now, towards a contradiction, suppose there is no $k \in \omega$ such that $\gamma_k \models \varphi$. Define $L_k := \{t \mid t \text{ is a } \mathcal{P}_k\text{–state with } t \models \gamma_k \text{ but } t \not\models \varphi\}$: our assumption amounts to saying that $L_k \neq \emptyset$ for all k. Then put $L := \{\emptyset\} \cup \biguplus_{k\in\omega} L_k$. Define a relation R on L by putting:

- $\emptyset R t$ iff $t \in L_0$;
- sRt iff $s \in L_k$, $t \in L_{k+1}$ and $t\upharpoonright_{\mathcal{P}_k} = s$.

Now, consider $t \in L_{k+1}$. This means that $t \models \gamma_{k+1}$ and $t \not\models \varphi$; as $\gamma_{k+1} \models \gamma_k$, we also have $t \models \gamma_k$. But then, since both γ_k and φ only use propositional letters from $\mathcal{P}_k$, by Proposition 4 we have $t\upharpoonright_{\mathcal{P}_k} \models \gamma_k$ and $t\upharpoonright_{\mathcal{P}_k} \not\models \varphi$, which means that $t\upharpoonright_{\mathcal{P}_k} \in L_k$.

From this it follows that (L, R) is a connected graph and thus clearly a tree with root $\emptyset$. Since L is a disjoint union of infinitely many non-empty sets, it must be infinite. On the other hand, by definition of R, all the successors of a state $s \in L_k$ are $\mathcal{P}_{k+1}$-states, and there are only finitely many of those as $\mathcal{P}_{k+1}$ is finite. Therefore, the tree (L, R) is finitely branching.

By König's lemma, a tree that is infinite and finitely branching must have an infinite branch. This means that there must be a sequence $\langle t_k \mid k \in \omega\rangle$ of states in L such that for any k, $t_{k+1}\upharpoonright_{\mathcal{P}_k} = t_k$. This naturally defines a $\mathcal{P}$-state that is the "limit" of the sequence. Precisely, this state is:

$$t = \{w \in \wp(\mathcal{P}) \mid \text{there are } w_k \in t_k \text{ with } w_{k+1}\upharpoonright_{\mathcal{P}_k} = w_k \text{ and } w = \bigcup_{k\in\omega} w_k\}$$

It is easy to check that for any k, $t\upharpoonright_{\mathcal{P}_k} = t_k$. Now, for any natural k, since $t\upharpoonright_{\mathcal{P}_k} = t_k \models \gamma_k$, by Proposition 4 we have $t \models \gamma_k$; hence, $t \models \Gamma$. On the other hand, for the same reason, since $t\upharpoonright_{\mathcal{P}_0} = t_0 \not\models \varphi$, also $t \not\models \varphi$.

But this contradicts the fact that $\Gamma \models \varphi$. So for some k we must have $\gamma_k \models \varphi$. □

Remark 3 (Decidability). *InqL is clearly decidable: to determine whether a formula φ is in InqL, by propositions 4 and 17 we only have to test whether $\mathcal{I}_{\{p_1,\ldots,p_n\}}$ supports φ, where $p_1, \ldots, p_n$ are the propositional letters in φ. This is a finite procedure since $\mathcal{I}_{\{p_1,\ldots,p_n\}}$ is finite and has only finitely many substates which have to be checked to determine support for implications.*

3.2 Disjunctive negative translation and expressive completeness

In this section, we observe that a formula can always be rewritten as a disjunction of negations, preserving logical equivalence with respect to inquisitive semantics. This observation will lead to a number of expressive completeness results. It will also play a crucial role later on in establishing completeness results, and in showing that InqL is isomorphic to the disjunctive-negative fragment of IPL.

We start by defining the *disjunctive negative translation* $\text{DNT}(\varphi)$ of a formula φ.

Definition 3.3 (Disjunctive negative translation)**.**

1. $\text{DNT}(p) = \neg\neg p$
2. $\text{DNT}(\bot) = \neg\neg\bot$
3. $\text{DNT}(\psi \vee \chi) = \text{DNT}(\psi) \vee \text{DNT}(\chi)$
4. $\text{DNT}(\psi \wedge \chi) = \bigvee\{\neg(\psi_i \vee \chi_j) \,|\, 1 \leq i \leq n,\, 1 \leq j \leq m\}$

   where:

   - $\text{DNT}(\psi) = \neg\psi_1 \vee \ldots \vee \neg\psi_n$
   - $\text{DNT}(\chi) = \neg\chi_1 \vee \ldots \vee \neg\chi_m$
5. $\text{DNT}(\psi \to \chi) = \bigvee_{k_1,\ldots,k_n}\{\neg\neg \bigwedge_{1\leq i\leq n}(\chi_{k_i} \to \psi_i) \,|\, 1 \leq k_j \leq m\}$

   where:

   - $\text{DNT}(\psi) = \neg\psi_1 \vee \ldots \vee \neg\psi_n$
   - $\text{DNT}(\chi) = \neg\chi_1 \vee \ldots \vee \neg\chi_m$

Proposition 21. *For any φ, $\varphi \equiv_{InqL} \text{DNT}(\varphi)$.*

Proof. By induction on φ. □

We skip over the details of the proof here. However, in section 3.6 we will see that, given some auxiliary results, a close examination of what is needed exactly to prove Proposition 21 instantly yields a range of interesting completeness results.

Note that the map DNT always returns a disjunction of negations, so we immediately have the following corollary.

Corollary 2. *Any formula is equivalent to a disjunction of negations.*

In particular, any formula is equivalent to a disjunction of assertions. This perfectly matches our intuitive understanding that meanings in inquisitive semantics are sets of *alternatives*, which are pairwise incomparable classical meanings ('incomparable' with respect to inclusion). Classical meanings are expressed by assertions (and thus always expressible by negations) while disjunction is the source of alternativehood, in the sense that a disjunction applied to two incomparable classical meanings yields the proposition consisting of those two classical meanings as alternatives.

Additionally, note that since any negation behaves classically in inquisitive semantics, in the scope of a negation we can always safely substitute classically equivalent subformulas. Since the set of connectives $\{\neg, \vee\}$ is complete in classical logic, given a formula $\varphi \equiv \neg\chi_1 \vee \cdots \vee \neg\chi_n$ we can always substitute each χ_k by a classically equivalent formula χ'_k using only disjunction and negation without altering the meaning of the formula, thus getting $\varphi \equiv \neg\chi'_1 \vee \cdots \vee \neg\chi'_n$. This proves the following corollary.

Corollary 3 (Expressive completeness of $\{\neg, \vee\}$)**.** *Any formula is equivalent to a formula containing only disjunctions and negations.*

Now consider an assertion χ. Since the set of connectives $\{\neg, \wedge\}$ is complete in classical logic, let χ' be a formula classically equivalent to χ which only contains negations and conjunctions: the classical equivalence of χ and χ' amounts to $|\chi| = |\chi'|$. Now, χ' is an assertion by Corollary 1, whence using Proposition 11 we have $[\chi] = \{|\chi|\} = \{|\chi'|\} = [\chi']$, i.e., $\chi \equiv \chi'$. Thus, we have the following corollary, stating that the assertive fragment coincides—up to equivalence—with the $\{\neg, \wedge\}$-fragment of the language.

Corollary 4 (Expressive completeness of $\{\neg, \wedge\}$ for assertions.)**.** *A formula is an assertion iff it is equivalent to a formula containing only conjunctions and negations.*

3.3 Inquisitive semantics and intuitionistic Kripke semantics

We now turn to the connection between inquisitive logic and intuitionistic logic.[2] This connection is suggested by the existence of a striking analogy between inquisitive and intuitionistic semantics. Both can be conceived of in terms of information and information growth. In inquisitive semantics, a formula is evaluated with respect to a state. Such a state can be thought of as an information state. Whether a certain state s supports a formula φ may depend not only on the information available in s, but also on the information that may *become* available. Formally, support is partly defined in terms of *subsets* of s. These subsets can be seen as possible future information states.

Similarly, in intuitionistic semantics, a formula is evaluated with respect to a point in a Kripke model, which can also be thought of as an information state. Each point comes equipped with a set of future points, called successors. Whether a point u in a model M satisfies a formula φ may depend not only on the information available at u, but also on the information that may become available. Formally, satisfaction at u is partly defined in terms of points in M that are accessible from u.

This informal analogy can be made precise: in fact, inquisitive semantics amounts to intuitionistic semantics on a suitable Kripke model.

Definition 3.4 (Kripke model for inquisitive semantics)**.** The Kripke model for inquisitive semantics is the model $M_I = \langle W_I, \supseteq, V_I \rangle$ where $W_I := \mathcal{S} - \{\emptyset\}$ is the set of all non-empty states and the valuation V_I is defined as follows: for any letter p, $V_I(p) = \{s \in W_I \mid s \models p\}$.

Observe that M_I is a Kripke model for intuitionistic logic. For, the relation $\supseteq$ is clearly a partial order. Moreover, suppose $s \supseteq t$ and $s \in V_I(p)$: this means that $s \models p$, and so by persistence $t \models p$, which amounts to $t \in V_I(p)$. So the valuation V_I is persistent. The next lemma shows that the two semantics coincide on every non-empty state.

Proposition 22 (Inquisitive support coincides with Kripke satisfaction on M_I)**.** *For every formula φ and every non-empty state s:*

$$s \models \varphi \iff M_I, s \Vdash \varphi$$

[2]Our investigation of this connection was inspired by Groenendijk (2008) and van Benthem (2009).

Proof. Straightforward, by induction on φ. The inductive step for implication uses the fact that an implication cannot be falsified by the empty state, as the latter supports all formulas, so that restricting the semantics to non-empty states does not make a difference. □

This simple observation already shows that the logic InqL contains intuitionistic propositional logic IPL. For suppose that $\varphi \notin$ InqL. Then there must be a non-empty state s such that $s \not\models \varphi$. But then we also have that $M_I, s \nVdash \varphi$, which means that $\varphi \notin$ IPL.

On the other hand, InqL is contained in classical propositional logic CPL, because any formula that is not a classical tautology is falsified by a singleton state in inquisitive semantics. So we have:

$$\mathsf{IPL} \subseteq \mathsf{InqL} \subseteq \mathsf{CPL}$$

Moreover, both inclusions are strict: for instance, $p \vee \neg p$ is in CPL but not in InqL, while $\neg\neg p \to p$ is in InqL but not in IPL.

Our next task is to investigate exactly where InqL sits between IPL and CPL. Ultimately, our result will be that there is a whole range of intermediate logics whose 'negative variant' coincides with InqL. In order to get to this result, let us first recall some basic features of intermediate logics, and define precisely what we mean by the negative variant of a logic.

3.4 Intermediate logics and negative variants

Recall that an *intermediate logic* is defined as a consistent set of formulae that contains IPL and is closed under the rules of modus ponens and uniform substitution, where *consistent* simply means 'not containing $\bot$' (Chagrov and Zakharyaschev 1997, p.109).

Intermediate logics ordered by inclusion form a complete lattice whose meet operation amounts to intersection and whose join operation, also called *sum*, is defined as follows: if Λ_i, $i \in I$ is a family of intermediate logics, then $\Sigma_{i\in I}\Lambda_i$ is the logic axiomatized by $\bigcup_{i\in I} \Lambda_i$, that is, the closure of $\bigcup_{i\in I} \Lambda_i$ under modus ponens and uniform substitution. The sum of *two* intermediate logic Λ and Λ' is denoted by $\Lambda + \Lambda'$.

In our investigation, we will meet several logics, beginning with InqL itself, that are *not* closed under uniform substitution. We shall refer to such logics as *weak intermediate logics*.

Definition 3.5. A weak intermediate logic is a set L of formulae closed under modus ponens and such that IPL $\subseteq L \subseteq$ CPL.

Weak intermediate logics ordered by inclusion form a complete lattice as well, where again meet is intersection and the join (or sum) of a family is the weak logic axiomatized by the union, i.e. the closure of the union under modus ponens.

If L is a weak intermediate logic, we write $\varphi \equiv_L \psi$ just in case $\varphi \leftrightarrow \psi \in L$.

Definition 3.6. Let K be a class of Kripke models (resp., frames). If Θ is a set of formulae and φ is a formula, we write $\Theta \models_K \varphi$ just in case any point in any model in K (resp., any point in any model based on any frame in K), that satisfies all formulas in Θ, also satisfies φ. We denote by $\mathsf{Log}(K)$ the set of formulae that are valid on each model (frame) in K, that is: $\mathsf{Log}(K) = \{\varphi \mid \models_K \varphi\}$.

It is straightforward to check that if K is a class of Kripke frames, $\mathsf{Log}(K)$ is an intermediate logic, while if K is a class of Kripke models, $\mathsf{Log}(K)$ is a *weak* intermediate logic.

Notation. For any formula φ, we denote by φ^n the formula obtained from φ by replacing any occurrence of a propositional letter with its negation.

Definition 3.7 (Negative variant of a logic)**.** If Λ is an intermediate logic, we define its negative variant Λ^n as:

$$\Lambda^n = \{\varphi \,|\, \varphi^n \in \Lambda\}$$

Remark 4. *For any intermediate logic Λ, its negative variant Λ^n is a weak intermediate logic including Λ.*

Proof. Fix an intermediate logic Λ. Since Λ is closed under uniform substitution, $\varphi \in \Lambda$ implies $\varphi^n \in \Lambda$ and so $\varphi \in \Lambda^n$. This shows $\Lambda \subseteq \Lambda^n$.

Moreover, if φ and $\varphi \to \psi$ belong to Λ^n, then both φ^n and $(\varphi \to \psi)^n = \varphi^n \to \psi^n$ are in Λ which is closed under modus ponens; therefore, $\psi^n \in \Lambda$, which means that $\psi \in \Lambda^n$. This shows that Λ^n is closed under modus ponens.

Finally, if $\varphi \in \Lambda^n$ then $\varphi^n \in \Lambda \subseteq \mathsf{CPL}$. Then, since CPL is substitution-closed, $\varphi^{nn} \in \mathsf{CPL}$ and therefore also $\varphi \in \mathsf{CPL}$, as the double negation law holds in CPL. This shows that $\Lambda^n \subseteq \mathsf{CPL}$ and therefore that Λ^n is indeed a weak intermediate logic. □

The following observation will turn out useful below.

Remark 5. *If a logic Λ has the disjunction property, then so does Λ^n.*

Proof. If $\varphi \vee \psi \in \Lambda^n$, then $\varphi^n \vee \psi^n \in \Lambda$; thus, by the disjunction property, at least one of φ^n and ψ^n is in Λ, which means that at least one of φ and ψ is in Λ^n. □

Definition 3.8 (Negative valuations)**.** Let F be an intuitionistic frame. A valuation V is called *negative* in case for any point w in F and for any proposition letter p:

$$(F, V), w \Vdash p \iff (F, V), w \Vdash \neg\neg p$$

We will call a model *negative* in case its valuation is negative. Observe that if M is a negative model, for any point w and formula φ we have $M, w \Vdash \varphi \iff M, w \Vdash \varphi^{nn}$.

Definition 3.9 (Negative variant of a model)**.** If $M = (W, R, V)$ is a Kripke model, we define the negative variant M^n of M to be model $M^n = (M, R, V^n)$ where

$$V^n(p) := \{w \in W \,|\, M, w \Vdash \neg p\}$$

that is, V^n makes a propositional letter true precisely where its negation was true in the original model.

A straightforward inductive proof yields the following result.

Proposition 23. *For any model M, any point w and formula φ:*

$$M, w \Vdash \varphi^n \iff M^n, w \Vdash \varphi$$

Remark 6. *For any model M, its negative variant M^n is a negative model.*

Proof. Take any point w of M and formula φ. According to the previous proposition and recalling that in intuitionistic logic triple negation is equivalent to single negation, we have $M^n, w \Vdash p \iff M, w \Vdash \neg p \iff M, w \Vdash \neg\neg\neg p \iff M^n, w \Vdash \neg\neg p$. □

Definition 3.10. Let K be a class of intuitionistic Kripke frames. Then we denote by $\mathbf{n}K$ the class of negative K-models, i.e., negative Kripke models whose frame is in K.

Proposition 24. *For any class K of Kripke frames, $\mathsf{Log}(\boldsymbol{n}K) = \mathsf{Log}(K)^n$.*

Proof. If $\varphi \notin \mathsf{Log}(K)^n$, i.e. if $\varphi^n \notin \mathsf{Log}(K)$, then there must be a K-model M (i.e., a model based on a K-frame) and a point w such that $M, w \nVdash \varphi^n$. But then, by Proposition 23 we have $M^n, w \nVdash \varphi$, and thus $\varphi \notin \mathsf{Log}(\mathbf{n}K)$ since M^n is a negative K-model.

Conversely, if $\varphi \notin \mathsf{Log}(\mathbf{n}K)$, let M be a negative K-model and w a point in M with $M, w \nVdash \varphi$. Then since M is negative, $M, w \nVdash \varphi^{nn}$. Therefore, by Proposition 23, $M^n, w \nVdash \varphi^n$. But M^n shares the same frame of M, which is a K-frame: so $\varphi^n \notin \mathsf{Log}(K)$, that is, $\varphi \notin \mathsf{Log}(K)^n$. □

The following result states that for any intermediate logic Λ, Λ^n is axiomatized by a system having Λ and all the atomic double negation formulas $\neg\neg p \to p$ as axioms, and modus ponens as unique inference rule.

Proposition 25. *If Λ is an intermediate logic, Λ^n is the smallest weak intermediate logic containing Λ and the atomic double negation axiom $\neg\neg p \to p$ for each propositional letter p.*

Proof. We have already observed (see Remark 3.4) that Λ^n is a weak intermediate logic containing Λ; moreover, for any letter p we have $\neg\neg\neg p \to \neg p \in \mathsf{IPL} \subseteq \Lambda$, so each atomic double negation formula is in Λ^n.

To see that Λ^n is *the smallest* such logic, let Λ' be another weak logic containing Λ and the atomic double negation axioms. Consider $\varphi \in \Lambda^n$: this means that $\varphi^n \in \Lambda$. But clearly, φ is derivable by modus ponens from φ^n and the atomic double negation axioms for letters in φ: hence, as Λ' contains Λ and the atomic double negation formulas and it is closed under modus ponens, $\varphi \in \Lambda'$. Thus, $\Lambda^n \subseteq \Lambda'$. □

With slight abuse of notation, we will henceforth identify Λ^n not only with a set of formulas, but also with the following derivation system:

Axioms:

- all formulas in Λ
- $\neg\neg p \to p$ for all proposition letters $p \in \mathcal{P}$

Rules:

- modus ponens

If Θ is a set of formulae and φ is a formula, we will write $\Theta \vdash_{\Lambda^n} \varphi$ in case φ is derivable from the set of assumptions Θ in the system Λ^n.

3.5 Disjunction Property + Disjunctive Negative Translation = InqL

We are now ready to connect some of the notions introduced in previous subsections. The following theorem characterizes InqL as the unique weak intermediate logic that has the disjunction property and preserves logical equivalence under disjunctive negation translation (as defined in section 3.2).

Theorem 7. *Let L be a weak intermediate logic. If $\varphi \equiv_L$ DNT(φ) for all formulas φ, then InqL $\subseteq L$. If, additionally, L has the disjunction property, then $L =$ InqL.*

Proof. Let L be a weak intermediate logic for which any formula φ is equivalent to DNT(φ). Suppose $\varphi \in$ InqL. Then DNT$(\varphi) \in$ InqL. Write DNT$(\varphi) = \neg\nu_1 \vee \cdots \vee \neg\nu_k$: since InqL has the disjunction property, we must have $\neg\nu_i \in$ InqL for some $1 \leq i \leq k$. Now, we know that IPL coincides with CPL as far as negations are concerned (Chagrov and Zakharyaschev 1997, p.47) and it follows from this that every two weak intermediate logics coincide as far as negations are concerned. So if $\neg\nu_i \in$ InqL, then also $\neg\nu_i \in L$. Hence, DNT$(\varphi) \in L$, and since $\varphi \equiv_L$ DNT(φ), also $\varphi \in L$. This shows that InqL $\subseteq L$.

Now suppose that L also has the disjunction property. Consider a formula $\varphi \in L$: since $\varphi \equiv_L$ DNT(φ) we have DNT$(\varphi) \in L$. But L has the disjunction property and therefore, using the same notation as above, $\neg\nu_i \in L$ for some i. Then again because all weak intermediate logics agree on negations, $\neg\nu_i \in$ InqL, whence DNT$(\varphi) \in$ InqL and also $\varphi \in$ InqL. This proves that $L \subseteq$ InqL. □

3.6 Axiomatizing inquisitive logic

Given that InqL is the only weak intermediate logic with the disjunction property that preserves logical equivalence under DNT, it is natural to ask next what exactly is required, in terms of axioms, in order to preserve logical equivalence under DNT. Answering this question will directly lead to an axiomatization of InqL.

In order to identify the relevant axioms, let us go back to the proof of Proposition 21, which stated that $\varphi \equiv_{\mathsf{InqL}}$ DNT(φ) for any φ. The proof is by induction on φ. The atomic case amounts to the validity of the atomic double negation axioms. The inductive step for disjunction is trivial, while the one for conjunction follows from the fact that IPL $\subseteq$ InqL, which means that intuitionistic equivalences (like instances of the distributive laws) hold in the inquisitive setting.

Finally, for the inductive step for implication we need—in addition to some intuitionistically valid equivalences—the following equivalence:

$$\left(\neg\chi \to \bigvee_{1\leq i\leq k} \neg\psi_i\right) \equiv_{\mathsf{InqL}} \bigvee_{1\leq i\leq k} (\neg\chi \to \neg\psi_i)$$

for all formulas $\chi, \psi_1, \ldots, \psi_k$. Since the right-to-left entailment already holds intuitionistically, what is needed more specifically is that any substitution instance of each of the following formulas be valid in InqL:

ND$_k$ $\quad (\neg p \to \bigvee_{1\leq i\leq k} \neg q_i) \quad \longrightarrow \quad \bigvee_{1\leq i\leq k} (\neg p \to \neg q_i)$

Thus—besides intuitionistic validities—we need all instances of $\mathtt{ND}_k$, $k \in \omega$, and all the atomic double negation axioms $\neg\neg p \to p$ in order to preserve logical equivalence under DNT. Any system containing those axioms and equipped with the modus ponens rule will be able to prove the equivalence between a formula φ and its translation $\textsc{dnt}(\varphi)$. This suffices to prove Proposition 26, where ND is the intermediate logic axiomatized by the axioms $\mathtt{ND}_k$, $k \in \omega$.

Proposition 26. *For any logic $\Lambda \supseteq \mathsf{ND}$ and any formula φ, $\varphi \equiv_{\Lambda^n} \textsc{dnt}(\varphi)$.*

This proposition immediately yields a whole range of intermediate logics whose negative variant coincides with inquisitive logic.

Theorem 8 (Completeness theorem)**.**
$\Lambda^n = \mathsf{InqL}$ for any logic $\Lambda \supseteq \mathsf{ND}$ with the disjunction property.

Proof. Let Λ be an extension of ND with the disjunction property. Then according to Proposition 26 we have $\varphi \equiv_{\Lambda^n} \textsc{dnt}(\varphi)$ for all φ; moreover, Λ^n has the disjunction property (see Remark 5). Hence by Theorem 7 we have $\Lambda^n = \mathsf{InqL}$. □

The logic ND has been studied by (Maksimova 1986). Among other things, she shows (1) that ND has the disjunction property, and (2) that the maximal intermediate logic with the disjunction property containing ND is ML, an intermediate logic introduced by Medvedev (1962), also known as 'the logic of finite problems'. Maximova also remarks that the logic KP, which was introduced by Kreisel and Putnam (1957) as the intermediate logic axiomatized by:

$$\mathtt{KP} \quad (\neg p \to q \vee r) \quad \longrightarrow \quad (\neg p \to q) \vee (\neg p \to r)$$

is one of the logics in between ND and ML which has the disjunction property. This immediately gives us three concrete axiomatizations of InqL.

Corollary 5. *$\mathsf{ND}^n = \mathsf{KP}^n = \mathsf{ML}^n = \mathsf{InqL}$.*

Medvedev's logic will be discussed in more detail in the next section, which investigates the schematic fragment of InqL. The completeness theorem established above will be further strengthened in section 5. There we will see that the negative variant of an intermediate logic Λ coincides with InqL *if and only if* $\mathsf{ND} \subseteq \Lambda \subseteq \mathsf{ML}$.

4 The schematic fragment of inquisitive logic

We have already remarked that inquisitive logic is not closed under uniform substitution; it is natural to ask, then, what the schematic fragment of InqL is. In this section we will address this issue and we will find that this fragment in fact coincides with Medvedev's logic of finite problems.

Definition 4.1 (Schematic fragment of InqL)**.** We denote by Sch(InqL) the set of formulae that are schematically valid in InqL, i.e., those formulae φ such that $\varphi^* \in \mathsf{InqL}$ for any substitution instance φ^* of φ.

Notice that Sch(InqL) is the greatest intermediate logic included in InqL.

Definition 4.2 (Medvedev frames)**.** A Medvedev frame consists of all the non-empty subsets of some finite set X, ordered by the superset relation. In other words, a Medvedev frame is a frame of the shape $(\wp(X) - \{\emptyset\}, \supseteq)$, where X is some finite set. The class of Medvedev frames will be denoted by **Med**.

Notice that the frame F_I underlying the Kripke model for inquisitive semantics is $(\wp\wp(\mathcal{P}) - \{\emptyset\}, \supseteq)$. So F_I is a Medvedev frame whenever the set of proposition letters $\mathcal{P}$ is finite.

Definition 4.3 (Medvedev logic)**.** Medvedev logic is the logic of the class **Med** of Medvedev frames: $\mathsf{ML} := \mathsf{Log}(\mathbf{Med})$. Notice that ML is an intermediate logic.

The following theorem establishes the main result of this section, namely that the schematic fragment of InqL coincides with Medvedev logic. There is, however, one subtlety that should be remarked: whereas we assumed so far that the set of atomic proposition letters $\mathcal{P}$ may be finite or countably infinite, it is at this stage important to assume that $\mathcal{P}$ is in fact countably infinite.

Theorem 9. *Sch(*InqL*)=*ML*.*

In order to prove this theorem, we need the following lemma. This lemma employs the notion of a *p-morphism*, which we assume to be familiar (see, for instance, Chagrov and Zakharyaschev 1997, p.30).

Lemma 1. *For any negative Medvedev model M, there exists a p-morphism η from M to the Kripke model for inquisitive logic M_I.*

Proof. Let $M = (W, R, V)$ be a negative Medvedev model. For any endpoint e of M, denote by i_e the valuation $i_e = \{p \in \mathcal{P} \mid e \in V(p)\}$ consisting of those letters true at e. For any point w in M, let E_w denote the set of endpoints accessible from w in M. Define the candidate p-morphism η as follows:

For any $w \in W$, $\eta(w) = \{i_e \mid e \in E_w\}$.

First, note that any point in M can see at least one endpoint. This means that for any $w \in W$, we have $E_w \neq \emptyset$, and therefore $\eta(w) \neq \emptyset$. This insures that $\eta(w) \in W_I$, so that the map $\eta : W \to W_I$ is well-defined. It remains to check that η is a p-morphism. Fix any $w \in W$:

- **Proposition Letters**. Take any proposition letter p. If $M, w \Vdash p$, then by persistence we have $M, e \Vdash p$ for any $e \in E_w$; this implies that $p \in i$ for any index $i \in \eta(w)$ and so $\eta(w) \models p$, whence $M_I, \eta(w) \Vdash p$.

  Conversely, suppose $M, w \nVdash p$. Then since the valuation V is negative, $M, w \nVdash \neg\neg p$, so there must be a successor v of w with $M, v \Vdash \neg p$. M is finite, so E_v is non-empty. Take a point $e \in E_v$. Then, by persistence, $M, e \Vdash \neg p$, whence $p \notin i_e$. But, by transitivity of R, we have that $e \in E_w$, so $i_e \in \eta(w)$. Thus $\eta(w) \not\models p$, whence $M_I, \eta(w) \nVdash p$.

- **Forth Condition**. Suppose wRv: then since our accessibility relation is transitive, $E_w \supseteq E_v$ and thus also $\eta(w) \supseteq \eta(v)$.

- **Back Condition**. Suppose $\eta(w) \supseteq t$: we must show that there is some successor v of w such that $\eta(v) = t$.

Since t is a non-empty subset of $\eta(w) = \{i_e \mid e \in E_w\}$, there must be some non-empty subset $E_* \subseteq E_w$ such that $t = \{i_e \mid e \in E_*\}$. We must show, then, that there is a successor v of w in M whose terminal successors are exactly those in E_*.

Recall that M is based on a frame that consists of all the non-empty subsets of some finite set X, ordered by the superset relation. In particular, all the endpoints in M are singleton subsets of X, and for any set of endpoints E, there is a point, namely $\bigcup E$, whose terminal successors are exactly the ones in E.

Thus, for v we can take $\bigcup E_*$. Then, $E_v = E_*$, and $\eta(v) = t$. □

Proof of Theorem 9. Suppose $\varphi \notin \mathrm{Sch}(\mathsf{InqL})$: then there is a substitution instance φ^* of φ such that $\varphi^* \notin \mathsf{InqL}$. But then it follows from Proposition 4 that φ^* can be falsified in a point of the model M_I for inquisitive semantics relative to the finite set of propositional letters P_{φ^*}; and since this model is a Medvedev model, $\varphi^* \notin \mathsf{ML}$. But then, as ML is closed under uniform substitution, also $\varphi \notin \mathsf{ML}$. This shows that $\mathsf{ML} \subseteq \mathrm{Sch}(\mathsf{InqL})$.
For the converse inclusion, suppose $\varphi(p_1, \ldots, p_n) \notin \mathsf{ML}$. This means that there is a model $M = (F, V)$, where F is a Medvedev frame, and a point w in this model, such that $M, w \nVdash \varphi$. Now, the idea is to use Lemma 1 to transfer this counterexample to the Kripke model M_I for inquisitive semantics. In order to do so, however, we need our starting model to be a *negative* Medvedev model. Our model is indeed a Medvedev model, but there is no reason why the valuation V should be negative. Therefore, what we want to do is replace V by a negative valuation $\widehat{V}$, and then simulate the behaviour of the propositional letters $p_1, \ldots, p_n$ with complex formulae $\psi_1, \ldots, \psi_n$.

In order to do this, associate any point u in M with a propositional letter q_u and define a new valuation $\widehat{V}$ as follows: for any point v and any proposition letter q_u, take $v \in \widehat{V}(q_u)$ if and only if $v \subseteq u$. For proposition letters q which are not of the shape q_u for some u, take $\widehat{V}(q) = \emptyset$. Then define $\widehat{M} = (F, \widehat{V})$.
Notice that the valuation $\widehat{V}$ is indeed negative. For, take any letter q_u and suppose that a certain point v is not in $\widehat{V}(q_u)$: then $v \nsubseteq u$, so we can take an element $x \in v - u$. Since $\{x\} \nsubseteq u$, $\{x\} \notin \widehat{V}(q_u)$, and therefore, since singletons are endpoints and thus behave classically, we have $\widehat{M}, \{x\} \Vdash \neg q_u$. Finally, since $\{x\} \subseteq v$, $\{x\}$ is a successor of v, and therefore $\widehat{M}, v \nVdash \neg\neg q_u$. So indeed $\widehat{M}$ is a negative Medvedev model, and Lemma 1 applies, yielding a p-morphism $\eta : \widehat{M} \to M_I$.
We now turn to the second task, namely, find a complex formula ψ_i that simulates in $\widehat{M}$ the behaviour of the atom p_i in M. For $1 \leq i \leq n$, define $\psi_i := \bigvee_{v \in V(p_i)} q_v$. We are going to show that for any point u:

$$M, u \Vdash p_i \iff \widehat{M}, u \Vdash \psi_i$$

If $M, u \Vdash p_i$, i.e. if $u \in V(p_i)$, then since $\widehat{M}, u \Vdash q_u$ we immediately have that $\widehat{M}, u \Vdash \bigvee_{v \in V(p_i)} q_v$. That is, $\widehat{M}, u \Vdash \psi_i$.
Conversely, if $\widehat{M}, u \Vdash \psi_i$, then there is a point $v \in V(p_i)$ such that $u \in \widehat{V}(q_v)$, which in turn, by definition of $\widehat{V}$, means that $u \subseteq v$. But then, by persistence,

$u \in V(p_i)$, that is, $M, u \Vdash p_i$. This proves the above equivalence. Now, it follows immediately that for any point u:

$$M, u \Vdash \varphi(p_1, \ldots, p_n) \iff \widehat{M}, u \Vdash \varphi(\psi_1, \ldots, \psi_n)$$

In particular, $\widehat{M}, w \nVdash \varphi(\psi_1, \ldots, \psi_n)$. Thus, using the p-morphism $\eta : \widehat{M} \to M_I$ provided by Lemma 1 we finally get that $M_I, \eta(w) \nVdash \varphi(\psi_1, \ldots, \psi_n)$. Therefore, $\varphi(\psi_1, \ldots, \psi_n) \notin \mathsf{InqL}$ and thus $\varphi(p_1, \ldots, p_n) \notin \mathrm{Sch}(\mathsf{InqL})$. □

Observe that the given proof in fact establishes something stronger than the equality $\mathrm{Sch}(\mathsf{InqL}) = \mathsf{ML}$. It shows that in order to falsify a formula $\varphi \notin \mathrm{Sch}(\mathsf{InqL})$ we do not have to look at arbitrary substitution instances of φ; it suffices to take into consideration substitutions of atomic proposition letters with arbitrarily large disjunctions of atoms. This yields the following corollary.

Corollary 6. *For any formula $\varphi(p_1, \ldots, p_n)$, the following are equivalent:*

1. $\varphi(p_1, \ldots, p_n) \in \mathsf{ML}$;
2. $\varphi(\bigvee_{1 \le i \le k} p_1^i, \ldots, \bigvee_{1 \le i \le k} p_n^i) \in \mathsf{InqL}$ *for all* $k \in \omega$;
3. $\varphi(\bigvee_{1 \le i \le k} \neg p_1^i, \ldots, \bigvee_{1 \le i \le k} \neg p_n^i) \in \Lambda$ *for all* $k \in \omega$, *where* Λ *is any intermediate logic with* $\Lambda^n = \mathsf{InqL}$.

We end this section with some notes on Medvedev's logic. This logic was first presented in (Medvedev 1962) as the logic arising from interpreting propositional formulas as *finite problems*. In (Medvedev 1966), the logic was characterized in terms of Kripke models as the logic of the class **Med**. The quest for an axiomatization of ML did not produce significant results until Maksimova et al. (1979) proved that ML is not finitely axiomatizable and indeed not axiomatizable with a finite number of propositional letters. The question of whether ML admits a recursive axiomatization (equivalently, of whether ML is decidable) is a long-standing open problem.

This makes the results we just established particularly interesting. For, in the first place we have seen that $\mathsf{ML} = \mathrm{Sch}(\mathsf{InqL}) = \mathrm{Sch}(\mathsf{KP}^n) = \mathrm{Sch}(\mathsf{ND}^n)$, which means that the systems KP^n and ND^n give pseudo-axiomatizations of ML: they derive 'slightly' more formulas than those in ML, but if we restrict our attention to the schematic validities, then we have precisely Medvedev's logic.

In the second place, Corollary 6 provides a connection between Medvedev's logic and other intermediate logics, among which the well-understood Kreisel-Putnam logic, that might pave the way for new attempts to solve the decidability problem for ML.

For instance—since both InqL and KP are decidable—if it were possible to find a finite bound b for the maximum number k of disjuncts that we need to use in order to falsify a non-schematically valid formula φ (possibly depending on the number of propositional letters in φ) then ML would be decidable. For, to determine whether $\varphi \in \mathsf{ML}$ it would then suffice to check whether the formula $\varphi(\bigvee_{1 \le i \le k} \neg p_1^i, \ldots, \bigvee_{1 \le i \le k} \neg p_n^i)$ is in KP for all $k \le b$, and this procedure can be performed in a finite amount of time.

It is possible to show (see Ciardelli 2009) that for formulas containing a single proposition letter p, such a bound exists and equals 2. As a consequence, the one-letter fragment of ML is decidable. This result is not quite new, since

Medvedev (1966) showed that the one-letter fragment of ML coincides with the well-known Scott logic, and this logic is known to be decidable—this follows, for instance, from Theorem 11.58 in (Chagrov and Zakharyaschev 1997, p.410)—but the argument presented here is new and could perhaps be generalized.

5 The range of intermediate logics whose negative variant is InqL

Using the result that the schematic fragment of InqL coincides with ML it is possible to strengthen the completeness result obtained in section 3.6: we can give a complete characterization of the range of intermediate logics whose negative variant coincides with InqL.

Theorem 10 (Range of intermediate logics whose negative variant is InqL)**.** *For any intermediate logic* Λ*:*

$$\Lambda^n = \mathsf{InqL} \quad \Longleftrightarrow \quad \mathsf{ND} \subseteq \Lambda \subseteq \mathsf{ML}$$

We articulate the proof of this theorem in two lemmata.

Lemma 2. *For any intermediate logic* Λ*, if* $\Lambda^n = \mathsf{InqL}$*, then* $\mathsf{ND} \subseteq \Lambda$*.*

Proof. By contraposition, suppose $\mathsf{ND} \not\subseteq \Lambda$. Then there is a number k for which the formula $\mathtt{ND}_k := (\neg p \to \bigvee_{1 \le i \le k} \neg q_i) \;\longrightarrow\; \bigvee_{1 \le i \le k} (\neg p \to \neg q_i)$ is not in Λ. Note that this formula is nothing but φ^n where φ denotes the formula

$$\left(p \to \bigvee_{1 \le i \le k} q_i\right) \;\longrightarrow\; \bigvee_{1 \le i \le k} (p \to q_i)$$

But then φ^{nn} cannot be in Λ. For if it were—Λ being closed under uniform substitution—φ^{nnn} should also be in Λ, and so should the equivalent formula φ^n. But φ^n is *not* in Λ. Thus, $\varphi^{nn} \notin \Lambda$, whence $\varphi^n \notin \Lambda^n$. On the other hand, $\varphi^n = \mathtt{ND}_k \in \mathsf{InqL}$. So, $\Lambda^n \neq \mathsf{InqL}$. □

Lemma 3. *For any intermediate logic* Λ*, if* $\Lambda^n = \mathsf{InqL}$*, then* $\Lambda \subseteq \mathsf{ML}$*.*

Proof. We know that Λ is always included in Λ^n, so if $\Lambda^n = \mathsf{InqL}$ we have $\Lambda = \mathrm{Sch}(\Lambda) \subseteq \mathrm{Sch}(\Lambda^n) = \mathrm{Sch}(\mathsf{InqL}) = \mathsf{ML}$, where the first equality uses the fact that Λ is closed under uniform substitution. □

Proof of Theorem 10. Let Λ be an intermediate logic. The operation of negative variant is obviously monotone, so if $\mathsf{ND} \subseteq \Lambda \subseteq \mathsf{ML}$ we have $\mathsf{InqL} = \mathsf{ND}^n \subseteq \Lambda^n \subseteq \mathsf{ML}^n = \mathsf{InqL}$, where the first and the last equality come from Corollary 5.

On the other hand, the previous two lemmata together show that $\Lambda^n = \mathsf{InqL}$ implies $\mathsf{ND} \subseteq \Lambda \subseteq \mathsf{ML}$. This completes the proof of the theorem. □

Note that as a corollary of this theorem we can easily derive a well-known result due to Maksimova (1986), which was already mentioned at the end of section 3.6.

Corollary 7. *If* $\Lambda \supseteq \mathsf{ND}$ *is a logic with the disjunction property, then* $\Lambda \subseteq \mathsf{ML}$*. In particular,* ML *is a* maximal *logic with the disjunction property.*

Proof. According to Theorem 8, if $\Lambda \supseteq \mathsf{ND}$ is a logic with the disjunction property, then $\Lambda^n = \mathsf{InqL}$ and thus, by Theorem 10, $\Lambda \subseteq \mathsf{ML}$. □

Finally, Theorem 10 can easily be extended to a *strong* completeness result.

Theorem 11 (Strong completeness)**.** *For any intermediate logic Λ, the following two conditions are equivalent:*

1. $\mathsf{ND} \subseteq \Lambda \subseteq \mathsf{ML}$
2. *For any set of formulas Θ and any formula φ:* $\quad \Theta \models_{\mathsf{InqL}} \varphi \iff \Theta \vdash_{\Lambda^n} \varphi$

Proof. First let Λ be an intermediate logic that satisfies condition 2. Then, in particular, $\mathsf{InqL} = \Lambda^n$, and therefore, by Theorem 10, $\mathsf{ND} \subseteq \Lambda \subseteq \mathsf{ML}$.

Now let us assume that Λ is an intermediate logic that satisfies condition 1, and show that $\vdash_{\Lambda^n}$ is sound and complete with respect to $\models_{\mathsf{InqL}}$. The soundness direction is straightforward since $\Lambda \subseteq \mathsf{InqL}$ (for, $\Lambda \subseteq \Lambda^n$ and $\Lambda^n = \mathsf{InqL}$), $\neg\neg p \to p \in \mathsf{InqL}$, and the set of formulas supported by a state is closed under modus ponens. For the completeness direction, suppose that $\Theta \models_{\mathsf{InqL}} \varphi$. Then, by compactness (Theorem 2), there are formulas $\theta_1, \ldots, \theta_k \in \Theta$ such that $\theta_1, \ldots, \theta_k \models_{\mathsf{InqL}} \varphi$, which by the deduction theorem amounts to $\theta_1 \wedge \cdots \wedge \theta_k \to \varphi \in \mathsf{InqL}$. Then by Theorem 10, $\theta_1 \wedge \cdots \wedge \theta_k \to \varphi \in \Lambda^n$, whence clearly $\Theta \vdash_{\Lambda^n} \varphi$. □

6 InqL as the disjunctive-negative fragment of IPL

As we already observed, the meanings of inquisitive semantics are sets of alternatives, where each alternative is a classical meaning. This essential feature of the semantics is mirrored on the syntactic, logical level by the fact that any formula φ is equivalent to a disjunction of negations $\textsc{dnt}(\varphi)$.

The completeness result in section 3 was based on the insight that preservation of logical equivalence under DNT is an essential feature of the logic InqL. But there is even more to say about DNT: in this section we will show that it constitutes a *translation* of InqL into IPL, in the following sense (cf. Epstein et al. 1995, chapter 10):

Definition 6.1 (Translations between logics)**.** Let L, L' be two logics arising from two entailment relations $\models_L$ and $\models_{L'}$. We say that a mapping t from formulas in the language of L to formulas in the language of L' is a *translation* from L to L' in case for any set of formulas Θ and any formula φ we have:

$$\Theta \models_L \varphi \iff t[\Theta] \models_{L'} t(\varphi)$$

where $t[\Theta] = \{t(\theta) \mid \theta \in \Theta\}$.

Moreover, we will show that the disjunctive-negative fragment of InqL coincides with the one of IPL, and that InqL is in fact isomorphic to the disjunctive-negative fragment of IPL through the translation DNT (just as CPL is isomorphic to the negative fragment of IPL through the translation mapping φ to $\neg\neg\varphi$).

Let us call a formula *disjunctive-negative* in case it is a disjunction of negations. The following proposition says that inquisitive entailment and intuitionistic entailment agree as far as disjunctive-negative formulas are concerned.

Proposition 27. *If φ is a disjunctive-negative formula and Θ a set of disjunctive-negative formulas, then:* $\Theta \models_{\mathsf{InqL}} \varphi \iff \Theta \models_{\mathsf{IPL}} \varphi$.

Proof. Consider an arbitrary set Θ of disjunctive-negative formulas and a disjunctive-negative formula $\varphi = \neg\xi_1 \vee \cdots \vee \neg\xi_k$. If $\Theta \models_{\mathsf{InqL}} \varphi$, then by compactness and the deduction theorem there must be $\theta_1, \ldots, \theta_n \in \Theta$ such that $\theta_1 \wedge \cdots \wedge \theta_n \to \varphi \in \mathsf{InqL}$.

Now since each θ_k is a disjunction of negations and since the distributive laws hold in intuitionistic logic, in IPL we can turn $\theta_1 \wedge \cdots \wedge \theta_n$ into a disjunction of conjunctions of negations. In turn, a conjunction of negations is equivalent to a negation in intuitionistic logic. So we can find formulas $\chi_1, \ldots, \chi_m$ such that $\theta_1 \wedge \cdots \wedge \theta_n \equiv_{\mathsf{IPL}} \neg\chi_1 \vee \cdots \vee \neg\chi_m$. But then $(\theta_1 \wedge \cdots \wedge \theta_n \to \varphi) \equiv_{\mathsf{IPL}} (\neg\chi_1 \vee \cdots \vee \neg\chi_m \to \varphi) \equiv_{\mathsf{IPL}} \bigwedge_{1 \leq i \leq m}(\neg\chi_i \to \varphi)$.

Equivalence in IPL implies equivalence in InqL, so $\theta_1 \wedge \cdots \wedge \theta_n \to \varphi \in \mathsf{InqL}$ implies that $\bigwedge_{1 \leq i \leq m}(\neg\chi_i \to \varphi) \in \mathsf{InqL}$, which in turn means that for each $1 \leq i \leq n$ we have $\neg\chi_i \to \varphi \in \mathsf{InqL}$. Writing out φ, this amounts to $\neg\chi_i \to \neg\xi_1 \vee \cdots \vee \neg\xi_k \in \mathsf{InqL}$. But since InqL contains the $\mathtt{ND}_k$ axioms, it follows that $\bigvee_{1 \leq j \leq k}(\neg\chi_i \to \neg\xi_j) \in \mathsf{InqL}$, and therefore, as InqL has the disjunction property, for some $1 \leq j \leq k$ we must have that $\neg\chi_i \to \neg\xi_j \in \mathsf{InqL} \subseteq \mathsf{CPL}$. Now, $\neg\chi_i \to \neg\xi_j \equiv_{\mathsf{IPL}} \neg\neg(\neg\chi_i \to \neg\xi_j)$, and since CPL and IPL agree about negations (Chagrov and Zakharyaschev 1997, p.47), also $\neg\chi_i \to \neg\xi_j \in \mathsf{IPL}$, whence *a fortiori* $\neg\chi_i \to \varphi \in \mathsf{IPL}$. Since this can be concluded for each i, we have $\bigwedge_{1 \leq i \leq m}(\neg\chi_i \to \varphi) \in \mathsf{IPL}$, and therefore also the equivalent formula $\theta_1 \wedge \cdots \wedge \theta_n \to \varphi$ must be in IPL. But then obviously $\Theta \models_{\mathsf{IPL}} \varphi$.
The converse implication is trivial, as InqL extends IPL. □

As a particular case of this proposition, let us remark that for any disjunctive-negative formula φ we have $\varphi \in \mathsf{InqL} \iff \varphi \in \mathsf{IPL}$.

Corollary 8. DNT *is a translation of* InqL *into* IPL.

Proof. We have to show that for any Θ and any φ:

$$\Theta \models_{\mathsf{InqL}} \varphi \iff \textsc{dnt}[\Theta] \models_{\mathsf{IPL}} \textsc{dnt}(\varphi)$$

where $\textsc{dnt}[\Theta] = \{\textsc{dnt}(\theta) \mid \theta \in \Theta\}$. It follows from Proposition 21 that $\Theta \models_{\mathsf{InqL}} \varphi \iff \textsc{dnt}[\Theta] \models_{\mathsf{InqL}} \textsc{dnt}(\varphi)$. But $\textsc{dnt}(\psi)$ is always a disjunctive-negative formula. So, by Proposition 27, $\textsc{dnt}[\Theta] \models_{\mathsf{InqL}} \textsc{dnt}(\varphi) \iff \textsc{dnt}[\Theta] \models_{\mathsf{IPL}} \textsc{dnt}(\varphi)$ and we are done. □

Observe that if the map t is a translation from a logic L to another logic L', then t naturally lifts to an embedding $\bar{t} : \mathcal{L}/_{\equiv_L} \to \mathcal{L}/_{\equiv'_L}$ of the Lindenbaum-Tarski algebra of L into the Lindenbaum-Tarski algebra of L', given by $\bar{t}([\psi]_{\equiv_L}) := [t(\psi)]_{\equiv_{L'}}$.[3]

Since we have seen that DNT is a translation from InqL to IPL, the map $\overline{\textsc{dnt}}$ defined by $\overline{\textsc{dnt}}([\psi]_{\equiv_{\mathsf{InqL}}}) = [\textsc{dnt}(\psi)]_{\equiv_{\mathsf{IPL}}}$ is an embedding of the Lindenbaum-Tarski algebra of InqL into the one of IPL. For the singleton set of propositional letters, this embedding is depicted in Figure 6.

Now, for any ψ, $\textsc{dnt}(\psi)$ is a disjunctive-negative formula. Conversely, consider a disjunctive-negative formula ψ. Since $\psi \equiv_{\mathsf{InqL}} \textsc{dnt}(\psi)$ but both ψ

[3] For more details on the issues of translations between logics, see (Epstein et al. 1995).

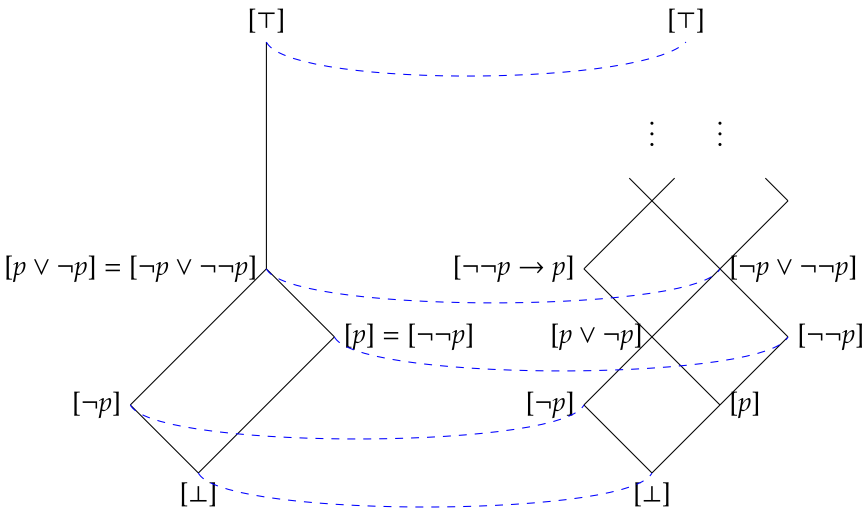


Figure 3: Embedding of the Lindenbaum-Tarski algebra of InqL (on the left), into the Lindenbaum-Tarski algebra of IPL (the Rieger-Nishimura lattice, on the right), for the singleton set of proposition letters $\mathcal{P} = \{p\}$.

and $\textsc{dnt}(\psi)$ are disjunctive-negative, it follows from Proposition 27 that $\psi \equiv_{\mathsf{IPL}} \textsc{dnt}(\psi)$; in other words, we have $[\psi]_{\equiv_{\mathsf{IPL}}} = [\textsc{dnt}(\psi)]_{\equiv_{\mathsf{IPL}}} = \overline{\textsc{dnt}}([\psi]_{\equiv_{\mathsf{IPL}}})$, so $[\psi]_{\equiv_{\mathsf{IPL}}}$ is in the image of the embedding $\textsc{dnt}$.

This shows that the image of the embedding $\textsc{dnt}$ is precisely the set of equivalence classes of disjunctive-negative formulas. In other words, just like CPL is isomorphic to the negative fragment of IPL, for InqL we have the following result.

Proposition 28. *InqL is isomorphic to the disjunctive-negative fragment of IPL.*

As a corollary of the well-known fact that CPL is isomorphic to the negative fragment of IPL we know that, for any n, there are exactly 2^{n+1} intuitionistically non-equivalent negative formulas in $\mathcal{L}_{\{p_1,\ldots,p_n\}}$, just as many as there are classically non-equivalent formulas in $\mathcal{L}_{\{p_1,\ldots,p_n\}}$.

Analogously, our result that InqL is isomorphic to the disjunctive-negative fragment of IPL comes with the corollary that there are exactly as many intuitionistically non-equivalent disjunctive-negative formulas in $\mathcal{L}_{\{p_1,\ldots,p_n\}}$ as there are inquisitively non-equivalent formulas in $\mathcal{L}_{\{p_1,\ldots,p_n\}}$.

The number of inquisitively non-equivalent formulas in $\mathcal{L}_{\{p_1,\ldots,p_n\}}$ is given by the number of distinct inquisitive meanings built up from indices in $\mathcal{I}_{\{p_1,\ldots,p_n\}}$. Such inquisitive meanings are nothing but antichains of the powerset algebra $\wp(\mathcal{I}_{\{p_1,\ldots,p_n\}})$. This algebra is isomorphic to $\wp(2^n)$, since $\mathcal{I}_{\{p_1,\ldots,p_n\}} = \wp(\{p_1,\ldots,p_n\})$ contains 2^n elements. Therefore, letting $D(n)$ denote the number of antichains of the powerset algebra $\wp(n)$, we have the following fact.

Corollary 9. *For any n, there are exactly $D(2^n)$ intuitionistically non-equivalent disjunctive-negative formulas in $\mathcal{L}_{\{p_1,\ldots,p_n\}}$.*

The numbers $D(n)$ are known as Dedekind numbers, and although no simple formula is known for their calculation, their values for small n have been computed and are available online, see for instance: `www.research.att.com/~njas/sequences/A014466`.

The number of inquisitive meanings in one propositional letter is 5, as displayed by the above picture; with two letters we have 167 meanings, and with three the number leaps to 56130437228687557907787.

7 Conclusions

We investigated a generalized version of inquisitive semantics, and the logic it gives rise to. In particular, we established that InqL coincides with the negative variant of any intermediate logic Λ such that ND $\subseteq \Lambda \subseteq$ ML, thus obtaining a range of sound and complete axiomatizations. We also showed that the schematic fragment of inquisitive logic coincides with ML, and proved that inquisitive logic is isomorphic to the disjunctive-negative fragment of intuitionistic logic.

References

K. Balogh. *Theme with variations: a context-based analysis of focus*. PhD thesis, University of Amsterdam, 2009.

A. Chagrov and M. Zakharyaschev. *Modal Logic*. Oxford University Press, 1997.

I. Ciardelli. A generalized inquisitive semantics. Term Paper, University of Amsterdam, 2008.

I. Ciardelli. Inquisitive semantics and intermediate logics. Master Thesis, University of Amsterdam, 2009.

R. Epstein, W. Carnielli, I. D'Ottaviano, S. Krajewski, and R. Maddux. *The Semantic Foundations of Logic, Volume 1: Propositional Logics*. Oxford University Press, 1995.

J. Groenendijk. Inquisitive semantics and dialogue management. ESSLLI course notes, `www.illc.uva.nl/inquisitive-semantics`, 2008.

J. Groenendijk. Inquisitive semantics: Two possibilities for disjunction. In P. Bosch, D. Gabelaia, and J. Lang, editors, *Seventh International Tbilisi Symposium on Language, Logic, and Computation*. Springer-Verlag, 2009.

J. Groenendijk and F. Roelofsen. Inquisitive semantics and pragmatics. In J. M. Larrazabal and L. Zubeldia, editors, *Meaning, Content, and Argument: Proceedings of the ILCLI International Workshop on Semantics, Pragmatics, and Rhetoric*. 2009. `www.illc.uva.nl/inquisitive-semantics`.

J. Groenendijk, M. Stokhof, and F. Veltman. Coreference and modality. In S. Lappin, editor, *Handbook of Contemporary Semantic Theory*, pages 179–216. Blackwell, Oxford, 1996.

G. Kreisel and H. Putnam. Eine Unableitbarkeitsbeweismethode für den intuitionistischen Aussagenkalkül. *Archiv für Mathematische Logik und Grundlagenforschung*, 3:74–78, 1957.

L. Maksimova. On maximal intermediate logics with the disjunction property. *Studia Logica*, 45:69–75, 1986.

L. Maksimova, V. Shetman, and D. Skvorcov. The impossibility of a finite axiomatization of Medvedev's logic of finitary problems. *Soviet Math. Dokl.*, 20:394–398, 1979.

S. Mascarenhas. Inquisitive semantics and logic. Master Thesis, University of Amsterdam, 2009.

J. T. Medvedev. Finite problems. *Soviet Math. Dokl.*, 3:227–230, 1962.

J. T. Medvedev. Interpretation of logical formulas by means of finite problems. *Soviet Math. Dokl.*, 7:857–860, 1966.

K. Sano. Sound and complete tree-sequent calculus for inquisitive logic. In *Proceedings of the Sixteenth Workshop on Logic, Language, Information, and Computation*, 2009. available via `www.illc.uva.nl/inquisitive-semantics`.

R. Stalnaker. Assertion. *Syntax and Semantics*, 9:315–332, 1978.

J. van Benthem. The information in intuitionistic logic. *Synthese*, 167(2):251–270, 2009.

Trust and the Dynamics of Testimony

Wesley H. Holliday

Department of Philosophy, Stanford University.
wesholliday@stanford.edu

Abstract

We propose a *dynamic testimonial logic* (DTL) to model communication and belief change among agents with different dispositions to trust each other as information sources. DTL is an extension of the dynamic epistemic logic approach to belief revision (of van Benthem 2007), with the addition of sources and trust. It is also in the spirit of the modal logic approach to trust (of Liau 2003), with the addition of dynamics for belief change. In the multi-agent framework of DTL, we can represent how communication by an information source leads other agents to revise their beliefs about the world, about the source's beliefs, and about the beliefs of other agents in the source's audience. We can also represent how an agent's uncertainty about whether another agent trusts a source can produce, after communication by the source, uncertainty about what the other agent believes, and how an agent can learn whom a source trusts from the source's communication. To capture these phenomena, we introduce a new class of *testimonial* models and model transformations, for which we give a complete axiomatization. Finally, we describe an application of DTL in modeling a special case of the phenomenon of *information cascade* discussed in the economics literature.

1 Judgment Aggregation to Information Cascades

As it is modeled formally, judgment aggregation is an instantaneous process: given a group of agents, each with an opinion on some proposition, an aggregation function takes their individual opinions and returns a group opinion, all at once (see List and Puppe 2009). Yet in many contexts—from courtrooms to committees—the protocol is to solicit individual opinions sequentially, not simultaneously. For one example in which the temporal dimension matters, consider what Sorensen (1984) calls the *epistemic bandwagon effect*:

> An expert's epistemic preferences can be justifiably influenced by his knowledge of another expert's preferences. Yet this provides the basis for an epistemic bandwagon effect. For the sake of simplicity, suppose there are three highly respectful experts, 1, 2, and

> 3, who prior to the roll-call vote are respectively in favour, indifferent, and opposed to a proposition. However, they only learn the others' preferences by their votes. If the roll-call vote is taken in order 1, 2, 3, expert 1 votes in favour. Having learned that another expert favours the proposition, the opinion of 2 is swayed and he too votes in favour. Having learned that two experts favour the proposition, 3 reverses his opinion (since he is highly respectful) and the proposition is unanimously favoured. However, if the roll-call vote is taken in order 3, 2, 1, incremental disclosure of preferences and high respect results in the proposition being unanimously opposed.... Disclosure order bias indicates that epistemic respect is trickier than has been supposed. (49-50)

Sorensen's epistemic bandwagon is a special case of what is known in the economics literature as an *information cascade*. An information cascade is a situation in which the preferences, predictions, decisions, etc., of agents are revealed sequentially, and agents acting later in a sequence follow patterns established earlier in the sequence, even when their private information would otherwise suggest deviating from the pattern. Economists have demonstrated information cascades in experiments with human subjects (Anderson and Holt 1997), and they have used the theory of information cascades to analyze phenomena including herd behavior in financial markets, momentum in political campaigns, and fads in medical practice (Bikhchandani et al. 1992). From this perspective, an epistemic bandwagon is an information cascade in which the actions performed sequentially are announcements of agent opinions and the cause of the cascade is epistemic respect among agents.

While Sorensen uses the notion of epistemic respect, a more standard notion is that of *epistemic trust*, understood as trust in another agent's judgment on the truth of a proposition. In general agents learn of each other's judgments via *testimony*, understood in the broad sense of "saying or affirming something in an apparent attempt to convey (correct) information" (Audi 1997, p. 405). A number of authors have stressed the importance in science and mathematics of epistemic trust in the testimony of others, not only between laypeople and experts but also among experts themselves (Hardwig 1985; 1991, Geist et al. 2010). In the multi-agent systems literature, formal models of epistemic trust have been developed using modal logic (Demolombe 2001; 2004, Liau 2003).

In this paper we introduce a *dynamic testimonial logic* (DTL) to model belief change over sequences of testimony among agents with different dispositions to trust each other as information sources. There are several motivations for the framework of DTL. First, DTL extends standard dynamic epistemic logic (DEL) (see van Ditmarsch et al. 2008) by explicitly representing *sources* of information, in such a way that the identity of a source affects how agents revise their beliefs in response to communication from the source. Second, DTL extends the DEL approach to belief revision (of van Benthem 2007) by representing the "missing link" between communication and belief revision, the *acceptance* of information. In DTL, whether an agent accepts information depends on whether the agent trusts the source. Third, while existing logics of epistemic trust are static logics, providing a snapshot view of the information and trust relations in a multi-agent setting, DTL is a dynamic logic, capable of modeling the role of trust in temporal phenomena such as information cascades. Finally, DTL models truly multi-agent belief revision, for it models not only how agents revise their beliefs about other agents' beliefs, but also how agents perform different types of belief revisions in response to the same informational event.

In the rest of Section 1 we provide the conceptual basis of DTL, drawing distinctions between *public announcement* and *testimony* and between the *doxastic* and *testimonial reliability* of agents. In Section 2 we review the logic of belief underlying DTL, the *conditional doxastic logic* of Baltag and Smets (2008). Turning to dynamics in Section 3, we review the approach to iterated belief revision of van Benthem's (2007) *dynamic logics of belief upgrade*. Given these foundations we introduce DTL in Section 4, adding *testimonial records* and *authority relations* to our models to capture agents' epistemic trust in the testimony of others. We then represent testimony dynamically by transformations on these enriched models. Finally, we describe an application of DTL in modeling epistemic bandwagons in Section 5, and we conclude with directions for further research in Section 6. The Appendix contains a complete axiomatization for DTL.

1.1 Trust and Authority in Testimony

Consider those experts on whose authority you would be willing to believe a proposition φ. We will say that you "trust the judgment" of these experts on φ. Among your trusted experts, some may be more authoritative for you than others. If expert 1 testifies that φ, expert 2 testifies that $\neg\varphi$, and you come to believe φ, then 1 is more authoritative for you than 2. If 2 were more authoritative, then you should have come to believe $\neg\varphi$. And if 1 and 2 were equally authoritative, you should not have changed your mind on φ either way, or you should have suspended judgment on φ altogether. The same points apply if 1 and 2 are groups of experts, rather than individuals.

If 1 is more authoritative for you than 2, then after 1 testifies that φ, you no longer trust 2 on $\neg\varphi$, in the sense that you will no longer believe $\neg\varphi$ on the authority of 2, something you might have done before 1 testified. However, if another expert 3 joins 2 in testifying that $\neg\varphi$, you may believe $\neg\varphi$ on the authority of 2 *together with* 3, though perhaps not on the authority of either of them individually. For our formalization we will assume that each agent has a ranking of the authority of other (sets of) agents, allowing incomparabilities (for how agents might rationally arrive at such rankings, see Goldman 2001).

Given our intuitive picture of trust and authority, our goal is to develop a model of testimony that addresses the following questions about what happens when an agent i testifies that φ. First, what do agents in i's audience come to believe about φ and related propositions? This question subdivides into two others: what determines whether agents in i's audience *accept* φ, and for those who accept φ, how do they revise their beliefs given this new information? Second, what do agents in i's audience come to believe *about i's beliefs*? Finally, what do agents come to believe *about the beliefs of other agents in i's audience*?

1.2 Testimony vs. Public Announcement

To identify the information provided by testimony, it is useful to compare testimony with *public announcement*, the classic case of an informational event in dynamic epistemic logic. For our purposes, the crucial difference between testimony and announcement is that while announcements are typically thought to come from an anonymous external source, testimony will always come from one of the agents within our model.

What difference does the individual source of testimony make? Let us make two assumptions about the kind of testimony in question. First, suppose testimony is *public*: the identity of the testifier and the content of the testimony is information available to all agents. Second, suppose testimony is heard under the *presumption of sincerity*: if an agent i testifies that φ, all other agents come to believe that i believes φ. As a consequence of these assumptions, when an agent i testifies that φ, other agents will acquire the information *that i believes* φ. But then what is the difference between a truthful *public announcement* that i believes φ and i's *own public testimony* that φ, if both provide the information that i believes φ?

One difference is that i's testimony provides *more* information:[1] it provides the information *that i is willing to publicly assert* φ. As we might say, i is willing to "go on the record" for φ. If i is the kind of agent who publicly asserts a proposition only if she has conducted a thorough inquiry into its truth, then the information that i is willing to publicly assert φ is vital information. A truthful public announcement (from no particular agent) that i believes φ does not provide this vital information. For it may be that i believes many propositions, while she only has the time and resources to investigate some few of them in such a way that she would be willing to make public assertions about them.

We can now make a distinction between two ways in which one agent might judge another to be "reliable" on the truth of a proposition. Agent j judges agent i to be *doxastically reliable* on φ just in case if j were to learn that i believes φ (and nothing stronger), then j would believe φ (cf. Demolombe 2004, on "competence"); j judges i to be *testimonially reliable* on φ just in case if j were to learn that i sincerely testified that φ, then j would believe φ. The important point is that judgments of doxastic and testimonial reliability may come apart. Suppose that i has expressed a general lack of understanding of economics. Then j might judge i's doxastic reliability on each economic proposition to be low. But suppose that j knows that i would publicly assert a proposition only if i had conducted a thorough inquiry into its truth. Then j might judge i's testimonial reliability on each economic proposition to be high; if i were to ever make a public assertion about economics, j would take it seriously.

As we have defined it, testimonial reliability concerns the reliability of an agent's *sincere* testimony. Yet an agent whose sincere testimony is highly reliable may often be insincere (cf. Cantwell 1998, p. 195). Perhaps i rarely says what she really believes about economics, due to peer pressure or controversy. Even so, when j is confident of i's sincerity, j might consider i's testimony to be highly reliable. We will see that the design of DTL reflects these distinctions.

2 Logic for Conditional Belief

In this section we define the logic of belief underlying DTL, the *conditional doxastic logic* (CDL) of Baltag and Smets (2008). CDL is equivalent to the

[1]At this point we are not considering the difference that public announcement is usually conceived as a source of "hard information" that *eliminates* possibilities, while testimony is better conceived as a source of "soft information" that reorders the relative plausibility of possibilities (see van Benthem 2007).

strongest logic for conditional belief introduced by Board (2004, Sec. 3.3), but we follow the presentation of Baltag and Smets. We focus on the semantics of CDL, referring the reader to the cited sources for axiomatizations.

Definition 2.1. Let At be a set of atomic sentence symbols and Agt a set of agent symbols. The language of CDL is defined by:

$$\varphi := p \mid \neg\varphi \mid \varphi \wedge \varphi \mid B_i^{\varphi}\varphi$$

where $p \in \mathsf{At}$, $i \in \mathsf{Agt}$.

We adopt the usual definitions of $\vee$, $\rightarrow$, and $\leftrightarrow$ in terms of $\neg$ and $\wedge$. The intended reading of $B_i^{\varphi}\psi$ is "i believes that ψ conditional on φ." Intuitively, the formula $B_i^{\varphi}\psi$ indicates that if i were to receive the information that φ (and nothing stronger), then i would believe ψ.[2]

For the semantics of CDL, we need some preliminary terminology. Where $\leq$ is a binary relation on a set W, a *comparability class* for $\leq$ is a set $C = \{w \in W \mid w \leq v \text{ or } v \leq w\}$ for some $v \in W$. The relation $\leq$ is a *well-preorder* on W if it is reflexive, transitive and every non-empty subset of W has a $\leq$-minimal element. Finally, $\leq$ is *locally well-preordered* on W if for each comparability class $C \subseteq W$, the restriction of $\leq$ to C is a well-preorder on C.

Definition 2.2. A *multi-agent plausibility model* is a triple $\mathcal{M} = \langle W, \leq, V\rangle$ where W is a non-empty set, $\leq$ is a family of locally well-preordered relations $\leq_i \subseteq W \times W$ for each $i \in \mathsf{Agt}$, and $V : \mathsf{At} \rightarrow \mathcal{P}(W)$.

As in epistemic logic, we think of each $w \in W$ as a possible state of the world (according to some agent), where agents may be uncertain about which is the *actual* state of the world. In CDL agents may also consider some states *more plausible* than others, as represented by the plausibility (pre-)ordering $\leq_i$ for each agent i. Following convention we read $w \leq_i v$ as "agent i considers state w at least as plausible as state v," so the *minimal* states in the ordering $\leq_i$ are the *most* plausible states for i. For comparability of states we write $w \sim_i v := w \leq_i v$ or $v \leq_i w$. Since $\sim_i$ is an equivalence relation it partitions W into equivalences classes, which for a given $w \in W$ we denote by $\sim_i(w) = \{v \in W \mid w \sim_i v\}$, called the *information cell* of w for i. Agents only compare states that they consider possible, so we take $\sim_i(w)$ to be the set of states that i considers possible according to her information at w. We read $w \sim_i v$ accordingly as "v is accessible for i from w." As usual, the valuation V sets the atomic facts at every state by mapping each $p \in \mathsf{At}$ to a set of states, (by the truth definition) the set of states satisfying p.

Definition 2.3. Given a model $\mathcal{M} = \langle W, \leq, V\rangle$ and state $w \in W$, we define $\mathcal{M}, w \models \varphi$ (φ is true in $\mathcal{M}$ at w) as follows:

- $\mathcal{M}, w \models p$ iff $w \in V(p)$
- $\mathcal{M}, w \models \neg\varphi$ iff $\mathcal{M}, w \not\models \varphi$
- $\mathcal{M}, w \models \varphi \wedge \psi$ iff $\mathcal{M}, w \models \varphi$ and $\mathcal{M}, w \models \psi$

[2] More accurately, i would believe that ψ *was* the case, before i received the information that φ. This qualification is necessary to make sense of satisfiable formulas such as $B_i^{\varphi}(\neg B_i\varphi \wedge \varphi)$.

- $\mathcal{M}, w \models B_i^{\varphi}\psi$ iff for all $v \in \min_{\leq_i}([\![\varphi]\!]_{\mathcal{M}} \cap \sim_i(w)) : \mathcal{M}, v \models \psi$

where we denote the set of most plausible states in $P \subseteq W$ by $\min_{\leq_i} P = \{v \in P \mid v \leq_i u \text{ for all } u \in P\}$ and the *truth set* of φ by $[\![\varphi]\!]_{\mathcal{M}} = \{u \in W \mid \mathcal{M}, u \models \varphi\}$. If the intended model is clear from the context, we write $[\![\varphi]\!]$ instead of $[\![\varphi]\!]_{\mathcal{M}}$.

If φ is true at a state w, we call w a "φ-state." Simply put, the truth definition for conditional belief says that i believes ψ conditional on φ iff all the most plausible φ-states for i are ψ-states. Given this definition, (unconditional) belief and knowledge are derived operators. Where $\top$ is any tautology and $\bot := \neg\top$, we define $B_i\varphi := B_i^{\top}\varphi$, read "$i$ believes φ," $K_i\varphi := B_i^{\neg\varphi}\bot$, read "$i$ knows φ," and $\hat{K}_i\varphi := \neg K_i\neg\varphi$, read "$i$ considers φ possible."

Multi-agent plausibility models contain information about what agents would believe upon learning various facts. They also contain information about what agents would believe upon learning about *other agents' beliefs*. Consider the formula $B_i^{B_k p} p \wedge B_i^{B_k \neg p} \neg p$ $(*)$, which is true if and only if in all the $B_k p$-states that i considers most plausible, k's belief is true, and likewise for the most plausible $B_k\neg p$-states. Intuitively, $(*)$ expresses that i takes k to be *doxastically reliable* on p, in the sense of Section 1.2. Various judgments of doxastic unreliability can be expressed in a similar way. We can even extend this observation to agents' beliefs about the relative doxastic reliability of other agents. For example, a formula such as $B_i^{B_j p \wedge B_k \neg p} p \wedge B_i^{B_j \neg p \wedge B_k p} \neg p$, which is consistent with $(*)$, expresses i's belief in the superior doxastic reliability of j relative to k on p.

Although multi-agent plausibility models contain information about agents' views of the doxastic reliability of other agents, they do not contain information about agents' views of the *testimonial* reliability of other agents. If we were to assimilate k's testimony that φ to a public announcement of $B_k\varphi$, then these models would be sufficiently rich to determine how agents' beliefs change in response to this "testimony." However, as discussed in Section 1.2, k's testimony that φ is not equivalent to a public announcement of $B_k\varphi$, in terms of the information provided. For this reason we will add additional structure to models for DTL in Section 4.

3 Logics for Iterated Belief Revision

Having defined the underlying logic of belief for DTL, we turn to the dynamics. To model belief revision, we follow the approach of van Benthem's (2007) *dynamic logics of belief upgrade*. These logics provide a formalization not only of "one-shot" belief revision, as CDL does, but also of two well-known *iterated* belief revision policies. Moreover, the same style of analysis can be used to provide complete logics for other iterated belief revision policies. Given the versatility of this approach, one is free to choose one's preferred belief revision policies and use the corresponding logic as a dynamic base for DTL.

Definition 3.1. Let At be a set of atomic sentence symbols and Agt a set of agent symbols. The language of belief upgrade is defined by:

$$\varphi := p \mid \neg\varphi \mid \varphi \wedge \varphi \mid B_i^{\varphi}\varphi \mid [\uparrow_i \varphi]\varphi$$

where $p \in \mathsf{At}$, $i \in \mathsf{Agt}$.

The intended reading of $[\uparrow_i \varphi]\,\psi$ is "after the revision of i's beliefs with the new information that φ, ψ is the case," or more concisely, "after i upgrades with φ, ψ is the case." While van Benthem uses the symbol $\uparrow$ for conservative upgrade in particular, we use $\uparrow$ for an arbitrary belief upgrade action, and we define three particular upgrade actions below.

Models for the logic of belief upgrade are the same multi-agent plausibility models as before.

Definition 3.2. Given a model $\mathcal{M} = \langle W, \leq, V\rangle$, the model $\mathcal{M} \uparrow_i \theta = \left\langle W, \leq^{\uparrow_i\theta}, V\right\rangle$ is obtained by changing the plausibility ordering $\leq_i$ to $\leq_i^{\uparrow\theta}$ as follows. *Conservative upgrade* (Boutilier 1996): in each information cell for i, the most plausible θ-states become most plausible overall, but otherwise the ordering remains the same. *Restrained upgrade* (Booth and Meyer 2006): in each information cell for i, each set of equi-plausible states is split, such that all θ-states in the set become more plausible than all $\neg\theta$-states in the set, but otherwise the ordering remains the same; then the most plausible θ-states become most plausible overall, but otherwise the ordering remains the same. *Lexicographic upgrade* (Nayak 1994): in each information cell for i, all θ-states become more plausible than all $\neg\theta$-states, but otherwise the ordering remains the same.

Definition 3.3. The truth definition for static formulas is that of CDL. The truth definition for belief upgrade is:

- $\mathcal{M}, w \vDash [\uparrow_i \theta]\,\varphi$ iff $\mathcal{M} \uparrow_i \theta, w \vDash \varphi$

According to the definition, to determine whether $[\uparrow_i \theta]\varphi$ is true at w in the initial model we simply check whether φ is true at w in the updated model.

For each of the three upgrade operations defined, one can give *reduction axioms* that allow the rewriting of any formula with upgrade operators as an equivalent formula in the static language of CDL. Given a complete axiomatization for CDL, these reduction axioms provide a complete axiomatization for the dynamic logics of upgrade. Van Benthem (2007) gives reduction axioms for lexicographic and conservative upgrade.[3] The analogous result for restrained upgrade, which we leave as an exercise for the reader, also holds.

In modeling belief change due to testimony, we wish to model not only how agents form new beliefs, but also how agents *suspend belief*. Suppose that agents j and k are equally authoritative in the eyes of i. If j testifies that φ and then k testifies that $\neg\varphi$, one policy for i would be to believe whoever testified first—in this case, agent j. A more sensible policy in a situation where equally authoritative sources conflict would be to suspend belief about φ. Alternatively, i might not perform any belief revision, ignoring the conflicting testimony of j and k. Yet conflicting testimony from *authoritative* sources does not seem to "cancel out" to provide i with no information. Something informative has occurred—two authoritative sources have testified for φ and $\neg\varphi$ respectively—and i's beliefs should reflect this.

Let us add a *suspension operator* $[\downarrow_i \varphi]$ to the language of the previous section. The intended reading of $[\downarrow_i \varphi]\,\psi$ is "after i suspends belief on φ, ψ is the case."

[3] In the multi-agent case, we must add the reduction axiom $[\uparrow_i \theta]\, B_j^{\varphi}\psi \leftrightarrow B_j^{[\uparrow_i\theta]\varphi}\,[\uparrow_i \theta]\,\psi$ for $j \neq i$ to those given by van Benthem.

In the belief revision literature the term "contraction" is more standard, but we prefer "suspension" for its suggestiveness in relation to testimony.

Definition 3.4. Given a model $\mathcal{M} = \langle W, \leq, V\rangle$, the model $\mathcal{M} \downarrow_i \theta = \left\langle W, \leq^{\downarrow_i \theta}, V\right\rangle$ is obtained by changing the plausibility ordering $\leq_i$ to $\leq_i^{\downarrow\theta}$ as follows. *Conservative suspension* (Ramachandran et al. 2009): in each information cell for i, the most plausible states and the most plausible $\neg\theta$-states becomes equally plausible and most plausible overall, but otherwise the old ordering remains the same.

Conservative suspension is the suspension counterpart of conservative upgrade. The pair has several natural properties, one of which is a kind of *confluence*: conservative upgrade with φ followed by conservative suspension with φ produces the same result as conservative upgrade with $\neg\varphi$ followed by conservative suspension with $\neg\varphi$. In our example above, this means that the order in which j and k testify does not make a difference to what i believes after she suspends belief on φ (or $\neg\varphi$), which seems intuitive. Other suspension policies are possible, but we will not consider them here.

Definition 3.5. The truth definition for belief suspension is:

- $\mathcal{M}, w \vDash [\downarrow_i \theta]\, \varphi$ iff $\mathcal{M} \downarrow_i \theta, w \vDash \varphi$

Reduction axioms for conservative suspension are easily obtained by analogy with those for conservative upgrade. A generalization of conservative suspension is the *multi-upgrade* operation $\downarrow_i \{\varphi_1, ..., \varphi_n\}$, defined as follows.

Definition 3.6. Given a model $\mathcal{M} = \langle W, \leq, V\rangle$, the model $\mathcal{M} \downarrow_i \{\theta_1, ..., \theta_n\} = \left\langle W, \leq^{\downarrow_i\{\theta_1,...,\theta_n\}}, V\right\rangle$ is obtained by changing the plausibility ordering $\leq_i$ to $\leq_i^{\downarrow\{\theta_1,...,\theta_n\}}$ as follows: in each information cell for i, the most plausible θ_k-states for each k ($1 \leq k \leq n$) become equally plausible and most plausible overall, but otherwise the ordering remains the same.

The advantage of multi-upgrade over simple (conservative) suspension is that with multi-upgrade an agent can suspend belief on θ while making sure to retain or gain belief in φ. The multi-upgrade $\downarrow_i \{\varphi \wedge \theta, \varphi \wedge \neg\theta\}$ accomplishes the desired effect,[4] for which we will find a use in the next section. Given the reduction axioms for conservative suspension, the reduction axioms for multi-upgrade are a straightforward generalization.

We end our discussion of upgrade and suspension by noting that in the multi-agent setting, belief revision via relation change is a kind of *public* belief revision; when one agent's plausibility relation changes, other agents may "notice" the change. To be precise, the following is a validity:

$$\vDash B_j\hat{K}_i\varphi \leftrightarrow [\uparrow_i \varphi]\, B_jB_i\varphi$$

We return to the issue of the publicity of belief revision in Section 4.1 below.

[4]Multi-upgrade accomplishes the desired effect provided there are states satisfying $\varphi \wedge \theta$ and states satisfying $\varphi \wedge \neg\theta$ in the information cell of interest. If there are no states satisfying $\varphi \wedge \neg\theta$, for example, then the multi-upgrade $\downarrow_i \{\varphi \wedge \theta, \varphi \wedge \neg\theta\}$ amounts to a conservative upgrade by $\varphi \wedge \theta$. To prevent this, one may wish to define multi-upgrade so that if there are no θ_k-states for some k ($1 \leq k \leq n$) in a given information cell, then the operation $\downarrow_i \{\theta_1, ..., \theta_n\}$ does nothing to that information cell. We will not require this safeguard, however, for we will only use multi-upgrade when it is guaranteed that θ_k-states exists for each k (see the definition of $\updownarrow\updownarrow_j^i\, \varphi$ in Section 4.4.2).

4 Dynamic Testimonial Logic

In this section we develop the framework of DTL.

4.1 Language of DTL

Definition 4.1. Let At be a set of atomic sentence symbols and Agt a *finite* set of agent symbols. The language of DTL is defined by:

$$\begin{array}{lll} \varphi_0 & := & p \mid \neg\varphi_0 \mid \varphi_0 \wedge \varphi_0 \\ \varphi & := & \varphi_0 \mid \neg\varphi \mid \varphi \wedge \varphi \mid U\varphi_0 \mid B_i^{\varphi}\varphi \mid \mathtt{rec}_i\varphi_0 \mid S \preceq_i^{\varphi_0} S' \mid [!_i\varphi_0]\varphi \end{array}$$

where $p \in \mathsf{At}$, $i \in \mathsf{Agt}$, and $S, S' \subseteq \mathsf{Agt}$.

The language of DTL includes several new types of formulas. The intended reading of $\mathtt{rec}_i\varphi$ is "i is (most recently) on the record as testifying in favor of φ." The intended reading of $S \preceq_i^{\varphi} S'$ is "S' is as (testimonially) authoritative as S on φ for i." We use the abbreviations $S \prec_i^{\varphi} S' := S \preceq_i^{\varphi} S' \wedge S' \not\preceq_i^{\varphi} S$ and $S \approx_i^{\varphi} S' := S \preceq_i^{\varphi} S' \wedge S' \preceq_i^{\varphi} S$. We also allow $\emptyset$ to occur in formulas: we read $\emptyset \prec_i^{\varphi} S$ as "S's testimony on φ is *authoritative* for i," $\emptyset \approx_i^{\varphi} S$ as "S's testimony on φ is *unauthoritative* for i," and $S \prec_i^{\varphi} \emptyset$ as "S's testimony on φ is *anti-authoritative* for i." The reason for this choice of terminology will be clear when we turn to the semantics. Finally, the intended reading of $[!_i\varphi]\psi$ is "after i (publicly) testifies that φ, ψ is the case." For simplicity in this version of DTL we consider only testimony on propositional formulas, so agents do not testify about the beliefs or authority of others. It is for this reason that only φ_0 formulas can appear inside testimony operators and in record and authority formulas. Finally, we have added the universal modality U for the purposes of our axiomatization, so we can express the equivalence of two φ_0 formulas in a model (see Appendix).

4.2 Semantics of DTL

Definition 4.2. A *testimonial model* $\mathcal{M} = \langle W, \leq, V, \mathsf{rec}, \preceq \rangle$ is a multi-agent plausibility model together with $\mathsf{rec} : \mathsf{Agt} \times W \to \mathcal{P}(\mathcal{P}(W))$ and a family $\preceq$ of relations $\preceq_{i,w}^{P} \subseteq \mathcal{P}(\mathsf{Agt}) \times \mathcal{P}(\mathsf{Agt})$ for each $i \in \mathsf{Agt}$, $w \in W$, and $P \subseteq W$.

The *testimonial record* rec records the set of propositions $\mathsf{rec}_i(w)$ to which i has testified at w, where a proposition is now understood as a *set of states* $P \subseteq W$, not a formula. The intuition is that when i testifies that φ, she is claiming that the actual state is among the φ-states. Hence she goes on the record for $[\![\varphi]\!]$ (at every state w, assuming the testimony is public).[5] The *authority relation* $\preceq_{i,w}^{P}$ encodes i's view at w of the relative authority of (sets of) agents on the proposition P. Since the authority relations are not necessarily total, we do not assume that all (sets of) agents are comparable in authority for an agent.

[5]This reflects our semantic perspective, from which we ignore syntactic differences of formulas that pick out the same proposition. We are not taking "on the record" in a literal sense, as a matter of *what the testifier said*. If we wished to keep track of such linguistic matters, we would have opted for a syntactic approach whereby agents go on the record for formulas rather than propositions.

Definition 4.3. A testimonial model is *legal* iff for every $i \in \mathsf{Agt}$, $S, S' \subseteq \mathsf{Agt}$, $w, v \in W$, and $P \subseteq W$:

1. $\preceq^P_{i,w}$ is a preorder on P.
2. $S \preceq^P_{i,w} S'$ iff $S \preceq^{W \setminus P}_{i,w} S'$.
3. If $S \preceq^P_{i,w} S'$ $(S \neq S')$, then there is a $v \in \sim_i (w)$ such that $v \in P$ and $\sim_k (v) \cap (W \setminus P) \neq \emptyset$ for all $k \in S$.
4. If $w \sim_i v$, then $\preceq^P_{i,w} = \preceq^P_{i,v}$ and $\mathsf{rec}_i(w) = \mathsf{rec}_i(v)$.

The first condition reflects the assumption that the relation of being *as authoritative as* is reflexive and transitive. The second condition states that authority relations are the same for a proposition and its complement, e.g., i considers j authoritative on whether there will be a recession next year if and only if i considers j authoritative on whether there will *not* be a recession next year. This condition could be dropped, at the expense of complicating the system.[6]

For the third condition, suppose that P is the truth set of some formula φ. Then the condition implies that if there is a group S' that is at least as authoritative as S on φ for i (at w), then i must consider it possible (at w) for the agents in S to sincerely testify for or against φ and yet be mistaken. That is, there must be an i-accessible state at which φ is true but no one in S knows φ and (by the second condition) an i-accessible state at which φ is false but no one in S knows $\neg\varphi$. For if the only i-accessible states at which φ is true (resp. false) are ones at which someone in S knows φ (resp. $\neg\varphi$), then i considers it impossible for the members of S to all sincerely testify against (resp. for) φ and yet be mistaken. But i must consider this possible if there is some S' that is as authoritative or more authoritative on φ than S. In the special case of $S = \emptyset$, the second condition implies that if there is a set S' of agents that is authoritative on φ for i (at w), then i must consider φ possible (at w).[7]

Finally, the fourth condition states that agents have knowledge of their own authority relations and testimonial records, reflecting the assumption that agents have introspective access to their views of other agents and memory of their own past testimony. Additional conditions on authority relations, which for the sake of generality we will not assume as part of the framework of

[6]Without the second condition, we would have to give $S \preceq^{\varphi}_i S'$ a more complicated reading than"S' is as authoritative as S on φ for i," since both $S \preceq^{\varphi}_i S'$ and $S' \preceq^{\neg\varphi}_i S$ could be true at the same time. One reason to drop the condition is that the principle that j is authoritative on φ for i if and only if j is authoritative on $\neg\varphi$ for i is subject to counterexamples. For example, since studies have shown that people tend to overestimate their driving ability relative to the average driver, if i knows this fact then i might consider j authoritative on whether j is *not* a very good driver but not consider j authoritative on whether j is a very good driver. However, in the case of *expert* testimony, the symmetric authority of the second condition seems plausible (cf. Liau 2003, on "symmetric trust").

[7]Given the second condition, the third condition also implies that if i "knows" φ, then i does not consider any set of agents authoritative on φ. For if i considers S authoritative on φ, then by the second condition i considers S authoritative on $\neg\varphi$, whence by the third condition there is an i-accessible state satisfying $\neg\varphi$, so i does not know φ. We could avoid this consequence by dropping the second condition, but it would not be worth the loss of simplicity in our system. Since we are not interested in agents who have already made up their minds about propositions, but rather in agents who come to believe or disbelieve propositions on the basis of testimony, it is not necessary to express that an agent may know φ and yet still consider others authoritative on φ.

DTL, may be desirable in certain modeling situations. For example, one might consider conditions on the authority of related sets of agents (cf. Cantwell 1998, p. 194). Possibilities include a uniformity condition of the form $S \preceq^P_{i,w} S' \Rightarrow S \cup S'' \preceq^P_{i,w} S' \cup S''$ and various right and left monotonicity conditions, such as $S \preceq^P_{i,w} S \cup S'$, $\emptyset \preceq^P_{i,w} S' \Rightarrow S \preceq^P_{i,w} S \cup S'$, and $S' \preceq^P_{i,w} \emptyset \Rightarrow S \cup S' \preceq^P_{i,w} S$. One might also consider conditions that connect authority relations for different propositions, e.g., $\left(S \preceq^P_{i,w} S' \text{ and } S \preceq^Q_{i,w} S'\right) \Rightarrow S \preceq^{P\cap Q}_{i,w} S'$ (cf. Liau 2003, p. 37). We leave such questions about the "logic of authority" aside in what follows.

Definition 4.4. The truth definition for the static part of DTL is:

- $\mathcal{M}, w \vDash U\varphi$ iff for all $v \in W$: $\mathcal{M}, v \vDash \varphi$
- $\mathcal{M}, w \vDash \texttt{rec}_i\varphi$ iff $[\![\varphi]\!] \in \mathsf{rec}_i(w)$
- $\mathcal{M}, w \vDash S \preceq^\varphi_i S'$ iff $S \preceq^{[\![\varphi]\!]}_{i,w} S'$

4.3 Defining Trust with Record and Authority

In DTL it is trust that determines whether one agent accepts the testimony of another. In this section we briefly discuss how trust may be defined in terms of the testimonial record and authority relations. Since the aim of this paper is to provide a general framework for modeling the role of trust in testimony, we will not fix "the" definition of trust in DTL. Rather, we will indicate one possibility among many. The question of how to best define trust in terms of the record and authority deserves a full treatment on its own (cf. Cantwell 1998).

Suppose that we have defined what it is for an agent to have testified *for* a proposition, $\mathsf{for}_i\varphi$, and *against* a proposition, $\mathsf{ags}_i\varphi$. Whether an agent is for or against a proposition will depend on which propositions she has testified to, according to the testimonial record, but let us leave the definitions open for a moment. Where $\boldsymbol{A}$ is the set of sequences of sets $\langle X_1, X_2, X_3\rangle$ such that $\{X_n \mid X_n \neq \emptyset\}$ is a *partition* of Agt, we might define a *trust formula* $T_{ji}\varphi$, read "j trusts the testimony of i on φ," as follows:

$$\alpha \wedge \bigvee_{\langle X,Y,Z\rangle\in \boldsymbol{A}} \Big(\bigwedge_{x\in X} K_j\texttt{for}_x\varphi \wedge \bigwedge_{y\in Y} K_j\texttt{ags}_y\varphi \wedge \bigwedge_{z\in Z}\neg\left(K_j\texttt{for}_z\varphi \vee K_j\texttt{ags}_z\varphi\right)\wedge\beta\Big)$$

where α and β are parameters. The parameter α determines the extent to which j must judge i doxastically reliable in order for j to trust i's testimony. The discussion in Section 1.2 suggest that only a minimal assumption of i's doxastic reliability is necessary, such as $\alpha := \hat{K}_jB_i\varphi \rightarrow \hat{K}_j(B_i\varphi \wedge \varphi)$. The parameter β determines the kind of trust defined. For example, for *weak trust* set $\beta := Y \setminus \{i\} \prec^\varphi_j X \cup \{i\}$. An agent *j weakly trusts* i's testimony on φ just in case if i were to join the group of agents whom j knows are *for* φ (and leave the group of agents whom j knows are *against*), then j would consider the group of agents who are for φ to be more authoritative on φ than the group of agents who are against. Stronger definitions of trust, which require additionally that j take i to be individually authoritative on φ, are also possible.

In addition to $T_{ji}\varphi$, we can define distrust $D_{ji}\varphi$, read"*j distrusts* the testimony of i on φ," and $A_{ji}\varphi$, read "j is *ambivalent* about i's testimony on φ," by changing α and β in the trust formula. For distrust set $\alpha := \hat{K}_j B_i\varphi \rightarrow \hat{K}_j(B_i\varphi \wedge \neg\varphi)$ and $\beta := X \cup \{i\} <^{\varphi}_{j} Y \setminus \{i\}$. For ambivalence set $\alpha := \hat{K}_j B_i\varphi \rightarrow \left(\hat{K}_j(B_i\varphi \wedge \varphi) \wedge \hat{K}_j(B_i\varphi \wedge \neg\varphi)\right)$ and $\beta := Y \setminus \{i\} \approx^{\varphi}_{j} X \cup \{i\} \wedge \neg\left(X \cup \{i\} \approx^{\varphi}_{j} \emptyset\right)$. The intuitions behind these definitions are easily grasped by analogy with the definition of trust. Note that the sense of "distrust" here is distrust in the judgment of i, which does not imply that j doubts the sincerity of i in testifying on φ.

Turning to $\mathsf{for}_i\varphi$ and $\mathsf{ags}_i\varphi$, the simplest option is to define $\mathsf{for}_i\varphi := \mathsf{rec}_i\varphi \wedge \neg\mathsf{rec}_i\neg\varphi$ and $\mathsf{ags}_i\varphi := \mathsf{rec}_i\neg\varphi \wedge \neg\mathsf{rec}_i\varphi$. Defining trust in this way, which we might call *narrow trust*, is sufficient for modeling "single-issue" testimonial sequences in which agents either testify for a single proposition or for its complement (or pass). For modeling sequences in which agents testify on multiple, related propositions, we may wish to consider a *wide trust* that depends on the authority not only of those who have testified that φ and those who have testified that $\neg\varphi$, but also of those who have testified that $\psi \wedge \varphi$, or ψ and $\psi \rightarrow \neg\varphi$, etc. In the interest of space, we leave aside a discussion of such wide trust here.

In addition to changing the definition of $\mathsf{for}_i\varphi$ and $\mathsf{ags}_i\varphi$, we might change the structure of the trust formula $T_{ji}\varphi$ itself in various ways. First, the operator B_j might be used instead of K_j in $T_{ji\varphi}$, so that j considers the authority not only of those whom she knows to be for or against φ, but also of those whom she believes to be for or against φ. However, for public testimony this makes little difference, since (as we will see in the next section) whenever an agent testifies that φ, all other agents come to *know* that the testifier is on the record for φ. Second, as we have defined $T_{ji}\varphi$, whether j trust i on φ depends only on the authority of i and the authority of those *on the record* for various propositions. Other policies are possible. For example, j may choose not to consider k's authority in favor of φ if although k testified that φ, j believes that k no longer believes φ.[8] Third and finally, as we have defined $T_{ji}\varphi$, j does not count his own authority in favor of or against φ. Yet we may wish to define $T_{ji}\varphi$ in such a way that whether j trusts i on φ depends on, among other things, j's current belief concerning φ and j's view of his own authority relative to others.

In what follows, we will assume that a definition of trust is given in terms of the testimonial record and authority relations. Our proposal for the dynamics of testimony does not depend on the details of the trust definition.

4.4 The Dynamics of Testimony

In this section, we define model transformations induced by the action of public testimony. The motivating idea is that if an agent j trusts another agent i on φ, then after i testifies that φ, j should come to believe φ. Moreover, if there is a *presumption of sincerity* in testimony, j should also come to believe that i believes φ. We begin by showing how to model the first part, j's belief revision concerning φ, by itself. We then show how to add the presumption of sincerity.

[8] By counting the authority of those agents who have testified that φ but who (j believes) no longer believe φ, one assumes that j judges the testimonial reliability of k in terms of how reliable k's sincere testimony has been in the past, even in cases where (j believes) k later gave up the belief that the testimony expressed. We could distinguish such pure testimonial reliability from a hybrid testimonial-doxastic reliability, judged by the reliability of an agent's testimony in just those cases in which the agent retained the belief that the earlier testimony expressed.

Throughout we let $\uparrow_i$ and $\downarrow_i$ stand for arbitrary belief upgrade and suspension operators respectively. Hence when we define model transformations below, we will actually be defining classes of model transformations, members of which differ with respect to the particular belief change operations used.

From Global to Local Belief Upgrades

In the simplest model of testimony, after an agent i testifies that φ, each agent j who trusts i's testimony on φ performs the belief upgrade $\uparrow_j \varphi$. Each agent who distrust i's testimony performs the upgrade $\uparrow_j \neg\varphi$, and each agent who is ambivalent about i's testimony performs the suspension $\downarrow_j \varphi$. The problem with this proposal is that $T_{ji}\varphi$ may be true at some states in the model and false at others—similarly for $D_{ji}\varphi$ and $A_{ji}\varphi$—reflecting other agents' uncertainty about j's attitude toward i. Suppose that $T_{ji}\varphi$ is true at the state at which we are evaluating formulas, so after i testifies that φ, j upgrades with $\uparrow_j \varphi$. Since belief upgrades work *globally* on the model, after the upgrade, $B_j\varphi$ may come to be true at a state in the model at which $T_{ji}\varphi$ is *false*. Hence agents who believe that j does *not* trust i on φ may nonetheless come to believe that j believes φ after i's testimony, a counterintuitive result.

The solution to this problem is to make global belief upgrades act locally. There is a technique for doing so, given by the following definition and lemma.

Definition 4.5. A formula φ is *introspectible* for an agent i in a model $\mathcal{M}$ iff for every information cell $C \subseteq W$ for i, $[\![\varphi]\!] \cap C \neq \emptyset \Rightarrow C \subseteq [\![\varphi]\!]$.

The introspectible formulas for i in $\mathcal{M}$ are those φ such that at any state in $\mathcal{M}$, if i considers φ possible, then i knows φ. Examples (for any model) include belief and knowledge formulas $B_i\psi$ and $K_i\psi$ and trust formulas $T_{ik}\psi$.

Lemma 1 (Localization)**.** Let $\mathcal{M}$ be a multi-agent plausibility model, $\psi_1, .., \psi_n$ a sequence of formulas, and $\varphi_1, ..., \varphi_n$ a sequence of introspectible formulas for i in $\mathcal{M}$ such that $[\![\varphi_j]\!] \cap [\![\varphi_k]\!] = \emptyset$ for $j \neq k$. Then there is a formula χ such that for every information cell $C \subseteq W$ for i with $[\![\varphi_k]\!] \cap C \neq \emptyset$, it holds that $\leq_i^{\mathcal{M}\uparrow_i\chi} \restriction C = \leq_i^{\mathcal{M}\uparrow_i\psi_k} \restriction C$.

Hence the effect of a single upgrade with the formula χ is that each comparability class containing a point that satisfies one of the φ_k is reordered locally just as it would be by a (global) upgrade by ψ_k. Intuitively, χ "localizes" revision by ψ_k to those parts of the model "targeted" by φ_k.

Proof. Take $\chi := \bigvee_{1\leq k\leq n} (\varphi_k \wedge \psi_k)$. Suppose C is an information cell for i with $[\![\varphi_k]\!] \cap C \neq \emptyset$. Then since φ_k is introspectible, $C \subseteq [\![\varphi_k]\!]$. It follows by the assumption of the lemma that $[\![\varphi_j]\!] \cap C = \emptyset$ for $j \neq k$. Hence $[\![\chi]\!] \cap C = [\![\varphi_k \wedge \psi_k]\!] \cap C = [\![\varphi_k]\!] \cap [\![\psi_k]\!] \cap C$. Given $C \subseteq [\![\varphi_k]\!]$ we also have $[\![\varphi_k]\!] \cap [\![\psi_k]\!] \cap C = [\![\psi_k]\!] \cap C$. Therefore $[\![\chi]\!] \cap C = [\![\psi_k]\!] \cap C$, which gives $\leq_i^{\mathcal{M}\uparrow_i\chi} \restriction C = \leq_i^{\mathcal{M}\uparrow_i\psi_k} \restriction C$. □

The following application of the Localization Lemma shows the utility of making a number of different local changes to a model with one global upgrade.

Definition 4.6 (Testimonial Upgrade)**.** Let $\updownarrow_j^i \varphi$ be the operation that performs the following sequence of relation changes:

$$\uparrow_j \left(T_{ji}\varphi \wedge \varphi\right) \vee \left(D_{ji}\varphi \wedge \neg\varphi\right), \downarrow_j A_{ji}\varphi \wedge \varphi$$

We propose that after i testifies that φ, j's plausibility ordering should change according to $\updownarrow^i_j\ \varphi$.[9] To understand the effect of the operation $\updownarrow^i_j\ \varphi$, note that each state in a testimonial model satisfies at most one of $T_{ji}\varphi$, $D_{ji}\varphi$, and $A_{ji}\varphi$, as these are mutually exclusive. Moreover, in a *legal* model, the states in any given information cell for j must all agree on which one of these formulas they satisfy, by the fourth condition of legality. Hence in any given information cell for j in which $T_{ji}\varphi$ is true, the above upgrade sequence will have the same effect *as if j upgraded with $\uparrow_j \varphi$ alone*, since all the states in the information cell satisfy $T_{ji}\varphi$ and none satisfy $D_{ji}\varphi$ or $A_{ji}\varphi$. Similarly, in any given information cell for j in which $D_{ji}\varphi$ is true, the upgrades will have the same effect as if j upgraded with $\uparrow_j \neg\varphi$ alone, and so on. The localization technique allows us in effect to do *different belief revision* for j in *different parts of the model*, depending on j's attitude toward i (trust, distrust, etc.) in different parts of the model. Hence if another agent k is uncertain about whether or not j trusts i, then after i's testimony, k will be uncertain about which belief revision j performed. We illustrate this phenomenon in the course of modeling the bandwagon effect in Section 5.

We are now ready to define the model transformation for public testimony with *no presumption of sincerity* (NPS).

Definition 4.7 (Public Testimony with NPS)**.** Given a testimonial model $\mathcal{M} = \langle W, \leq, V, \mathsf{rec}, \preceq\rangle$, the model $\mathcal{M}!_i\varphi = \left\langle W, \leq^{!_i\varphi}, V, \mathsf{rec}^{!_i\varphi}, \preceq\right\rangle$ is defined as follows: $\mathsf{rec}^{!_i\varphi}_j(w) = \mathsf{rec}_j(w)$ for $j \neq i$, $\mathsf{rec}^{!_i\varphi}_i(w) = (\mathsf{rec}_i(w) \setminus \{W \setminus [\![\varphi]\!]\}) \cup \{[\![\varphi]\!]\}$, and $\leq^{!_i\varphi}$ is obtained from $\leq$ by the sequence of operations $\updownarrow^i_{j\in\mathsf{Agt}\setminus\{i\}}\ \varphi$.

The definition states that when i testifies that φ, we put i on the record *at each state w* for $[\![\varphi]\!]$, reflecting the public nature of the testimony. Moreover, we take i off the record for $W \setminus [\![\varphi]\!]$, reflecting the assumption that by testifying that φ, i is implicitly retracting past testimony against φ. We could also define a general *retraction* operation allowing agents to retract past testimony for some ψ without having to go on the record for $W \setminus [\![\psi]\!]$, but we will not do so here.

The notation $\updownarrow^i_{j\in\mathsf{Agt}\setminus\{i\}}\ \varphi$ indicates that every agent j other than i performs the individual testimonial upgrade $\updownarrow^i_j\ \varphi$. (Since the agents are not upgrading by doxastic formulas, the order in which the upgrades occur does not matter.) Note that if $\mathcal{M}$ is legal, then the updated model $\mathcal{M}!_i\varphi$ is also legal, since belief upgrade and suspension do not change the relations $\preceq_j$ or $\preceq^P_{j,w}$ and since the operation $!_i$ makes the same changes to $\mathsf{rec}_j(w)$ and $\mathsf{rec}_j(v)$ for $w \sim_j v$.[10]

Definition 4.8. The truth definition for the testimony operator is:

[9]Note that the definition of $\updownarrow^i_j\ \varphi$ is not supposed to reflect j's *mental representation* of the information given by i's testimony; j need not think of the information received from i's testimony as "either I trust j on φ and φ, or...." The point is rather to identify a change to j's plausibility ordering that produces an intuitively correct effect on the testimonial model, for it is the testimonial model that reflects the epistemic situation of j after i's testimony.

[10]Since it is permitted in a testimonial model for $\mathsf{rec}_j(w)$ and $\mathsf{rec}_j(v)$ to differ for $w \nsim_j v$, we might also consider testimony that is less than fully public, which leaves some agents uncertain about the content of the testimony. This would require more complex model transformations, and private testimony "behind closed doors" may require changes to the definition of a testimonial model (cf. van Ditmarsch et al. 2008, Sec. 6.9).

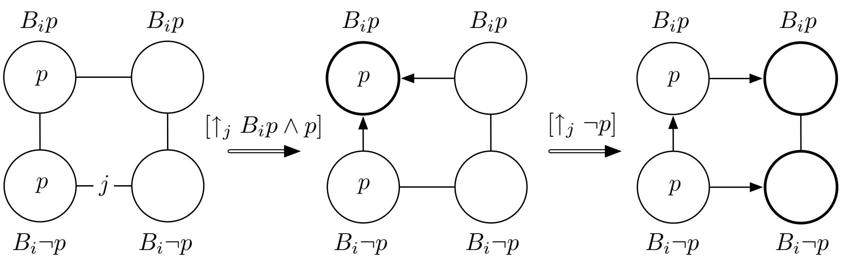


Figure 1: A failure to preserve belief in the testifier's sincerity

- $\mathcal{M}, w \models [!_i\varphi]\, \psi$ iff $\mathcal{M}, w \models$ `pre` implies $\mathcal{M}!_i\varphi, w \models \psi$

where the precondition `pre` $:= B_i\varphi$ with the *assumption of sincerity* (AS) and `pre` $:= \top$ without. In addition to the box modality $[!_i\varphi]$, we can define a dual diamond modality $\langle !_i\varphi \rangle$ by replacing 'implies' by 'and' in the definition above.

Modeling the Presumption of Sincerity

Given a *presumption of sincerity* (PS) in testimony, if *j* trusts *i* on φ and *i* testifies that φ, *j* should come to believe not only that φ but also that *i* believes φ. Our first question is whether we should model *j*'s belief revision in a single step, in which *j* performs the upgrade $\uparrow_j B_i\varphi \wedge \varphi$, or in two steps, in which *j* first performs $\uparrow_j B_i\varphi$ and then $\uparrow_j \varphi$ or *vice versa*. Technically, we could model the revision either way. Conceptually, it seems preferable to model it in two steps.

Suppose that agent *j* receives testimony from *i* that φ, followed by testimony from *k* that $\neg\varphi$. While *j* considers *i* authoritative on φ, *j* considers *k* still more authoritative. However, *i* does not consider *k* authoritative at all, and *j* knows this. Intuitively, after the testimonies *j* should believe that *i* believes φ, that *k* believes $\neg\varphi$, and that $\neg\varphi$. However, if we model *j*'s first belief revision with the single upgrade $\uparrow_j B_i\varphi \wedge \varphi$, we may not obtain the desired result. Instead, after *k*'s testimony *j* may lose his belief that $B_i\varphi$, formed given his presumption of *i*'s sincerity, even though *j knows* that *i* does not trust *k* on $\neg\varphi$. Figure 1[11] illustrates how this counterintuitive result may occur. To simplify, we do not draw arrows for *i* or represent *k* in the model at all. However, a more complex model with *k* represented would give the same result. Note that the loss of *j*'s belief in $B_i\varphi$ after the second upgrade is independent of the kind of belief upgrades used.

If we wish to model *j*'s belief revision after *i*'s testimony in two steps, the obvious candidates are the sequences of upgrades $\uparrow_j B_i\varphi, \uparrow_j \varphi$ and $\uparrow_j \varphi, \uparrow_j B_i\varphi$. Both sequences avoid the problem of *j* too easily losing his belief that $B_i\varphi$.[12] However, both sequences also violate an intuitive constraint on *j*'s

[11] In the following Figures, circles represent states and lines represent plausibility orderings, labelled for each agent, with arrows pointing toward more plausible states and arrowless lines indicating equi-plausibility. Every state is equi-plausible with itself for each agent, but we omit the reflexive loops, as well as arrows implied by transitivity. Atomic sentences true at a state are indicated inside the circle representing the state; all other atomic sentences are false at the state.

[12] Unless $\uparrow_j B_i\varphi$ is a conservative upgrade, in which case *j* may fail to believe $B_i\varphi$ after $\uparrow_j B_i\varphi, \uparrow_j \varphi$ (even if *j* considers it possible that $B_i\varphi \wedge \varphi$) or one upgrade after $\uparrow_j \varphi, \uparrow_j B_i\varphi$, if as before *k* testifies that $\neg\varphi$ and *j* performs the upgrade $\uparrow_j \neg\varphi$ (even if *j* considers it possible that $B_i\varphi \wedge \neg\varphi$). Conservative

belief revision in response to i's testimony, namely that the revision should not promote any states satisfying $\varphi \wedge \neg B_i\varphi$ to become more plausible. The only reason j is now upgrading with φ is that j took i to have sincerely testified for φ, so it does not make sense for j to promote a φ-state at which i does not believe φ. Yet the sequences $\uparrow_j B_i\varphi, \uparrow_j \varphi$ and $\uparrow_j \varphi, \uparrow_j B_i\varphi$ can clearly have this effect.

We can respect the constraint on state promotion as follows. If j trusts i's testimony on φ, then we use the sequence $\uparrow_j B_i\varphi, \uparrow_j B_i\varphi \wedge \varphi$ to model j's belief revision after i's testimony that φ. If j distrusts i's testimony, we use $\uparrow_j B_i\varphi, \uparrow_j B_i\varphi \wedge \neg\varphi$, and if j is ambivalent about i's testimony, we use $\uparrow_j B_i\varphi, \downarrow_j \{B_i\varphi \wedge \varphi, B_i\varphi \wedge \neg\varphi\}$. (Note here the essential use of multi-upgrade.) Otherwise j performs only the upgrade $\uparrow B_i\varphi$, in which case whether j comes to believe φ will depend on whatever beliefs j has about i's doxastic reliability.

Definition 4.9 (Testimonial Upgrade with PS)**.** Let $\gamma := B_i\varphi \wedge \varphi$, $\gamma' := B_i\varphi \wedge \neg\varphi$. The operation $\Updownarrow^i_j \varphi$ performs the following sequence of relation changes:

$$\uparrow_j B_i\varphi, \uparrow_j \left(T_{ji}\varphi \wedge \gamma\right) \vee \left(D_{ji}\varphi \wedge \gamma'\right), \downarrow_j \left\{A_{ji}\varphi \wedge \gamma, A_{ji}\varphi \wedge \gamma'\right\}$$

Definition 4.10 (Public Testimony with PS)**.** Given a testimonial model $\mathcal{M} = \langle W, \leq, V, \mathsf{rec}, \preceq \rangle$, the model $\mathcal{M}!_i\varphi = \left\langle W, \leq^{!_i\varphi}, V, \mathsf{rec}^{!_i\varphi}, \preceq \right\rangle$ is defined as follows: $\mathsf{rec}^{!_i\varphi}_j(w)$ is determined as in Definition 4.7, and $\leq^{!_i\varphi}$ is obtained from $\leq$ by the sequence of operations $\Updownarrow^i_{j \in \mathsf{Agt}\setminus\{i\}} \varphi$.

The truth definition for testimony with PS is the same as with NPS.

4.5 Basic DTL Validities

Given the semantics developed in the previous section, a number of interesting validities holds in DTL for the class of legal models. The validities below do not depend on the definitions of $T_{ji}\varphi$, $D_{ji}\varphi$, and $A_{ji}\varphi$, except that the definitions guarantee the validity of $T_{ji}\varphi \to \alpha$ for the parameter α from Section 4.3, that is,

(0) $\vDash T_{ji}\varphi \to \left(\hat{K}_j B_i\varphi \to \hat{K}_j\left(B_i\varphi \wedge \varphi\right)\right)$

and likewise for $D_{ji}\varphi$ and $A_{ji}\varphi$ with their respective values of α.

We have organized selected validities according to the questions about testimony with which we began in Section 1.1. The semantic conditions for which each formula is valid are indicated in parentheses to the right of the formula.

What do agents in i's audience come to believe about φ?

(1.1) $\vDash T_{ji}\varphi \to [!_i\varphi] B_j\varphi$ (NPS or AS)
(1.2) $\vDash D_{ji}\varphi \to [!_i\varphi] B_j\neg\varphi$ (NPS or AS)
(1.3) $\vDash A_{ji}\varphi \to [!_i\varphi] \neg B_j\varphi \wedge \neg B_j\neg\varphi$ (NPS or AS)

The semantic requirement of NPS or AS indicates that if we have the presumption of sincerity, then we must also have the assumption of sincerity. The reason (1.1) holds with NPS is that if j trusts i on φ, then by the third condition on legal models there is a state accessible for j from the current state

upgrade has been criticized for similar failures to preserve belief revisions (Booth and Meyer 2006).

that satisfies φ. Therefore, when i testifies that φ and j upgrades with φ, j's upgrade will be successful. The reason (1.1) holds for AS is that if i believes φ, then j considers it possible that $B_i\varphi$ (given reflexivity for $\leq_j$), and if moreover j trusts i on φ, then j also considers it possible that $B_i\varphi \wedge \varphi$ by (0) above. Hence when i testifies that φ and j upgrades with $B_i\varphi \wedge \varphi$ for PS or with φ for NPS, j's upgrade will be successful. The explanations for (1.2) and (1.3) are similar. Note that if instead of (0) we had (0′) $\vDash T_{ji}\varphi \to \hat{K}_j (B_i\varphi \wedge \varphi)$, then (1.1) would always hold (and similarly for (1.2)-(1.3) with the analogous changes for $D_{ji}\varphi$ and $A_{ji}\varphi$).

What do agents in i's audience come to believe about i's beliefs?

(2.1) $\vDash [!_i\varphi]\, B_jB_i\varphi$ (AS, PS)
(2.2) $\vDash \hat{K}_jB_i\varphi \to \langle !_i\varphi\rangle\, B_jB_i\varphi$ (NAS, PS)

and similarly for $D_{ji}\varphi$ and $A_{ji}\varphi$.

The reason (2.1) holds for AS and PS is that given PS, after i testifies that φ, j will upgrade with $B_i\varphi$; but given AS, i believes φ, so j considers $B_i\varphi$ possible (by reflexivity for $\leq_j$ again), which guarantees the success of j's upgrade. Without AS, it must be built into the antecedent of (2.2) that j considers $B_i\varphi$ possible.

What do agents come to believe about the beliefs of other agents in i's audience?

(3.1) $\vDash K_kT_{ji}\varphi \to [!_i\varphi]\, K_kB_j\varphi$ (NPS or AS)
(3.2) $\vDash [!_i\varphi]\, B_kT_{ji}\varphi \to [!_i\varphi]\, B_kB_j\varphi$ (NPS or AS)

and similarly for $D_{ji}\varphi$ and $A_{ji}\varphi$.

The reason (3.1) holds is that if k knows that j trusts i on φ, then no matter how k revises her beliefs in response to i's testimony, k will still know that j trusts i on φ after i testifies that φ, i.e., $\vDash K_kT_{ji}\varphi \to [!_i\varphi]\, K_kT_{ji}\varphi$. By (1.1) every state satisfying $T_{ji}\varphi$ will also satisfy $B_j\varphi$ after i testifies that φ, so k will know $B_j\varphi$ after i testifies that φ. The explanation of (3.2) is similar. The reason that (3.1) does not hold with B_k in place of K_k is that while k may believe that j trusts i on φ before i testifies that φ, after i testifies that φ and k revises her beliefs accordingly, k may no longer believe that j trusts i on φ, i.e., $\nvDash B_kT_{ji}\varphi \to [!_i\varphi]\, B_kT_{ji}\varphi$.[13]

5 Application: Information Cascades

In this section we give an application of DTL in modeling the example of an information cascade introduced in Section 1, Sorensen's (1984) epistemic

[13] Could such a situation plausibly come about? We will suggest the structure that such a situation may have, leaving it to the reader to provide a concrete example. Suppose that k believes that j trusts i on φ, but k also believes that i does not believe φ. Further suppose that k has a conditional belief to the effect that if i *does* believe φ, then another agent l believes $\neg\varphi$. Finally, suppose that k has a conditional belief to the effect that if l believes $\neg\varphi$, then j believes that l believes $\neg\varphi$ and j does not trust anyone on φ. Given these assumptions, when i testifies that φ, k will come to believe that l believes $\neg\varphi$ and hence that j does not trust i on φ.

bandwagon effect. Given the intended readings of the DTL formulas, let the initial premises about the three experts in Sorensen's scenario be:

$$\bigwedge_{i \in \mathsf{Agt}} \left(\neg \mathsf{rec}_i p \wedge \neg \mathsf{rec}_i \neg p\right)$$

1. $B_1 p$ $\qquad \emptyset \approx_1^p \{2\} \approx_1^p \{3\} <_1^p \{2,3\}$
2. $\neg B_2 p \wedge \neg B_2 \neg p$ $\qquad \emptyset <_2^p \{1\} \approx_2^p \{3\} <_2^p \{1,3\}$
3. $B_3 \neg p$ $\qquad \emptyset \approx_3^p \{1\} \approx_3^p \{2\} <_3^p \{1,2\}$

Let us assume that $[!_i \varphi]$ is the operator for testimony with the presumption of sincerity, defined as in Definition 4.10, where in Definition 4.9 $T_{ji}\varphi$ is the narrow, weak trust of Section 4.3 and $\Uparrow_i \varphi$ is any of the upgrade operations given in Section 3. Under these assumption we define a new abbreviation $\langle ?_i \varphi \rangle \psi$, read "after i testifies with her opinion on φ, ψ is the case" as follows:

$$(\neg B_i \varphi \wedge \neg B_i \neg \varphi \wedge \langle !_i \top \rangle \psi) \vee \langle !_i \varphi \rangle \psi \vee \langle !_i \neg \varphi \rangle \psi$$

With this definition, the following formulas represent the outcomes of Sorensen's two testimonial sequences:

$$\langle ?_1 p \rangle \langle ?_2 p \rangle \langle ?_3 p \rangle B_1 p \wedge B_2 p \wedge B_3 p$$

$$\langle ?_3 p \rangle \langle ?_2 p \rangle \langle ?_1 p \rangle B_1 \neg p \wedge B_2 \neg p \wedge B_3 \neg p$$

These formulas are derivable in DTL from the premises above under the stated assumptions, but we will not give the derivation here. Instead, we will show how to model the bandwagon effect semantically.

Model $\mathcal{M}_0$ in Figure 2 represents some of the relevant first and second-order beliefs of the three experts in the initial situation, focusing on the perspective of expert 3. We assume that the authority relations given above for the three agents hold everywhere in $\mathcal{M}_0$, except that in all of the states on the right side of the model, 2 does not consider 1 authoritative on p. The shaded circle on the left represents the "actual state." At the actual state, expert 1 believes p, expert 3 believes $\neg p$, and expert 2 is undecided. There is also information about each expert's beliefs about the beliefs and authority relations of the others, which we have added to Sorensen's original scenario. For example, 3 knows that 1 believes p, while 2 is uncertain about what 1 believes, and 3 knows this about 2. Finally, 3 is uncertain about whether or not 2 considers 1 authoritative on p.

When expert 1 testifies that p, two things happen. First, experts 2 and 3 both upgrade with $B_1 p$, given the presumption of sincerity. Since 3 already knows that 1 believes p, nothing changes for 3. But 2's plausibility ordering does change, reflected in the transition from $\mathcal{M}_0$ to $\mathcal{M}_1$ in Figure 3. Second, since on the left side of the model 2 considers 1 authoritative on p, and since no one else has testified to the contrary, on the left side of the model 2 *trusts* 1 on p. Hence when 2 performs the upgrade $\Uparrow_2 (T_{21} p \wedge B_1 p \wedge p) \vee (D_{21} \varphi \wedge B_1 p \wedge \neg p)$, this upgrade changes the model as if the upgrade $B_1 p \wedge p$ occurred on the left side of the model only, as reflected in the transition from $\mathcal{M}_1$ to $\mathcal{M}_2$. (The testimony operation $!_1 p$ transforms $\mathcal{M}_0$ to $\mathcal{M}_2$ in one step, but we have broken up the transformation for the purpose of explanation.) Note that 3's initial uncertainty in $\mathcal{M}_0$ about whether 2 *trusts* 1 on p leads to uncertainty for 3 in $\mathcal{M}_2$

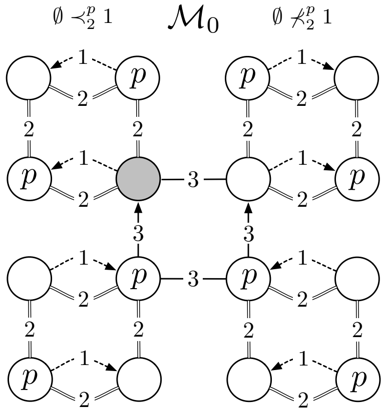


Figure 2: The initial epistemic situation of the three experts

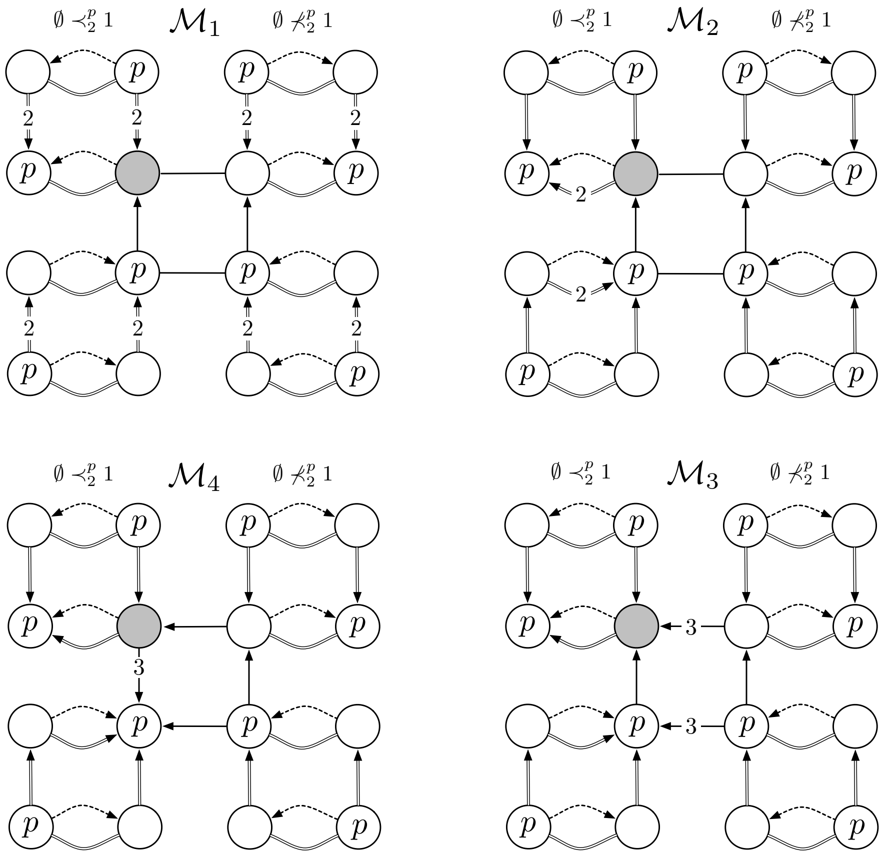


Figure 3: An epistemic bandwagon effect (clockwise from upper left)

about whether 2 *believes* p after 1's testimony. By assumption, 3 does not trust 1 on p, so there are no further changes to 3's ordering.

Having been convinced of p by the testimony of 1, 2 now testifies that p. Once again, two things happen. First, 1 and 3 upgrade with B_2p, given the presumption of sincerity. Nothing changes for 1, but 3's plausibility ordering does change, as reflected in the transition from $\mathcal{M}_2$ to $\mathcal{M}_3$. Not only does 3 come to believe that 2 believes p, but also 3 comes to believe *that* 2 *trusts* 1 *on* p, illustrating how an agent can learn from another agent's testimony whom the testifier trusts. Second, since 3 considers 1 and 2 together *jointly* authoritative on p, 3 now trusts 2 on p. Hence when 3 upgrades with $\Uparrow_3$ $(T_{32}p \wedge B_2p \wedge p) \vee (D_{32}\varphi \wedge B_2p \wedge \neg p)$, 3 comes to believe p, as reflected in the transition from $\mathcal{M}_3$ to $\mathcal{M}_4$. At this point, the epistemic bandwagon effect has occurred. All agents (falsely) believe p and believe that the others believe p.

Our model not only reflects how bandwagons start, but also suggest one way to stop them. Since expert 3 knew that expert 2 was initially undecided about p, and since the only informational event to occur before 2's testimony was 1's testimony, when 2 testified that p, 3 came to believe that 2 trusts 1 on p. If 3 were more sophisticated, 3 might have reasoned that since 1's testimony *influenced* 2 on p, the joint authority of 1 and 2 on p should not be greater than the authority of 1 alone, in which case 3 would not have come to believe p after the testimony of 2. To capture this reasoning formally, we would have to represent *influence* explicitly in our model, and we would have to provide a mechanism by which agents can change their authority relations. These additions are beyond the scope of this paper, but they suggest that it may be possible to do more than modeling the bad news about bandwagons.

6 Conclusion

We have proposed a dynamic testimonial logic (DTL) to model belief change over sequences of testimony among agents with different dispositions to trust each other as information sources. DTL adds to standard DEL the semantic structures of *testimonial records* and *authority relations* and the dynamic action of *testimony*. In the framework of DTL, we have shown how to define epistemic trust in terms of the record and authority and how to use the technique of "localizing" belief upgrades to simultaneously do different belief revisions for an agent in different parts of a model, determined by the agent's attitude (trust, distrust, or ambivalence) in different parts of the model toward an information source. We have also shown how to capture the presumption of sincerity in testimony with the right choice of belief upgrades. Finally, our DTL model of the epistemic bandwagon showed, first, how an agent's uncertainty about whom another agent trusts can lead to uncertainty about what the other agent believes and, second, how an agent may learn from testimony about whom a testifier trusts. The Appendix contains a complete axiomatization for DTL.

For future work, there are a number of possible extensions to the framework of DTL. These include working out a more sophisticated definition of trust, adding trust and testimony on doxastic formulas, representing testimony that is not fully public, and modeling agents' reasoning about how testimony has influenced the beliefs of others. In an extended version of DTL, we might model not only how bandwagons start, but also how to stop them.

Acknowledgements I wish to thank Johan van Benthem, Tomohiro Hoshi, Thomas Icard, and Eric Pacuit for helpful feedback on this paper. I also benefited from discussion of this work at the 2009 Logical Dynamics Workshop at Stanford University and at the 2009 LORI-II Workshop at Southwest University in Chongqing, China. I thank the organizers and participants at those events.

A Appendix

Theorem 1. *Together with an axiomatization of CDL (plus the universal modality for propositional formulas) and reduction axioms for a pair of belief upgrade and suspension operators, the following axiom system for DTL is sound and complete for the class of legal testimonial models.* The static axioms are the following:

(R1) $U(\varphi \leftrightarrow \psi) \rightarrow (\mathtt{rec}_i\varphi \leftrightarrow \mathtt{rec}_i\psi)$ (R2) $\mathtt{rec}_i\varphi \rightarrow K_i\mathtt{rec}_i\varphi$
(A1) $U(\varphi \leftrightarrow \psi) \rightarrow \left(S \preceq_i^{\varphi} S' \leftrightarrow S \preceq_i^{\psi} S'\right)$ (A2) $S \preceq_i^{\varphi} S$
(A3) $\left(S \preceq_i^{\varphi} S' \wedge S' \preceq_i^{\varphi} S''\right) \rightarrow S \preceq_i^{\varphi} S''$ (A4) $S \preceq_i^{\varphi} S' \leftrightarrow S \preceq_i^{\neg\varphi} S'$
(A5) $S \preceq_i^{\varphi} S' \rightarrow \hat{K}_i\left(\varphi \wedge \bigwedge_{j\in S} \neg K_j\varphi\right)$ (A6) $S \preceq_i^{\varphi} S' \rightarrow K_i S \preceq_i^{\varphi} S'$

The additional reduction axioms for belief upgrade and suspension in DTL are:

(B1) $[\pi, \varphi]\,\alpha \leftrightarrow \alpha$ for $\pi := \uparrow_i, \downarrow_i$ and $\alpha := \mathtt{rec}_j\psi, S \preceq_j^{\psi} S', U\psi$

The reduction axioms for the testimony operator with AS and PS are:

(T1) $[!_i\varphi]\,\mathtt{rec}_i\psi \leftrightarrow \left(B_i\varphi \rightarrow ((\mathtt{rec}_i\psi \wedge \neg U(\psi \leftrightarrow \neg\varphi)) \vee U(\psi \leftrightarrow \varphi))\right)$
(T2) $[!_i\varphi]\,\mathtt{rec}_j\psi \leftrightarrow \left(B_i\varphi \rightarrow \mathtt{rec}_j\psi\right)$ for $j \neq i$
(T3) $[!_i\varphi]\,\alpha \leftrightarrow (B_i\varphi \rightarrow \alpha)$ for $\alpha := p, S \preceq_i^{\psi} S', U\psi$
(T4) $[!_i\varphi]\,\neg\psi \leftrightarrow \left(B_i\varphi \rightarrow \neg\,[!_i\varphi]\,\psi\right)$
(T5) $[!_i\varphi]\,(\psi \wedge \psi') \leftrightarrow \left([!_i\varphi]\,\psi \wedge [!_i\varphi]\,\psi'\right)$
(T6) For θ and ψ that do not contain testimony operators:
$[!_i\varphi]\,B_j^{\theta}\psi \leftrightarrow \left(B_i\varphi \rightarrow r\left(\left[\updownarrow\updownarrow_{k\in X}^{i}\,\varphi\right] B_j^{\theta^*}\psi^*\right)\right)$

where $X = \left\{k \in \mathsf{Agt} \mid k \neq i \text{ and } k \text{ occurs in } B_j^{\theta}\psi\right\}$, $\left[\updownarrow\updownarrow_{k\in X}^{i}\,\varphi\right]$ abbreviates a string of upgrade and suspension operators corresponding to the relation changes in Definition 4.9, and r is the *reduction function* which, given a formula in the language of upgrade and suspension, uses the appropriate reduction axioms (as in van Benthem 2007) to return an equivalent, reduced formula in the language of CDL plus rec and $\preceq$. The formulas θ^* and ψ^* are obtained from θ and ψ respectively by replacing, for all α, each occurrence of $\mathtt{rec}_i\alpha$ in θ and ψ by an occurrence of $(\mathtt{rec}_i\alpha \wedge \neg U(\alpha \leftrightarrow \neg\varphi)) \vee U(\alpha \leftrightarrow \varphi)$.

We have given the reduction axioms for testimony with the *assumption of sincerity* and *presumption of sincerity*. For no assumption of sincerity, simply drop the precondition of $B_i\varphi$ in the conditionals of (T1) – (T4) and (T6). For no presumption of sincerity, replace $\left[\updownarrow\updownarrow_{k\in X}^{i}\,\varphi\right]$ in (T6) by the abbreviation $\left[\updownarrow_{k\in X}^{i}\,\varphi\right]$.

For CDL we assume the axioms of Baltag and Smets (2008, p. 37) or the equivalent system BRSIC of Board (2004, Sec. 3.3). For $U\varphi$ with φ propositional we take the S5 axioms plus $U\varphi \to K_i\varphi$ (cf. Blackburn et al. 2001, p. 415ff.).

Soundness. (R1) and (A1) hold in virtue of the truth definition for record and authority formulas. (R2) and (A6) hold in virtue of the fourth condition on legal models, (A2) and (A3) in virtue of the first, and (A4) and (A5) in virtue of the second and third respectively. (B1) holds because upgrades and suspensions do not change the record, authority relations, or universal propositional facts.

(T1) – (T6) hold by definition of the testimony operation on models. (T1) says that after i testifies that φ, i is on the record for ψ iff i was already on the record for ψ and ψ is not equivalent to $\neg\varphi$ in the model (for if they are equivalent, then i would have been taken off the record for ψ when she testified that φ) or ψ is equivalent to φ (in which case i was added to the record for ψ when she testified that φ). (T2) says that for agents other than i, the record does not change after i's testimony. (T3) reflects the fact that testimony does not change atomic facts, authority relations, or universal propositional facts. (T4) – (T5) give standard properties of dynamic operators.

(T6) captures the effect of i's testimony that φ on agents' beliefs, which by definition is determined by the sequence of testimonial upgrades $\updownarrow\!\updownarrow^{i}_{k\in X}\varphi$. Note that we only consider what happens to the beliefs of those agents whose symbols appear in $B^{\theta}_{j}\psi$, since the beliefs of others do not matter for evaluating the formula. Following the definition of the testimony operation, we do not change the testifier i's beliefs.

The reason for the change from θ and ψ to θ^* and ψ^* is that θ and ψ may contain a formula $\mathtt{rec}_i\alpha$, the truth value of which may change after i's testimony. Hence we use the same idea as in (T1) and express what must be true in the original model in order for $\mathtt{rec}_i\alpha$ to be true in the model updated by $!_i\varphi$. However, if θ or ψ contains testimony operators, the replacement of θ and ψ by θ^* and ψ^* may not achieve the correct result. For example, if ψ is $[!_i\neg\varphi]\,\mathtt{rec}_i\neg\varphi$, then ψ^* is $[!_i\neg\varphi]\,(\mathtt{rec}_i\neg\varphi \wedge \neg U(\neg\varphi \leftrightarrow \neg\varphi)) \vee U(\neg\varphi \leftrightarrow \varphi)$, and while ψ is valid, ψ^* is unsatisfiable. We avoid this problem by the restriction in (T6) that θ and ψ do not contain testimony operators.

Since we can reduce DTL formulas by applying the reduction axioms from the "inside out," eliminating testimony operators from subformulas first, the restriction on θ and ψ does not prevent us from reducing any DTL formula. Because θ and ψ are subformulas of $B^{\theta}_{j}\psi$, any testimony operators will be eliminated from them by the time we get to $[!_i\varphi]\,B^{r(\theta)}_{j}r(\psi)$.

Completeness. Using the reduction axioms (R1) – (T6), every formula of DTL with dynamic operators is reducible to an equivalent formula in the static part of the language. It therefore suffices to show completeness for the static part of DTL, which we will now sketch. Following the standard strategy, we show that if φ is not refutable by the axioms of DTL given above, then φ is satisfiable in a legal testimonial model. To produce the satisfying model we use the canonical model construction for CDL (Board 2004, Proof of Theorem 2, p. 77), but with two differences. The first difference is that although we construct maximally consistent sets (MCSs) from the subformulas of φ (which may now include formulas of the form $U\psi$, $\mathtt{rec}_i\psi$, and $S \preceq^{\psi}_{i} S'$) in the same way as for CDL, we do not take the domain of the canonical model to contain *all* such subformula-generated MCSs. Instead, where Γ_φ is a MCS containing φ

and R_U is a relation on MCSs such that $\Sigma R_U \Delta$ iff $\{\psi \mid U\psi \in \Sigma\} = \{\psi \mid U\psi \in \Delta\}$, we take for the domain of the canonical model the set of subformula-generated MCSs Γ such that $\Gamma_\varphi R_U \Gamma$. The second difference is that we must also construct a testimonial record and authority relations in the canonical model for DTL.

Definition A.1. The canonical testimonial model based on φ is the model $\mathcal{M}_\varphi = (W, \leq, V, \mathsf{rec}, \preceq)$ with W defined as above, $\leq_i$ defined as for CDL by Board (2004), V defined as usual by $V(p) = \{\Gamma \in W \mid p \in \Gamma\}$, and $\mathsf{rec}_i(\Gamma)$ and $\preceq^P_{i,w}$ defined by:

- $\mathsf{rec}_i(\Gamma) = \{P \subseteq W \mid \text{there is an } \alpha \text{ with } \mathtt{rec}_i\alpha \in \Gamma \text{ and } [\![\alpha]\!] = P\}$
- $S \preceq^P_{i,\Gamma} S'$ iff there is an α with $S \preceq^\alpha_i S' \in \Gamma$ and $[\![\alpha]\!] = P$

Lemma 2 (Truth)**.** $\mathcal{M}_\varphi, \Gamma \vDash \psi \Leftrightarrow \psi \in \Gamma$

Proof. By induction on ψ. We mention only the cases for $\mathtt{rec}_i\chi$ and $S \preceq^\chi_i S'$, leaving it as an exercise to the reader to check that Board's (2004) proof for the case of $B^\psi_i \chi$ works with our modified canonical model, given the axioms for U.

If $\mathcal{M}_\varphi, \Gamma \vDash \mathtt{rec}_i\chi$, then $[\![\chi]\!] \in \mathsf{rec}_i(\Gamma)$ by the definition of truth. It follows from the definition of $\mathcal{M}_\varphi$ that there is a α such that (i) $\mathtt{rec}_i\alpha \in \Gamma$ and (ii) $[\![\alpha]\!] = [\![\chi]\!]$. From (ii) we have $\mathcal{M}_\varphi, \Gamma \vDash U(\chi \leftrightarrow \alpha)$ by the definition of truth and hence (iii) $U(\chi \leftrightarrow \alpha) \in \Gamma$ by the truth lemma for CDL plus U. It follows that $\mathtt{rec}_i\chi \in \Gamma$, for otherwise $\neg\mathtt{rec}_i\chi \in \Gamma$ by the maximality of Γ, in which case Γ is inconsistent by (R1) given (i) and (iii). In the other direction, if $\mathtt{rec}_i\chi \in \Gamma$, then $[\![\chi]\!] \in \mathsf{rec}_i(\Gamma)$ by the definition of $\mathcal{M}_\varphi$ and hence $\mathcal{M}_\varphi, \Gamma \vDash \mathtt{rec}_i\chi$ by the definition of truth. The case for $S \preceq^\chi_i S'$ is analogous, using (A1) instead of (R1). □

Lemma 3 (Canonicity)**.** *$\mathcal{M}_\varphi$ is a legal testimonial model.*

Proof. $\mathcal{M}_\varphi$ satisfies the first legality condition given axioms (A2) and (A3), the second given (A4), the third given (A5), and the fourth given (R2) and (A6). □

Since the φ with which we began is by assumption not refutable by the DTL axioms, it is contained in one of the maximally consistent sets $\Gamma \in W$. Hence by the Truth Lemma $\mathcal{M}_\varphi, \Gamma \vDash \varphi$, so by the Canonicity Lemma φ is satisfiable in a legal testimonial model, our desired result.

References

L. R. Anderson and C. A. Holt. Information cascades in the laboratory. *The American Economic Review*, 87(5):847–862, 1997.

R. Audi. The place of testimony in the fabric of knowledge and justification. *American Philosophical Quarterly*, 34(4):405–422, 1997.

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

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

S. Bikhchandani, D. Hirshleifer, and I. Welch. A theory of fads, fashion, custom, and cultural change as information cascades. *The Journal of Political Economy*, 100(5):992–1026, 1992.

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

O. Board. Dynamic interactive epistemology. *Games and Economic Behavior*, 49: 49–80, 2004.

R. Booth and T. Meyer. Admissible and restrained revision. *Journal of Artificial Intelligence Research*, 26:127–151, 2006.

C. Boutilier. Iterated revision and minimal change of conditional beliefs. *Journal of Philosophical Logic*, 25:263–305, 1996.

J. Cantwell. Resolving conflicting information. *Journal of Logic, Language, and Information*, 7:191–220, 1998.

R. Demolombe. To trust information sources: a proposal for a modal logical framework. In C. Castelfranchi and Y.-H. Tan, editors, *Trust and Deception in Virtual Societies*, pages 111–124. Kluwer Academic Publishers, 2001.

R. Demolombe. Reasoning about trust: A formal logical framework. In C. D. Jensen, S. Poslad, and T. Dimitrakos, editors, *iTrust 2004*, volume 2995 of *Lectures Notes in Computer Science*, pages 291–303. Springer, 2004.

H. van Ditmarsch, W. van der Hoek, and B. Kooi. *Dynamic Epistemic Logic*. Springer, 2008.

C. Geist, B. Löwe, and B. V. Kerkhove. Peer review and knowledge by testimony in mathematics. In B. Löwe and T. Müller, editors, *Philosophy of Mathematics: Sociological Aspects and Mathematical Practice*. College Publications, London, 2010.

A. I. Goldman. Experts: Which ones should you trust? *Philosophy and Phenomenological Research*, 63(1):85–110, 2001.

J. Hardwig. Epistemic dependence. *The Journal of Philosophy*, 82(7):335–349, 1985.

J. Hardwig. The role of trust in knowledge. *The Journal of Philosophy*, 88(12): 693–708, 1991.

C.-J. Liau. Belief, information acquisition, and trust in multi-agent systems—a modal logic formulation. *Artificial Intelligence*, 149:31–60, 2003.

C. List and C. Puppe. Judgement aggregation: a survery. In P. Anand, P. Pattaniak, and C. Puppe, editors, *Oxford Handbook of Rational and Social Choice*. Oxford University Press, 2009.

A. C. Nayak. Iterated belief change based on epistemic entrenchment. *Erkenntnis*, 41:353–390, 1994.

R. Ramachandran, A. C. Nayak, and M. A. Orgun. Two approaches to iterated belief contraction. In D. Karagiannis and Z. Jin, editors, *KSEM 2009*, volume 5914 of *Lecture Notes in Computer Science*, pages 78–89. Springer, 2009.

R. Sorensen. Problems with electoral evaluation of expert opinions. *The British Journal for the Philosophy of Science*, 35(1):47–53, 1984.

Joint Revision of Belief and Intention

Thomas Icard, Eric Pacuit and Yoav Shoham

Stanford University; Tilburg University; Stanford University.
icard@stanford.edu, E.J.Pacuit@uvt.nl, shoham@stanford.edu

Abstract

We present a formal semantical model to capture action, belief and intention, based on the "database perspective" Shoham (2009). We then provide postulates for belief and intention revision, and state a representation theorem relating our postulates to the formal model. Our belief postulates are in the spirit of the AGM theory. The intention postulates stand in rough correspondence with the belief postulates.

1 Introduction and Motivation

While there is an extensive literature developing logical models to reason about changing *informational* attitudes (eg., belief, knowledge, certainty), other mental states have received less attention[1]. However, this is changing with recent articles introducing dynamic logics of intention van der Hoek et al. (2007), Herzig and Lorini (2008)[2]. These papers take as a starting point logical frameworks derived from Cohen and Levesque's seminal paper Cohen and Levesque (1990) aimed at formalizing Bratman's planning theory of intention Bratman (1987). In this paper we take a different angle on intentions, focusing on intention revision as it relates to, and is intertwined with, belief revision.

We view the problem of intention revision as a database management problem (see Shoham (2009) for more on the conceptual underpinnings of this standpoint). At any given moment, an agent must keep track of a number of facts about the current situation. This includes beliefs about the current state, beliefs about possible future states, which actions are available now and in the future, and also what the agent plans to do at future moments. It is important that all of this information be *jointly consistent* at any given moment and furthermore that it can be *modified* as needed while maintaining consistency.

In the following we introduce a simple logic that formally models such a "database". That is, *consistency* in this logic is meant to represent not only that the agent's beliefs are consistent and the agent's future plan is consistent, but

[1] A notable exception is work on logics of preferences and preference change. See van Benthem (2009) for a survey of recent work.

[2] See also a recent discussion of "goal dynamics" in Castelfranchi and Paglieri (2007).

also that the agent's beliefs and intentions together form a *coherent* picture of what may happen, and of how the agent's own actions will play a role in what happens. Many of the BDI-style logics emanating from Cohen and Levesque (1990) can be viewed as addressing this issue (Rao and Georgeff 1992, Meyer et al. 1999, are two examples). Our primary contribution in this article (in line with the recent articles on dynamic BDI logics mentioned above) is to focus also on how the database is to be modified, and in the process to provide a clear picture of how intentions and beliefs relate.

What can cause an agent's database to change? In this paper, we focus on two main sources:

1. The agent makes some observation, e.g. from sensory input. If the new observation is inconsistent with the agent's beliefs, these beliefs will have to be revised to accommodate it. While we recognize the classical AGM theory Alchourrón et al. (1985) is not without problems, in particular when it comes to iterated revision,[3] our account of belief revision simply adopts this framework. The goal is thus to give general conditions on a single revision with new information *that the agent has already committed to incorporating*.

2. The agent forms a new intention. Here we focus on *future directed* intentions, understood as time-labelled actions that might make up a plan. Analogously to belief revision, it is assumed the agent has already committed to a new intention, so it must be accommodated by any means short of revising beliefs. The force of the theory is in restricting how this can be accomplished. To be more precise, we purport to model an intelligent database, which receives instructions from some planner (e.g. a STRIPS-like planner) that is itself engaged in some form of practical reasoning. The job of the database is to maintain consistency and coherence between intentions and beliefs.

This simple description, however, obscures some important subtleties in the interaction between beliefs and intentions, subtleties we would also like to capture.

The following will serve as a running example. Suppose an agent intends to drive to the city at 6:00 this evening. Upon adopting this intention, the agent will come to have new beliefs based on the predicted success of this intention, e.g. that he will be in the city by 7:00. These further beliefs are important in the course of further planning, for instance, what he will do in the city. The intention is also supported by the absence of certain beliefs. It would be irrational to form this intention if the agent believed his car was not working and this was the only means of getting there. Likewise, even if originally the agent thought his car might be working, upon learning that it is not and lacking other ideas of how to get there, the intention to go to the city should be dropped. Yet, by dropping this intention that was based on the now-dropped belief, other beliefs, including the belief that he will be in the city by 7:00, should also be dropped, which may in turn force other intentions and beliefs to be dropped. And so on.

[3]Though, see Darwiche and Pearl (1997), Boutilier (1996) for postulates for iterated revision.

To deal with these subtleties, we treat intention-contingent beliefs, or simply *contingent beliefs*, and concrete "physical" beliefs, or *non-contingent beliefs*, separately. Non-contingent beliefs concern the world as it is, independent of the agent's future plans, but including what (sequences of) actions will be physically possible. Thus, in addition to non-contingent atomic facts, the agent will have beliefs about what the preconditions and postconditions of actions are, and about which sequences of actions might be possible. Our treatment of contingent beliefs is similar to the notion of *weak belief* in van der Hoek et al. (2007), but differs in an important respect. We assume that the postconditions of intended actions are believed in this stronger sense, but that the preconditions need not be believed. The intuition behind this decision is that, from the perspective of a planner, the postconditions of intended actions are justifiably believed *merely by the fact that the agent has committed to bringing them about*. In this way, these beliefs are *contingent* on the success of the agent's plans. The preconditions, on the other hand, may still present a practical problem yet to be solved by the planner. To say that they are believed underrates the fact that they are not *directly justified* by any intended future action. Hence, contingent beliefs are simply derived from the agent's non-contingent beliefs by adding the postconditions (and all consequences) of any intended actions. These kinds of beliefs might also be called "optimistic" beliefs, since the agent assumes the success of the action without ensuring the preconditions hold.

In this way, our account avoids the potentially infinite regress alluded to above by allowing belief revision to trigger intention revision, but restricting intention revision to trigger belief revision only in this stronger, derivative sense of contingent belief. Our postulates will reflect this fact.

In the next section, we describe the belief and intention revision postulates on an informal level before going into formal details and definitions. We then define the logic underlying the database, as a simple temporal logic with transitions labeled by actions. The models of this logic are then used to give a semantic characterization of our revision operations, which are shown in the next section to represent our main postulates. Finally in the last section before the conclusion, we define a notion of contingent beliefs, as described above, and provide postulates for revision of these beliefs, as derived from the separate postulates for beliefs and intentions.

2 A Preview of the Postulates

The main aim of our framework is to implement the "database perspective" on intentions in the form of a dual logical theory of belief and intention revision. In this section, before going into the formalities of our framework, we offer a preview of the revision postulates that we will be working toward for the rest of the paper. Relevant definitions of key terms like *coherence* will come later.

If B is a set of non-contingent beliefs (i.e. a set of formulas, which by definition does not depend on intentions) and I is a set of intentions (which shall be action/time pairs (a, t), including an empty pair ϵ), we shall define a class of intention revision operators $\circ$ that adhere to the following restrictions when $\langle B, I\rangle \circ (a, t) = \langle B', I'\rangle$ for some proposed new intention (a, t).

1. $\langle B', I'\rangle$ is coherent;

2. If $\langle B, \{(a,t)\}\rangle$ is coherent, then $(a,t) \in I'$;

3. If $\langle B, I \cup \{(a,t)\}\rangle$ is coherent, then $I \cup \{(a,t)\} \subseteq I'$;

4. $I' \subseteq I \cup \{(a,t)\}$;

5. $B' = B$.

Revision of non-contingent beliefs in AGM is in many ways analogous to intention revision. However, in a sense, intention revision is subordinate to belief revision. By 5 above, intention revision does not change the (con-contingent) belief set. But it is dependent on the belief set. Conversely, belief revision should not be dependent on the intention set, but it should in general change the intention set. To deal with this, we assume that implicit in any belief revision operator $*$ is an underlying intention revision operator $\circ^*$. We will define a class of belief revision operators that satisfy the following postulates, where again $\langle B, I\rangle * \phi = \langle B', I'\rangle$.

1. $\langle B', I'\rangle = \langle B', I\rangle \circ^* \epsilon$, where $\circ^*$ satisfies the aforementioned intention revision postulates (ensuring coherence);

2. ϕ is consistent, iff $\phi \in B'$;

3. If $\neg\phi \notin B$, then $Cl(B \cup \{\phi\}) = B'$;

4. If ϕ and ψ are equivalent and $\langle B, I\rangle * \psi = \langle B'', I''\rangle$, then $B' = B''$;

5. $B' = Cl(B')$;

6. If $\neg\psi \notin B'$ and $\langle B, I\rangle * \psi = \langle B'', I''\rangle$, then we have $Cl(B' \cup \{\psi\}) \subseteq B''$;

7. If $\langle B, I''\rangle * \phi = \langle B'', I'''\rangle$, then $B' = B''$.

Essentially, these postulates can be seen as (a slight variation of) AGM plus the intention revision postulates above.

For the rest of the paper we shall make precise how the postulates are to be represented, and in the last section investigate how these postulates look for contingent beliefs.

3 Logical Preliminaries

Our aim in this section is to develop a simple logical system that will represent the database describing the agent's beliefs about the current moment and future moment and actions that may be performed. We start with a number of simplifying assumptions about time, actions and states. First of all, we assume time is discrete and infinite in both directions, let $\mathbb{Z}$ denote the set of time-points, or moments. Nothing we say crucially depends on this assumption. Second, at each moment, some subset of the set of atomic sentences $\mathsf{Prop} = \{p, q, r, \ldots\}$ are true (intuitively, the generated propositional language describes different *ground facts* about the current state of affairs). Third, there is a finite set of primitive action symbols $\mathsf{Act} = \{a, b, c, \ldots\}$

Entries in the database will be represented by the formal language $\mathcal{L}$ given by the following grammar:

$$\phi \ := \ p_t \mid pre(a)_t \mid post(a)_t \mid Do(a)_t \mid \Box\phi \mid \phi \wedge \phi \mid \neg\phi$$

with $p \in \mathsf{Prop}, a \in \mathsf{Act}$, and $t \in \mathbb{Z}$. Intuitively, p_t means that the atomic formula p is true at time t and $Do(a)_t$ means the agent will do (or did) action a at time t. To every action and every time we associate formulas $pre(a)_t$ and $post(a)_{t+1}$, which we treat as distinguished propositional variables, and are understood as the preconditions and postconditions of a at time t. The modal operator is interpreted as historic necessity. The other boolean connectives and the dual modal operator $\Diamond$ are defined as usual.

Definition 3.1 (Paths)**.** Let P be the set

$$\mathcal{P}(\mathsf{Prop} \cup \{pre(a), post(a) : a \in \mathsf{Act}\}).$$

A *path* $\pi : \mathbb{Z} \to (P \ \times \ \mathsf{Act})$ assigns to each time t the set of proposition-like formulas true at that time, and the next action a on the path. Let $\pi(t)_1$ denote the left projection and $\pi(t)_2$ denotes the right projection. A path is called *appropriate* if the following obtains:

$$\text{If } \pi(t)_2 = a, \text{ then } post(a) \in \pi(t+1)_1.$$

There is a natural equivalence relation on a set Π of paths: we write $\pi \sim_t \pi'$ if for all $t' \leq t$, $\pi(t') = \pi'(t')$. Intuitively, $\pi \sim_t \pi'$ if π and π' represent the same situation up to time t. We extend the definition of *appropriate* to sets of paths by declaring Π to be appropriate if all paths $\pi \in \Pi$ are appropriate and moreover satisfy the following condition:

$$\text{If } pre(a) \in \pi(t)_1, \text{ then there is some } \pi' \sim \pi \text{ such that } \pi'(t)_2 = a.$$

Definition 3.2 (Truth Definition)**.** The truth relation $\models_\Pi$ is defined relative to some underlying appropriate set of paths Π. For convenience we leave off the relativizing subscript.

$\pi, t \models \alpha_{t'}$, iff $\alpha \in \pi(t')_1$, with $\alpha \equiv p, pre(a)$, or $post(a)$.

$\pi, t \models Do(a)_{t'}$, iff $\pi(t')_2 = a$.

$\pi, t \models \Box\phi$, iff for all $\pi' \in \Pi$, if $\pi \sim_t \pi'$ then $\pi', t \models \phi$.

$\pi, t \models \phi \wedge \psi$, iff $\pi, t \models \phi$ and $\pi, t \models \psi$.

$\pi, t \models \neg\phi$, iff $\pi, t \not\models \phi$.

The usual logical notions of satisfiability and validity are defined as usual. We next present a simple sound and complete logic where consistent sets are meant to represent the agent's database describing the "view" of the current situation. The proof of this theorem is by standard techniques.

Theorem 1 (The logic $\mathbf{L}_{Path}$ of paths)**.** *The following logic is sound and strongly complete with respect to the class of all appropriate sets of paths. We call this logic* $\mathbf{L}_{Path}$.

1. *Propositional Tautologies;*
2. **S5** *axioms and rules for* $\Box$ ($\Box\phi \to \phi$, $\Box\phi \to \Box\Box\phi$, $\Diamond\phi \to \Box\Diamond\phi$ *and Necessitation: from* ϕ *infer* $\Box\phi$*);*
3. $\bigvee_{a \in \mathsf{Act}} Do(a)_t$;
4. $Do(a)_t \to \bigwedge_{b \neq a} \neg Do(b)_t$;
5. $Do(a)_t \to post(a)_{t+1}$;
6. $pre(a)_t \to \Diamond Do(a)_t$;
7. *Modus Ponens.*

4 Modeling Revision

Beliefs in our framework are represented by sets of $\mathbf{L}_{Path}$-consistent formulas of $\mathcal{L}$, or equivalently, as (appropriate) sets of paths. Given a set of formulas B, we can consider the set of paths on which all formulas of B hold at time 0,[4] denoted $\rho(B)$. Conversely, given a set of paths Π, we let $\beta(\Pi)$ be defined as the set of formulas valid at 0 in all paths in Π.[5] We will use this correspondence in the representation theorem. For now we restrict our attention to sets of paths, and in particular we will represent beliefs by the minimal set under a total preorder on paths. Intentions in our models will simply be action/time pairs.

The fact that postconditions of actions always hold on a path, but that preconditions may not, is a direct implementation of our proposal that preconditions, unlike postconditions, need not be believed when an action is intended. Even if all of the paths in some (minimal) set include action a being taken at time t, it need not be that the preconditions also hold along all paths at t. We might therefore think of our belief model as, in some sense, one of "optimistic" or "imaginary" beliefs. On the other hand, we do put a slightly weaker requirement on sets of paths, that the preconditions hold on *some* path in the set. Where again I is a set of pairs (a, t), we require that the joint preconditions of all intended actions not be *disbelieved* by the agent. This is our notion of coherence.

Definition 4.1 (Coherence)**.** The pair (Π, I) is said to be *coherent* (at time 0) if there is some path $\pi \in \Pi$,

$$\pi, 0 \models \Diamond \bigwedge_{(a,t) \in I} pre(a)_t.$$

Intuitively, intentions cohere with beliefs if the agent considers it possible to carry out all of the intended actions. This is a kind of minimal requirement on *rational balance* between the two mental states.

Remark 1. A word is in order concerning this choice of coherence conditions. Consider our example of the agent that intends to go to the city at 6:00. As we

[4] As our framework is absent of operations that move time forward, we may assume it is "always" time 0.

[5] In general, $\beta(\rho(B)) = B$, but $\rho(\beta(\Pi)) \neq \Pi$.

pointed out, it is not actually necessary that the agent believe his car is working; only that he does not believe his car is not working.

Anticipating our treatment of contingent beliefs, we can also ask, what can be our agent's working assumptions about the future, upon adopting this intention? In so far as the agent is *committing* himself to this action, we may assume that he *will* go to the city at 6:00. If we then consider the subset of paths in our belief set on which this action is taken at 6:00, the postconditions will hold along all of them. However, to allow that the preconditions may not yet be believed, we admit paths on which the preconditions do not strictly hold. We only require that they hold on *some* path in the set, so that the agent cannot stray too far from reality.

Indeed, this is arguably closer to how we reason about future actions. We often commit to actions without explicitly considering the path that will lead us there. Eventually this decision will have to be made, but there is nothing incoherent about glossing over these details at the current moment. Our example agent should assume he will be in the city by 7:00 and can continue making plans about what he will do in the city once he is there. But he should not assume the preconditions will hold until he has made further, specific plans for bringing them about. This topic will be revisited in the penultimate section.

From here on we assume a coherent pair (Π, I), and define revision operations on these sets that preserve coherence. These operations will be used to represent our revision postulates. Selection functions, defined here, are simply the intention revision postulates given in the first section, under a different guise.

Definition 4.2 (Selection Function)**.** A *selection function* γ is a function that assigns an intention set to a tuple consisting of a set of paths, an intention set and a pair (a, t) satisfying the following conditions. If $\gamma(\Pi, I, (a, t)) = I'$ then,

1. (Π, I') is coherent;
2. If $(\Pi, \{(a, t)\})$ is coherent, $(a, t) \in I'$;
3. If $(\Pi, I \cup \{(a, t)\}$ is coherent, then $I' = I \cup \{(a, t)\}$.
4. $I' \subseteq I \cup \{(a, t)\}$.

In the simple case of the empty intention pair ϵ, this reduces merely to requiring coherence.

Definition 4.3 (Belief Sets)**.** Suppose Π is an appropriate set of paths. If we define a total preorder $\leq$ on Π, then the *belief set* of $(\Pi, \leq)$ is the set $\{\pi \in \Pi : \pi \leq \pi'$ for all $\pi' \in \Pi\}$. We denote this by $\min_{\leq}(\Pi)$, or just $\min(\Pi)$ when the ordering is understood from context.

Definition 4.4 (Belief Intention Model)**.** A belief-intention model is a triple $(\Pi, \leq, I, \gamma)$ where Π is a set of paths, $\leq$ is a total preorder on Π, I is a finite set of pairs (a, t) with $a \in \mathsf{Act}$ and $t \in \mathbb{Z}^+$, $(\min(\Pi), I)$ is coherent and γ is a selection function.

Definition 4.5 (Adding an Intention)**.** Let $(\Pi, \leq, I, \gamma)$ be a belief-intention model. Adding the intention (a, t) results in the model $(\Pi, \leq, I', \gamma')$ where we have $I' = \gamma(\min(\Pi), I, (a, t))$ and $\gamma' = \gamma$. We denote this model by $(\Pi, \leq, I, \gamma) \bullet (a, t)$.[6]

Definition 4.6 (Adding a Belief)**.** Let $(\Pi, \leq, I, \gamma)$ be a belief-intention model. Adding a (consistent) belief ϕ results in the model $(\Pi, \leq', I', \gamma')$, where $\gamma' = \gamma$, $I' = \gamma(\min_{\leq'}(\Pi), I, \epsilon)$, and $\leq'$ is defined so that $\pi \leq \pi'$, if and only if one of the following holds:

1. $\pi, 0 \vDash \phi$ and $\pi', 0 \nvDash \phi$;
2. $\pi, 0 \vDash \phi$ and $\pi', 0 \vDash \phi$, and $\pi \leq \pi'$;
3. $\pi, 0 \nvDash \phi$ and $\pi', 0 \nvDash \phi$, and $\pi \leq \pi'$.

This is the so-called *lexicographic* reordering operation, familiar from the belief revision and dynamic epistemic logic literatures. We denote the new belief-intention model by $(\Pi, \leq, I, \gamma) \star \phi$.

Remark 2. Lexicographic reordering is only one of many possible choices one could make here, and we adopt it only for concreteness. When we go on in future work to consider the problem of iterated revision, this decision will become more important. For now, it is sufficient to choose any revision policy that obeys the AGM postulates, as belief revision *per se* is not our central concern.

5 Representation of Revision Postulates

We are now ready to represent the postulates in full detail. In the following let $Cl(X)$ denote the closure of a set X of $\mathcal{L}$ formulas under consequence in $\mathbf{L}_{Path}$. And if I is a finite set of pairs (a, t), with $a \in \mathsf{Act}$ and $t \in \mathbb{Z}^+$, define,

$$Cohere_I := \Diamond \bigwedge_{(a,t) \in I} pre(a)_t.$$

Definition 5.1 (Belief Intention Base)**.** A belief intention base is a pair $\langle B, I \rangle$, where:

- B is a consistent set of formulas such that $Cl(B) = B$.
- I is a finite set of pairs (a, t).

Definition 5.2 (Coherence)**.** A belief-intention base $\langle B, I \rangle$ is *coherent* just in case $\neg Cohere_I \notin B$.

We then have the following obvious correspondence.

Lemma 1. *$\langle B, I \rangle$ is coherent, iff $(\rho(B), I)$ is coherent.*

Now having provided all of the necessary formal details, we repeat our postulates for intention and belief revision.

[6]Notice that this setup allows the possibility that $\gamma' \neq \gamma$, so that after revision the selection function itself can change. Of course this would only become interesting in the iterated case

Definition 5.3 (Intention Revision)**.** Suppose $\langle B, I\rangle \circ (a, t) = \langle B', I'\rangle$. The operator $\circ$ is called *proper* if the following conditions obtain.

1. $\langle B', I'\rangle$ is coherent;
2. If $\langle B, \{(a, t)\}\rangle$ is coherent, then $(a, t) \in I'$;
3. If $\langle B, I \cup \{(a, t)\}\rangle$ is coherent, then $I \cup \{(a, t)\} \subseteq I'$;
4. $I' \subseteq I \cup \{(a, t)\}$;
5. $B' = B$.

The first postulate simply says that intention revision should restore coherence. The second postulate says that the new intention (a, t) takes precedence over all other currently held intentions; it should be added if it is possible to maintain coherence, even if this means discarding current intentions. The third postulate, taken together with the fourth postulate, says that if it is possible to maintain coherence by simply adding the new intention, then this is the only change that is made. The fourth in addition guarantees that, unlike in the case of belief revision below, no extraneous intentions are ever added.[7] Finally, the fifth postulate says that non-contingent beliefs do not change with intention revision.

Recall that we assume every belief revision operator $*$ is given with its own intention revision operator $\circ^*$, so that a belief revision may trigger an intention revision.

Definition 5.4 (Belief Revision)**.** Suppose $\langle B, I\rangle * \phi = \langle B', I'\rangle$. The operator $*$ is called *proper* if the following conditions obtain.

1. $\langle B', I'\rangle = \langle B', I\rangle \circ^* \epsilon$, where $\circ^*$ is proper;
2. ϕ is consistent, iff $\phi \in B'$;
3. If $\neg\phi \notin B$, then $Cl(B \cup \{\phi\}) = B'$;
4. If $\mathbf{L}_{Path} \vdash \phi \leftrightarrow \psi$ and $\langle B, I\rangle * \psi = \langle B'', I''\rangle$, then $B' = B''$;
5. $B' = Cl(B')$;
6. If $\neg\psi \notin B'$ and $\langle B, I\rangle * \psi = \langle B'', I''\rangle$, then $Cl(B' \cup \{\psi\}) \subseteq B''$;
7. If $\langle B, I''\rangle * \phi = \langle B'', I'''\rangle$, then $B' = B''$.

Postulate 1 simply says that if intention revision is necessary to retain coherence, this revision is itself proper. Postulate 2 is a slight variation of the AGM success postulate, which we adopt on a par with intention revision postulate 2. In this setting it only makes sense to adopt a new belief if it is non-contradictory. Postulates 3-6 fill out the rest of the AGM theory, and postulate 7 says that the underlying intention set is irrelevant to belief revision.

We can now represent these postulates in terms of the belief intention models of Definition 4.4.

[7]This postulate, in particular, could be lifted depending on the application. Since we are modeling a database, we do not want the database to engage in any kind of planning. Adding new intentions when old intentions become inconsistent amounts to planning.

Theorem 2 (Representation Theorem). *For every belief intention base $\langle B, I\rangle$, with proper revision functions $*$ and $\circ^*$, there is a belief intention model $(\Pi, \leq, I, \gamma)$, such that:*

1. $\rho(B) = min_{\leq}(\Pi)$;
2. *I is the same set in the base and in the model;*
3. *For all $\phi \in \mathcal{L}$: If $(\Pi, \leq, I, \gamma) \star \phi = (\Pi, \leq', I', \gamma')$ and $\langle B, I\rangle * \phi = \langle B', I''\rangle$, then,*

$$\rho(B') = min_{\leq'}(\Pi), \text{ and } I' = I''.$$

The proof of this theorem simply rides on the proof of the representation theorem for AGM in terms of the "system of spheres" interpretation Grove (1988), with the intention revisions simply going along for the ride.

6 Intention-Contingent Beliefs

Definition 6.1. A contingent belief set B^I is derived from a belief-intention base $\langle B, I\rangle$ in the following way:

$$B^I = Cl(B \cup \{Do(a)_t : (a,t) \in I\}).$$

That is, one believes everything that was already believed non-contingently, and moreover that any actions the agent has committed to will in fact be carried out, in addition to everything that follows from this assumption, including that the postconditions of all intended actions will hold. In fact, B^I itself gives rise to a well defined belief base. This lemma follows directly from Definition 5.2 and the logic $\mathbf{L}_{Path}$.

Proposition 1. *If $\langle B, I\rangle$ is a coherent belief-intention base, then B^I is consistent.*

Notably the reverse direction of Proposition 1 does not hold. This is because of the nonparallel we have drawn between believing in preconditions and believing in postconditions (see Remark 1).

Now that we may treat B^I as a kind of belief base in its own right, we can consider what the postulates on belief and intention revision look like on the single set. The following proposition shows how the revision operators in Definitions 5.3 and 5.4 manifest themselves in the set of contingent beliefs. We give the postulates solely in terms of the set B^I itself (with no mention of the set B from which it is derived). Some information is lost with this restriction, including the distinction between non-contingently believed formulas and formulas that were added because of intentions. But arguably, this represents the kind of information the planner would solicit from the database. We shall write $B^I \circ (a,t)$ for the set $B'^{I'}$ where $\langle B, I\rangle \circ (a,t) = \langle B', I'\rangle$, and likewise for $B^I * \phi$. We make no claim to completeness here, but verification of soundness is straightforward.

Proposition 2. *The following postulates hold for any a, t, and ϕ, assuming $\circ$ and $*$ are proper.*

Intention Revision

1. $B^I \circ (a, t)$ is consistent;
2. If $\neg Cohere_{I\cup\{(a,t)\}} \notin B^I$, then $B^I \circ (a, t) = Cl(B^I \cup \{post(a)_{t+1}\})$;
3. If $\phi \notin B$ and $post(a)_{t+1} \rightarrow \phi \notin B^I$, then $\phi \notin B^I \circ (a, t)$;
4. If $\phi \in B^I$ and $\phi \wedge \bigwedge_{(b,u)\in I} \neg post(b)_{u+1}$ is consistent, then $\phi \in B^I \circ (a, t)$.

Belief Revision:

1. $B^I * \phi$ is consistent.
2. If $\neg\phi \notin B^I$ and $\phi \rightarrow \neg Cohere_I \notin B^I$, then $B^I * \phi = Cl(B^I \cup \{\phi\})$;
3. If ϕ is consistent, $\phi \in B^I * \phi$.
4. $B^I * \phi = Cl(B^I * \phi)$;
5. If ϕ and ψ are $\mathbf{L}_{Path}$-equivalent, then $B^I * \phi = B^I * \psi$.

These postulates closely mirror those for intention revision and belief revision separately. Take first the postulates for intention revision. 1 says that the new set should be consistent. 2 says that the new intention should simply be added if it is possible to do so and still maintain consistency (that is, coherence of the underlying belief-intention base). 3 ensures that no extraneous beliefs result from adding a new intention. And 4 guarantees that beliefs unrelated to the intention set, in particular those in B that have nothing to do with I, should remain, i.e. that an intention revision should not change the non-contingent beliefs.

It is interesting to note that when considering B^I, a proper belief revision operator $*$ will not, strictly speaking, satisfy all of the AGM postulates. For example, we see in Belief Revision Postulate 2 the need for an extra condition. Even if $\neg\phi \notin B^I$, we may not simply take the closure of B^I and ϕ, since adding ϕ may trigger removal of intentions, which in turn may trigger removal of beliefs from B^I. So postulate 3 of Definition 5.4 requires the extra hypothesis that $\neg Cohere_I$ does not follow in B^I from ϕ. Otherwise, the postulates follow the same spirit as the AGM postulates we had for belief intention bases. Postulate 1 perfectly mirror the corresponding postulates for intention revision, ensuring consistency, and simple addition in the case that the new belief can be consistently added. Postulates 3-5 are directly inherited from the AGM postulates we had in Definition 5.4.

It would be possible to obtain even more detailed postulates, were we to label formulas in B^I by their "justifications". For example, $post(a)_{t+1}$ could be in B^I either because it is believed non-contingently or because (a, t) is in I. Labeling formulas in this way would amount to separating B and I as we do in belief-intention bases, so we leave this possibility aside. B^I allows for a slightly simpler, if also conflated, picture of how beliefs and intentions conspire to give rise to contingent beliefs.

7 Related Work

Starting with Cohen and Levesque's classic paper Cohen and Levesque (1990), many logical systems have been developed for reasoning about informational and motivational attitudes, including intentions, in a dynamic environment (see, Meyer and Veltman 2007 and van der Hoek and Wooldridge 2003, for surveys). The central issues in this literature are (i) how to characterize the process of intention *generation*, i.e. certain kinds of practical reasoning, and (ii) how to model the *persistence* of the agents' intentions over time (see, Herzig and Lorini 2008, for a survey of the philosophical and logical literature surrounding these two issues). The problem addressed in this paper, namely how an agent should revise beliefs and intentions together given new information or a change of plans, has received relatively little attention (cf. Georgeff and Rao 1995; van der Hoek et al. 2007; Lorini et al. 2009; Roy 2009; Shoham 2009).[8]

Broadly speaking, the logical framework we use in this paper falls into the category of the so-called "BDI logics" mentioned above, in the sense that we model an agent using the mental states of *belief* and *intention* (we leave out *desires*). We do not have the space to go into a detailed comparison with the many different BDI approaches. Instead we highlight some key details about our logical system that will help place it in this literature. Our semantics (Definition 3.1) is closest to the branching-time models of Rao and Georgeff Rao and Georgeff (1992). However, one important difference is that we focus on the intention to perform an action *at a specific moment in time*. The benefits of this are discussed at length in Shoham (2009). Our treatment also shares some features with van der Hoek et al. (2007), which also proposes a formal model of intention and belief revision. Some of the basic intuitions are similar (eg., contingent beliefs are quite similar to their *weak beliefs* – however, see above), but there are also fundamental differences. Van der Hoek, Jamroga and Wooldridge extend a BDI logic with a dynamic modal operator describing what is true after the agent makes an observation. Thus, intention revision and belief revision are characterized in the formal language as validities in their logic.[9] More importantly, we differ on a number of basic conceptual issues. For example, in this paper, plans are not explicitly part of the framework, but, as a feature of the database perspective, are conceived of in the background as a recipe describing precisely what actions the agent will perform at specific moments in time. In their framework a plan describes what needs to be true in order to fulfill some desire, and consequently they focus on the problem of revising intentions and beliefs in the presence of new information and less on the effect adopting new intentions has on beliefs.

8 Conclusions and Future Work

We have presented a framework for reasoning about joint revision of beliefs and intentions. Already in the case of a single revision a number of subtle issues arise. We have chosen to address these issues by adopting a particular stance on what intentions are and how they relate to beliefs, which we have called the *database perspective* Shoham (2009). By viewing the problem of joint belief

[8]This list contains only papers that focus on logical systems that explicitly represent how an agent's intentions (and other mental attitudes) can change in the presence of new information. Indeed, philosophers and computer scientists have discussed a number of issues relevant to the problem we study in this paper. A complete survey of such issues is outside the scope of this paper (Shoham 2009, has pointers to some relevant papers).

[9]See van Benthem (2004) for a comparison between these two modeling styles for belief change, *vis-à-vis* AGM-style postulates versus modal languages with model change operators.

and intention revision as a database management problem, we have been able to bypass some of the more vexing problems about intention familiar from the philosophical literature, while at the same time confronting some basic logical problems of practical significance.

In a sense, one can see the AGM framework for belief revision as identifying what the problem of belief revision is in the first place. The standard postulates can be taken as *constitutive* of a particular kind of doxastic action, according to which the agent has committed to believing some new piece of information and must integrate this new belief with old beliefs. The interesting questions, on this view, arise when we ask how this simple picture can be embellished, to deal with iterated belief revision, interaction with other mental states and actions, and so on.[10] In the same way, one can view our treatment of joint intention and belief revision in this paper as a proposal to define what the problem is about, and to propose a framework in which further questions can be fruitfully asked and explored. Indeed, there are many directions from here that should be explored. A few of the main directions would include:

- We have mentioned several times the problem of iterated revision. This is an important and difficult topic that already comes up with belief revision by itself, and is of great interest both practically and theoretically. A large literature already exists on this problem (see e.g. Darwiche and Pearl (1997), Boutilier (1996)), but there is still further work to be done (c.f. Stalnaker (2009)).

- In this paper only atomic actions are considered. However, agents typically reason with more elaborate representations of plans, and these more elaborate representations would undoubtedly interact with beliefs in subtle ways. For example, it may not be immediately clear how our definition of coherence should be adapted to a setting in which one has conditional intentions (e.g. 'Action a, if ϕ, b otherwise'). But such intentions are crucial for agents planning in uncertain environments.

- Other mental attitudes, like goals, desires and preferences, we have left out completely. This is not because we assume they are unimportant, but rather because we want to focus in on these particular issues that arise in the interaction between belief and intention. To be sure, other interesting issues surface when belief and intention are treated together with other attitudes.

We think these are all exciting and important questions, and there are many more (see Shoham (2009) for a longer list). They are all left for future work.

Acknowledgements A version of this paper will also appear in the Proceedings of Knowledge Representation and Reasoning Icard et al. (2010). We would like to thank three anonymous reviewers for helpful feedback. Also, thanks to the organizers of the LIRA yearbook for including this work in the present volume.

[10] A view like this is taken, for example, in Stalnaker (2009).

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.

C. Boutilier. Iterated revision and minimal change of conditional beliefs. *Journal of Philosophical Logic*, 25(3), 1996.

M. Bratman. *Intention, Plans and Practical Reason*. Harvard University Press, 1987.

C. Castelfranchi and F. Paglieri. The role of beliefs in goal dynamics: Proglegomena to a constructive theory of intentions. *Synthese*, 155:237 – 263, 2007.

P. R. Cohen and H. Levesque. Intention is choice with committment. *Artificial Intelligence*, 42(3):213 — 261, 1990.

A. Darwiche and J. Pearl. On the logic of iterated belief revision. *Artificial Intelligence*, 89:1–29, 1997.

M. P. Georgeff and A. S. Rao. The semantics of intention maintenance for rational agents. In *Proceedings of the Fourteenth International Joint Conference on Artificial Intelligence (IJCAI-95)*, pages 704–710, 1995.

A. Grove. Two modellings for theory change. *Journal of Philosophical Logic*, 17, 1988.

A. Herzig and E. Lorini. A logic of intention and attempt. *Synthese*, 163(1):45 – 77, 2008.

T. Icard, E. Pacuit, and Y. Shoham. Joint revision of belief and intention. In F. Lin and U. Sattler, editors, *Proceedings of Knowledge Representation and Reasoning (KR & R)*, 2010.

E. Lorini, M. Dastani, H. van Ditmarsch, A. Herzig, and J.-J. Meyer. Intention and assignments. In X. He, J. Horty, and E. Pacuit, editors, *Logic, Rationality and Interaction*, volume 5834 of *Lecture Notes in Computer Science*. Springer, 2009.

J.-J. Meyer and F. Veltman. *Handbook of Modal Logic*, chapter Intelligent Agents and Common Sense Reasoining. Elsevier, 2007.

J.-J. Meyer, W. van der Hoek, and B. van Linder. A logical approach to the dynamics of commitments. *Artificial Intelligence*, 113:1 – 40, 1999.

A. S. Rao and M. Georgeff. Modeling rational agents within a BDI-architecture. In R. Fikes and E. Sandewall, editors, *Proceedings of Knowledge Representation and Reasoning (KR & R)*, 1992.

O. Roy. A dynamic-epistemic hybrid logic for intentions and information changes in strategic games. *Synthese*, 171:291 – 320, 2009.

Y. Shoham. Logical theories of intention and the database perspective. *Journal of Philosophical Logic*, 38(6), 2009.

R. Stalnaker. Iterated belief revision. *Erkenntnis*, 70, 2009.

J. van Benthem. Dynamic logic for belief revision. *Journal of Applied Non-Classical Logics*, 14, 2004.

J. van Benthem. For better or for worse: Dynamic logics of preference change. In *Preference Change*, volume 42 of *Theory and Decision Library A*, pages 57 – 84. Springer, 2009.

W. van der Hoek and M. Wooldridge. Towards a logic of rational agency. *Logic Journal of the IGPL*, 11(2):135 – 160, 2003.

W. van der Hoek, W. Jamroga, and M. Wooldridge. Towards a theory of intention revision. *Synthese*, 155:265 – 290, 2007.

Cooperation in Normal Modal Logics

Cédric Dégremont and Lena Kurzen

ILLC, Universiteit van Amsterdam
cedric.uva@gmail.com, l.m.kurzen@uva.nl

Abstract

This paper develops a unified perspective on modal logics for cooperation of agents that have preferences. We consider different families of normal modal cooperation logics – one of them containing the normal simulation of Pauly's Coalition Logic – and prove embedding results that clarify the relations between them. We show how game theoretical and social choice theoretical notions can be interpreted on three different classes of models, and identify via invariance results the expressive power required for expressing these notions. Explicit definability results in extended modal languages are given for each notion and class of models. Complexity results for extended modal logics are then used to obtain upper bounds on the complexity (model checking and satisfiability) of modal logics expressing the notions. This way, our analysis shows how demanding certain game theoretical and social choice theoretical notions are in terms of complexity and expressive power. Our analysis shows how the choice of models (simple models with coalitional power as a primitive vs. more complex power based or action based models) effects the expressive power and complexity required to express the notions. For instance, we found opposite results for different classes of models as to whether strict or non-strict stability notions are easier to express.

1 Introduction

Cooperation of agents is a crucial concept in many fields such as philosophy, social science and computer science. Various modal logic (ML) frameworks have been developed for formal reasoning about cooperation in multi-agent systems. These logics focus on different aspects of cooperation and differ in what they take as primitives.

Coalition Logic (CL) Pauly (2002) focusses on coalitional power. It uses formulas of the form $\langle\!\![C]\!\!\rangle\phi$ saying that coalition C has a joint strategy to ensure that ϕ. Another class of cooperation logics is motivated by the aim to make coalitional power more explicit. This is done by explicitly representing the actions or strategies by which coalitions can achieve some result Walther et al. (2007), Borgo (2007), Gerbrandy and Sauro (2007). Another concept that plays a crucial role when reasoning about strategic interaction of (groups of) agents is that of *preferences*, which has also received much attention from modal logicians Girard (2008). In general, logics for reasoning about cooperation are often evaluated w.r.t their power to express concepts from game theory (GT) and social choice theory (SCT), and w.r.t. to their computational complexity.

1.1 Aim

One aim of this work is to determine how much expressive power and complexity is needed for ML's to be able to express GT and SCT concepts. Of course, this depends on the models under consideration. We analyze three classes of models, each of them modelling coalitional power from a different perspective. Additionally, we clarify the results by also analyzing the relation between the models, and their relation to other frameworks from the literature.

Through the comparison, we are able to determine how demanding different notions are on each class of models. Thus, our results help to make design choices when developing ML's for cooperation since we know the impact of certain choices on the complexity and expressive power required to express GT and SCT notions. Additionally, we clarify the relationship between complexity and expressive power results of existing cooperation logics by showing how different such logics can be embedded into each other.

1.2 Methodology

Our methodology is as follows. First of all, we focus on classes of models/logics for cooperation with natural model theoretical properties. Here, we consider normal ML's that model coalitional power in different ways. We extend them with a representation of agents' preferences as total preorders over the states. We analyze the relation between them and also their relation to existing frameworks by giving embedding results. Then we focus on a set of notions that are of interest when reasoning about cooperation of agents that have preferences, and give natural interpretations of them in each of the models. Next, we determine the expressive power required by these notions by checking under which operations on models these properties are invariant. Using characterization results for extended modal logics, we then obtain extended modal languages that can express the notions. Among these, we choose the ones with the lowest expressive power and give explicit definability results for the notions. Using known complexity results for extended ML's, we also obtain upper bounds (UB) on the complexity of ML's (satisfiability (SAT) and model checking (MC) (combined complexity)) that can express each notion.

The remainder of this paper is structured as follows: In Section 2, we present three classes of models for reasoning about cooperation, and give the interpretations of extended modal languages on these models. Section 3 then gives our main results. The first three subsections give embedding results showing how the different modal cooperation frameworks we consider relate to each other, and to cooperation frameworks from the literature. Then, in Section 3.4 we give invariance and explicit definability results for several game theoretical and social choice theoretical properties, and also upper bounds on the complexity of modal logics being able to express them. Section 4 concludes this work.

2 Three ways of modelling cooperation

This section presents the classes of models considered in this work. They correspond to models discussed in the literature, each focusing on different aspects of cooperation. We deliberately consider simplifying models or generalizations in order to avoid additional complexity due to assumptions on the models. This allows us to distinguish more clearly how the notions themselves are demanding and to evaluate from a high level perspective wich models are most appropriate for reasoning about which aspects of cooperation.

The first class of models, coalition labelled transition systems, Dégremont and Kurzen (2009) focuses on agents' preferences and their interaction with cooperation; it greatly simplifies the computation of coalitional powers, which are directly represented as accessibility relations. The second class, *action-based coalitional models*, gives a natural account of coalitional power by representing it in terms of actions that agents can perform. The third class Broersen et al. (2007) are *power-based coalitional models*. Its focus lies on reasoning about and computing coalitional power itself, encoding coalitions' choices as partitions of the state space. In all classes, preferences are represented as total preorders ($\mathtt{TPO}$) over the states.

2.1 The Models

Our models are based on a finite set of agents $\mathtt{N}$. j ranges over $\mathtt{N}$. PROP is the set of propositional letters and NOM a set of nominals, which is disjoint from PROP. A nominal is true in exactly one state. We let $p \in$ PROP and $i \in$ NOM.

Coalition-labelled transition systems.

A simple way to use Kripke models for reasoning about coalitional power is to use *sequential* systems, with an accessibility relation for each coalition. Then a group has the power to make the system move into exactly the states accessible by the relation. These models are really a generalization of Segerberg (1989) models to coalitional interaction. Since they tend to simplify greatly the interpretation of coalitional powers, it allows us to focus on the expressive power required by the notions themselves and by reasoning about preferences.

Definition 2.1 ($\wp(\mathtt{N})$-$\mathtt{LTS}$)**.** A $\wp(\mathtt{N})$-$\mathtt{LTS}$ (Labeled Transition Systems indexed by a finite set of coalitions $\wp(\mathtt{N})$) is of the form $\langle W, \mathtt{N}, \{ \xrightarrow{C} \mid C \subseteq \mathtt{N}\}, \{ \leq_j \mid j \in \mathtt{N}\}, V\rangle$, where $W \neq \emptyset$, $\mathtt{N} = \{1, \ldots, n\}$ for some $n \in \mathbb{N}$, $\xrightarrow{C} \subseteq W \times W$ for each $C \subseteq \mathtt{N}$, $\leq_j \subseteq W \times W$ for each $j \in \mathtt{N}$, and $V : \text{PROP} \cup \text{NOM} \to \wp(W)$, $|V(i)| = 1$ for each $i \in \text{NOM}$.

W is a set of states and $w \xrightarrow{C} v$ says that coalition C can change the state from w into v. Other interpretations are possible, e.g. group preferences. $\leq$ is a $\mathtt{TPO}$, and $w \leq_j v$ means that j finds v at least as good (a.l.a.g) as w. $w \in V(p)$ means that p is true at w.

Action-based coalitional models.

In action-based coalitional models, coalitional power is represented using actions. Agents can perform certain actions; this then changes the current state. The general idea is similar to that underlying some existing logics for cooperation, e.g. Borgo (2007), Walther et al. (2007).

Definition 2.2 (ABC)**.** A $\mathtt{N},(\mathtt{A_j})_{\mathtt{j}\in\mathtt{N}}$-ABC (action-based coalitional model indexed by a finite set of agents N and a collection of finite sets of actions $(A_j)_{j\in\mathtt{N}}$) is of the form $\langle W,\mathtt{N},\{\xrightarrow{j,a} \mid j\in\mathtt{N}, a\in A_j\},\{\leq_j \mid j\in\mathtt{N}\},V\rangle$, where $W\neq\emptyset$, $\mathtt{N}=\{1,\dots,n\}$, for some $n\in\mathbb{N}$; for each $j\in\mathtt{N}$ A_j is a finite set, $\xrightarrow{j,a}\subseteq W\times W$ for each $j\in\mathtt{N}, a\in A_j$, $\leq_j\subseteq W\times W$ is a TPO for each $j\in\mathtt{N}$, and $V:\textsc{prop}\cup\textsc{nom}\to\wp(W)$, $|V(i)|=1$ for each $i\in\textsc{nom}$. Given a relation $R\subseteq W\times W$, we write $R[w]:=\{v\in W\mid wRv\}$.

$\xrightarrow{j,a}[w]\subseteq X$ means that at w, j can guarantee by doing a that the next state is in X. Thus, at w, j can guarantee that the next state will be in X iff for some set Y,$X\supseteq Y\in\{\xrightarrow{j,a}[w]\mid a\in A_j\}$; ($Y$ is then in the *exact power* of j at w). Finally, we take powers to be additive: powers of coalitions arise from individual's powers. W.l.o.g. let $C=\{1,\dots,|C|\}$. Then, at w, $C\subseteq\mathtt{N}$ can guarantee that the next state will be in X iff for some set Y we have $X\supseteq Y\in\{\bigcap_{j\in C}\xrightarrow{j,a_j}[w]\mid(a_1,\dots,a_{|C|})\in\times_{j\in C}A_j\}$; ($Y$ is in the exact power of C at w). V and $\leq_j$ are as for $\wp(\mathtt{N})-\mathtt{LTS}$.

We say that an ABC model $\mathcal{M}$ is *reactive*, if the following two conditions are fulfilled.

1. for any action profile $(a_j)_{j\in\mathtt{N}}\in\times_{j\in N}(A_j)$, and for each state w, $(\bigcap_{j\in\mathtt{N}}\xrightarrow{j,a_j}[w])\neq\emptyset$, i.e. for every collective choice the system specifies at least one possible next state.

2. in each state, each agent has at least one available action, i.e. for each $j\in\mathtt{N}$ and each $w\in W$, there is some $a_j\in A_j$ such that $\xrightarrow{j,a_j}[w]\neq\emptyset$.

We say that an ABC model $\mathcal{M}$ is N-*determined* if for each $w\in W$, whenever there is some agent $j\in\mathtt{N}$ and some action $a_j\in A_j$ such that $v\in\xrightarrow{j,a_j}[w]$, then there is an action profile $(a_j)_{j\in\mathtt{N}}\in\times_{j\in\mathtt{N}}(A_j)_j$ such that $(\bigcap_{j\in\mathtt{N}}\xrightarrow{j,a_j}[w])=\{v\}$.

If an ABC model is both N-*determined* and reactive, we say that it is a $\mathtt{ABC}^{NR}$ model.

Power-based coalitional models.

These models generalize those of CL's normal simulation NCL Broersen et al. (2007), and additionally have a preference relation for each agent.

Definition 2.3 (PBC-Model)**.** A $\wp(\mathtt{N})$-PBC-model (power based coalitional model indexed by a finite set of coalitions $\wp(\mathtt{N})$) is a tuple $\langle W,\mathtt{N},\{\sim_C\mid C\subseteq\mathtt{N}\},F_{\mathbf{X}},\{\leq_j\mid j\in\mathtt{N}\},V\rangle$, where $W\neq\emptyset$, each $\sim_C\subseteq W\times W$ is an equivalence relation, $F_{\mathbf{X}}:W\to W$ is a total function, $\leq_j\subseteq W\times W$ is a TPO for each $j\in\mathtt{N}$, and $V:\textsc{prop}\cup\textsc{nom}\to\wp(W)$, $|V(i)|=1$ for each $i\in\textsc{nom}$.

$F_{\mathbf{X}}$ determines the system's actual course of action: if we are in w, then $F_{\mathbf{X}}(w)$ is the next state. The equivalence relation $\sim_C$ describes C's lack of power: $w \sim_C v$ means that it is not in the power of C to decide between w and v and thus neither whether we move to $F_{\mathbf{X}}(w)$ or $F_{\mathbf{X}}(v)$. C can on the other hand choose an equivalence class $[w]_{\sim_C}$ and thus restrict the set of possible next states to $F_{\mathbf{X}}[[w]_{\sim_C}]$.

The models of NCL are PBC models with additional properties.

Definition 2.4 (NCL-Independence)**.** For every $C \subseteq \mathtt{N}$, $\sim_\emptyset \subseteq (\sim_C \circ \sim_{\overline{C}})$.

Definition 2.5 (NCL-Model)**.** An NCL model is a PBC model satisfying the following conditions:

1. For all $C, D \subseteq \mathtt{N}$, if $D \subseteq C$, *then* $\sim_C \subseteq \sim_D$.

2. NCL-Independence.

3. $\sim_{\mathtt{N}} = id = \{(w, v) \in W \times W \mid w = v\}$.

Next, we mention an important model from the literature whose relation to the previous models will be discussed.

Alternating-time temporal models.

An alternating transition system Alur et al. (1998) is of the form $\langle W, \mathtt{N}, \delta, V\rangle$ where $W \neq \emptyset$, $\delta : W \times \mathtt{N} \to \wp(\wp(W))$ and satisfies the following properties.

- *Consistency*. For all $C \subseteq \wp(\mathtt{N})$, for all $w \in W$, for all $(X_j)_{j\in C} \in \times_{j\in C}\delta(w, j)$ we have $(\bigcap_{j\in C} X_j) = \emptyset$.

2.2 Extended modal languages

For each type of models, we introduce a language from which we will actually consider different fragments. These languages will later be used to define different GT and SCT notions.

Language interpreted on $\wp(\mathtt{N}) - \mathtt{LTS}$.

$$\alpha ::= \leq_j \ \mid C \mid \alpha \cap \alpha \mid \overline{\alpha}$$

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

where $j \in \mathtt{N}$, $C \in \wp(\mathtt{N}) - \{\emptyset\}$, $p \in$ PROP, $i \in$ NOM, $x \in$ SVAR. SVAR is a countable set of variables.

Semantics. Programs α are interpreted as relations. Formulas are interpreted with an assignment $g : \textsc{svar} \to W$. We skip booleans.

$$
\begin{array}{lll}
\mathcal{M}, w, g \Vdash p & \text{iff} & w \in V(p) \\
R_{\leq_j} & = & \leq_j \\
R_C & = & \xrightarrow{C} \\
R_{\beta \cap \gamma} & = & R_\beta \cap R_\gamma \\
R_{\overline{\beta}} & = & (W \times W) \setminus R_\beta
\end{array}
\qquad
\begin{array}{lll}
\mathcal{M}, w, g \Vdash i & \text{iff} & w \in V(i) \\
\mathcal{M}, w, g \Vdash x & \text{iff} & w = g(x) \\
\mathcal{M}, w, g \Vdash \langle \alpha \rangle \phi & \text{iff} & \exists v : wR_\alpha v \\
 & & \text{and } \mathcal{M}, v, g \Vdash \phi \\
\mathcal{M}, w, g, \Vdash @_i \phi & \text{iff} & \mathcal{M}, v, g \Vdash \phi \\
 & & \text{for } V(i) = \{v\} \\
\mathcal{M}, w, g, \Vdash @_x \phi & \text{iff} & \mathcal{M}, g(x), g \Vdash \phi \\
\mathcal{M}, w, g, \Vdash \downarrow x.\phi & \text{iff} & \mathcal{M}, w, g[x := w] \Vdash \phi
\end{array}
$$

Language interpreted on ABC models.

The basic language for ABC is defined as follows (the extension with hybrid and boolean modal logic formulas is as for $\wp(\mathtt{N}) - \mathtt{LTS}$). Note it would be more correct to talk about a family of languages, since they are indexed by the set of actions that are assigned to these agents, thus by a collection of sets $(A_j)_{j \in \mathtt{N}}$.

$$\alpha ::= \leq_j \ | \ a_j \ | \ \alpha^{-1} \ | \ \alpha \cap \alpha \ | \ \overline{\alpha}$$

$$\phi ::= \ p \ | \ i \ | \ x \ | \neg\phi \ | \ \phi \wedge \phi \ | \ \langle \alpha \rangle \phi \ | \ @_i \phi \ | \ @_x \phi \ | \ \downarrow x.\phi$$

where $j \in \mathtt{N}$, $a_j \in A_j$ (the set of actions available to j) and $p \in \textsc{prop}$.

$$
\begin{array}{lll}
R_{a_j} & = & \xrightarrow{j,a} \\
R_{\alpha^{-1}} & = & \{(v, w) | w R_\alpha v\}
\end{array}
$$

We only give a few clauses to give the intuition.

$$
\begin{array}{lll}
\mathcal{M}, w, g \Vdash \langle a_j \rangle \phi & \text{iff} & \exists v : w \xrightarrow{j,a} v \text{ and } \mathcal{M}, v, g \Vdash \phi \\
\mathcal{M}, w, g \Vdash \langle \leq_j \rangle \phi & \text{iff} & \exists v : w \leq_j v \text{ and } \mathcal{M}, v, g \Vdash \phi \\
\mathcal{M}, w, g \Vdash \langle \alpha \rangle \phi & \text{iff} & \exists v : wR_\alpha v \text{ and } \mathcal{M}, v, g \Vdash \phi
\end{array}
$$

We will make use of someshortcuts when writing big disjunctions or unions. For $C \subseteq \mathtt{N}$, we let $\vec{C} := \times_{j \in C} A_j$. For an action profile $\vec{a_j} = (a_j)_{j \in N} \in \vec{C}$ we often write $\bigcap \vec{a_j}$ to stand for $\bigcap_{j \in C} a_j$. As an example, for the language indexed by $A_1 = T_1, M_1, B_1$ and $A_2 = L_2, R_2$ instead of writing $[T_1 \cap L_2]p \vee [M_1 \cap L_2]p \vee [B_1 \cap L_2]p \vee [T_1 \cap R_2]p \vee [M_1 \cap R_2]p \vee [B_1 \cap R_2]p$, we often write $\bigvee_{\vec{a_j} \in \vec{\{1,2\}}} [\cap \vec{a_j}]p$.

Language interpreted on PBC/NCL models.

The language $\mathcal{L}_{\mathtt{NCL}}$ for PBC and NCL is given as defined in Broersen et al. (2007), extended with hybrid and boolean modal logic formulas as for $\wp(\mathtt{N}) - \mathtt{LTS}$.

$$\alpha ::= \leq_j \ \Big| \ \alpha^{-1} \ \Big| \ \alpha \cap \alpha$$

$$\phi ::= \ p \ \Big| \ i \ \Big| \ x \ \Big| \ \neg\phi \ \Big| \ \phi \wedge \phi \ \Big| \ \langle C \rangle \phi \ \Big| \ \mathbf{X}\phi \ \Big| \ \langle \alpha \rangle \phi \ \Big| \ @_i \phi \ \Big| \ @_x \phi \ \Big| \ \downarrow x.\phi$$

where $j \in \mathtt{N}$ (the set of agents), $C \in \wp(\mathtt{N})$ and $p \in \textsc{prop}$.

$$\begin{array}{lll} \mathcal{M}, w, g \Vdash \langle C \rangle \phi & \text{iff} & \exists v : w \sim_C v \text{ and } \mathcal{M}, v, g \Vdash \phi \\ \mathcal{M}, w, g \Vdash \mathbf{X}\phi & \text{iff} & \mathcal{M}, F_{\mathbf{X}}(w), g \Vdash \phi \\ \mathcal{M}, w, g \Vdash \langle \leq_j \rangle \phi & \text{iff} & \exists v : w \leq_j v \text{ and } \mathcal{M}, v, g \Vdash \phi \\ \mathcal{M}, w, g \Vdash \langle \alpha \rangle \phi & \text{iff} & \exists v : w R_\alpha v \text{ and } \mathcal{M}, v, g \Vdash \phi \end{array}$$

Language interpreted on alternating transition systems.

The basic language $\mathcal{L}_{\mathtt{ATL}}$ is defined as follows:

$$\phi ::= p \mid \neg\phi \mid \phi \wedge \phi \mid \langle\langle C \rangle\rangle \mathbf{X}\phi$$

where p ranges over PROP and C over $\wp(\mathtt{N})$.

For finite sequences λ, let $Last(\lambda)$ be the last element of λ. Let W^+ stand for the set of non-empty finite sequences. For $j \in \mathtt{N}$, let a strategy for j be a function $f_j : W^+ \to \wp(W)$, such that for each finite sequence of states λ, $f_j(\lambda) \in \delta(j, Last(\lambda))$. Let a collective strategy for C be a collection $F_C = (f_j)_{j \in C}$. Finally $out(w, F_C) = \{\lambda | \lambda[0] = w \& \forall i \geq 0 (w_{i+1} \in \bigcap_{j \in C} f_j(\lambda_i))\}$ where λ_i is the prefix of λ of length $i+1$. $\mathcal{L}_{\mathtt{ATL}}$ is interpreted as follows:

$$\mathcal{M}, w \Vdash \langle\langle C \rangle\rangle \mathbf{X}\phi \quad \text{iff} \quad \exists F_C : \forall \lambda \in out(w, F_C) : \mathcal{M}, \lambda[1] \Vdash \phi$$

The axiomatization of ATL is given in Table 1.

$(\bot)$	$\vdash \neg\langle\langle C \rangle\rangle \mathbf{X}\bot$
$(\top)$	$\vdash \langle\langle C \rangle\rangle \mathbf{X}\top$
(Σ)	$\vdash \neg\langle\langle \emptyset \rangle\rangle \mathbf{X}\neg\phi \to \langle\langle \Sigma \rangle\rangle \mathbf{X}\phi$
(S)	$\vdash (\langle\langle C_1 \rangle\rangle \mathbf{X}\phi \wedge \langle\langle C_2 \rangle\rangle \mathbf{X}\psi) \to \langle\langle C_1 \cup C_2 \rangle\rangle \mathbf{X}(\phi \wedge \psi)$
	for $C_1 \cap C_2 = \emptyset$
$(\langle\langle C \rangle\rangle \mathbf{X}mon)$	$\vdash \phi \to \psi$ implies $\vdash \langle\langle C \rangle\rangle \mathbf{X}\phi \to \langle\langle C \rangle\rangle \mathbf{X}\psi$

Table 1: Axiomatization of ATL.

This section has introduced all the frameworks for which we would like to investigate how demanding reasoning about cooperation is. The languages defined above will be used to express SCT and GT notions on the respective models. We will also clarify the relationship between ATL and classes of models we considered.

3 Results

In this section, we give our main results obtained when analyzing the different ways of modelling cooperation presented before. First we determine the relation between the classes of models; then we analyze how demanding different concepts from GT and SCT are on the models.

3.1 Towards a unified perspective on cooperation logics

This section gives translations between the different frameworks for cooperation. We start by analyzing coalitional power as modelled in PBC, NCL and CL and investigate relations between standard assumptions on coalitional power.

3.2 On the relation between PBC and NCL models

We say that C can *force* a set X at w iff at w it can guarantee that the next state is in X. Put differently, C can force X if some subset of X is in the exact power of C at w. Let us discuss reasonable assumptions about the coalitional powers that reflect the independence of agents and are generally assumed in the literature (cf. Pauly (2002), Broersen et al. (2007), Belnap et al. (2001)). To be precise, we distinguish between two assumptions and show how they relate. Let $P_C(w)$ be the collection of exact powers of C at w; informally, $P_C(w)$ contains the possible sets of states coalition C can choose from at w. Let $\overline{C} = \mathtt{N} \setminus C$ and $\overline{X} = W \setminus X$. Independence of coalitions says that whatever choices two disjoint coalitions make, there will be a next state resulting from these choices.

Definition 3.1 (Independence of coalitions (IC))**.** $\forall w$, if $C \cap D = \emptyset$ then $\forall X \in P_C(w)\ \forall Y \in P_D(w):\ X \cap Y \neq \emptyset$.

The next condition says that if C can ensure that the next state is in X, $\overline{C}$ cannot ensure that is not.

Definition 3.2 (Condition about complementary coalitions (CCC))**.** $\forall w$, $\forall X$, if $\exists X'$ such that $X \supseteq X' \in P_C(w)$, then there is no Y such that $\overline{X} \supseteq Y \in P_{\overline{C}}(w)$.

Coalition monotonicity says that if a coalition can achieve something then so can all supersets of it.

Definition 3.3 (Coalition monotonicity (CM))**.** $\forall w\ \forall X$, if $C \subseteq D$ and $\exists Y$ such that $X \supseteq Y \in P_C(w)$, then $\exists Z$ such that $X \supseteq Z \in P_D(w)$.

Fact 1. IC implies CCC.

Proof. Take some w and write $P(C)$ for $P_C(w)$. Assume that $X \supseteq X' \in P(C)$ (a). Now assume for contradiction that for some Y, $\overline{X} \supseteq Y \in P(\overline{C})$ (b). Since $C \cap \overline{C} = \emptyset$, by IC $Y \cap X' \neq \emptyset$ (c). Then by (c) and (b), $\overline{X} \cap X' \neq \emptyset$ (d). But (d) and (a) implies $\overline{X} \cap X \neq \emptyset$, a contradiction. □

Fact 2. CCC + CM implies IC.

Proof. Assume that $C \cap D = \emptyset$, i.e. $D \subseteq \overline{C}$ (e). Let $X \in P(C)$ (f), $Y \in P(D)$ (g). Assume for contradiction that $X \cap Y = \emptyset$ (h), i.e. $Y \subseteq \overline{X}$ (i). By (f) and CCC, there is no Z such that $\overline{X} \supseteq Z \in P(\overline{C})$ (j). But then by (j), (e) and CM there is no Z such that $\overline{X} \supseteq Z \in P(D)$ (k). But (k) contradicts (g). □

Note that on PBC models, CCC is actually the following:

$$\begin{aligned} \forall w \quad & [\forall X \text{ if } \exists v(v \in \sim_{\emptyset} [w] \text{ and } \sim_C [v] \subseteq X), \text{ then} \\ & \neg\exists t(t \in \sim_{\emptyset} [w] \text{ and } \sim_{\overline{C}} [t] \subseteq \overline{X})]. \end{aligned} \tag{1}$$

For NCL, Broersen et al. (2007) take the condition of NCL-Independence (Definition 2.4), which has has a natural modal axiomatization.

On power based models the following holds:

Fact 3. CCC is equivalent to NCL-Independence.

Proof. From right to left. Assume that CCC does not hold. Then there are w, X and v such that $v \in R_\emptyset[w]$ (1) and $R_C[v] \subseteq X$ (2), and also there is some t such that $t \in R_\emptyset[w]$ (3) and $R_{\overline{C}}[v] \subseteq \overline{X}$ (4). But by (2) and (4), $R_C[v] \cap R_{\overline{C}}[v] = \emptyset$ (5). By (1) and (3), $v, t \in R_\emptyset[w]$ but since $R_\emptyset$ is an equivalence relation, $(v, t) \in R_\emptyset$ (6). But (6) and (5) together imply the negation of NCL-Independence.

From left to right. Assume that NCL-Independence does not hold. Then there are w and z such that $z \in R_\emptyset[w]$ (7) and $(w, z) \notin R_C \circ R_{\overline{C}}$ (8). By (8) and the fact that for every C, R_C is an equivalence relation, $R_C[w] \cap R_{\overline{C}}[z] = \emptyset$ (9). Then $R_{\overline{C}}[z] \subseteq (W \setminus R_C[w])$ (10). Assume for contradiction that CCC holds. Instantiating w by w and X by $R_C[w]$ in (1) we get: if $\exists v(v \in R_\emptyset[w] \wedge R_C[v] \subseteq R_C[w])$, then $\neg\exists t(t \in R_\emptyset[w] \wedge R_{\overline{C}}[t] \subseteq \overline{R_C[v]})]$ (11). By reflexivity of $R_\emptyset$, $(w \in R_\emptyset[w] \wedge R_C[w] \subseteq R_C[w])$ (12). By (11) and existential generalization of (12), $\neg\exists t(t \in R_\emptyset[w] \wedge R_{\overline{C}}[t] \subseteq \overline{R_C[v]})]$ (13). But from (7) and (10), $(z \in R_\emptyset[w] \wedge R_{\overline{C}}[z] \subseteq \overline{R_C[v]})]$ (14). But (14) contradicts (13), thus CCC does not hold. □

To briefly summarize, we have shown that independence of coalitions (IC) implies the condition about complementary coalitions; and that on the other hand, if we we have CCC and additionally coalition monotonicity, then IC also holds. Comparing IC and NCL-Independence, we have seen that they are actually equivalent.

3.3 On the relation between NCL and CL

Let us analyze the relation between CL and its normal simulation NCL. First, we will briefly recall the semantics of CL. For the details we refer the reader to Pauly (2002).

Definition 3.4 (CL-Model)**.** A CL-model is a pair $((\mathtt{N}, W, E), V)$ where N is a set of agents, $S \neq \emptyset$ is a set of states, $E : W \to (\wp(\mathtt{N}) \to \wp(\wp(W)))$ is called an effectivity structure. It satisfies the conditions of **playability**:

- Liveness: $\forall C \subseteq \mathtt{N} : \emptyset \notin E(C)$,
- Termination: $\forall C \subseteq \mathtt{N} : W \in E(C)$,
- N-maximality. $\forall X \subseteq W : (W \setminus X \notin E(\emptyset) \Rightarrow X \in E(N))$
- Outcome monotonicity. $\forall X \subseteq X' \subseteq W, C \subseteq \mathtt{N} : (X \in E(C) \Rightarrow X' \in E(C))$,
- Superadditivity. $\forall X_1, X_2 \subseteq W, C_1, C_2 \subseteq \mathtt{N} : ((C_1 \cap C_2 = \emptyset \;\&\; X_1 \in E(C_1) \;\&\; X_2 \in E(C_2)) \Rightarrow X_1 \cap X_2 \in E(C_1 \cup C_2))$.

$V : \textsc{prop} \to \wp(W)$ is a propositional valuation function.

The language $\mathcal{L}_{\text{CL}}$ of CL is a standard modal language with a modality $\langle\!\langle C \rangle\!\rangle$ for each $C \subseteq \mathtt{N}$. The intended meaning of $\langle\!\langle C \rangle\!\rangle\phi$ is "coalition C has the power to achieve that ϕ". The semantics is as follows:

$$M, w \vDash \langle\!\langle C \rangle\!\rangle\phi \text{ iff } [\![\phi]\!]_M \in E(w)(C).$$

In what follows, we will write $E_w(C)$ for $E(w)(C)$.

Let us now give a brief overview of NCL.

In Broersen et al. (2007), a translation τ from $\mathcal{L}_{\mathtt{CL}}$ to $\mathcal{L}_{\mathtt{NCL}}$ is given such that for all $\phi \in \mathcal{L}_{\mathtt{CL}}$, ϕ is satisfiable in an CL model iff $\tau(\phi)$ is satisfiable in an NCL model. τ is defined as follows: $\tau(p) = p, \tau(\langle\!\lbrack C\rbrack\!\rangle\, \phi) = \langle\emptyset\rangle[C]\mathbf{X}\tau(\phi)$. The main result is then that ϕ is a theorem of CL iff $\tau(\phi)$ is one of NCL. Via completeness of CL and soundness of NCL, it follows that whenever $\tau(\phi)$ is satisfied in an NCL model, then there is a CL model that satisfies ϕ.

We want to make this result more explicit and constructive in order to get a clear view of how the two frameworks are related. We show how to translate pointed NCL models (M, w) into CL models $f(M, w)$ such that for all $\phi \in \mathcal{L}_{\mathtt{CL}}$, $(M, w) \Vdash \tau(\phi)$ iff $f(M, w) \Vdash \phi$.

Proposition 1. For all $\phi \in \mathcal{L}_{\mathtt{CL}}$, if $\tau(\phi)$ is satisfiable in a pointed model M, w of NCL, then ϕ is satisfiable in a model $f(M, w)$ of CL.

Proof. We define f as follows. For $M = \langle W, \mathtt{N}, \{\sim_C \mid C \subseteq \mathtt{N}\}, F_X, \{\leq_j \mid j \in \mathtt{N}\}, V\rangle$, $f(M) := \langle \mathtt{N}, (W, E), V\rangle$, where

$$E_w(C) := \{\{Y | Y \supseteq F_X[[w']_{\sim_C}]\} | w' \in [w]_{\sim_\emptyset}\}.$$

First, we show that E is playable and thus $f(M)$ is a CL model. Liveness follows from the totality of F_X. Termination follows from the closure of $E_w(C)$ under supersets. For N-maximality, let $X \subseteq W$ such that $W \setminus X \notin E_w(\emptyset)$. Then there is some $w' \in [w]_{\sim_\emptyset}$ such that $F_X(w') \in X$. Since $X \supseteq F_X[\{w'\}] = \{F_X(w')\}$, $X \in E_w(\mathtt{N})$. Outcome-monotonicity follows from the closure of $E_w(C)$ under supersets. For Superadditivity, let $X_1, X_2 \subseteq W, C_1, C_2 \subseteq \mathtt{N}$, such that $C_1 \cap C_2 = \emptyset$. Assume that $X_1 \in E_w(C_1)$ and $X_2 \in E_w(C_2)$. Then for all $i \in \{1, 2\}$, $\exists w_i \in [w]_{\sim_\emptyset}$ such that $X_i \supseteq F_X[[w_i]_{\sim_{C_i}}]$. We have that $E_w(C_1 \cup C_2) = \{\{Y | Y \supseteq F_X[[w']_{\sim_{C_1 \cup C_2}}]\} | w' \in [w]_{\sim_\emptyset}\}$. Thus, we have to show that $\exists w^+ \in [w]_{\sim_\emptyset} : X_1 \cap X_2 \supseteq F_X[[w^+]_{\sim_{C_1 \cup C_2}}]$. We have that $w_1 \sim_\emptyset w_2$. Thus, $w_1 \sim_{C_1} \circ \sim_{\overline{C_1}} w_2$ and since $C_1 \cap C_2 = \emptyset$ and thus $C_2 \subseteq \overline{C_1}$, $\sim_{\overline{C_1}} \subseteq \sim_{C_2}$. Then $w_1 \sim_{C_1} \circ \sim_{C_2} w_2$. Thus, $\exists w^+ : w_1 \sim_{C_1} w^+$ and $w^+ \sim_{C_2} w_2$. Thus, $w^+ \in [w_1]_{\sim_{C_1}} \cap [w_2]_{\sim_{C_2}}$ and therefore $[w^+]_{\sim_{C_1}} = [w_1]_{\sim_{C_1}}$ and $[w^+]_{\sim_{C_2}} = [w_2]_{\sim_{C_2}}$. Since $\sim_{C_1 \cup C_2} \subseteq (\sim_{C_1} \cap \sim_{C_2})$, $[w^+]_{\sim_{C_1 \cup C_2}} \subseteq [w^+]_{\sim_{C_1}} \cap [w^+]_{\sim_{C_2}}$. Hence, $F_X[[w^+]_{\sim_{C_1 \cup C_2}}] \subseteq X_1 \cap X_2$, and thus $X_1 \cap X_2 \subseteq E_w(C_1 \cup C_2)$.

This shows that $f(M)$ is a CL model. Now, we show by induction that for all $\phi \in \mathcal{L}_{\mathtt{CL}}$, for an NCL model M, $M, w \Vdash \tau(\phi)$ iff $f(M, w) \Vdash \phi$. The only interesting case is $\phi := \langle\!\lbrack C\rbrack\!\rangle\, \psi$. Let $M, w \Vdash \langle\emptyset\rangle[C]\mathbf{X}\tau(\psi)$. Then there is some $w' \in [w]_{\sim_\emptyset}$ such that for all $w'' \in [w']_{\sim_C}$, $M, F_X(w'') \Vdash \tau(\psi)$. By induction hypothesis, $f(M, F_X(w'')) \Vdash \psi$. Now, $[\![\psi]\!]_{f(M,w)} \in E_w(C)$ follows from the fact that for all $w'' \in [w']_{\sim_C}, f(M, F_X(w'')) \Vdash \psi$. For the other direction, let $f(M, w) \Vdash \langle\!\lbrack C\rbrack\!\rangle\, \psi$. Then, there is some $X \in E_w(C)$ such that $X \subseteq [\![\psi]\!]_{f(M,w)}$. By definition of $f(M, w)$, there is some $w' \in [w]_{\sim_\emptyset}$ such that $X \supseteq F_X[[w']_{\sim_C}]$. Since by inductive hypothesis, $[\![\tau(\psi)]\!]_{M,w} = [\![\psi]\!]_{f(M,w)}$, $X \subseteq [\![\tau(\psi)]\!]_{M,w}$. Hence, $M, w \Vdash \langle\emptyset\rangle[C]\mathbf{X}\tau(\psi)$. □

So, we have shown how to transform NCL models into corresponding CL models, thus shedding some light on the relation between the two frameworks.

On the relation between ABC and ATL.

We will give a translation $tr : \mathcal{L}_{\texttt{ATL}} \rightarrow \mathcal{L}_{\texttt{ABC}}$ such that for any $\phi \in \mathcal{L}_{\texttt{ATL}}$ there is a pointed ATL model $\mathcal{M}, w$ such that $\mathcal{M}, w \Vdash \phi$ iff there is an ABC model $\mathcal{M}', v$ such that $\mathcal{M}', v \Vdash tr(\phi)$. More precisely given a pointed alternating transition systems $\mathcal{M}, w$ with $\mathcal{M} = \langle W, \texttt{N}, \delta, V \rangle$ and $\mathcal{M}, w \Vdash \phi$, we will show how to construct an ABC model $TR(\mathcal{M})$ such that $\mathcal{M}, w \Vdash \phi$ iff $TR(M), f(v) \Vdash tr(\phi)$ where f maps states in the domain of $\mathcal{M}$ to states in the domain of $TR(M)$.

First, we recall an important result that we will need.

Theorem 1 (Goranko and van Drimmelen (2006))**.** Every satisfiable formula $\phi \in \mathcal{L}_{\texttt{ATL}}$ is satisfiable in a finite Concurrent Game System.

Without going into details, Concurrent Games Systems (CGS) are almost the same as alternating transition systems and it is thus easy to give a transformation in both directions for which satisfiability is invariant and for which the size of the target model is bounded:

Corollary 1. Every satisfiable formula $\phi \in \mathcal{L}_{\texttt{ATL}}$ is satisfiable in a finite alternating transition system.

Proof. Starting with some pointed alternating transition system M, w such that $\mathcal{M}, w \Vdash \phi$ we can construct a (possibly infinite) pointed CGS that satisfies this formula. By Goranko and van Drimmelen's theorem, there is a finite pointed CGS satisfying ϕ, which we can translate back as a finite alternating transition system. □

Henceforth, we assume a finite domain of ATL models.

Transforming ATL models into ABC models.

We give a procedural definition of our transformation. First we copy the state space W, the valuation V and the set of agents N. Now, for all pairs $(w, i) \in W \times \texttt{N}$, $\delta(w, i)$ is finite. We label each element in $\delta(w, i)$ with an action name $a^1_{w,i}, \ldots, a^{|\delta(w,i)|}_{w,i}$. Let *Label* be this labeling function. Now for each set of states $X^w_i \in \delta(w, i)$ and for each $v \in X^w_i$ we add the pair (w, v) to $\xrightarrow{i, a^k_{w,i}}$ where $a^k_{w,i}$ is the appropriate label, i.e. $a^k_{w,i} = Label(w, i, X^w_i)$. We define a function $f : Dom(\mathcal{M}) \rightarrow Dom(TR(\mathcal{M}))$ mapping a state to itself.

Translating $\mathcal{L}_{\texttt{ATL}}$ into $\mathcal{L}_{\texttt{ABC}}$

The translation is model-dependent. Given an ATL model $\mathcal{M}$, we define for each agent $j \in \texttt{N}$ a set of actions $A_j = \bigcup_{w \in |\mathcal{M}|} \bigcup_{X^w_j \in \delta(w,j)} Label(w, j, X^w_j)$. The finiteness of of A_j follows from the fact that $|W|$ is finite and $\delta(w, i)$ is finite.

The translation $tr : \mathcal{L}_{\texttt{ATL}} \rightarrow \mathcal{L}_{\texttt{ABC}}$ is recursively defined as follows:

$$
\begin{array}{lcl}
tr(p) & := & p \\
tr(\neg\phi) & := & \neg tr(\phi) \\
tr(\phi \wedge \psi) & := & tr(\phi) \wedge tr(\psi) \\
tr(\langle\langle C \rangle\rangle \mathbf{X}\phi) & := & \bigvee_{\vec{c} \in \times_{j \in C} A_j} [\bigcap_{a_j \in |\vec{c}|} a_j] tr(\phi), C \neq \emptyset \\
tr(\langle\langle \emptyset \rangle\rangle \mathbf{X}\phi) & := & [\bigcup_{j \in \texttt{N}} \bigcup_{a_j \in A_j} a_j] tr(\phi)
\end{array}
$$

Lemma 1. For all $\phi \in \mathcal{L}_{\texttt{ATL}}$ if there exists a pointed ATL model $\mathcal{M}, w$ such that $\mathcal{M}, w \Vdash \phi$ then there exists a reactive $\texttt{ABC}^{\texttt{N}}$ model $\mathcal{M}', v$ such that $M', v \Vdash tr(\phi)$

Proof. We prove that for all $\phi \in \mathcal{L}_{\texttt{ATL}}$ there exists a pointed ATL model $\mathcal{M}, w$ such that $\mathcal{M}, w \Vdash \phi$ iff $TR(M), f(v) \Vdash tr(\phi)$. From left to right. Let an ATL model $\mathcal{M}, w$ be such that $\mathcal{M}, w \Vdash \phi$. We build the action based model according to the above procedure. It is easy to see that satisfiability is preserved for propositional letters and booleans. If $\mathcal{M}, w \Vdash \langle\langle C\rangle\rangle \mathbf{X}\psi$ then there is a collection of strategies (one strategy for each agent in C) $(f_j)_{j\in C}$ such that, for all histories $\lambda \in out(w, (f_j)_{j\in C})$, $\mathcal{M}, \lambda[1] \Vdash \psi$. This means that there is some collection of choices $(X_j)_{j\in C} \in \times_{j\in C}\delta(w, j)$ such that for all $v \in \bigcap_{j\in C}(X_j)_{j\in C}$, $\mathcal{M}, v \Vdash \psi$ (1). Now consider the collection of actions $(Label(w, j, X_j))_{j\in C}$ corresponding to these choices at w. Call it $\vec{c}$. Now assume that $s \in Dom(TR(\mathcal{M}))$ is such that for all $j \in C$, $w \xrightarrow{j,a} s$ whenever $a = \vec{c}(j)$. By construction, $s \in \bigcap_{j\in C}(X_j)_{j\in C}$ and thus by (1) $\mathcal{M}, s \Vdash \psi$. But then $TR(\mathcal{M}), s \Vdash tr(\psi)$ (2). Thus $TR(\mathcal{M}), s \Vdash [\bigcap_{a_j\in|\vec{c}|} a_j]tr(\psi)$. Then we conclude that $TR(\mathcal{M}), s \Vdash \bigvee_{\vec{c}\in\times_{j\in C}A_j}[\bigcap_{a_j\in|\vec{c}|} a_j]tr(\psi)$. □

Lemma 2. $\forall\phi \in \mathcal{L}_{\texttt{ATL}}$: if $\models_{\texttt{ATL}} \phi$ then $\models_{\texttt{ABC}^{NR}} tr(\phi)$.

Proof. We follow the methodology of Broersen et al. (2007), by proving that the translation of all axioms and rules of a complete axiomatization of ATL are valid on the class of reactive ABC frames.

• $(\bot)$. $t(\neg\langle\langle C\rangle\rangle\mathbf{X}\bot) = \neg\bigvee_{\vec{c}\in\times_{j\in C}A_j}[\bigcap_{a_j\in|\vec{c}|} a_j]\bot$. But by the first clause of the definition of reactive ABC models, for any collection of actions for each agent $\vec{a} \in \times_{j\in\texttt{N}}(a_j)_j \in \texttt{N}$ and each state w, $\bigcap_{(a_j)_j\in|\vec{a}|}(\xrightarrow{j,a} [w]) \neq \emptyset$, thus $\mathcal{M}, w \nVdash \bigvee_{\vec{c}\in\times_{j\in C}A_j}[\bigcap_{a_j\in|\vec{c}|} a_j]\bot$.

• $(\top)$. $t(\langle\langle C\rangle\rangle\mathbf{X}\top) = \bigvee_{\vec{c}\in\times_{j\in C}A_j}[\bigcap_{a_j\in|\vec{c}|} a_j]\top$. But for any action profile $\vec{c}$, $[\bigcap_{a_j\in|\vec{c}|} a_j]\top$, so we have just have to ensure that each agent has at least one available action so that $\times_{j\in C}A_j$ is not empty, but this follows from the second clause of the definition of a reactive ABC model.

• (N). $t(\neg\langle\langle\emptyset\rangle\rangle\mathbf{X}\neg\phi \to \langle\langle\texttt{N}\rangle\rangle\mathbf{X}\phi) = (\neg[\cup_{j\in\texttt{N}} \cup_{a_j\in A_j} a_j]\neg tr(\phi)) \to (\bigvee_{\vec{c}\in\times_{j\in\texttt{N}}A_j}[\cap_{a_j\in|\vec{c}|}a_j]tr(\phi)$. Assume that $\mathcal{M}, v \Vdash (\neg[\cup_{j\in\texttt{N}} \cup_{a_j\in A_j} a_j]\neg tr(\phi))$. Then there is some agent $j \in \texttt{N}$ and some action $a_j \in A_j$ such that $\xrightarrow{j,a} [w] \cap ||\phi||^{\mathcal{M}} \neq \emptyset$. Let $v \in\xrightarrow{j,a} [w] \cap ||\phi||^{\mathcal{M}}$. By N-determinacy there is some action profile $\vec{a} \in \times_{j\in\texttt{N}}(a_j)_j \in \texttt{N}$ such that $\bigcap_{(a_j)_j\in|\vec{a}|}(\xrightarrow{j,a} [w]) = \{v\}$. But then $\mathcal{M}, v \Vdash \bigvee_{\vec{c}\in\times_{j\in\texttt{N}}A_j}[\bigcap_{a_j\in|\vec{c}|} a_j]tr(\phi)$.

• (S). $t((\langle\langle C_1\rangle\rangle\mathbf{X}\phi \wedge \langle\langle C_2\rangle\rangle\mathbf{X}\psi) \to \langle\langle C_1 \cup C_2\rangle\rangle\, \mathbf{X}\,(\phi \wedge \psi)) = [\bigvee_{\vec{c}\in\times_{j\in C_1}A_j}[\cap_{a_j\in|\vec{c}|}a_j]tr(\phi) \wedge \bigvee_{\vec{c}\in\times_{j\in C_2}A_j}[\cap_{a_j\in|\vec{c}|}a_j]tr(\psi)] \to [\bigvee_{\vec{c}\in\times_{j\in C_1\cup C_2}A_j}[\cap_{a_j\in|\vec{c}|}a_j]tr(\phi \wedge \psi)$ for $C_1 \cap C_2 = \emptyset$. There are two cases. Case 1: there is no successor for the collection of actions corresponding to the two witness collections of actions for C_1 and C_2. But then $[\bigcap_{a_j\in|\vec{c}|} a_j]tr(\phi \wedge \psi)$ trivially holds for this collection. Case 2: there is a successor but by construction it is both in $||\phi||$ and $||\psi||$. Thus, take this collection as our new witness.

• $(\langle\langle C\rangle\rangle\mathbf{X}mon)$. $\vdash \phi \to \psi$ implies $\vdash \langle\langle C\rangle\rangle\mathbf{X}\phi \to \langle\langle C\rangle\rangle\mathbf{X}\psi$. Straightforward. □

Proposition 2. For all $\phi \in \mathcal{L}_{\texttt{ATL}}$ there exists a pointed ATL model $\mathcal{M}, w$ such that $\mathcal{M}, w \Vdash \phi$ iff there exists an $\texttt{ABC}^{NR}$ model $\mathcal{M}', v$ such that $M', v \Vdash tr(\phi)$.

Proof. Immediate from the two preceding lemmas. □

3.4 Complexity and Expressivity for Expressing Different Notions

This section summarizes the main results that we obtained when investigating how much expressive power and complexity is required for expressing each of the notions on each class of models. As mentioned earlier, we obtain out results by determining under which operations on models (frames) certain properties from GT and SCT are invariant (closed). For the definitions of these operations and the underlying characterization results, the reader is referred to Blackburn et al. (2001), Cate (2005) We start with the simplest notions of coalitional power and preferences.

Simple coalitional power and preference.

The property of a coalition C having the power to ensure that in the next state p will be the case turns out to be invariant under bisimulations on $\wp(\mathtt{N}) - \mathtt{LTS}$ and on $\mathtt{PBC}, \mathtt{NCL}$. Thus, it can be expressed using the respective basic multi-modal languages, i.e. by $\langle C \rangle p$ and $\langle \emptyset \rangle [C] \mathbf{X}$p, respectively. Since the complexity of MC and SAT of these logics is known, for $\wp(\mathtt{N}) - \mathtt{LTS}$ and $\mathtt{PBC}$, we thus get PSPACE and P as upper bounds on SAT and MC of logics expressing the notion. For $\mathtt{NCL}$, the respective upper bounds are NEXPTIME and P. On $\mathtt{ABC}$ models on the other hand, saying that a coalition can achieve something involves the intersection of the relations for the actions for the agents. It is not invariant under bisimulation but under ∩-bisimulation; therefore it can be expressed in the basic language extended with intersection: $\bigvee_{\vec{a_j} \in \vec{C}} [\bigcap \vec{a_j}] p$. The upper bounds on SAT and MC that we obtain are then again PSPACE and P, respectively.

	Invariance	Formula	UB for MC, SAT
$\wp(\mathtt{N}) - \mathtt{LTS}$	Bisimulation	$\langle C \rangle p$	P, PSPACE
$\mathtt{ABC}$	∩-Bisimulation	$\bigvee_{\vec{a_j} \in \vec{C}} [\bigcap \vec{a_j}] p$	P, PSPACE
$\mathtt{PBC}$	Bisimulation	$\langle \emptyset \rangle [C] \mathbf{X} p$	P, PSPACE

Table 2: *"C can ensure that in the next state p is true."*

The simplest preference notion that we consider is that of an agent finding some state at least as good in which p is true. Since, in all our models preferences are represented in the same way and the preference fragments of the different languages we consider are the same, we get the same results for this notion on all the models.

	Inv.	Formula	UB for MC, SAT
$\wp(\mathtt{N}) - \mathtt{LTS}, \mathtt{ABC}, \mathtt{PBC}$	Bis.	$\langle \leq_j \rangle p$	P, PSPACE

Table 3: *"j finds a state a.l.a.g. where p is true."*.

Coalition C **can make agent** j **happy.** The basic combination of coalitional power and individual preference is the ability of a coalition to ensure that the next state will be one that is at least as good for some agent. This property turns out to be easiest to express on $\wp(\mathtt{N})-\mathtt{LTS}$, since here it is invariant under $\cap$-bisimulation. For ABC and PBC on the other hand, we have to express that the states accessible by one relation are a subset of the states accessible by another relation. This is not invariant under any bisimulations but under taking generated submodels and disjoint unions.

	Formula	MC, SAT
$\wp(\mathtt{N})\mathtt{LTS}$	$\langle C\cap \leq_j\rangle\top$	P,PSPACE
ABC	$\bigvee_{\vec{a_j}\in\vec{C}}(\downarrow x.[\bigcap \vec{a_j}](\downarrow y.@_x\langle\leq_j\rangle y))$	PSPACE, Π^0_1
PBC	$\downarrow x.\langle\emptyset\rangle[C]\mathbf{X}\downarrow y.@_x\langle\leq_j\rangle y$	PSPACE, Π^0_1

Table 4: *"C can move the system into a state a.l.a.g. for j"*

Nash-stability. Nash-stability says that no single agent has the power to make the system move into a state that is *strictly* better for him.

	Formula	UB for SAT
$\wp(\mathtt{N})-\mathtt{LTS}$	$\bigwedge_{j\in\mathtt{N}}\downarrow x.[j\cap\leq_j]\langle\leq_i\rangle x$	Π^0_1
ABC	$\bigwedge_{j\in\mathtt{N}}\bigwedge_{a_j\in A_j}\downarrow x.\langle a_j\rangle\langle\leq\rangle x$	EXPTIME
PBC	$\bigwedge_{j\in\mathtt{N}}\downarrow x.[\emptyset]\langle\{j\}\rangle\mathbf{X}\langle\leq\rangle x$	EXPTIME

Table 5: *"The current state is Nash stable."*

On all these models, Nash-stability is invariant under taking generated submodels and disjoint unions, and can be expressed in a modal logic with model checking problem (combined complexity) in PSPACE.

Strong Nash-stability. Strong Nash-stability says that no single agent has the power to make the system move into a state that is *a.l.a.g.* for him. Note that since we take preferences as TPOs, if a state is strongly Nash-stable, it is Nash-stable.

	Formula	UB for SAT
$\wp(\mathtt{N})-\mathtt{LTS}$	$\wedge_{j\in\mathtt{N}}[i\cap\leq_j]\bot$	P, PSPACE
ABC	$\neg\vee_{j\in\mathtt{N}}\vee_{a_j\in A_j}\downarrow x.[a_j]\langle\leq^{-1}\rangle x$	PSPACE, Π^0_1
PBC	$\neg\vee_{j\in\mathtt{N}}\downarrow x.\langle\emptyset\rangle[\{j\}]\mathbf{X}\langle\leq^{-1}\rangle x$	PSPACE, Π^0_1

Table 6: *"The current state is strongly Nash stable."*

On $\wp(\mathtt{N})-\mathtt{LTS}$, strong Nash-stability is invariant under $\cap$-bisimulation. On ABC and PBC only under GSM and DU.

Comparing the results for Nash-stability and strong Nash-stability, we can see that on $\wp(\mathtt{N})-\mathtt{LTS}$ strong Nash-stability is easier to express than Nash-stability whereas on ABC and PBC we get opposite results.

4 Conclusion

Our embeddings results show that action- and power-based models, together with coalition-labelled transition systems, constitute three natural families of cooperation logics with different primitives. The main open problem is to extend action-based models to reason about transitive closure in order to simulate more powerful logics such as $\mathtt{ATL}^*$.

Our invariance results have shown that many social choice-theoretical and game-theoretical notions are not invariant under bounded morphic images, in many cases it is only a matter of allowing the underlying logics to reason about the intersection of two relations. In fact, being able to express intersection seems crucial when reasoning about cooperation of agents in normal ML's.

Our definability results together with known upper bounds on combined complexity of model checking and satisfiability have shown that whether strong or weak stability or efficiency notions are less demanding crucially depends on the choice of primitives. In action- and power-based models expressing the latter type notions turns out to be easier, while in coalition labelled transition systems the situation is just the opposite. This has to do with whether coalitional power can be expressed in a simple way, and thus whether the intersection of relations is sufficient or whether we need to express something like a subset relation.

Our definability results made use of very big conjunctions and disjunctions. When taking conjunctions/disjunctions over all coalitions, they will be exponentially related to the number of agents. The consequences we draw about the upper bounds on the complexity of satisfiability or of combined complexity of model checking is thus to be balanced by the fact that we generally use very big conjunctions or disjunctions that might well be exponential if we take the number of agents as a parameter for the complexity results.

Our invariance results indicate that our definability results are tight to some extent. Indeed, this shows that within a large family of extended modal languages with a natural model-theoretical characterization we could not improve on them. It follows that upper bounds are accurate to some extent. Naturally, it is always possible to design ad hoc logics to express exactly the notion of interest. It leads us to the question of lower bounds. Can we use results from the computational social choice literature to obtain lower bounds on the *data* complexity of model-checking a logic that can express some notion? In general, the difficulty is that usually the results from this literature take e.g. the number of resources (and/or number of agents) as primitives, while the data complexity of a modal logic is usually taken relatively to its state space, which is in general exponentially bigger than the number of resources. It is natural to expect that the interesting hardness results would be for logarithmic complexity classes.

Acknowledgements The authors would like to thank the participants of the *Logics for Dynamics of Information and Preferences Mini-workshop*, the ESSLLI workshop *Logical Methods for Social Concepts* and the *Workshop on Logic, Rationality and Interaction* for their helpful comments on previous versions.

References

R. Alur, T. A. Henzinger, and O. Kupferman. Alternating-time temporal logic. In *Compositionality: The Significant Difference*, volume 1536 of *Lecture Notes in Computer Science*, pages 23–60. Springer-Verlag, 1998.

N. Belnap, M. Perloff, and M. Xu. *Facing the future: Agents and choices in our indeterminist world*. Oxford University Press, Oxford, 2001.

P. Blackburn, M. de Rijke, and Y. Venema. *Modal Logic*. CUP, 2001.

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

J. Broersen, A. Herzig, and N. Troquard. Normal Coalition Logic and its conformant extension. In D. Samet, editor, *TARK'07*, pages 91–101. PUL, 2007. URL http://www.irit.fr/~Andreas.Herzig/P/Tark07ncl.html.

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

C. Dégremont and L. Kurzen. Modal logics for preferences and cooperation: Expressivity and complexity. In *Knowledge Representation for Agents and Multi-Agent Systems, Post-proceedings of KRAMAS 2008*, volume 5605 of *Lecture Notes in Artificial Intelligence*. Springer-Verlag, 2009.

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)*, 2007.

P. Girard. *Modal Logic for Preference Change*. PhD thesis, Stanford, 2008.

V. Goranko and G. van Drimmelen. Complete axiomatization and decidability of alternating-time temporal logic. *Theor. Comput. Sci.*, 353(1-3):93–117, 2006.

M. Pauly. A modal logic for coalitional power in games. *JLC*, 12(1):149–166, 2002.

K. Segerberg. Bringing it about. *Journal of Philosophical Logic*, 18(4):327–347, 1989.

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.

Specifying strategies in terms of their properties

R. Ramanujam and Sunil Simon

The Institute of Mathematical Sciences, Chennai, India; ILLC, University of Amsterdam.
jam@imsc.res.in, s.e.simon@uva.nl

Abstract

Logical analyses of games typically consider players' strategies as atomic objects and the reasoning is about existence of strategies, rather than about strategies themselves. This works well with the underlying assumption that players are rational and possess unbounded computational abilities. However, in many practical situations players have limited computational resources. Thus a prescriptive theory which provides advice to players needs to view strategies as relations constraining players' moves rather than view them as complete functions.

We propose a syntactic framework in which strategies are constructed in a structural manner and show how explicit reasoning of strategies can be achieved. We also look at how structurally specified strategies can be adapted to the case where the game itself has compositional structure. We suggest that rather than analyzing the composite game, one needs to compose game-strategy pairs in order to reason about strategic response of players. We consider a propositional dynamic logic whose programs are regular expressions over such game-strategy pairs and present a complete axiomatization of the logic.

1 Introduction

The central innovation introduced by game theory is its strategic dimension. A player's environment is not neutral, and she expects that other players will try to outguess her plans. Reasoning about such expectations and strategizing one's own response accordingly constitutes the main logical challenge of game theory. Games are defined by sets of rules which specify what moves are available to each player, and according to her own preferences over the possible outcomes, every player plans her strategy. If the game is rich enough, the player has access to a wide range of strategies, and the choice of what strategy to employ in a game situation depends not only on the player's understanding of how the game can proceed from then on, but also based on his expectation of what strategies other players are following.

While this observation holds true for much of game playing, game theory largely consists of reasoning *about* games rather than reasoning *in* games. It is assumed that the entire structure of the game is laid out in front of us, and we reason from above, predicting how rational players would play, and such predictions are summarised into assertions on existence of equilibria. In an ideal world where players have unbounded computational abilities and where rationality is common knowledge, such predictions would be realistic. Players could strategize based on all possible behaviours of others and if optimal strategies exist then they will always be able to deduce these strategies. However, in reality, players are bounded memory agents having limited computational abilities. Much of the theory analysing solution concepts in games assumes that players are rational, have unbounded computational resources and talks only about the existence of stable strategy profiles.

These comments hold true even for finite duration games with perfect information. The classic example of such a game is the game of chess. Using the backward induction algorithm, Zermelo (1913) argued that chess is determined, i.e. either there exists a pure strategy for one of the two players (white or black) guaranteeing that she will always win or each one of the two players has a strategy guaranteeing at least a draw. However, neither do we know which of the three alternatives is the correct one, nor a winning strategy if it exists. For games like Hex, it is known that the first player can force a win Gale (1979) but nonetheless a winning strategy is not known. Theoretically a finite game like chess or hex is not very interesting since the winner can be determined in time linear in the size of the game tree using the backward induction procedure.

As is apparent, existence results are of not much help in advising players on how to play. The situation gets worse in the case of games with overlapping objectives where solution concepts in general look for equilibrium strategy profiles where none of the players gain by unilaterally deviating. In general, such games can have multiple equilibrium profiles and it is not clear which equilibrium the players would try to attain. An equilibrium selection theory was proposed by Harsanyi and Selten (1987) to deal with such situations. The theory models the uncertainty of each player in terms of a belief hierarchy which specifies a player's beliefs about what others play, about what they believe she and others play and so on, ad infinitum. The theory thus makes use of unbounded iteration of beliefs and it is hardly clear whether this matches in any way the reasoning done by players when they actually play a game.

And yet, as Aumann and Dreze (2005) point out, game theory started by trying to develop a prescriptive theory for rational agents. The seminal work of von Neumann and Morgenstern envisaged game theory as constituting advice for players in game situations, so that strategies may be synthesized accordingly. While this was summarily achieved for two person zero sum games, advice functions for multi-player games with overlapping objectives have been hard to come by. Aumann and Dreze (2005) argue that such a prescriptive game theory must account for the beliefs and expectations each player has about strategies followed by other players. The interactive element is crucial and a rational player should then play so as to maximize his utility, given how he thinks the others will play.

We suggest that any prescriptive theory which takes into account the limited computational abilities of players needs to consider strategies as partial plans rather than complete ones. Or in other words, strategies need to be considered as relations constraining players' moves, rather than functions prescribing them uniquely. Thus rather than viewing them as atomic objects, strategies need to be viewed as structured objects built in some compositional fashion. This calls for a syntactic grammar for composition of partial strategies and it also suggests that logical languages designed to reason about composition of programs could provide valuable insight in developing a similar framework for strategies.

Logical study of games have been extensively analysed in the literature. The work on alternating temporal logic Alur et al. (2002) considers selective quantification over paths that are possible outcomes of games in which players and an environment alternate moves. The emphasis is on the existence of a strategy for a coalition of players to force an outcome. In Harrenstein et al. (2003) and van der Hoek et al. (2005), logics are developed to describe equilibrium concepts and for strategic reasoning. Chatterjee et al. (2007) looks at a logic where quantification over strategy terms is part of the logical formalism and study its relationship with alternating temporal logic and other variants. All of the above mentioned logics have the common property that the game arena is taken to be fixed and a functional notion of strategy is adopted. Strategies are taken to be atomic objects whereby the logical structure present within the strategy is not taken into account for analysis.

The idea of taking into account the structure available within strategies and making assertions about a specific strategy leading to a specified outcome is, of course, not new. van Benthem (2001; 2002) uses dynamic logic to describe games as well as strategies. When dealing with finite extensive form games, this approach of describing the complete strategy explicitly in a dynamic logic framework is appropriate, however the technique does not generalise satisfactorily to unbounded duration games.

On the other hand, propositional game logic Parikh (1985), the seminal work on logical aspects of game theory, talks of existence of strategies, but builds composite structure into games. Goranko (2003) looks at an algebraic characterisation of games and presents a complete axiomatization of identities of the basic game algebra. Pauly (2001) has built on this to provide interesting relationships between programs and games, and to describe coalitions to achieve desired goals. Goranko (2001) relates Pauly's coalition logics with work done in alternating temporal logic. In this line of work, the game itself is structurally built from atomic objects. However, the reasoning done is about existence of strategies and not reasoning *with* strategies: the ability of a player to strategize in response to the opponent's actions. Ghosh (2008) presents a complete axiomatisation of a logic describing both games and strategies in a dynamic logic framework, but again the assertions are about atomic strategies.

2 Preliminaries

Game tree

Let N be a finite set of players, we use i to range over this set. let Σ_i for $i \in N$ be a finite set of action symbols which represent moves of players and $\Sigma = \bigcup_{i \in N}$. For technical convenience, we restrict our attention to two player games, i.e. we take $N = \{1, 2\}$. We use the notation i and $\bar{\imath}$ to denote the players where $\bar{\imath} = 2$ when $i = 1$ and $\bar{\imath} = 1$ when $i = 2$.

Let $(S, \Longrightarrow, s_0)$ be a finite tree rooted at s_0 on the set of vertices S and $\Longrightarrow$: $(S \times \Sigma) \rightarrow S$. An *extensive form game tree* is given by $T = (S, \Longrightarrow, s_0, \lambda)$ where S is the set of game positions and s_0 is the initial game position. For a game position $s \in S$, let $\vec{s} = \{s' \in S \mid s \xrightarrow{a} s'$ for some $a \in \Sigma\}$. A game position s is a leaf node (or terminal node) if $\vec{s} = \emptyset$, let S^{leaf} denote the set of all leaf nodes of T. The turn function $\lambda : S \rightarrow N$ associates each game position with a player.

Technically we need player labelling only at the non-leaf nodes. However, for the sake of uniform presentation, we do not distinguish between leaf nodes and non-leaf nodes as far as player labelling is concerned.

Figure 1(a) shows an example game tree. Here nodes are labelled with the players and edges represents the actions. A *play* in T is a finite path $\rho : s_0 \overset{a_0}{\Longrightarrow} s_1 \cdots \overset{a_k}{\Longrightarrow} s_k$ where s_k is a leaf node.

A *strategy* for player i, is a subtree of T where for each player i node, there is a unique outgoing edge and for player $\bar{\imath}$, every move is included. Figure 1(b) shows a strategy for player i in the game tree Figure 1(a). For $i \in N$, let Ω^i denote the set of all strategies for player i in the game. For a tree T, let *frontier*(T) denote the set of all leaf nodes of T.

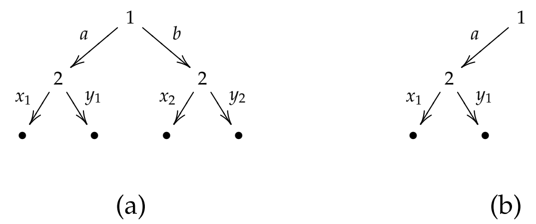


Figure 1: Game and strategy.

The formulas of the logic refer to extensive form game trees. One convenient way of representing the tree is to specify it in the following syntax.

Syntax for game trees:

Let *Nodes* be a finite set. The finite game structure is specified using the syntax:

$$\mathrm{G} := (i, x) \mid \Sigma_{a_m \in J}((i, x), a_m, t_{a_m})$$

where $J \subseteq \Sigma_i$, $x \in Nodes$, $i \in N$ and $t_{a_m} \in \mathrm{G}$.

Given $g \in \mathrm{G}$ we define the tree T_g generated by g inductively as follows.

- $g \equiv (i, x)$: $T_g = (S_g, \Longrightarrow_g, \lambda_g, s_{g,0})$ where $S_g = \{s_x\}$, $\lambda_g(s_x) = i$ and $s_{g,0} = s_x$.
- $g \equiv ((i, x), a_1, t_{a_1}) + \cdots + ((i, x), a_k, t_{a_k})$: Inductively we have trees $T_1, \ldots T_k$ where for $j : 1 \leq j \leq k$, $T_j = (S_j, \Longrightarrow_j, \lambda_j, s_{j,0})$. Define $T_g = (S_g, \Longrightarrow_g, \lambda_g, s_{g,0})$ where
  - $S_g = \{s_x\} \cup S_{T_1} \cup \ldots \cup S_{T_k}$ and $s_{g,0} = s_x$.
  - $\lambda_g(s_x) = i$ and for all j, for all $s \in S_{T_j}$, $\lambda_g(s) = \lambda_j(s)$.

  The edge relation is the union of the edge relation on the individual tree along with the edges $s_x \overset{a_j}{\Longrightarrow}_g s_{j,0}$ for $j : 1 \leq j \leq k$.

3 Strategy specification

We give a syntax to specify strategies in a structured manner. Atomic strategy formulas specify, for a player, what conditions she tests for before making a move. We consider the case when these conditions are simply boolean formulas. Composite strategy specifications are built from atomic ones using connectives (without negation). We use an implication of the form: "if the opponent's play conforms to a strategy π then play σ". This connective is crucial to capture the notion of players strategizing in response to opponents actions.

For a countable set of propositions P^i, let $\Psi(P^i)$ be the boolean formulas over P^i built using the following syntax:

$$\Psi(P^i) := p \in P^i \mid \neg\psi \mid \psi_1 \vee \psi_2.$$

For $i \in N$, let $Strat^i(P^i)$ be the set of strategy specifications given by the following syntax:

$$Strat^i(P^i) := [\psi \mapsto a]^i \mid \sigma_1 + \sigma_2 \mid \sigma_1 \cdot \sigma_2 \mid \pi \Rightarrow \sigma$$

where $\pi \in Strat^{\bar{\imath}}(P^1 \cap P^2)$, $\psi \in \Psi(P^i)$ and $a \in \Sigma_i$.

The idea is to use the above constructs to specify properties of strategies. For instance the interpretation of a player i specification $[p \mapsto a]^i$ will be to choose move "a" for every i node where p holds. Consider the game given in Figure 1 (a). Suppose the proposition p holds at the root, then the strategy depicted in Figure 1 (b) conforms to the specification $[p \mapsto a]^1$.

The specification $\pi \Rightarrow \sigma$ says, at any node player i sticks to the specification given by σ if on the history of the play, all moves made by $\bar{\imath}$ conform to π. In strategies, this captures the aspect of players actions being responses to the opponent's moves. The opponent's complete strategy may not be available, the player makes a choice taking into account the apparent behaviour of the opponent on the history of play.

Let $\Sigma_i = \{a_1, \ldots, a_m\}$, we use the abbreviation $null^i = [\top \mapsto a_1] + \cdots + [\top \mapsto a_m]$. The intuitive meaning is, any strategy of player i conforms to $null^i$.

Semantics:

Given a state u and a valuation $V : u \to 2^P$, the truth of a formula $\psi \in \Psi(P^i)$ is defined as follows:

- $u \models p$ iff $p \in V(u)$.
- $u \models \neg\psi$ iff $u \not\models \psi$.
- $u \models \psi_1 \vee \psi_2$ iff $u \models \psi_1$ or $u \models \psi_2$.

We consider game trees along with a valuation function $V : S \to 2^P$. Given a strategy μ of player i and a node $s \in \mu$, let $\rho_s : s_0 a_0 s_1 \cdots s_m = s$ be the unique path in μ from the root node to s. For all $j : 0 \leq j < m$, let $out_{\rho_s}(s_j) = a_j$ and $out_{\rho_s}(s)$ be the unique outgoing edge in μ at s. For a strategy specification $\sigma \in Strat^i(P^i)$, we define when μ *conforms* to σ (denoted $\mu \models_i \sigma$) as follows:

- $\mu \models_i \sigma$ iff for all player i nodes $s \in \mu$, we have $\rho_s, s \models_i \sigma$

where we define $\rho_s, s_j \models_i \sigma$ for any s_j in ρ_s as,

- $\rho_s, s_j \models_i [\psi \mapsto a]^i$ iff $s_j \models \psi$ implies $out_{\rho_s}(s_j) = a$.
- $\rho_s, s_j \models_i \sigma_1 + \sigma_2$ iff $\rho_s, s_j \models_i \sigma_1$ or $\rho_s, s_j \models_i \sigma_2$.
- $\rho_s, s_j \models_i \sigma_1 \cdot \sigma_2$ iff $\rho_s, s_j \models_i \sigma_1$ and $\rho_s, s_j \models_i \sigma_2$.
- $\rho_s, s_j \models_i \pi \Rightarrow \sigma$ iff for all player $\bar{\imath}$ nodes $s_k \in \rho_s$ such that $k \leq j$, if $\rho_s, s_k \models_{\bar{\imath}} \pi$ then $\rho_s, s_j \models_i \sigma$.

Above, $\pi \in Strat^{\bar{\imath}}(P^1 \cap P^2)$ and $\psi \in \Psi(P^i)$.

4 Reasoning about strategies

We present a logic to reason about strategies with respect to a single extensive form game tree g. Strategy specifications are employed in the formulas of the logic to partially specify strategies rather than giving a complete description.

Syntax:

Let $g \in$ G be an extensive form game tree. The syntax of the logic is given by:

$$\Phi := p \in P \mid \neg\alpha \mid \alpha_1 \vee \alpha_2 \mid \langle(g, \sigma)\rangle\gamma$$

where $i \in N$, $\sigma \in Strat^i(P^i)$ and $\gamma \in \Psi(P)$.

The intuitive meaning of $\langle(g, \sigma)\rangle\gamma$ is: in the game g, the player has a strategy conforming to the specification σ which ensures γ. Since we are considering a fixed game g, this implies that γ holds at all the leaf node of the appropriate strategy. The restriction of γ to boolean formulas over the set of propositions is due to this reason. Nesting of the modality $\langle(g, \sigma)\rangle$ does not make sense for a fixed game. At a later stage we will look at composing games at which point γ can be taken to be any arbitrary formula.

Semantics:

The model $M = (T_g, V)$ where $T_g = (S, \Longrightarrow, s_0, \lambda)$ is the extensive form game tree associated with g and V is the valuation function $V : S \to 2^P$.

The truth of a formula $\alpha \in \Phi$ in a model M and a position s (denoted $M, s \models \alpha$) is defined as follows:

- $M, s \models p$ iff $p \in V(s)$.
- $M, s \models \neg\alpha$ iff $M, s \not\models \alpha$.
- $M, s \models \alpha_1 \vee \alpha_2$ iff $M, s \models \alpha_1$ or $M, s \models \alpha_2$.
- $M, s \models \langle(g, \sigma)\rangle\gamma$ iff $\exists\mu \in \Omega^i$ such that $\mu \models_i \sigma$ and for all $s' \in frontier(\mu)$, $M, s' \models \gamma$.

The formula $\langle(g, \sigma)\rangle\gamma$ says that there exists a strategy for player i conforming to μ such that all the leaf nodes satisfy γ. The dual $[(g, \sigma)]\gamma$ says that for all strategies of player i conforming to σ, there exists a leaf node which satisfy γ.

Strategy comparison

Consider the formula $\langle(g, null^i)\rangle\gamma$. The formula asserts that player i can ensure the reward γ no matter what player $\bar{\imath}$ does. This makes no reference to *how* player i may achieve this objective, and thus is similar to assertions in most game logics. Now consider the formula $\langle(g, \sigma)\rangle\gamma$. This says something stronger: that there exists a strategy μ satisfying σ for player i such that irrespective of what player $\bar{\imath}$ plays, γ is guaranteed. Here, the mechanism μ used by player i to ensure γ is specified by the property σ.

The extensive form game tree g merely defines the rules of how the game progresses and terminates. However, to compare strategies of players, we need to specify the objectives. For $i \in N$, let R_i be a finite set of rewards for player i, $\preceq^i \subseteq R_i \times R_i$, be a preference ordering on R_i and let $R = R_1 \times R_2$. Let the payoff function $payoff : S^{leaf} \to R$ associate each leaf node with a reward. For a leaf node s, and $payoff(s) = (r_1, r_2)$, let $payoff(s)[i]$ denote the i'th component of r, i.e. $payoff(s)[1] = r_1$ and $payoff(s)[2] = r_2$.

In order to refer to rewards of the players in formulas of the logic, we use special propositions to code them up. This is similar to the approach adopted in Bonanno (2002). Without loss of generality assume that $r_1^1 \preceq^1 r_1^2 \preceq^1 \cdots \preceq^1 r_1^l$. Let $\Theta_1 = \{\theta_1^1, \ldots, \theta_1^l\}$ be a set of special propositions used to encode the rewards in the logic, i.e. θ_1^j corresponds to the reward r_1^j. Likewise for player 2, corresponding to the set R_2, we have a set of propositions Θ_2. The valuation function satisfies the condition:

- For all states s, for $i \in N$, $\{\theta_i^1, \ldots, \theta_i^j\} \subseteq V(s)$ iff $payoff(s)[i] = r_i^j$.

The preference ordering on the rewards for each player is simply inherited from the implication available in the logic.

Coming to the notion of strategy comparison, we say that σ is better for player i than σ' if the following condition holds: irrespective of what player $\bar{\imath}$ plays if there exists a strategy μ' satisfying σ' such that θ_i is guaranteed, then there also exists a strategy μ satisfying σ which guarantees θ_i. This can be expressed by the formula,

$$BT^i(\sigma, \sigma') \equiv \bigwedge_{\theta_i \in \Theta_i} (\langle(g, \sigma')\rangle\theta_i \supset \langle(g, \sigma)\rangle\theta_i)$$

Given a finite set of strategy specifications Υ^i for player i, we say that σ is the best strategy if the following holds:

$$Best^i(\sigma) \equiv \bigwedge_{\sigma' \in \Upsilon^i} BT^i(\sigma, \sigma')$$

Note that in the case of a finite extensive form game tree, we can code up the game positions uniquely using propositions. In this case, it is possible to represent a complete strategy in terms of a strategy specification. At each game position, it specifies a unique action. Suppose the number of player i game positions are k and the proposition $p_i^1, \ldots p_i^k$ uniquely identifies all of these positions, then the specification representing a complete strategy would have the form $\sigma \equiv [p_i^1 \mapsto a_1] \cdots [p_i^k \mapsto a_k]$. In this particular scenario, the notion of strategy comparison and best strategy reduces to the classical notions by taking the set Υ^i to be the set of all strategies for player i.

5 Composition of game - strategy pairs

In the previous section we looked at strategies being defined by their properties. Strategy specifications are structurally built and the reasoning performed was with respect to one fixed extensive form game tree. Instead of working with a single game, we can look at complex games arising out of composition of these atomic games. In this context, we argue that reasoning about game - strategy pairs and their composition is more useful than composing games and analysing strategies separately. Here we present a logic to reason about game - strategy pairs. Both strategy specification and game structure is embedded into the syntax of the logic.

The logic

The logic is a simple dynamic logic where we take regular expressions over game-strategy pairs as programs in the logic.The formulas of the logic can then be used to specify the result of a player following a particular strategy in a specified game enabled at a state.

Syntax:

The syntax of the logic is given by:

$$\Phi := p \in P \mid \neg\alpha \mid \alpha_1 \vee \alpha_2 \mid \langle\xi\rangle\alpha$$

where $\xi \in \Gamma$, the set Γ consists of game strategy pairs which is defined below.

The construct ξ represents regular expressions over game-strategy pairs (g, σ). For the atomic construct (g, σ) the intuitive meaning of the formula $\langle g, \sigma\rangle\alpha$ is: in game g player i has a strategy μ conforming to the specification σ such that α holds at all leaf nodes reached by following μ. In other words, the strategy μ ensures the outcome α.

Game strategy pairs:

Syntax for game strategy specification pair is given by:

$$\Gamma := (g, \sigma) \mid \xi_1; \xi_2 \mid \xi_1 \cup \xi_2 \mid \xi^*$$

where $g \in \mathrm{G}$, $\sigma \in Strat^i(P^i)$.

The atomic construct (g, σ) as mentioned in the earlier section, specifies that in game g a strategy conforming to specification σ is employed. Game strategy pairs are then composed using standard dynamic logic connectives. $\xi_1 + \xi_2$ would mean playing ξ_1 or ξ_2. Sequencing in our setting is does not mean the usual relational composition of games. Rather, it is the composition of game strategy pairs of the form $(g_1, \sigma_1); (g_2, \sigma_2)$. This is where the extensive form game tree interpretation makes the main difference. Since the strategy specifications are intended to be partial, a pair (g, σ) gives rise to a set of possibilities and therefore composition over these trees need to be performed. ξ^* is the iteration of the ';' operator.

Model:

The formulas of the logic express properties about game trees and strategies which are composed using tree regular expressions. These formulas are to be interpreted on game positions and they make assertions about the frontier of the game trees which results from the pruning performed as dictated by the strategy specification. Therefore the models of the logic are game trees, but this can potentially be an infinite set of finite game trees. Alternatively, we can think of these game trees as being obtained from unfoldings of a Kripke structure. As we will see later, the logic cannot distinguish between these two.

A model $M = (W, \longrightarrow, \lambda, V)$ where W is the set of states (or game positions), the relation $\longrightarrow \subseteq W \times \Sigma \times W$, player labelling $\lambda : W \to N$ and $V : W \to 2^P$.

The truth of a formula $\alpha \in \Phi$ in a model M and a position w (denoted $M, w \models \alpha$) is defined as follows:

- $M, w \models p$ iff $p \in V(w)$.
- $M, w \models \neg\alpha$ iff $M, w \not\models \alpha$.
- $M, w \models \alpha_1 \vee \alpha_2$ iff $M, w \models \alpha_1$ or $M, w \models \alpha_2$.
- $M, w \models \langle\xi\rangle\alpha$ iff $\exists(w, X) \in R_\xi$ such that $\forall w' \in X$ we have $M, w' \models \alpha$.

In the semantics of $\langle\xi\rangle\alpha$, the state w can be thought of as the starting game position and X, the set of leaf nodes of the game. We require that the player has a strategy confirming to the specification to ensure that α holds in all of the leaf nodes.

For $\xi \in \Gamma$, we have $R_\xi \subseteq W \times 2^W$. To define the relation formally, let us first assume that R is defined for the atomic case, namely when $\xi = (g, \sigma)$. The semantics for composite game strategy pairs is given as follows:

- $R_{\xi_1;\xi_2} = \{(u, X) \mid \exists Y = \{v_1, \ldots, v_k\}$ such that $(u, Y) \in R_{\xi_1}$ and $\forall v_j \in Y$ there exists $X_j \subseteq X$ such that $(v_j, X_j) \in R_{\xi_2}$ and $\bigcup_{j=1,\ldots,k} X_j = X\}$.
- $R_{\xi_1 \cup \xi_2} = R_{\xi_1} \cup R_{\xi_2}$.
- $R_{\xi^*} = \bigcup_{n \geq 0} (R_\xi)^n$.

In the atomic case when $\xi = (g, \sigma)$ we want a pair (u, X) to be in R_ξ if the game g is enabled at state u and there is a strategy conforming to the specification σ such that X is the set of leaf nodes of the strategy. In order to make this precise, we will require the following notations and definitions.

Restriction on trees:

For $w \in W$, let T_w denote the tree unfolding of M starting at w. Given a state w and $g \in$ G, let $T_w = (S_M^w, \Longrightarrow_M, \lambda_M, s_w)$ and $T_g = (S_g, \Longrightarrow_g, \lambda_g, s_{g,0})$. The restriction of T_w with respect to the game g (denoted $T_w \upharpoonright g$) is the subtree of T_w which is generated by the structure specified by T_g. The restriction is defined inductively as follows: $T_w \upharpoonright g = (S, \Longrightarrow, \lambda, s_0, f)$ where $f : S \to S_g$. Initially $S = \{s_w\}$, $\lambda(s_w) = \lambda_M(s_w)$, $s_0 = s_w$ and $f(s_w) = s_{g,0}$.

For any $s \in S$, let $f(s) = t \in S_g$. Let $\{a_1, \ldots, a_k\}$ be the outgoing edges of t, i.e. for all $j : 1 \leq j \leq k$, $t \xrightarrow{a_j}_g t_j$. For each a_j, let $\{s_j^1, \ldots, s_j^m\}$ be the nodes in S_M^w

such that $s \stackrel{a_j}{\Longrightarrow}_M s_j^l$ for all $l : 1 \leq l \leq m$. Add nodes $s_j^1, \ldots, s_j^m$ to S and the edges $s \stackrel{a_j}{\Longrightarrow} s_j^l$ for all $l : 1 \leq l \leq m$. Also set $\lambda(s_j^l) = \lambda_M(s_j^l)$ and $f(s_j^l) = t_j$.

We say that a game g is enabled at w (denoted *enabled*(g, w)) if the tree $T_w \restriction g = (S, \Longrightarrow, \lambda, s_0, f)$ has the following property:

- $\forall s \in S$, $\lambda(s) = \lambda_g(f(s))$ and $\vec{s} = \overrightarrow{f(s)}$.

For a game tree T, let $\Omega^i(T)$ denote the set of strategies of player i on the game tree T and *frontier*(T) denote the set of all leaf nodes of T.

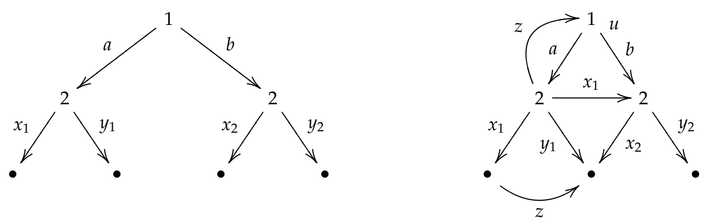
Atomic game-strategy pair:

For atomic game-strategy pair $\xi = (g, \sigma)$ we define R_ξ as follows:
Let g be the game with a single node $g = (i, x)$,

- $R_{(g,\sigma)} = \{(u, \{u\})\}$ if *enabled*(g, u) holds, for all $i \in N$, for all $\sigma \in Strat^i(P^i)$.

For $g = ((i, x), a_1, t_{a_1} + \ldots + (i, x), a_k, t_{a_k})$

- $R_{(g,\sigma)} = \{(u, X) \mid$ *enabled*(g, u) and $\exists \mu \in \Omega^i(T_u \restriction g)$ such that $\mu \models_i \sigma$ and *frontier*$(\mu) = X\}$.



(a): T_g (b): T_w

Figure 2: Extensive form game tree and Kripke structure M

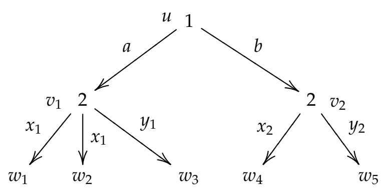


Figure 3: Restriction of T_w to T_g at u

Example 1. Let the extensive form game g be the one given in Figure 2(a) and the Kripke structure M be as shown in Figure 2(b). For the node u of the structure the restriction $T_u \restriction g$ is shown in Figure 3. This is the maximal subtree of T_u according to the structure dictated by g. For instance at node v_1 there are two x_1 labelled edges present in M and therefore both have to be included in $T_u \upharpoonright g$ as well.

Now consider the player 1 strategy specification $\sigma = null^1$. At node u, the choice "a" can ensure player 1 the states $\{w_1, w_2, w_3\}$ and the choice "b" can ensure the states $\{w_4, w_5\}$. Therefore we have the relation $R_{(g,\sigma)} = \{(u, \{w_1, w_2, w_3\}), (u, \{w_4, w_5\}), (v_1, \{w_1, w_2, w_3\}), (v_2, \{w_4, w_5\})\}$.

Suppose $M, u \models p$ and consider the specification $\sigma = [p \mapsto a]^1$. Since p holds at the root, player 1 is restricted to make the choice "a" at u. Hence the relation in this case would be $R_{(g,\sigma)} = \{(u, \{w_1, w_2, w_3\}), (v_1, \{w_1, w_2, w_3\}), (v_2, \{w_4, w_5\})\}$.

Example 2. For a game g and a specification σ of player i, the formula $\langle(g, \sigma)\rangle\alpha$ asserts that the game g is enabled and player i has a strategy in g conforming to σ to ensure α.

The logic is also powerful enough to assert the non-existence of strategies for a player with respect to ensuring an outcome α. For a game g, consider the formula

- $\alpha' = \langle(g, null^i)\rangle\top \wedge \neg\langle(g, null^i)\rangle\alpha$.

The first conjunct $\langle(g, null^i)\rangle\top$ asserts the fact that game g is enabled. Given that g is enabled, the only way $\neg\langle(g, null^i)\rangle\alpha$ can be true is if player i does not have a strategy conforming to $null^i$ which ensures α. Recall that any strategy of player i conforms to $null^i$. Thus α' holds at a state u iff player i does not have a strategy at u that ensures the objective α.

6 Axiom system

We now present an axiomatization of the valid formulas of the logic. For a set $A = \{a_1, \dots, a_k\} \subseteq \Sigma$, we will use the notation $\Re(i, x, A)$ to denote the game $((i, x), a_1, t_{a_1} + \cdots + (i, x), a_k, t_{a_k})$.

We also make use of the following abbreviations:

- Let $g_a^i = ((i, x), a, (j, y))$ and $g_a^{\bar{\imath}} = ((\bar{\imath}, x), a, (j, y))$,
  - $\langle a\rangle\alpha = \langle(g_a^i, [\top \mapsto a]^i) \cup (g_a^{\bar{\imath}}, [\top \mapsto a]^{\bar{\imath}}\rangle\alpha$.

It can be easily verified that this gives the standard interpretation for $\langle a\rangle\alpha$, i.e. $\langle a\rangle\alpha$ holds at a state u iff there is a state w such that $u \xrightarrow{a} w$ and α holds at w.

For game g, we use the formula $g^{\surd}$ to denote that the game structure g is enabled. This is defined as:

- For $g = (i, x)$, let $g^{\surd} = \top$.

- For $g = \Re(i, x, A)$, let
  - $g^{\surd} = \mathbf{turn}_i \wedge (\bigwedge_{j=1,\dots,k}(\langle a_j\rangle\top \wedge [a_j]t_{a_j}^{\surd}))$.

The axiom schemes

(A1) Propositional axioms:

(a) All the substitutional instances of tautologies of PC.

(b) $turn_i \equiv \neg turn_{\bar{\imath}}$.

(A2) Axiom for single edge games:

(a) $\langle a \rangle(\alpha_1 \vee \alpha_2) \equiv \langle a \rangle \alpha_1 \vee \langle a \rangle \alpha_2$.

(A3) Dynamic logic axioms:

(a) $\langle \xi_1 \cup \xi_2 \rangle \alpha \equiv \langle \xi_1 \rangle \alpha \vee \langle \xi_2 \rangle \alpha$.

(b) $\langle \xi_1 ; \xi_2 \rangle \alpha \equiv \langle \xi_1 \rangle \langle \xi_2 \rangle \alpha$.

(c) $\langle \xi^* \rangle \alpha \equiv \alpha \vee \langle \xi \rangle \langle \xi^* \rangle \alpha$.

(A4) $\langle g, \sigma \rangle \alpha \equiv g^{\surd} \wedge push(g, \sigma, \alpha)$.

Inference rules

$$(MP)\ \frac{\alpha,\ \ \alpha \supset \beta}{\beta} \qquad (NG)\ \frac{\alpha}{[a]\alpha}$$

$$(IND)\ \frac{\langle \xi \rangle \alpha \supset \alpha}{\langle \xi^* \rangle \alpha \supset \alpha}$$

Axiom (A2a) does not hold for general game strategy pairs (i.e. $\xi \in \Gamma$). In particular $\langle \xi \rangle(\alpha_1 \vee \alpha_2) \supset \langle \xi \rangle \alpha_1 \vee \langle \xi \rangle \alpha_2$ is not valid. However (A2a) is sound since $\langle a \rangle$ asserts properties about a single edge.

Since the relation R is synthesised over tree structures, the interpretation of sequential composition is quite different from the standard one. Consider the usual relation composition semantics for $R_{\xi_1;\xi_2}$, i.e. $R_{\xi_1;\xi_2} = \{(u, X) | \exists Y$ such that $(u, Y) \in R_{\xi_1}$ and for all $v \in Y$, $(v, X) \in R_{\xi_2}\}$. It is easy to see that under this interpretation the formula $\langle \xi_1 \rangle \langle \xi_2 \rangle \alpha \supset \langle \xi_1 ; \xi_2 \rangle \alpha$ is not valid.

Proposition 1. The formula $\langle \xi_1 ; \xi_2 \rangle \alpha \equiv \langle \xi_1 \rangle \langle \xi_2 \rangle \alpha$ is valid.

Proof. Suppose $\langle \xi_1 ; \xi_2 \rangle \alpha \supset \langle \xi_1 \rangle \langle \xi_2 \rangle \alpha$ is not valid. Then there exists M and u such that $M, u \models \langle \xi_1 ; \xi_2 \rangle \alpha$ and $M, u \not\models \langle \xi_1 \rangle \langle \xi_2 \rangle \alpha$. Since $M, u \models \langle \xi_1 ; \xi_2 \rangle \alpha$, from semantics we have there exists $(u, X) \in R_{\xi_1;\xi_2}$ such that $\forall w \in X$, $M, w \models \alpha$. From definition of R, $\exists Y = \{v_1, \ldots, v_k\}$ such that $(u, Y) \in R_{\xi_1}$ and $\forall v_j \in Y$ there exists $X_j \subseteq X$ such that $(v_j, X_j) \in R_{\xi_2}$ and $\bigcup_{j=1,\ldots,k} X_j = X$. Therefore we get $\forall v_k \in Y$, $M, v_k \models \langle \xi_2 \rangle \alpha$ and hence from semantics, $M, u \models \langle \xi_1 \rangle \langle \xi_2 \rangle \alpha$. This gives the required contradiction.

Suppose $\langle \xi_1 \rangle \langle \xi_2 \rangle \alpha \supset \langle \xi_1 ; \xi_2 \rangle \alpha$ is not valid. Then there exists M and u such that $M, u \models \langle \xi_1 \rangle \langle \xi_2 \rangle \alpha$ and $M, u \not\models \langle \xi_1 ; \xi_2 \rangle \alpha$. We have $M, u \models \langle \xi_1 \rangle \langle \xi_2 \rangle \alpha$ iff there exists $(u, Y) \in R_{\xi_1}$ such that $\forall v_k \in Y$, $M, v_k \models \langle \xi_2 \rangle \alpha$. $M, v_k \models \langle \xi_2 \rangle \alpha$ iff there exists $(v_k, X_k) \in R_{\xi_2}$ such that $\forall w_k \in X_k$, $M, w_k \models \alpha$. Let $X = \bigcup_k X_k$, from definition of R we get $(u, X) \in R_{\xi_1;\xi_2}$. Hence from semantics $M, u \models \langle \xi_1 ; \xi_2 \rangle \alpha$. □

Definition of *push*:

For all $i \in N$, $g \in \mathrm{G}$, $\sigma \in Strat^i(P^i)$ and $\alpha \in \Phi$, we define $push(g, \sigma, \alpha)$ as follows. We have various cases depending on the structure of g.

The case when g is an atomic game, i.e. $g = (i, x)$, for all $i \in N$ and $\sigma \in Strat^i(P^i)$ we have,

(C1) $push(g, \sigma, \alpha) = \alpha$.

Suppose $g = \Re(i, x, A)$ for $A = \{a_1, \ldots, a_k\}$. For each $a_m \in A$ let $g_{a_m} = ((i, x), a_m, (j_m, y_m))$, where (j_m, y_m) is the root of t_{a_m}.

For $\pi = [\psi \mapsto a]^{\bar{\imath}}, \pi_1 + \pi_2, \pi_1 \cdot \pi_2 \in Strat^{\bar{\imath}}(P^{\bar{\imath}})$.

(C2) $push(g, \pi, \alpha) = \bigwedge_{a_m \in A} [a_m]\langle t_{a_m}, \pi\rangle\alpha$.

(C3) $push(g, \sigma \Rightarrow \pi, \alpha) =$
$\bigwedge_{a_m \in A} (\langle g_{a_m}, \sigma\rangle\top \supset [a_m]\langle t_{a_m}, \sigma \Rightarrow \pi\rangle\alpha) \wedge \neg\langle g_{a_m}, \sigma\rangle\top \supset [a_m]\langle t_{a_m}, null^{\bar{\imath}}\rangle\alpha)$.

(C4) $push(g, [\psi \mapsto a]^i, \alpha) = (\psi \supset \langle a\rangle\langle t_a, [\psi \mapsto a]^i\rangle\alpha)$
$\wedge(\neg\psi \supset (\bigvee_{a_m \in A} \langle a_m\rangle\langle t_{a_m}, [\psi \mapsto a]^i\rangle\alpha))$.

(C5) $push(g, \sigma_1 \cdot \sigma_2, \alpha) = \bigvee_{a_m \in A} (\langle g_{a_m}, \sigma_1\rangle\langle t_{a_m}, \sigma_1 \cdot \sigma_2\rangle\alpha$
$\wedge\langle g_{a_m}, \sigma_2\rangle\langle t_{a_m}, \sigma_1 \cdot \sigma_2\rangle\alpha)$.

(C6) $push(g, \sigma_1 + \sigma_2, \alpha) = \bigvee_{a_m \in A} (\langle g_{a_m}, \sigma_1\rangle\langle t_{a_m}, \sigma_1 + \sigma_2\rangle\alpha$
$\vee\langle g_{a_m}, \sigma_2\rangle\langle t_{a_m}, \sigma_1 + \sigma_2\rangle\alpha)$.

(C7) $push(g, \pi \Rightarrow \sigma, \alpha) = \bigvee_{a_m \in A}(\langle g_{a_m}, \sigma\rangle\langle t_{a_m}, \pi \Rightarrow \sigma\rangle\alpha)$.

The soundness of axiom (A4) can be verified by analysing the various cases listed above.

7 Completeness

To show completeness, we prove that every consistent formula is satisfiable. Let α_0 be a consistent formula, and $CL(\alpha_0)$ denote the subformula closure of α. Let $\mathcal{AT}(\alpha_0)$ be the set of all maximal consistent subsets of $CL(\alpha_0)$, referred to as atoms. We use u, w to range over the set of atoms. Each $u \in \mathcal{AT}$ is a finite set of formulas, we denote the conjunction of all formulas in u by $\widehat{u}$. For a nonempty subset $X \subseteq \mathcal{AT}$, we denote by $\widetilde{X}$ the disjunction of all $\widehat{u}, u \in X$. Define a transition relation on $\mathcal{AT}(\alpha_0)$ as follows: $u \xrightarrow{a} w$ iff $\widehat{u} \wedge \langle a\rangle\widehat{w}$ is consistent. The valuation V is defined as $V(w) = \{p \in P \mid p \in w\}$ and $\lambda(w) = i$ iff $turn_i \in w$. The model $M = (W, \longrightarrow, \lambda, V)$ where $W = \mathcal{AT}(\alpha_0)$. Once the Kripke structure is defined, the game theoretic semantics given earlier defines the relation $R_{(g,\sigma)}$ on $W \times 2^W$ for $g \in \mathrm{G}$ and a strategy specification σ.

Lemma 1. For all $g \in \mathrm{G}$, for all $i \in N$ and $\sigma \in Strat^i(P^i)$, for all $X \subseteq W$ and for all $u \in W$ the following holds:

1. if $(u, X) \in R_{(g,\sigma)}$ then $\widehat{u} \wedge \langle g, \sigma\rangle\widetilde{X}$ is consistent.

2. if $\widehat{u} \wedge \langle g, \sigma\rangle\widetilde{X}$ is consistent then there exists $X' \subseteq X$ such that $(u, X') \in R_{(g,\sigma)}$.

Proof. By induction on the structure of (g, σ).

For atomic game $g = (i, x)$, from axiom (A4) case (C1) we get $\langle(i, x), \sigma\rangle\alpha \equiv turn_i \wedge \alpha$. The lemma follows from this quite easily. For the case when g is a single edge, i.e. $g = ((i, x), a, (j, y))$, it is easy to see that the lemma holds.

Let $g = \Re(i, x, A)$ for $A = \{a_1, \ldots, a_k\}$.
$\sigma = [\psi \mapsto a]^i$:

Suppose $(u, X) \in R_{(g,\sigma)}$, since *enabled*$(g, u)$ holds we have there exists sets $Y_1, \ldots, Y_k$ such that for all $j : 1 \le j \le k$, for all $w_j \in Y_j$ we have $u \xrightarrow{a_j} w_j$. Since u is an i node, any strategy of i will pick a unique edge at u. We have the following two cases:

- $M, u \models \psi$: From semantics, the strategy should choose a w_a such that $u \xrightarrow{a} w_a$ and $(w_a, X) \in R_{(t_a,\sigma)}$. By induction hypothesis, we have $\widehat{w_a} \wedge \langle t_a, \sigma\rangle\widetilde{X}$ is consistent. Hence $\widehat{u} \wedge \langle a\rangle\langle t_a, \sigma\rangle\widetilde{X}$ is consistent.

- $M, u \not\models \psi$: The strategy can choose any w_j such that $u \xrightarrow{a_j} w_j$ and $(w_j, X) \in R_{(t_j,\sigma)}$. By induction hypothesis, $\widehat{w_j} \wedge \langle t_j, \sigma\rangle\widetilde{X}$ is consistent. Hence $\widehat{u} \wedge \langle a_j\rangle\langle t_j, \sigma\rangle\widetilde{X}$ is consistent.

From axiom (A4) case (C4) we get $\widehat{u} \wedge \langle g, \sigma\rangle\widetilde{X}$ is consistent.

Suppose $\widehat{u} \wedge \langle g, \sigma\rangle\widetilde{X}$ is consistent. From axiom (A4) it follows that there exists sets $Y_1, \ldots, Y_k$ such that for all $j : 1 \le j \le k$, for all $w_j \in Y_j$ we have $u \xrightarrow{a_j} w_j$ and hence *enabled*(g, u) holds. Let $X = \{v_1, \ldots, v_m\}$. We have the following two cases:

- if $M, u \models \psi$: then from case (C4), $\widehat{u} \wedge (\langle a\rangle\langle t_a, \sigma\rangle\widetilde{X})$ is consistent. Hence we get there exists w_a such that $u \xrightarrow{a} w_a$ and $\widehat{w_a} \wedge \langle t_a, \sigma\rangle\widetilde{X})$ is consistent. By induction hypothesis there exists $X' \subseteq X$ such that $(w_a, X') \in R_{(t_a,\sigma)}$ and by definition of R we have $(u, X') \in R_{(g,\sigma)}$.

- if $M, u \not\models \psi$: then from case (C4), $\widehat{u} \wedge \bigvee_{a_j \in A} \langle a_j\rangle\langle t_j, \sigma\rangle\widetilde{X}$. Therefore there exists w_j such that $u \xrightarrow{a_j} w_j$ and $\widehat{w_j} \wedge \langle t_j, \sigma\rangle\widetilde{X}$ is consistent. By induction hypothesis there exists $X' \subseteq X$ such that $(w_j, X') \in R_{(t_j,\sigma)}$ and therefore we have $(u, X') \in R_{(g,\sigma)}$.

$\sigma = [\psi \mapsto a]^{\bar{\imath}}, \pi_1 + \pi_2, \pi_1 \cdot \pi_2 \in Strat^{\bar{\imath}}(P^{\bar{\imath}})$:

Suppose $(u, X) \in R_{(g,\pi)}$, since *enabled*(g, u) holds, we have there exists $Y_1, \ldots, Y_k$ such that for all $j : 1 \le j \le k$, for all $w_j \in Y_j$, we have $u \xrightarrow{a_j} w_j$. Since u is an i node, any strategy τ of $\bar{\imath}$ conforming to π will have all the branches at u (by definition of strategy). Therefore we get for all w_j with $u \xrightarrow{a_j} w_j$, there exists $X_j \subseteq X$ such that $(w_j, X_j) \in R_{(t_j,\pi)}$ and $X = \bigcup_{j=1,\ldots,k} X_j$. By induction hypothesis and the fact that $X_j \subseteq X$, we have $\widehat{w_j} \wedge \langle t_j, \pi\rangle\widetilde{X}$ is consistent. Hence from axiom (A4) case (C2), we conclude that $\widehat{u} \wedge \langle g, \sigma\rangle\widetilde{X}$ is consistent.

Suppose $\widehat{u} \wedge \langle g, \pi \rangle \widetilde{X}$ is consistent. From axiom (A4) we get that $\widehat{u} \wedge g^{\surd}$ is consistent. This implies that there exists sets $Y_1, \ldots, Y_k$ such that for all $j : 1 \leq j \leq k$, for all $w_j \in Y_j$ we have $u \xrightarrow{a_j} w_j$ and hence *enabled*(g, u) holds. From case (C2), we have $\widehat{u} \wedge (\bigwedge_{a_j \in A} [a_j]\langle t_{a_j}, \pi \rangle \alpha)$ is consistent. Therefore for all j such that $u \xrightarrow{a_j} w_j$, we have $w_j \wedge \langle t_{a_j}, \pi \rangle \widetilde{X}$ is consistent. By induction hypothesis there exists $X'_j \subseteq X$ such that $(w_j, X'_j) \in R_{(t_{a_j}, \pi)}$. Let $X' = \bigcup_{j=1,\ldots,k} X'_j$, by definition of R we have $(u, X') \in R_{(g,\pi)}$.

The cases when $\sigma = \sigma_1 \cdot \sigma_2, \sigma_1 + \sigma_2, \pi \Rightarrow \sigma_1$ follows easily from axiom (A4) cases (C5) and (C6). Since the root of g is an i node the case when $\sigma = \pi \Rightarrow \sigma_1$, also follows from case (C7) and the induction hypothesis.

The interesting case is when the root of g is an i node and when the specification is $\sigma_1 \Rightarrow \pi$.

Let $g = \Re(i, x, A)$ where $A = \{a_1, \ldots, a_k\}$ and $\sigma = \sigma_1 \Rightarrow \pi$.
Suppose $(u, X) \in R_{g,\sigma}$ since *enabled*(g, u) holds, its easy to show that $\widehat{u} \wedge g^{\surd}$ is consistent. For a strategy τ of player $\bar{\imath}$ to satisfy $\sigma_1 \Rightarrow \pi$ on g, it should make sure of the following:

- for each edge $a_j \in A$, if $u \xrightarrow{a_j} w_j$ conforms with σ_1 then the strategy on t_j should satisfy π.

- for each edge $a_j \in A$, if $u \xrightarrow{a_j} w_j$ does not conform with σ_1 then any strategy can be employed on the game t_j.

From the above observations and axiom (A4) case (C3), we get $\widehat{u} \wedge \langle g, \sigma_1 \Rightarrow \pi \rangle \widetilde{X}$ is consistent.

Part 2 of the lemma again follows from (C3) and a similar argument. □

Lemma 2. For all $\xi \in \Gamma$, for all $X \subseteq W$ and $u \in W$, if $\widehat{u} \wedge \langle \xi \rangle \widetilde{X}$ is consistent then there exists $X' \subseteq X$ such that $(u, X') \in R_\xi$.

Proof. By induction on the structure of ξ.

- $\xi = (g, \sigma)$: Suppose $\widehat{u} \wedge \langle g, \sigma \rangle \widetilde{X}$ is consistent. From lemma 1 item 2, it follows that there exists $X' \subseteq X$ such that $(u, X') \in R_\xi$.

- $\xi = \xi_1 \cup \xi_2$: By axiom (A3a) we get $\widehat{u} \wedge \langle \xi_1 \rangle \widetilde{X}$ is consistent or $\widehat{u} \wedge \langle \xi_2 \rangle \widetilde{X}$ is consistent. By induction hypothesis there exists $X_1 \subseteq X$ such that $(u, X_1) \in R_{\xi_1}$ or there exists $X_2 \subseteq X$ such that $(u, X_2) \in R_{\xi_2}$. Hence we have $(u, X_1) \in R_{\xi_1 \cup \xi_2}$ or $(u, X_2) \in R_{\xi_1 \cup \xi_2}$.

- $\xi = \xi_1; \xi_2$: By axiom (A3b), $\widehat{u} \wedge \langle \xi_1 \rangle \langle \xi_2 \rangle \widetilde{X}$ is consistent. Hence $\widehat{u} \wedge \langle \xi_1 \rangle (\bigvee (\widehat{w} \wedge \langle \xi_2 \rangle \widetilde{X}))$ is consistent, where the join is taken over all $w \in Y = \{w \mid w \wedge \langle \xi_2 \rangle \widetilde{X}$ is consistent $\}$. So $\widehat{u} \wedge \langle \xi_1 \rangle \widetilde{Y}$ is consistent. By induction hypothesis, there exists $Y' \subseteq Y$ such that $(u, Y') \in R_{\xi_1}$. We also have that for all $w \in Y$, $\widehat{w} \wedge \langle \xi_2 \rangle \widetilde{X}$ is consistent. Therefore we get for all $w_j \in Y' = \{w_1, \ldots, w_k\}$, $\widehat{w_j} \wedge \langle \xi_2 \rangle \widetilde{X}$ is consistent. By induction hypothesis, there exists $X_j \subseteq X$ such that $(w_j, X_j) \in R_{\xi_2}$. Let $X' = \bigcup_{j=1,\ldots,k} X_k \subseteq X$, we get $(u, X') \in R_{\xi_1;\xi_2}$.

- $\xi = \xi_1^*$: Let Z be the least set containing X and closed under the condition: for all w, if $\widehat{w} \wedge \langle\xi_1\rangle\widetilde{Z}$ is consistent, then $w \in Z$. By definition of Z and induction hypothesis, we get for all $w \in Z$, there exists $X_w \subseteq X$ such that $(w, X_w) \in R_{\xi_1^*}$. It is also easy to see that $\vdash \widetilde{X} \supset \widetilde{Z}$. Using standard techniques, it is also easy to show that $\vdash \langle\xi_1\rangle\widetilde{Z} \supset \widetilde{Z}$.

  Applying the induction rule (IND), we have $\vdash \langle\xi_1^*\rangle\widetilde{Z} \supset \widetilde{Z}$. By assumption, $\widehat{u} \wedge \langle\xi_1^*\rangle\widetilde{X}$ is consistent. So $\widehat{u} \wedge \langle\xi_1^*\rangle\widetilde{Z}$ is consistent. Hence $\widehat{u} \wedge \widetilde{Z}$ is consistent and therefore $u \in Z$. Thus we have $(u, X') \in R_{\xi_1^*}$ for some $X' \subseteq X$.

□

Lemma 3. For all $\langle\xi\rangle\alpha \in CL(\alpha_0)$, for all $u \in W$, $\widehat{u} \wedge \langle\xi\rangle\alpha$ is consistent iff there exists $(u, X) \in R_\xi$ such that $\forall w \in X$, $\alpha \in w$.

Proof. (⇒) Follows from lemma 2 by considering the set $X_\alpha = \{w \in W \mid \alpha \in w\}$. (⇐) Suppose $\exists (u, X) \in R_\xi$ such that $\forall w \in X$, $\alpha \in w$. We need to show that $\widehat{u} \wedge \langle\xi\rangle\alpha$ is consistent, this is done by induction on the structure of ξ.

- The case when $\xi = (g, \sigma)$ follows easily from lemma 1 and $\xi = \xi_1 \cup \xi_2$ follows from the induction hypothesis and axiom (A3a).

- $\xi = \xi_1; \xi_2$: Since $(u, X) \in R_{\xi_1;\xi_2}$, there exists $Y = \{v_1, \ldots, v_k\}$, there exists sets $X_1, \ldots, X_k \subseteq X$ such that $\bigcup_{j=1,\ldots,k} X_j = X$, for all $j : 1 \leq j \leq k$, $(v_j, X_j) \in R_{\xi_2}$ and $(u, Y) \in R_{\xi_1}$. By induction hypothesis, for all j, $\widehat{v_j} \wedge \langle\xi_2\rangle\alpha$ is consistent. Since v_j is an atom and $\langle\xi_2\rangle\alpha \in CL(\alpha_0)$, we get $\langle\xi_2\rangle\alpha \in v_j$. Again by induction hypothesis we have $\widehat{u} \wedge \langle\xi_1\rangle\langle\xi_2\rangle\alpha$ is consistent. Hence from (A3b) we have $\widehat{u} \wedge \langle\xi_1; \xi_2\rangle\alpha$ is consistent.

- $\xi = \xi_1^*$: If $u \in X$ then $\vdash \widehat{u} \supset \widetilde{X}$. We have $\vdash \widetilde{X} \supset \alpha$ and hence we get $\widehat{u} \wedge \alpha$ is consistent. From axiom (A3c) we have $\widehat{u} \wedge \langle\xi_1^*\rangle\alpha$ is consistent.

  Else we have $(u, X) \in R_{\xi_1;\xi_1^*}$. Let $Z_0 = X$ and $Z_{n+1} = Z_n \cup \{w \mid (w, Z') \in R_{\xi_1}, Z' \subseteq Z_n\}$. Take the least m such that $u \in Z_m$. We have for all $w \in Z_{m-1}$, $\vdash \widehat{w} \supset \langle\xi_1^*\rangle\widetilde{X'}$ for some $X' \subseteq X$. We also have $(u, Z'_m) \in R_{\xi_1}$ for some $Z'_m = \{v_1, \ldots, v_k\} \subseteq Z_m$. Let $X_1, \ldots, X_k \subseteq X$ such that $\forall j : 1 \leq j \leq k$, we have $(v_j, X_j) \in R_{\xi_1^*}$ and $X' = \bigcup_{j=1,\ldots,k} X_j$. By an argument similar to the previous case we can show that $\widehat{u} \wedge \langle\xi_1\rangle\langle\xi_1^*\rangle\widetilde{X'}$ is consistent. Hence we get $\widehat{u} \wedge \langle\xi_1; \xi_1^*\rangle\alpha$ is consistent. Therefore from axiom (A3c) we have $\widehat{u} \wedge \langle\xi_1^*\rangle\alpha$ is consistent.

□

Theorem 1. For all $\beta \in CL(\alpha_0)$, for all $u \in W$, $M, u \models \beta$ iff $\beta \in u$.

The theorem follows from lemma 3 by a routine inductive argument.

Decidability:

Since the size of the action set $|\Sigma|$ is constant, the size of $CL(\alpha_0)$ is linear in $|\alpha_0|$. Atoms are maximal consistent subsets of $CL(\alpha_0)$, hence $|\mathcal{AT}(\alpha_0)|$ is exponential in the size of α_0. From the completeness theorem we get that for a formula α_0, if α_0 is satisfiable then it has a model of exponential size, i.e. $|W| = O(2^{|\alpha_0|})$. For all game strategy pairs ξ occurring in α_0, the relation R_ξ can be computed in time exponential in the size of the model. Therefore it follows that the logic is decidable in nondeterministic double exponential time.

8 Extensions

Concurrency as introduced in game logic van Benthem et al. (2007) can be represented in our framework with the addition of the operator $\xi_1 \times \xi_2$ in the syntax of game strategy pairs. For instance, $(g_1, \sigma_1) \times (g_2, \sigma_2)$ would mean that the game g_1 is played with a strategy conforming to σ_1 and concurrently, the game g_2 is played with a strategy conforming to σ_2. The semantics can be defined in the usual manner:

- $R_{\xi_1 \times \xi_2} = \{(u, X) \mid X = X_1 \cup X_2$ such that $(u, X_1) \in R_{\xi_1}$ and $(u, X_2) \in R_{\xi_2}\}$.

It is easy to see that the completeness theorem also follows with the addition of the following axiom.

- $\langle \xi_1 \times \xi_2 \rangle \alpha \equiv \langle \xi_1 \rangle \alpha \wedge \langle \xi_2 \rangle \alpha$.

Test operator

The test operator as in dynamic logic can also be added into the syntax of game strategy pairs. For $\beta \in \Phi$, the interpretation of $\beta? \in \Gamma$ would be to test whether β holds at the particular state and if yes, continue else fail. The semantics can be given as:

- $R_{\beta?} = \{(u, \{u\}) \mid M, u \models \beta\}$.

The test operator gives the ability of checking for certain conditions and then deciding which game to proceed with. This construct is particularly interesting in our framework, since unlike programs we have players in the game. For instance, let π denote the strategy specification of player 2 and σ the specification of player 1. The formula $(g_1, \pi); win_2?; (g_2, \sigma)$ says that in g_1 if player 2 by employing a strategy conforming to π can ensure win_2 then proceed with the game g_2 where player 1 plays σ. Note that if the test fails then g_2 is not played. This is in contrast to the tests performed in a strategy specification. In a specification if the test fails then the player is free to choose any action.

With the addition of the following axiom, the completeness theorem goes through.

- $\langle \beta? \rangle \alpha \equiv \beta \wedge \alpha$

References

R. Alur, T. A. Henzinger, and O. Kupferman. Alternating-time temporal logic. *Journal of the ACM*, 49:672–713, 2002.

R. J. Aumann and J. H. Dreze. When all is said and done, how should you play and what should you expect? `http://www.ma.huji.ac.il/raumann/pdf/dp_387.pdf`, March 2005.

G. Bonanno. Modal logic and game theory: Two alternative approaches. *Risk Decision and Policy*, 7:309–324, December 2002.

K. Chatterjee, T. A. Henzinger, and N. Piterman. Strategy logic. In *Proceedings of the 18th International Conference on Concurrency Theory (CONCUR)*, volume 4703 of *Lecture Notes in Computer Science*, pages 59–73. Springer, 2007.

D. Gale. The game of hex and Brouwer fixed-point theorem. *The American Mathematical Monthly*, 86:818–827, 1979.

S. Ghosh. Strategies made explicit in dynamic game logic. In *Proceedings of the Workshop on Logic and Intelligent Interaction, ESSLLI 2008*, pages 74–81, 2008.

V. Goranko. Coalition games and alternating temporal logics. In *Proceedings of 8th conference on Theoretical Aspects of Rationality and Knowledge*, pages 259–272, 2001.

V. Goranko. The basic algebra of game equivalences. *Studia Logica*, 75(2): 221–238, 2003.

P. Harrenstein, W. van der Hoek, J.-J. Meyer, and C. Witteven. A modal characterization of Nash equilibrium. *Fundamenta Informaticae*, 57:2-4:281–321, 2003.

J. C. Harsanyi and R. Selten. *A general theory of equilibrium selection in games*. MIT Press, 1987.

R. Parikh. The logic of games and its applications. *Annals of Discrete Mathematics*, 24:111–140, 1985.

M. Pauly. *Logic for Social Software*. PhD thesis, University of Amsterdam, October 2001.

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

J. van Benthem. Extensive games as process models. *Journal of Logic Language and Information*, 11:289–313, 2002.

J. van Benthem, S. Ghosh, and F. Liu. Modelling simultaneous games with concurrent dynamic logic. In *A Meeting of Minds, Proceedings of the Workshop on Logic, Rationality and Interaction*, pages 243–258, 2007.

W. van der Hoek, W. Jamroga, and M. Wooldridge. A logic for strategic reasoning. *Proceedings of the Fourth International Joint Conference on Autonomous Agents and Multi-Agent Systems (AAMAS 05)*, pages 157–164, 2005.

E. Zermelo. Über eine Anwendung der Mengenlehre auf die Theorie des Schachspiels,. In *Proceedings of the Fifth Congress Mathematicians*, pages 501–504. Cambridge University Press, 1913.

An Update Operator for Strategic Ability

Paolo Turrini, Jan Broersen, Rosja Mastop and John-Jules Ch. Meyer

Utrecht University, Utrecht University, Delft University of Technology, Utrecht University.
paolo@cs.uu.nl, r.j.mastop@tudelft.nl, broersen@cs.uu.nl, jj@cs.uu.nl

Abstract

Coalition Logic does not explicitly talk about the effects of a coalitional move on the strategic ability of the remaining players, while in Game Theory reasoning patterns involving this concept often occur. To fill this gap, we study an update operator for strategic ability update in coalition structures. Its formal connections with the update operators known from Dynamic Epistemic Logic will be discussed. This paper is the follow-up of the presentation *Updating Coalition Structures: some issues and some results*, held as a Logic and Interactive RAtionality Seminar on June 8th 2009, and it has been published under the present title in *Proceedings of the Second International Workshop on Logic, Rationality, and Interaction* (LORI 2009) Chongqing, China, October 8-11, 2009.

1 Introduction

Ever since the work of Rohit Parikh on the logic of games Parikh (1985) the research on the characterization of game-theoretical notions in terms of a logical language has grown rapidly. In Cooperative Game Theory for instance results on the correspondence between strategic games and neighbourhood models - such as Pauly Representation Theorem for Coalition Logic Pauly (2001) or the completeness of Alternating-Time Temporal Logic (ATL) Goranko and van Drimmelen (2006) - have opened the possibility of studying cooperative interactions by means of modal logic. ATL and Coalition Logic reason on what coalitions can achieve by cooperating, however they do not explicitly describe what the effects of a given coalitional action or strategy are on the moves of the remaining players. Game Theory instead deals with reasoning structures, as for instance that of Dominant Strategy Equilibrium Osborne and Rubinstein (1994), in which players consider all the possible reactions of their opponents and choose the best strategy given all such reactions.

As affirmed in van Benthem (2007), p.1:

> Much of game theory is about the question whether strategic

i \ j	C	D
C	$(3,3)$	$(0,4)$
D	$(4,0)$	$(1,1)$

Table 1: A Prisoners' Dilemma

> equilibria exist. But there are hardly any explicit languages for defining, comparing, or combining strategies as such - the way we have them for actions and plans, maybe the closest intuitive analogue to strategies. True, there are many current logics for describing game structure - but these tend to have existential quantifiers saying that "players have a strategy" for achieving some purpose, while descriptions of these strategies themselves are not part of the logical language.

In order to capture the reasoning structure behind Dominant Strategy Equilibrium and many other solution concepts, we intuitively need a language able to talk about strategic ability update and consequently to make the role of strategic ability explicit. Updates are not new to the realm of modal logics. Formalizations of dynamics of information flow, like Dynamic Epistemic Logic van Ditmarsch et al. (2007) (DEL), reason about how agents' knowledge is updated after an epistemic event, for instance a public announcement, takes place.

Logics for strategic ability using a model update have already been studied, ranging from the use of counterfactuals in CATL van der Hoek et al. (2005), to the action expressions used in Coalition Action Logic Borgo (2007) and the first order strategy terms in Strategy Logic Chatterjee et al. (2007). Nevertheless all these extensions use arbitrary strategy terms that do not allow to reduce strategy execution to strategic ability. The reduction of the language of Public Announcement Epistemic Logic to Epistemic Logic is instead one of the most elegant results in Dynamic Epistemic Logic.

The idea of this paper is to extend the *update paradigm* of public announcements to account for the changes that moves in a game induce on players' strategic ability and to study strategies reducing them to the choice structures under which they can be executed.

1.1 Motivating Example

To provide a clearer intuition of the notion of strategic ability update, we resort to the well known gametheoretical example of the Prisoners' Dilemma Osborne and Rubinstein (1994), that is an interactive situation in which the advantages of cooperation are overruled by the incentive for individual players to defect. In Table 1 a Prisoners' Dilemma is described, where players i and j, that we assume to be rational, can choose between a cooperative move C and a defective move D, yielding an outcome (x_i, x_j), x_k being the payoff for each $k \in \{i, j\}$. If we focus on player i we can observe that, after the choice C by j, the choice D becomes preferable to the choice C - yielding $(4,0)$ instead of $(3,3)$ - and the same holds in case j moved D - yielding $(1,1)$ instead of $(0,4)$. Our rationality assumption warrants player i to reason on the updates of his own choices brought about by player j, and to select his best response in each such scenario.

Our aim is to formally capture the reasoning structure of players in strategic interaction, in which players consider the best action to take, *given* what their opponents do. This should not be confused with the reasoning patterns in extensive games, in which players reason on the best action to take *after* their opponents have moved, neither with the notion of ability to guarantee an outcome *independently* of what the other players do, which is the typical reading of the operators in the various game logics. To make these intuitions precise we will provide a semantics for the notion of game restriction induced by the moves of the players in a strategic interaction. We will work on cooperative structures, where players can form coalitions to achieve their goals Aumann and Peleg (1960). In our treatment we will focus on coalitional ability, abstracting away from players' preferences.

The paper is structured as follows: in the first part we introduce Coalition Logic, that we use to model strategic ability; in the second part we introduce an operator to talk about the model transformations induced by the choices of coalitions: the subgame operator. Finally we give reduction axioms for the subgame operator and discuss the links with Public Announcement Logic.

2 Coalition Logic and Strategic Ability

In Game Theory players may be able to force the interaction to end up in an outcome satisfying certain properties. An abstract representation of this notion is given by the dynamic effectivity function, first described in Pauly (2001), which we adopt to model strategic ability.

Definition 2.1 (Dynamic Effectivity Function).
Given a finite set of agents Agt and a set of states W, a *dynamic effectivity function* is a function $E : W \to (2^{Agt} \to 2^{2^W})$.

Any subset of Agt will henceforth be called a *coalition*. The elements of W are called *states* or *worlds*; the sets of states $X \in E(w)(C)$ are called the *choices* of coalition C in state w. The set $E(w)(C)$ is called the *choice set* of C in w. The complement of a set X is indicated as $\overline{X}$ and calculated relative to the expected domain. A dynamic effectivity function can be seen as a "formal description of the power structure in a society" Abdou and Keiding (1991); it assigns, in each world, to every coalition a set of sets of states that represents the strategic ability of that coalition. Intuitively, if $X \in E(w)(C)$, C is said to be able from w to *force* the interaction to end up in some member of X. Every effectivity function has the property of **outcome monotonicity**: for all $X \subseteq W, Y \subseteq W, w \in W, C \in 2^{Agt}$, if $X \in E(w)(C)$ and $X \subseteq Y$, then $Y \in E(w)(C)$. Said in other words, if a coalition is able to force the the interaction to end up in some member of X then is also able to force the interaction to end up in some member of any supersets of X. Together with outcome monotonicity we will assume the properties of **regularity**: if $X \in E(w)(C)$, then $\overline{X} \notin E(w)(\overline{C})$; and **closed-worldness**: $E(w)(\emptyset) = \{W\}$. Regularity means that disjoint coalitions do not make choices that contradict each other, while closed-worldness requires the empty coalition not to influence the interaction. For an in depth discussion on the desirability of these properties see the results in Broersen et al. (2008).

2.1 Models and Language

The models we refer to are structures of the form

$$\langle W, E, V \rangle$$

where W is a nonempty set of states, E an outcome monotonic, regular and closed-world effectivity function, $V : W \to 2^P$ a valuation function that assigns to each state a subset of a countable set of atomic propositions P, to be interpreted as true at that state. The formulas for the basic language are of the form

$$p|\neg\phi|\phi \wedge \psi|[C]\phi|A\phi$$

where p is any atomic proposition in P, $[C]\phi$ is the coalitional operator expressing the fact that coalition C can force or bring about the formula ϕ; $A\phi$ is the global modality, which talk about a formula that holds in every world in the model. Their interpretation is standard Pauly (2001) Blackburn et al. (2001) van Benthem (September, 2006) and it is given as follows:

$$\begin{array}{rll} M, w \models p & \text{iff} & p \in V(w) \\ M, w \models \neg\phi & \text{iff} & \text{not } M, w \models \phi \\ M, w \models \phi \wedge \psi & \text{iff} & M, w \models \phi \text{ and } M, w \models \psi \\ M, w \models [C]\phi & \text{iff} & \phi^M \in E(w)(C) \\ M, w \models A\phi & \text{iff} & M, v \models \phi, \text{ for all } v \in W \end{array}$$

where $\phi^M = \{w \in W | M, w \models \phi\}$ is the *truth set* of ϕ.

What we can say in Coalition Logic The Prisoners' Dilemma can intuitively be rewritten as a coalition model. Here coalition $\{i\}$ can force that $\{i\}$ defects and can force that $\{i\}$ cooperates, but $\{i\}$ cannot force that $\{j\}$ cooperates (and equivalently it cannot force that $\{j\}$ defects). In any world w, we have therefore that $PD, w \models [\{i\}](\text{ i defects }) \wedge \neg[\{i\}](\text{ j defects })$. On the other hand we cannot express what i can do given that j defects. This would mean i to have a strategy forcing that i defects and j defects and a strategy forcing that i cooperates and j defects. This at the model level is $PD, w \models [\{i\}](\text{ i defects and j defects }) \wedge [\{i\}](\text{ i cooperates and j defects })$. By the property of outcome monotonicity, we would then get $PD, w \models [\{i\}](\text{ j defects })$, which is at odds with our initial statement. The reason of this limitation is to be found in the interpretation of the coalition logic operator, that expresses what a coalition can achieve *independently* of what its opponents do. Reasoning about how the strategic ability (to force some outcome) of a coalition depends on the possible moves of its opponents requires that we can express in our language that a coalition can force some outcome *given* what its opponents do.

3 Strategic Ability Update

To model strategic ability update we introduce an operator $[C \downarrow \psi]\phi$ whose informal reading is: "after coalition C chooses ψ, ϕ holds". We define the dual $\langle C \downarrow \psi \rangle\phi$ as an abbreviation of $\neg[C \downarrow \psi]\neg\phi$. Intuitively what we do is to talk about the model *restrictions* that are caused by the possible move ψ of coalition

C. For this reason it will be called *the subgame operator*. Its formal interpretation goes as follows:

$$M, w \models [C \downarrow \psi]\phi \Leftrightarrow \psi^M \in E(w)(C) \text{ implies } M \downarrow_{(C,\psi^M,w)}, w \models \phi$$

The interpretation of the operator has a conditional reading: if a coalition C has a certain choice ψ^M at w, then the model where this choice is actually executed makes a certain proposition ϕ true. The capacity of C to choose ψ^M is seen here as a precondition for C to actually execute ψ^M.

The restricted models $M \downarrow_{(C,\psi^M,w)}$ are so defined:

$$M \downarrow_{(C,\psi^M,w)} \doteq \langle W, E \downarrow_{(C,\psi^M,w)}, V \rangle$$

They inherit the domain and the valuation function from the original coalition model while they update the coalitional relation [1] $E \downarrow_{(C,\psi^M,w)}$ in the following way:

$$\begin{array}{ll} E \downarrow_{(C,\psi^M,w)} (w)(D) \doteq (\{\psi^M\})^{\text{sup}} & \text{for } D \cap C \neq \emptyset \\ E \downarrow_{(C,\psi^M,w)} (w)(D) \doteq (E(w)(D) \sqcap \psi^M)^{\text{sup}} & \text{for } D \cap C = \emptyset \text{ and } D \neq \emptyset \\ E \downarrow_{(C,\psi^M,w)} (w')(D) \doteq E(w')(D) & \text{for } w' \neq w \text{ or } D = \emptyset \end{array}$$

where for a set of sets $\mathcal{X}$, $(\mathcal{X})^{\text{sup}} = \{X \subseteq W |$ there is $Y \in \mathcal{X}$ and $Y \subseteq X \subseteq W\}$. In words, $()^{\text{sup}}$ is the superset closure of a set of sets. Moreover taken two sets of sets $\mathcal{X}, \mathcal{P}$, $\mathcal{X} \sqcap \mathcal{P} = \{\xi \cap \psi | \xi \in \mathcal{X} \text{ and } \psi \in \mathcal{P}\}$.

The way the relation is updated deserves some comment. A distinction is made between the strategic ability update of the players who made a certain choice ϕ and all the other players. After coalition C has made a choice ϕ, all the coalitions involving agents belonging to C are given $(\phi^M)^{\text{sup}}$ as a choice set. This view maintains that a coalition comprising players in a coalition that has already formed cannot further influence the outcome of the game. This fact implies that the subgame operator is not coalition monotonic, in the sense given in Pauly (2001), that is bigger coalitions need not have bigger power. Said in other words, we do not allow players to make a choice within a certain coalition and then, at the same time, to make a choice within different coalitions. The models of reference are strategic games, in which strategies are decided in the beginning once and for all Osborne and Rubinstein (1994). The other (nonempty) coalitions instead *truly update* their choice set having it restricted by the choice of C. Restriction is implemented in this case by intersecting the effectivity function with the move that has been carried out. If for instance C chooses to force ψ and $\overline{C}$ were able to decide on ξ, then given the choice by C, $\overline{C}$ is able to force $\xi \wedge \psi$. The coalitional relation at worlds different from the one where the choice is made remains instead unchanged. This means that the update is local. Again, the references are strategic games, where the sequential structure of strategies is substantially ignored. Notice that by the last condition the empty coalition never gains power. In sum the strategic ability update is governed by three principles: the **irrelevance of hybrid coalitions**,

[1] Here the word *functional relation* would be more appropriate. In fact the Effectivity Function behaves as a relation in a Neighbourhood model and our restriction uniquely associates to an Effectivity Function the restriction imposed by a coalitional choice.

that does not allow members of the coalition that moved to further influence the interaction, the **restriction of opponents' choices**, that truly updates the effectivity function of the coalitions opposing the one that moved, and the **locality of the update**, that leaves the coalitional power at different worlds untouched.

The following relevant fact can be easily verified:

Proposition 1. For every C,w, $\psi^M \in E(w)(C)$, we have that $E \downarrow_{(C,\psi^M,w)}$ is outcome monotonic, regular and closed-world.

The proposition represents the basis for our reduction results. Whatever update is carried out a model is obtained that obeys the properties that have been assumed for coalition models.

Even though the interpretation of the update operator may look complex, its structural behaviour is rather simple. The validities in Table 2 allow to translate every sentence where the operator is occurring to a sentence where the operator is not occurring, provided an appropriate law for substitution of equivalent formulas (as $R5$ in the Table). Resemblance to Public Announcement Logic is no coincidence. The axioms reduce in fact the update operator to the global modality and the coalition logic operator. So the operator adds no expressivity to the language and completeness of the language with the update operator follows from the completeness of the language without it. A completeness proof for Closed-World coalition logic, where the global modality interacts with the coalition logic modality by means of the axiom $[\emptyset]\phi \leftrightarrow A\phi$ is provided in Broersen et al. (2008).

3.1 Back to the game

With the new operator it becomes possible to formalize the conditional aspect of strategic reasoning. In the structure PD we have that $PD, w \models [\{i\} \downarrow \text{i defects}]([\{j\}](\text{j defects and i defects}) \wedge [\{j\}](\text{j cooperates and i defects}))$. Nothing changes at the level of grand coalition, since $PD \models [\emptyset \downarrow \phi][Agt]\psi \leftrightarrow [Agt]\psi$.

4 Discussion: Choices and Announcements

Public Announcement Logic formalizes the effect of the announcement of a true formula in each agent's a epistemic relation $R(a)$, defined as a partition on a domain W. The standard operator $[\phi]\psi$ says that ψ holds after ϕ is announced. Its semantics is given as follows:

$$M, w \models [\phi]\psi \Leftrightarrow M, w \models \phi \text{ implies } M|\phi, w \models \psi$$

where $M|\phi = (W', R'(a), V')$ takes these values:

- $W' = \phi^M$
- $R'(a) = R(a) \cap (W \times \phi^M)$
- $V'(p) = V(p) \cap \phi^M$

	Axioms
	Regularity
A1	$[C]\phi \rightarrow \neg[\overline{C}]\neg\phi$
	Closed-Worldness
A2	$[\emptyset]\phi \leftrightarrow A\phi$
	Global Modality Axioms
A3	$\phi \rightarrow E\phi$
A4	$EE\phi \rightarrow E\phi$
A5	$\phi \rightarrow AE\phi$
A6	$A(\phi \rightarrow \psi) \rightarrow (A\phi \rightarrow A\psi)$
	Strategic Ability Update Axioms
A7	$[C \downarrow \xi]p \leftrightarrow ([C]\xi \rightarrow p)$
A8	$[C \downarrow \xi]\neg\phi \leftrightarrow ([C]\xi \rightarrow \neg[C \downarrow \xi]\phi)$
A9	$[C \downarrow \xi](\phi \wedge \psi) \leftrightarrow ([C \downarrow \xi]\phi \wedge [C \downarrow \xi]\psi)$
A10	$[C \downarrow \xi]A\phi \leftrightarrow ([C]\xi \rightarrow A\phi)$
A11	$[C \downarrow \xi][D]\phi \leftrightarrow ([C]\xi \rightarrow [D](\xi \rightarrow \phi))$ (for $D \cap C = \emptyset$ and $D \neq \emptyset$)
A12	$[C \downarrow \xi][D]\phi \leftrightarrow A(\xi \rightarrow \phi)$ (for $D \cap C \neq \emptyset$)
A13	$[C \downarrow \xi][D]\phi \leftrightarrow ([C]\xi \rightarrow [D]\phi)$ (for $D = \emptyset$)
	Rules
R1	$\phi \wedge (\phi \rightarrow \psi) \Rightarrow \psi$
R2	$\phi \rightarrow \psi \Rightarrow [C]\phi \rightarrow [C]\psi$
R3	$\phi \Rightarrow A\phi$
R4	$\phi \Rightarrow [C \downarrow \xi]\phi$
R5	$\phi \leftrightarrow \psi \Rightarrow [C \downarrow \xi]\theta \leftrightarrow [C \downarrow \xi]\theta[\phi/\psi]$

Table 2: Proof System

	Axioms
	Public Announcement Axioms
A1	$[\phi]p \leftrightarrow (\phi \rightarrow p)$
A2	$[\phi]\neg\psi \leftrightarrow (\phi \rightarrow \neg[\phi]\psi)$
A3	$[\phi](\xi \wedge \psi) \leftrightarrow ([\phi]\xi \wedge [\phi]\psi)$
A4	$[\phi]\Box_a\psi \leftrightarrow (\phi \rightarrow \Box_a[\phi]\psi)$
	Rules
R1	$\xi \wedge (\xi \rightarrow \psi) \Rightarrow \psi$
R2	$\xi \Rightarrow [\phi]\xi$

Table 3: Proof System for Public Announcement Logic

The model restriction of public announcement *throws worlds away*. In fact, as shown for instance in van Benthem and Liu (2004), public announcements can be defined by only updating the epistemic relation. A reduction can be shown in which every sentence from the modal language with the $S5$ knowledge relation and the public announcement operator can be translated into a sentence from the same language without the public announcement operator occurring in it. We report the reduction axioms in Table 3.

If we compare the public announcement operator to the subgame operator, we can observe the structure of the two axiom systems is very similar in the atomic and boolean case, but very different in the modal case. A subtle difference can be though observed in the atomic clause. If Public Announcement Logic reduces the atomic announcement to an implication between atoms ($[q]p \leftrightarrow (q \rightarrow p)$), the subgame operator reduces it to an implication between an atom and a choice ($[C \downarrow q]p \leftrightarrow ([C]q \rightarrow p)$). This fact witnesses that we are really reducing strategy execution to strategic ability. The appendix will make it clear that the similarity of the logics applies to the proof techniques as well, that are at least for the basic cases identical to those of Public Announcement Logic van Ditmarsch et al. (2007). The specific differences are given, once again, by the way the coalitional relation is updated.

5 Conclusion and future work

We have built a logic for strategic ability update, where we can represent the effects of a coalitional choice on the players' strategic ability, extending the *update paradigm* of Dynamic Epistemic Logic to account for the dynamics of strategic ability in Coalition Structures. Our framework explicitly expresses how a coalitional move modifies the ability of all the players involved in the interaction, providing a useful framework for capturing coalitional reasoning in strategic settings. Our results are limited to Coalition Logic. Further study is needed to analyze whether the same characterizations are possible in different frameworks for strategic ability, for instance the Consequentialist-STIT framework, ATL and the full Game Logic. Further work can also be done in characterizing within this framework a number of other gametheoretical concepts like Nash Equilibrium and the Core for Cooperative Games without transferable utility.

A Proofs for Reduction Axioms

Atomic and Boolean Cases

$$[C \downarrow \xi]p \leftrightarrow ([C]\xi \rightarrow p)$$

Take arbitrary M, w. $M, w \models [C \downarrow \xi]p \Leftrightarrow M, w \models [C]\xi$ implies that $M \downarrow_{(C,\xi^M,w)}, w \models p \Leftrightarrow M, w \models [C]\xi$ implies that $M, w \models p \Leftrightarrow M, w \models [C]\xi \rightarrow p$. Q.E.D.

$$[C \downarrow \xi]\neg\phi \leftrightarrow ([C]\xi \rightarrow \neg[C \downarrow \xi]\phi)$$

Take arbitrary M, w. $M, w \models [C \downarrow \xi]\neg\phi \Leftrightarrow M, w \models [C]\xi$ implies that $M \downarrow_{(C,\xi^M,w)}, w \models \neg\phi \Leftrightarrow M, w \models [C]\xi$ implies that $(M, w \models [C]\xi$ and $M \downarrow_{(C,\xi^M,w)}, w \models \neg\phi) \Leftrightarrow M, w \models [C]\xi$ implies that not$(M, w \models [C]\xi$ implies $M \downarrow_{(C,\xi^M,w)}, w \not\models \neg\phi) \Leftrightarrow M, w \models [C]\xi$ implies that not$(M, w \models [C]\xi$ implies $M \downarrow_{(C,\xi^M,w)}, w \models \phi) \Leftrightarrow M, w \models [C]\xi$ implies that $M, w \not\models [C \downarrow \xi]\phi \Leftrightarrow M, w \models [C]\xi \rightarrow \neg[C \downarrow \xi]\phi$ Q.E.D.

$$[C \downarrow \xi](\phi \wedge \psi) \leftrightarrow ([C \downarrow \xi]\phi \wedge [C \downarrow \xi]\psi)$$

Take arbitrary M, w. $M, w \models [C \downarrow \xi](\phi \wedge \psi) \Leftrightarrow M, w \models [C]\xi$ implies that $M \downarrow_{(C,\xi^M,w)}, w \models \phi \wedge \psi \Leftrightarrow M, w \models [C]\xi$ implies that $(M \downarrow_{(C,\xi^M,w)}, w \models \phi$ and $M \downarrow_{(C,\xi^M,w)}, w \models \psi) \Leftrightarrow (M, w \models [C]\xi$ implies that $M \downarrow_{(C,\xi^M,w)}, w \models \phi)$ and $(M, w \models [C]\xi$ implies that $M \downarrow_{(C,\xi^M,w)}, w \models \psi) \Leftrightarrow (M, w \models [C \downarrow \xi]\phi)$ and $(M, w \models [C \downarrow \xi]\psi) \Leftrightarrow M, w \models ([C \downarrow \xi]\phi \wedge [C \downarrow \xi]\psi)$ Q.E.D.

Interaction with Global Modality

$$[C \downarrow \xi]A\phi \leftrightarrow ([C]\xi \rightarrow A\phi)$$

Take an arbitrary M, w. $M, w \models [C \downarrow \xi]A\phi \Leftrightarrow M, w \models [C]\xi$ implies that $M \downarrow_{(C,\xi^M,w)}, w \models A\phi \Leftrightarrow M, w \models [C]\xi$ implies that $M \downarrow_{(C,\xi^M,w)}, w \models [\emptyset]\phi \Leftrightarrow M, w \models [C]\xi$ implies that $M, w \models [\emptyset]\phi \Leftrightarrow M, w \models [C]\xi$ implies that $M, w \models A\phi \Leftrightarrow M, w \models [C]\xi \rightarrow A\phi$

Interaction with Coalition Modality

$$[C \downarrow \xi][D]\phi \leftrightarrow ([C]\xi \rightarrow [D](\xi \rightarrow \phi))(\text{ for } D \cap C = \emptyset \text{ and } D \neq \emptyset)$$

Proof by contraposition.

$\Leftarrow$: Suppose, for some $D \neq \emptyset$, that $[C]\xi \rightarrow [D](\xi \rightarrow \phi)$ and $M, w \not\models [C \downarrow \xi][D]\phi$ for some C such that $(C \cap D) = \emptyset$. The semantic clauses then tell us that (if $\xi^M \in E(w)(C)$ then $(\xi \rightarrow \phi)^M \in E(w)(D)$) and $\xi^M \in E(w)(C)$ and $\phi^M \notin E'(w)(D)$. [I write E' for $E \downarrow_{(C,\xi^M)}$.] By modus ponens $(\xi \rightarrow \phi)^M \in E(w)(D)$.

By the definition of update, $E'(w)(D) = (E(w)(D) \sqcap \xi^M)^{\text{sup}}$. So, $((\xi \rightarrow \phi)^M \cap \xi^M) \in E'(w)(D)$. By elementary set theory this just says that $\phi^M \in E'(w)(D)$. Contradiction.

$\Rightarrow$: Suppose, for some $D \neq \emptyset$, that $M, w \models [C \downarrow \xi][D]\phi$ and $M, w \not\models [C]\xi \rightarrow [D](\xi \rightarrow \phi)$ for some C such that $(C \cap D) = \emptyset$. The semantic clauses then tell us that (if $\xi^M \in E(w)(C)$ then $\phi^M \in E'(w)(D)$) and $\xi^M \in E(w)(C)$ and $(\xi \rightarrow \phi)^M \notin E(w)(D)$. By modus ponens we are assuming that $\phi^M \in E'(w)(D)$ and $(\xi \rightarrow \phi)^M \notin E(w)(D)$.

By the definition of update, $E'(w)(D) = (E(w)(D) \sqcap \xi^M)^{\sup}$. Because $\phi^M \in E'(w)(D)$, there must be some $X \in E(w)(D)$, such that $(X \cap \xi^M) \subseteq \phi^M$. By elementary set theory, it must be the case that $X \subseteq (\xi \rightarrow \phi)^M$.

Hence, by outcome monotonicity of E, if $X \in E(w)(D)$, then $(\xi \rightarrow \phi)^M \in E(w)(D)$. Contradiction.

$$[C \downarrow \xi]([D]\phi \leftrightarrow A(\xi \rightarrow \phi))(\text{ for } D \cap C \neq \emptyset)$$

Proof. Take arbitrary M, w, and arbitrary $\xi^M \in E(w)(C)$. Consider a coalition D with $D \cap C \neq \emptyset$. We have that $E \downarrow_{(C,\xi^M,w)} (w)(D) = (\xi^M)^{\sup}$ by semantics. This means that $\xi^M \subseteq \phi^M$ iff $\phi^M \in E \downarrow_{(C,\xi^M,w)} (w)(D)$. It is easy to conclude that $M, w \models [C \downarrow \xi]([D]\phi \leftrightarrow A(\xi \rightarrow \phi))$. Notice that this also means $M, w \models [C \downarrow \xi][D]\phi \leftrightarrow A(\xi \rightarrow \phi)$. Q.E.D.

$$[C \downarrow \xi][D]\phi \leftrightarrow ([C]\xi \rightarrow [D]\phi)(\text{ for } D = \emptyset)$$

It follows directly from the semantics of the update operator for the case of $D = \emptyset$. Q.E.D.

References

J. Abdou and H. Keiding. *Effectivity Functions in Social Choice*. Kluwer Academic Publishers, 1991.

R. J. Aumann and B. Peleg. Von Neumann-Morgenstern solutions to cooperative games without side payments. *Bulletin of the American Mathematical Society*, 66:173–179, 1960.

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

S. Borgo. Coalitions in action logic. In M. M. Veloso, editor, *IJCAI*, pages 1822–1827, 2007.

J. Broersen, R. Mastop, J.-J. C. Meyer, and P. Turrini. A Logic for Closed-World Interaction. In S. Hölldobler, C. Lutz, and H. Wansing, editors, *Proceedings of the 11th European Conference on Logics in Artificial Intelligence (JELIA)*, volume 5293 of *Lecture Notes in Computer Science*, pages 89–99. Springer, 2008. ISBN 978-3-540-87802-5.

K. Chatterjee, T. Henzinger, and N. Piterman. Strategy logic. Technical Report UCB/EECS-2007-78, University of California, Berkeley, May 2007. URL `http://chess.eecs.berkeley.edu/pubs/312.html`.

V. Goranko and G. van Drimmelen. Complete axiomatization and decidability of alternating-time temporal logic. *Theor. Comput. Sci.*, 353(1-3):93–117, 2006.

M. Osborne and A. Rubinstein. *A course in Game Theory*. The MIT Press, 1994.

R. Parikh. The logic of games and its applications. In *Selected papers of the international conference on "foundations of computation theory" on Topics in the theory of computation*, pages 111–139, New York, NY, USA, 1985. Elsevier North-Holland, Inc. ISBN 0-444-87647-2.

M. Pauly. *Logic for Social Software*. ILLC Dissertation Series, 2001.

J. van Benthem. Where is logic going, and should it? *Topoi*, 25:117–122, September, 2006.

J. van Benthem. In praise of strategies. Research Report, http://www.illc.uva.nl/Publications/ResearchReports/PP-2008-03.text.pdf, 2007.

J. van Benthem and F. Liu. Dynamic logic of preference upgrade. *Journal of Applied Non-Classical Logics*, 14, 2004.

W. van der Hoek, W. Jamroga, and M. Wooldridge. A logic for strategic reasoning. In *AAMAS '05: Proceedings of the fourth international joint conference on Autonomous agents and multiagent systems*, pages 157–164, New York, NY, USA, 2005. ACM. ISBN 1-59593-093-0. doi: http://doi.acm.org/10.1145/1082473.1082497.

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

The Strategic Equivalence of Games with Unawareness

Tomohiro Hoshi and Alistair Isaac

Stanford University, Stanford University
thoshi@stanford.edu, aisaac@stanford.edu

Abstract

An equivalence relation is more transparent when complemented by an explicit specification of the set of transformations under which equivalence is preserved. In the case of extensive games equivalent with respect to their strategic normal form, the set of transformations was provided by Thompson (1952). We extend Thompson's result to the case of games with unawareness (Feinberg (2009))

1 Introduction

When are two games equivalent? Given a set of formal objects, an equivalence relation partitions this set into subsets. Any two members of a subset are then equivalent with respect to the defined partition. For example, one can partition the set of extensive games into subsets which share a corresponding reduced normal form strategic game. Any two extensive games which fall into the same equivalence class are equivalent with respect to *strategic structure*, but not in general equivalent with respect to *temporal structure*.

Thompson (1952) defines a set of four transformations on extensive games which preserve the underlying strategic structure. He proves that any two extensive games which share strategic structure can be transformed into one another by some sequence of his transformations. This paper extends Thompson's transformations and his result to games with unawareness (Feinberg (2009)). In games with unawareness, players may be ignorant of aspects of game structure (including moves available to them, other players, etc.).

The *epistemic structure* of a game with unawareness is modeled by a set of standard games, each of which is indexed by a *view*. Each game in this set represents the perspective of some player on the awareness of others. Effectively, Feinberg has made the hierarchy of "higher-order expectations" discussed in Lewis (1969) explicit by assigning a game to each. In the limiting case, where all players are aware of the complete game structure (and aware that others are aware, ad infinitum) we have *common knowledge*, and games with unawareness simply reduce to standard strategic and extensive games. Awareness is just one way of cashing out the informational relationship between agents and the world. As such, it exhibits compelling parallels with belief and knowledge; however, we should be careful not to reduce awareness to either of these concepts. Unlike knowledge, awareness is not factive; unlike belief, awareness directly constrains the agent's perception of the game.

What is added to an equivalence notion by explicitly specifying the transformations under which it remains invariant? One answer is *transparency*. Consider, for example, geometry, in which we have many different notions of equivalence (e.g. topological equivalence, affine equivalence, similarity, congruence)—what is the relationship between these equivalence notions? Felix Klein's Erlangen program sought to answer this question by organizing geometries into a hierarchy in accordance with the transformations under which their objects remain invariant (Klein (1893)). Topological structure remains invariant under arbitrary stretching of the plane, rotations, and translations. Affine structure is invariant under uniform stretching in a single direction, rotations, translations, and dilations. Similarity is invariant under translations, rotations, and uniform dilation. Finally, congruence is invariant only under translation and rotation. Stated in terms of permissible transformations, the relationship between these various equivalence relations (and the corresponding geometries) becomes more transparent; we can clearly see, for example, that congruence is a restriction of similarity which takes size to be meaningful.

We also find the close relationship between equivalence and transformations in logic, which illustrates a second benefit to making transformational rules explicit: *precision*. Proof systems provide a set of permissible transformations over syntactic objects. Two syntactic objects which share truth value in all circumstances are logically equivalent. In order to ensure precision of our logical system, we demand proofs of its *soundness* and *completeness*. Soundness demonstrates that the transformations are indeed permissible, i.e. if one sentence can be transformed into another *and vice versa*, they are indeed logically equivalent. Completeness demonstrates that the transformations are adequate to take any sentence into any other which is logically equivalent. Only once we have proved both soundness and completeness do we have a satisfactory account of a logical system. Thompson's proof demonstrates the soundness and completeness of his transformations for strategic equivalence. He proves that they preserve underlying strategic form (soundness), and that for any games which do share strategic form, they can be reached by applications of his transformations (completeness). Just as the local manipulations of deductive rules give us insight into the nature of truth (insight not necessarily provided by direct model checking), transformations on extensive games can give us insight into the nature of strategic structure.

In the context of game theory, much ink has been spilt over the correct analysis of game equivalence. Our view is that there is room in game theory for many notions of equivalence; ideally, they will eventually be organized into a hierarchy in terms of the increasingly strict transformations under which game structure remains invariant. From this perspective Thompson's transformations provide a fruitful starting point for the development of more refined equivalence notions. Contributions to this project can be found scattered throughout the literature. Kohlberg and Mertens (1986) supplement Thompson's transformations with two which respectively introduce superfluous chance moves

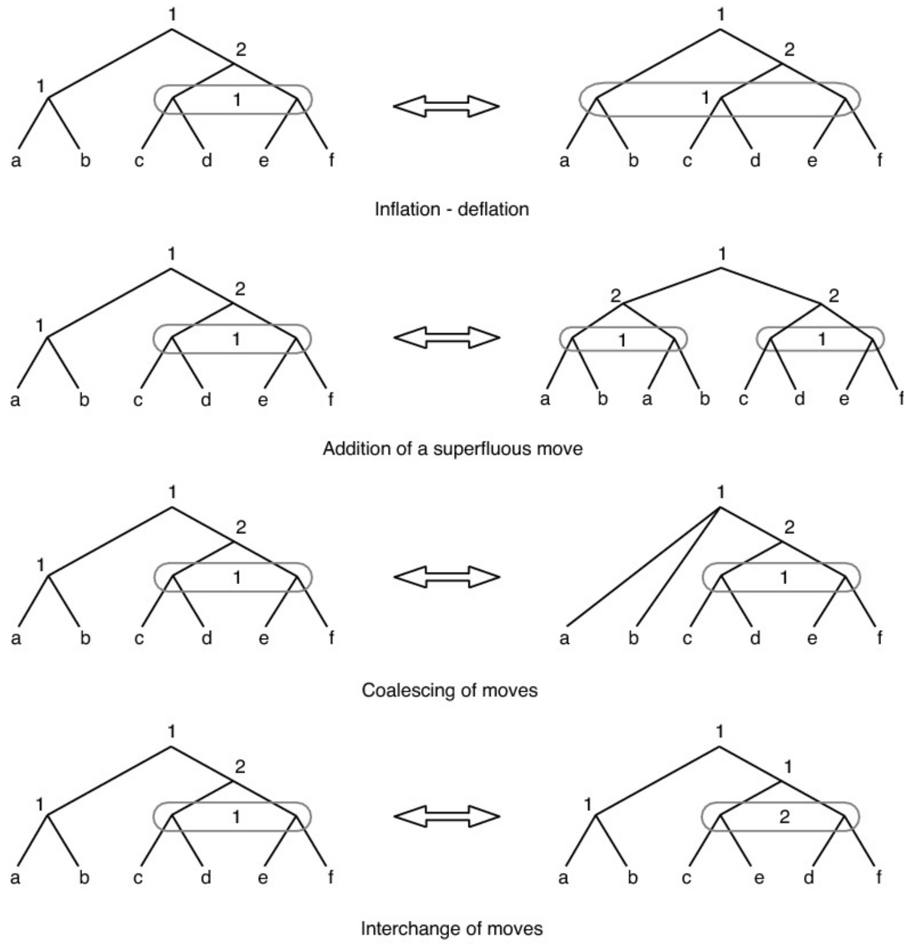


Figure 1: The four Thompson Transformations (after Thompson (1952), fig. 3)

and remove chance moves. This extends Thompson's result to games with chance players (see discussion and refinement in de Bruin (1999)). (Since this problem has been solved, we omit discussion of chance players in our presentation below.) Elmes and Reny (1994) address the worry that the transformation *inflation-deflation* (see Figure 1) does not preserve perfect information. They demonstrate that a modification of *addition of a superfluous move* allows any two games *with perfect information* which share a corresponding strategic form to be transformed into each other without appealing to *inflation-deflation*. Finally, Bonanno (1992) investigates the notion of game equivalence which arises from applications of *interchange of moves* only; such games are equivalent with respect to their *set-theoretic form*. Bonanno's motivation is the preservation of temporal structure without introducing an (arbitrary?) ordering over moves which, from an informational standpoint, are simultaneous. Our hope with this paper is to contribute to this general endeavor by proving soundness and completeness for transformations on a richer structure than standard extensive games.

Section 2 outlines Thompson's basic result and extends it with a new transformation, *coalescing of players*. The insight behind this addition is that players who share payoffs are strategically equivalent. Such redundant players do not usually appear in standard game applications, but they arise in a natural way when considering games with unawareness. Section 3 introduces strategic and extensive games with unawareness, closely following the treatment of Feinberg (2009). In Section 3.3 we extend the concept of strategic equivalence to extensive games with unawareness. Two extensive games with unawareness are strategically equivalent if they share a reduced strategic form with unawareness (modulo coalescing of players). The insight here is that the awareness of a player may change over the course of a temporally extended game. Such players must be treated as distinct from a strategic standpoint because they literally perceive themselves to be playing different games. In situations where the awareness of a player does not change, however, we prefer to treat them as a single strategic agent, and this is ensured by coalescing of players. Finally, Sections 4 and 5 give the main result of the paper. We first introduce transformations over extensive games with unawareness which correspond to those provided by Thompson in the standard case. We prove that two extensive games with unawareness are equivalent with respect to their reduced strategic form if and only if they can be transformed into each other by some sequence of these transformations.

2 The Equivalence of Extensive Games

In this section, we summarize Thompson's work on game equivalence and develop some basic technical apparatus. Our presentation of Thompson's result will be minimal and readers are referred to Thompson (1952) for more details (see also Osborne and Rubinstein (1994), Section 11.2).

We start with basic definitions for strategic games and extensive games.

Definition 2.1 (Strategic Games)**.** A *strategic game* (*SG*) is a tuple $g = \langle I, \{A_i\}_{i\in I}, \{u_i\}_{i\in I}\rangle$ such that:

- I is a set of players.

- A_i is the set of actions available to each player.
- u_i is a utility function that associates payoffs with action profiles $(a_i)_{i \in I}$ in $\Pi_{i \in I} A_i$.

For strategic games g, g', we say g is a *restriction* of g', if the sets of players and actions in g are subsets of those in g' and the utility function in g is a restriction of that in g' with respect to the set of actions in g.

Also two strategic games g, g' are *isomorphic* if there is an isomorphic map between g and g' in the standard sense. In our notation, isomorphism is characterized by the following definition.

Definition 2.2 (Isomorphism on SG)**.** Let $g = \langle I, \{A_i\}_{i \in I}, \{u_i\}_{i \in I} \rangle$ and $g' = \langle I', \{A'_{i'}\}_{i' \in I'}, \{u'_{i'}\}_{i' \in I'} \rangle$. g is *isomorphic* to g', written as $g \cong g'$, if there are functions ρ, α, v such that

1. $\rho : I \to I'$ is bijective.
2. $\alpha : \bigcup_{i \in I} A_i \to \bigcup_{i' \in I'} A'_{i'}$ is bijective and, for all i and $a \in A_i$, $\alpha(a) \in A'_{\rho(i)}$.
3. $v : \mathbb{R} \to \mathbb{R}$ is such that $v(u_i((a_i)_{i \in I})) = u'_{\rho(i)}((\alpha(a_i))_{i \in I})$.

If a given set of functions, ρ, α, v, satisfy the above conditions, we say they *isomorphically map* g to g'.

Definition 2.3 (Extensive Games)**.** An *extensive game* (EG) is a tuple $G = \langle (W, \prec), I, \{A_i\}_{i \in I}, \{F_i\}_{i \in I}, \{u_i\}_{i \in I} \rangle$ where:

1. $(W, \prec)$ is a finite tree with a disjoint union of vertices $W = \bigcup_{i \in I} V_i \cup Z$ where V_i denotes the set of player i's decision points and Z is the set of terminal vertices and the order $w' \prec w$ denotes w' occurs before w on the tree. We denote the set of immediate successors of $w \in W$ by $Succ(w)$. Also we write $w \preceq w'$, if $w \prec w'$ or $w = w'$.
2. I is a set of players.
3. A_i maps (w, w'), where $w' \in Succ(w)$, to the action that i can play at w which leads to w'. It is required that $u \neq v$ implies $A_i(w, u) \neq A_i(w, v)$. We define $A_i(w, \cdot) = \{A_i(w, v) | v \in Succ(w)\}$. We define $\mathbf{A}_i = \{A_i(w, w') \mid w' \in Succ(w) \text{ and } w \in V_i\}$.
4. F_i partitions the set V_i and induces the function f_i that maps $w \in V_i$ to the information set $f_i(w) \in F_i$ that contains w. If $w \in f(w')$, we say w is indistinguishable from w': otherwise, w is distinguishable from w' for i.
5. It is required that $w' \in f_i(w) \in F_i$ implies $A_i(w', \cdot) = A_i(w, \cdot)$.
6. $u_i : Z \to \mathbb{R}$ is the payoffs for the player i defined on the terminal vertices.

For extensive games G, G', we say G is a *restriction* of G' if information sets, sets of nodes, players, and actions in G are subsets of those in G' and the tree relation $\prec$ and the utility function in G are restrictions of those in G' with respect to the sets of nodes and actions in G.

We also say that two extensive games, G, G', are *isomorphic* if there is an isomorphic map between G and G' in the standard sense. In our notation:

Definition 2.4 (Isomorphism on EG)**.** Let $G = \langle (W, \prec), I, \{A_i\}_{i\in I}, \{F_i\}_{i\in I}, \{u_i\}_{i\in I}\rangle$ with $W = \bigcup_{i\in I} V_i \cup Z$ and $G' = \langle (W', \prec'), I', \{A'_i\}_{i\in I'}, \{F'_i\}_{i\in I'}, \{u'_i\}_{i\in I'}\rangle$ with $W' = \bigcup_{i'\in I'} V'_{i'} \cup Z'$ be extensive games. G is isomorphic to G', written as $G \cong G'$, if there are $\rho, \alpha, \upsilon, \phi, \iota$ such that

1. $\alpha : I \to I'$ is bijective.
2. $\phi : \bigcup_{i\in I} V_i \cup Z \to \bigcup_{i'\in I'} V'_{i'\in I'} \cup Z'$ is bijective and
   (a) for all $i \in I$ and v$\in V_i$, $\phi(\mathrm{v}) \in V'_{\rho(i)}$
   (b) for all $v \in Z$, $\phi(v) \in Z'$
   (c) for all v, w, if $v \prec w$, then $\phi(v) \prec' \phi(w)$
3. $\alpha : \bigcup_{i\in I} \mathbf{A}_i \to \bigcup_{i'\in I'} \mathbf{A}_{i'}$ is bijective and, for all $a = A_i(w, v)$, $\alpha(a) = A'_{\rho(i)}(\phi(w), \phi(v))$.
4. $\iota : \bigcup_{i\in I} F_i \to \bigcup_{i'\in I'} F'_{i'}$ is bijective and, for all $f \in F_i$, $\iota(f) = \{\phi(\mathrm{v}) \mid \mathrm{v} \in f\} \in F'_{\rho(i)}$.
5. $\upsilon_i : \mathbb{R} \to \mathbb{R}$ is such that, for all $z \in Z$, $\upsilon(u_i(z)) = u'_{\rho(i)}(\phi(z))$.

If a given set of functions, $\rho, \alpha, \upsilon, \phi, \iota$, satisfy the above conditions, we say that they *isomorphically map* G to G'.

A *strategy* s_i *of a player* $i \in I$ in an extensive game $\langle (W, \prec), I, \{A_i\}_{i\in I}, \{F_i\}_{i\in I}, \{u_i\}_{i\in I}\rangle$ is a function that assigns to every $f \in F_i$ an action in $A_i(w, \cdot)$. We denote the set of i's strategies in an extensive game by S_i. A strategy profile $\mathbf{s}$ of players in I is a sequence $(\mathbf{s}_i)_{i\in I}$ where $\mathbf{s}_i \in S_i$. We denote the set of strategy profiles $\Pi_{i\in I} S_i$ by S. Given $\mathbf{s} \in S$, $\mathbf{s}_i$ is the strategy for i in the strategy profile $\mathbf{s}$. The *outcome* $O(\mathbf{s})$ *of* $\mathbf{s} \in S$ is the terminal node that results when each player i plays the corresponding game by following s_i, i.e. $O(\mathbf{s}) = w_1 \dots w_K$ such that, for each k $(1 \le k \le K)$, $w_k \in V_i$ for some $i \in I$ and $A_i(w_k, w_{k+1}) = s_i(f(w_k))$.

Definition 2.5 (Strategic Form)**.** The strategic form of an extensive game $G = \langle (W, \prec), I, \{A_i\}_{i\in I}, \{F_i\}_{i\in I}, \{u_i\}_{i\in I}\rangle$ is a strategic game $sf(G) = \langle I, S, \{U_i\}_{i\in I}\rangle$, where $U_i : S \to \mathbb{R}$ is such that $U_i(\mathbf{s}) = u_i(O(\mathbf{s}))$ for $\mathbf{s} \in S$.

Let $G = \langle (W, \prec), I, \{A_i\}_{i\in I}, \{F_i\}_{i\in I}, \{u_i\}_{i\in I}\rangle$. For $i \in I$ and $s, t \in S_i$, sR_it (s is equivalent to t for i) if for all $\mathbf{s}, \mathbf{t} \in S$, if $\mathbf{s}_i = s$, $\mathbf{t}_i = t$, and $\mathbf{s}_j = \mathbf{t}_j$ for all $j \in I - \{i\}$, then $u_i(O(\mathbf{s})) = u_i(O(\mathbf{t}))$. Denote the equivalence class under R_i that contains $s \in S_i$ by $\bar{s}$ and the set of equivalence classes of S_i under R_i by $\bar{S}_i$. Let $\bar{S} = \Pi_{i\in I}\bar{S}_i$. Given $\mathbf{s} \in S$, write $\bar{\mathbf{s}}$ for the sequence $(\bar{\mathbf{s}}_i)_{i\in I}$, where each $\bar{\mathbf{s}}_i$ is in $\bar{S}_i$.

Definition 2.6 (Reduced Strategic Form)**.** The *reduced strategic form* $red(G)$ of an extensive game $G = \langle (W, \prec), I, \{A_i\}_{i\in I}, \{F_i\}_{i\in I}, \{u_i\}_{i\in I}\rangle$ is the strategic game $\langle I, \bar{S}, \{\bar{U}_i\}_{i\in I}\rangle$ where $\bar{U}_i(\bar{\mathbf{s}}) = u_i(O(\mathbf{s}))$ for every $\bar{\mathbf{s}} \in \bar{S}$ with $\mathbf{s} \in S$.

Thus the reduced strategic form of an extensive game does not have redundancy in the sense that no two strategies of a player result in the same payoff against all strategies of the other players. The reduced strategic form can be obtained from the strategic form by simply removing such redundant strategies.

Definition 2.7 (Strategic Equivalence)**.** An extensive game G_1 is *strategically equivalent* to an extensive game G_2, written as $G_1 \approx G_2$, if $red(G_1) \cong red(G_2)$.

Thompson (1952) considers two extensive games to be equivalent if they are strategically equivalent. He introduces a set of four local manipulations which are sufficient to transform an extensive game into any strategically equivalent game. We will not give formal definitions of the Thompson transformations here, but examples of each are depicted graphically in Figure 1. See Thompson (1952) for more details. (See also our discussion of the extended versions of Thompson's transformations in Definitions 4.4–4.6 and 4.8.)

Definition 2.8 (Transformability)**.** An extensive game G_1 is transformable into G_2, written as $G_1 \sim G_2$, if there is a sequence of extensive games, $G_1^*, \ldots, G_n^*$, such that $G_1^* = G_1$, $G_n^* \cong G_2$, and G_i^* $(1 \leq i \leq n-1)$ is a result of applying one of Thompson's transformation rules to G_{i+1}^*.

It is easy to see the four transformation rules preserve strategic equivalence. Therefore,

Lemma 1. Suppose G_1 is transformed into G_2 by one of Thompson's four transformations. Then $G_1 \approx G_2$. Moreover, if $G_1 \sim G_2$, then $G_1 \approx G_2$.

For the other direction, we need the following lemmas.

Lemma 2. For every extensive game G, there is an extensive game G' such that $G \sim G'$ and $sf(G') \cong red(G')$.

Let us say G is in *canonical form*, if $sf(G) \cong red(G')$.

Lemma 3. Let G_1, G_2 be extensive games in canonical form. $red(G_1) \cong red(G_2)$ iff there is an extensive game G^* such that $G_1 \cong G^*$ with $G_2 \sim G^*$.

Theorem 1 (Thompson (1952))**.** For two extensive games, G_1, G_2, $G_1 \approx G_2$ iff $G_1 \sim G_2$.

Proof. For the left-to-right direction, suppose $G_1 \approx G_2$. By Lemma 2, G_1 and G_2 can be transformed into their canonical forms, G_1', G_2'. If $red(G_1) \cong red(G_2)$, then $red(G_1') \cong red(G_2')$ (Lemma 1). The desired claim immediately follows from Lemma 3. □

This result can be extended slightly based on the following consideration. Strategic equivalence is a plausible analysis of game equivalence if we take strategies to be differentiated by their payoffs. Strategies which produce the exact same payoffs are treated as equivalent and all but one of them as redundant. Analogously, we may also take players to be differentiated by their payoffs. Since all strategic considerations apply equally to such players, we may remove "redundant" players without changing the essential strategic structure of the game.

For illustration, consider the games A and B in Figure 2. The only difference between these two games is that they are played by different sets of players: A is played by 1 and 2; B is played by 1, 2, and 3. Despite the different numbers of players, there is a sense in which these two games are the same. Note that in B, players 2 and 3 share the same utility function. Since their payoffs are the same, any strategic analysis of player 3's performance at the rightmost node in B must suggest the same action as that applied to player 2's performance at the rightmost node in A. This general point applies at every decision point

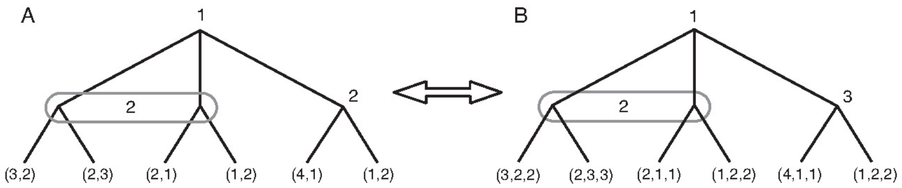


Figure 2: Coalescing of Players

in an extensive game; so long as players with identical payoffs are held to the same standard of rationality, they will seek to ensure the same outcomes obtain. These considerations license a stronger notion than strategic equivalence and the introduction of a new transformation.

Let $G = \langle (W, \prec), I, \{A_i\}_{i\in I}, \{F_i\}_{i\in I}, \{u_i\}_{i\in I}\rangle$ be an extensive game. We define an equivalence relation R_I on I so that, for every $i, j \in I$, $iR_I j$ iff $u_i = u_j$. We write $\bar{I}(i)$ for the equivalence class under R_I containing i and $\bar{I}$ for the set of equivalence classes. The following defines a new transformation that *coalesces* players with the same utility function.

Definition 2.9 (Coalescing of Players)**.** Let G_1 and G_2 be two extensive games with $G_1 = \langle (W, \prec), I, \{A_i\}_{i\in I}, \{F_i\}_{i\in I}, \{u_i\}_{i\in I}\rangle$. We define the equivalence relation $\sim_0$ so that $G_1 \sim_0 G_2$ if $G_2 \cong G^* = \langle (\bigcup_{\mathbf{j}\in I^*} V^*_{\mathbf{j}} \cup Z^*, \prec^*), I^*, \{A^*_{\mathbf{i}}\}_{\mathbf{i}\in I^*}, \{F^*_{\mathbf{i}}\}_{\mathbf{i}\in I^*}, (u^*_{\mathbf{i}})_{\mathbf{i}\in I^*}\rangle$, where

1. $I^* = \bar{I}$
2. $V^*_{\mathbf{i}} = \bigcup_{j\in\mathbf{i}} V_j$
3. $Z^* = Z$
4. $\prec^* = \prec$
5. $F^*_{\mathbf{i}} = \bigcup_{j\in\mathbf{i}} F_j$
6. $A^*_{\mathbf{i}} = \bigcup_{j\in\mathbf{i}} A_j$

Essentially, this transformation treats different players with the same utility function as the same player and, conversely, the same player at distinguishable points of a game as distinct players with identical utility functions.

We now need a new notion of reduced strategic form, corresponding to this extended sense of equivalence.

Definition 2.10 (p-Reduced Strategic Form)**.** Let $G = \langle (W, \prec), I, \{A_i\}_{i\in I}, \{F_i\}_{i\in I}, \{u_i\}_{i\in I}\rangle$ be an extensive game. Let the reduced strategic form of G be $red(G) = \langle I, \bar{S}, \{\bar{U}_i\}_{i\in I}\rangle$. For every $\bar{\mathbf{s}} = (\bar{\mathbf{s}}_i)_{i\in I} \in \bar{S}$, define $\bar{\mathbf{s}}^p = (\bigcup_{j\in\mathbf{i}} s_j)_{\mathbf{i}\in\bar{I}}$. Define $\bar{S}^p = \{\bar{\mathbf{s}}^p \mid \bar{\mathbf{s}} \in S\}$. The p-reduced strategic form $red^p(G)$ of G is a strategic game $\langle \bar{I}, \bar{S}^p, \{\bar{U}^p_{\mathbf{i}}\}_{\mathbf{i}\in\bar{I}}\rangle$ where $\bar{U}^p_{\mathbf{i}}(\bar{\mathbf{s}}^p) = \bar{U}_i(\mathbf{s})$.

Next we extend the notions of transformability and equivalence based on the above definitions. Let G_1 and G_2 be two extensive games.

Definition 2.11 (p-Transformability)**.** G_1 is p-transformable into G_2, written as $G_1 \sim^p G_2$ if there is a sequence of extensive games, $G^*_1, \ldots, G^*_n$, such that $G^*_1 = G_1$, $G^*_n \cong G_2$, and G^*_i $(1 \le i \le n-1)$ is a result of applying one of Thompson's transformation rules or coalescing of players to G^*_{i+1}.

Definition 2.12 (p-Equivalence)**.** G_1 is *p-equivalent* to G_2, written as $G_1 \approx^p G_2$ if $red^p(G_1) \cong red^p(G_2)$.

Theorem 2. For all extensive games, G_1, G_2, $G_1 \approx^p G_2$ iff $G_1 \sim^p G_2$.

Proof. For the right-to-left direction, suppose $G_1 \sim^p G_2$. Then there is a sequence $G_1^*, \ldots, G_k^*$ such that $G_1 = G_1^*$, $G_2 = G_k^*$ and, for all i $(1 \leq i \leq k)$ G_i^* is transformable into G_{i+1}^* by one application of some transformation rule X. If X is one of Thompson's four rules, then $red(G_i) \cong red(G_{i+1})$ (Lemma 1). If X is coalescing of players, it is clear from Definition 2.9 and 2.10 that $red^p(G_i) \cong red(G_{i+1})$ and thus to $red^p(G_{i+1})$.

For the left-to-right direction, let I be the set of players in G_1. By Lemma 2, G_1 is transformable by Thompson's four rules into some game G' such that $sf(G') = red(G')$. Take an extensive game G'' such that $G' \sim^p G''$ where the set of players in G'' is $\bar{I}$. It is clear that $sf(G'') = red^p(G'')$. The rest of the proof follows that for Thompson's theorem. □

3 Games with Unawareness

Our aim is to find corresponding equivalence notions for games with unawareness. We start with basic definitions. Our presentation is minimal; for a full justification and explication of the definitions, readers are referred to Feinberg (2009).

Let X be a non-empty set. X^* is the set of finite sequences of elements in X. We denote the empty sequence in X by λ. When a sequence v is an initial segment of a sequence u, we write $v \leq u$. If $v \leq u$ but $u \not\leq v$, we write $v \prec u$. Also $v\hat{\ }u$ is the concatenation of the sequences, v and u, in that order. When there is no danger of confusion, we will use set theoretical notation for sequences. For instance, for a sequence v and $\mathrm{v} \in X$, $\mathrm{v} \in v$ just means that v appears in the sequence v.

3.1 Strategic Games with Unawareness (SGU)

The key idea of strategic games with unawareness is to assign strategic games to sequences of players, which are called *views*. More precisely, given a strategic game $g = \langle I, \{A_i\}_{i \in I}, \{u_i\}_{i \in I} \rangle$, a *view* is a sequence of agents $v \in I^*$. A strategic game γ with unawareness is a collection of restrictions of g that are assigned in some coherent way (specified in the definition below) to views in a given set $\mathcal{V} \subseteq I^*$. We say a game assigned to a view is the game that view *is aware of* or the view *perceives*. Each game then constitutes the *perspective* of the view to which it is assigned. For example, the game assigned to the view 12 is the game that player 1 perceives player 2 to be aware of; the game assigned to 121 is that which player 1 perceives player 2 to perceive that player 1 is aware of.

Definition 3.1 (Strategic Games with Unawareness)**.** Let $g = \langle I, \{A_i\}_{i \in I}, \{u_i\}_{i \in I} \rangle$. Given a set of views $\mathcal{V} \subseteq I^*$ that includes λ, a collection of strategic games, $\gamma = \{g_v\}_{v \in \mathcal{V}}$, is a *strategic game with unawareness* (SGU), if $g_\lambda = g$ (we call this game *the master game* of γ) and the following conditions are satisfied, where $g_v = \langle I_v, \{(A_i)_v\}_{i \in I_v}, \{(u_i)_v\}_{i \in I_v} \rangle$:

C1 For every $v \in \mathcal{V}$, $v\hat{\ }\mathrm{v} \in \mathcal{V}$ iff $\mathrm{v} \in I_v$.

C2 For every $v\hat{}\tilde{v} \in \mathcal{V}$,

1. $v \in \mathcal{V}$
2. $\emptyset \neq I_{v\hat{}\tilde{v}} \subseteq I_v$
3. $\emptyset \neq (A_i)_{v\hat{}\tilde{v}} \subseteq (A_i)_v$ for all $i \in I_{v\hat{}\tilde{v}}$

C3 If $v\hat{}\mathrm{v}\hat{}\tilde{v} \in \mathcal{V}$, then $g_{v\hat{}\mathrm{v}\hat{}\tilde{v}} = g_{v\hat{}\mathrm{v}\hat{}\mathrm{v}\hat{}\tilde{v}}$ and $v\hat{}\mathrm{v}\hat{}\mathrm{v}\hat{}\tilde{v} \in \mathcal{V}$.

C4 For every action profile $(a)_{v\hat{}\tilde{v}} = \{(a_j)\}_{j \in I_{v\hat{}\tilde{v}}}$, there exists a completion to an action profile $(a)_v = \{a_j, a_k\}_{j \in I_{v\hat{}\tilde{v}}, k \in I_v \setminus I_{v\hat{}\tilde{v}}}$ such that

$$(u_i)_{v\hat{}\tilde{v}}((a)_{v\hat{}\tilde{v}}) = (u_i)_v((a)_v).$$

Intuitively, **C1** says that players must assign perspectives on the game to all the players they can see. **C2** ensures that perspectives only include aspects of the game that can be seen from that view; if I don't know that X is an option, then I can't perceive you as seeing X as an option. **C3** stipulates common knowledge of reflexivity; players are always aware of what they are aware of (and other players know this and assign perspectives accordingly). Finally **C4** is a consistency condition; it ensures that if a player cannot see the entire game, the outcomes that player can see do all obtain in the master game.

Definition 3.2 (Isomorphism on SGU). Let $\gamma = \{g_v\}_{v \in \mathcal{V}}$ and $\gamma' = \{g_{v'}\}_{v' \in \mathcal{V}'}$ be SGU's with g, g' their master games. We say γ is *isomorphic* to γ', written as $\gamma \cong \gamma'$, if there are functions, ρ, α, υ between g and g', as specified in Definition 2.2, such that $g \cong g'$, and:

1. $\mathcal{V}' = \{\rho(\mathrm{v}_1) \dots \rho(\mathrm{v}_k) \mid \mathrm{v}_1 \dots \mathrm{v}_k \in \mathcal{V}\}$
2. for each $g_v \in \gamma$, the restrictions of ρ, α, υ with respect to g_v isomorphically map g_v to $g_{\rho(v)}$, where $\rho(v) = \rho(i_1) \dots \rho(i_n)$ with $v = i_1 \dots i_n$.

When a given set of functions, ρ, α, υ, satisfy the above conditions, we say that they *isomorphically map* γ to γ'.

3.2 Extensive Games with Unawareness (EGU)

Strategic games with unawareness assigned strategic games to sequences of players. Since the awareness of a player may change throughout the course of a game, this strategy will not work in the context of extensive games with unawareness. Instead, we assign extensive games to sequences of decision points. This allows different states of awareness of a single player to be distinguished by the decision point at which they occur. Let $G = \langle (W, \prec), I, \{A_i\}_{i \in I}, \{F_i\}_{i \in I}, \{u_i\}_{i \in I} \rangle$ be an extensive game. Then $W = \bigcup_{i \in I} V_i \cup Z$, where each V_i is the set of player i's decision points. The *set of viewpoints* is the set of all decision points $V = \bigcup_{i \in I} V_i$. The set of *views* is the set of all finite sequences of viewpoints is $\bar{V} = \bigcup_{n=0}^{\infty} V^{(n)}$. We denote views by italicized letters, such as v, u, etc. and viewpoints by unitalicized letters, such as v, u, etc.

An extensive game with unawareness is a collection of restrictions of G that are assigned in a coherent way to views in $\mathcal{V} \subseteq \bar{V}$. We say a game assigned to a view is the game that view *is aware of* or the view *perceives*, as in the case of SGU's. The game assigned to v^u, where $\mathrm{v} \in V_1$ and $\mathrm{u} \in V_2$, is the game that player 1 at v perceives player 2 at u to be aware of; the game assigned to v^u, where v, u $\in V_1$ is the game that player 1 at v perceives himself to be aware of at u. When it is clear, we will say a viewpoint v perceives a game, etc., ignoring the distinction between viewpoints and players.

Definition 3.3 (Extensive Games with Unawareness)**.** Let $G = \langle (W, <), I, \{A_i\}_{i\in I}, \{F_i\}_{i\in I}, \{u_i\}_{i\in I}\rangle$. Given a set of views $\mathcal{V} \subseteq \bar{V}$ that includes λ, a collection of extensive games, $\Gamma = \{G_v\}_{v\in\mathcal{V}}$ is an *extensive game with unawareness* (EGU) if $G_\lambda = G$ (we call this game *the master game*) and the following conditions are satisfied, where $G_v = \langle (W_v, \prec_v), I_v, \{(A_i)_v\}_{i\in I_v}, \{(F_i)_v\}_{i\in I_v}, \{(u_i)_v\}_{i\in I_v}\rangle$:

CE1 For every $v \in \mathcal{V}$, $\mathrm{v} \in V_i$, $v\hat{\ }\mathrm{v} \in \mathcal{V}$ iff $i \in I_v$, $\mathrm{v} \in (V_i)_v$

CE2 For every $v\hat{\ }\bar{v} \in \mathcal{V}$,

1. $v \in \mathcal{V}$
2. $\emptyset \neq W_{v\hat{\ }\bar{v}} \subseteq W_v$
3. $\emptyset \neq I_{v\hat{\ }\bar{v}} \subseteq I_v$.

CE3 For every $v\hat{\ }\bar{v} \in \mathcal{V}$, $i \in I_{v\hat{\ }\bar{v}}$, and $w \in (V_i)_{v\hat{\ }\bar{v}}$,

1. $(V_i)_{v\hat{\ }\bar{v}} = (V_i)_v \cap (W_{v\hat{\ }\bar{v}})\backslash Z_{v\hat{\ }\bar{v}}$
2. $(F_i)_{v\hat{\ }\bar{v}} = \{f \cap (W_{v\hat{\ }\bar{v}}\backslash Z_{v\hat{\ }\bar{v}}) | f \in (F_i)_v\}$
3. $(A_i)_{v\hat{\ }\bar{v}}(w, w') = (A_i)_v(w, w'')$ for the unique successor w'' of w in W_v such that $w'' \leq w'$, where w' is the successor of w in $W_{v\hat{\ }\bar{v}}$.

CE4 If $v\hat{\ }\mathrm{v}\hat{\ }\bar{v} \in \mathcal{V}$ with $\mathrm{v} \in V_i$, then:

1. $f_i(\mathrm{v}) \cap (V_i)_{v\hat{\ }\mathrm{v}} \neq \emptyset$, and for any $\tilde{\mathrm{v}} \in f_i(\mathrm{v})$, $G_{\mathrm{v}} = G_{\tilde{\mathrm{v}}}$
2. for every sequence $\tilde{v}$ all of whose elements are from $f_i(\mathrm{v}) \cap (V_i)_{v\hat{\ }\mathrm{v}}$, $G_{v\hat{\ }\mathrm{v}\hat{\ }\bar{v}} = G_{v\hat{\ }\tilde{v}\hat{\ }\bar{v}}$, and
3. $v\hat{\ }\tilde{v}\hat{\ }\bar{v} \in \mathcal{V}$.

CE5 Let $v\hat{\ }\bar{v} \in \mathcal{V}$. For every terminal node $w \in Z_{v\hat{\ }\bar{v}}$, there exists a node $w' \in Z_v$ such that $w \prec w'$ and $(u_i)_{v\hat{\ }\bar{v}}(w) = (u_i)_v(w')$.

We denote G_λ by $G = \langle (W, <), I, \{A_i\}_{i\in I}, \{F_i\}_{i\in I}, \{u_i\}_{i\in I}\rangle$.

Conditions **CE1**, **CE2**, and **CE3** do essentially the same work for EGU as **C1** and **C2** did for SGU, ensuring that nodes assign views to all nodes they can see and do so in a consistent manner. **CE5** corresponds to **C4**, ensuring that payoffs in restricted games are always consistent with outcomes which might obtain in the master game. A remark is in order concerning condition **CE4**. Feinberg (2009) has a weaker condition, which states that $v\hat{\ }\mathrm{v}\hat{\ }\bar{v} \in \mathcal{V}$ with $\mathrm{v} \in V_i$ implies $f_i(\mathrm{v}) \cap (V_i)_{v\hat{\ }\mathrm{v}} \neq \emptyset$, and if $\tilde{\mathrm{v}} \in f_i(\mathrm{v}) \cap (V_i)_{v\hat{\ }\mathrm{v}}$ then $v\hat{\ }\mathrm{v}\hat{\ }\tilde{\mathrm{v}}\hat{\ }\ldots\hat{\ }\tilde{\mathrm{v}}\hat{\ }\bar{v} \in \mathcal{V}$ and

$$G_{v\hat{\ }\mathrm{v}\hat{\ }\bar{v}} = G_{v\hat{\ }\mathrm{v}\hat{\ }\tilde{\mathrm{v}}\hat{\ }\bar{v}} = \cdots = G_{v\hat{\ }\mathrm{v}\hat{\ }\tilde{\mathrm{v}}\hat{\ }\ldots\hat{\ }\tilde{\mathrm{v}}\hat{\ }\bar{v}}$$

Discussing the condition, Feinberg writes "This definition follows the interpretation of an information set as representing indistinguishable information and indistinguishable awareness" (22). However, the condition is still weak for the stated purpose, since it allows for the possibility that two nodes indistinguishable in the master game assign different perspectives to the game *so long as neither can see each other*. If this possibility obtained, however, it would violate the intuitive restriction Feinberg states that indistinguishable nodes should exhibit the exact same awareness of game structure. Our stronger condition conforms more closely to **C3**, stipulating not only that nodes within an information set must have the same perspective (reflexivity), but also that all players are aware of this property and assign perspectives accordingly (common knowledge of reflexivity). This stronger condition is needed for the proof of Theorem 4.

Definition 3.4 (Isomorphism on EGU)**.** Let $\Gamma = \{G_v\}_{v \in \mathcal{V}}$ and $\Gamma' = \{G_{v'}\}_{v' \in \mathcal{V}'}$ be EGU's with G, G' their master games. We say Γ is *isomorphic* to Γ', written as $\Gamma \cong \Gamma'$, if there are functions, $\rho, \alpha, \upsilon, \phi, \iota$ between G and G', as specified in Definition 2.4, such that $G \cong G'$, and:

1. $\mathcal{V}' = \{\phi(\mathrm{v}_1) \ldots \phi(\mathrm{v}_k) \,|\mathrm{v}_1 \ldots \mathrm{v}_k \in \mathcal{V}\}$
2. for each $G_v \in \Gamma$, the restrictions of $\rho, \alpha, \upsilon, \phi, \iota$ with respect to G_v isomorphically map G_v to $G_{\phi(v)}$, where $\phi(v) = \phi(\mathrm{v}_1) \ldots \phi(\mathrm{v}_n)$ with $v = \mathrm{v}_1 \ldots \mathrm{v}_n$.

When a given set of functions, $\rho, \alpha, \upsilon, \phi, \iota$, satisfy the above conditions, we say that they *isomorphically map* Γ to Γ'.

3.3 Strategic Forms and Equivalence on EGU

When are two EGU's strategically equivalent? We must begin by defining the SGU which corresponds to a given EGU. At a first pass, we might reduce all extensive games in the EGU to their corresponding strategic forms. If one adopts this approach, the strategic form of an EGU Γ will be the collection of strategic forms of extensive games in Γ that are assigned to views in G. However, views in an EGU are sequences of *viewpoints*, whereas views in an SGU are sequences of *players*. A natural move here is to induce sequences of players from sequences of viewpoints in terms of the associated players. Unfortunately, this strategy still does not work, since different viewpoints in a view can be associated with a single player. For illustration, consider the possibility that v and u are distinct (distinguishable) viewpoints for player 1, but $G_{\mathrm{v}\hat{\ }\mathrm{u}}$ and $G_{\mathrm{u}\hat{\ }\mathrm{v}}$ do not share the same strategic form. In an SGU, we cannot associate two distinct games with the corresponding sequence of players, $1\hat{\ }1$.

Therefore, for the strategic form of an EGU Γ, we consider distinguishable viewpoints of player i (in the master game G) as being played by distinct players with the same utility function as i. This decision mirrors the conceptual point that, if a player's state of awareness changes during the game, then the strategic analysis of that player must change as well. Furthermore, if player 1 perceives player 2's state of awareness as changing (even if it does not actually change), player 1 must play as if the two states of player 2's awareness are strategically distinct. (For a related discussion, see the analysis of solution concepts for games with unawareness in Feinberg (2009).)

Given an EGU $\Gamma = \{G_v\}_{v \in \mathcal{V}}$, define the equivalence relation R_{V_i} so that, for all v, u $\in V_i$, $\mathrm{u} R_{V_i} \mathrm{v}$ iff $G_{v\hat{\ }\mathrm{v}\hat{\ }\tilde{v}} = G_{v\hat{\ }\mathrm{u}\hat{\ }\tilde{v}}$ for all $v\hat{\ }\mathrm{v}\hat{\ }\tilde{v}, v\hat{\ }\mathrm{u}\hat{\ }\tilde{v} \in \mathcal{V}$. The relation R_{V_i} partitions the set V_i of player i's viewpoints. We denote by $\bar{V}_i(\mathrm{v})$ the equivalence class under R_{V_i} that contains v $\in V_i$. Note that, by **CE4** in Definition 3.3, every information set $f \in F_i$ is such that $f \subseteq \bar{V}_i(\mathrm{v})$ for some v. The following operation relabels players in an EGU with the equivalence classes $\bar{V}_i(\mathrm{v})$.

Definition 3.5 (*p*-Normal Form)**.** Let $\Gamma = \{G_v\}_{v \in \mathcal{V}}$ be an EGU and for each $v \in \mathcal{V}$, $G_v = \langle (\bigcup_{i \in I_v} (V_i)_v \cup Z, \prec_v), I_v, \{(A_i)_v\}_{i \in I_v}, \{(F_i)_v\}_{i \in I_v}, \{(u_i)_v\}_{i \in I_v} \rangle$. The *p-normal form* Γ^p of Γ is a collection $\{G_v^p\}_{v \in \mathcal{V}}$, where $G_v^p = \langle (\bigcup_{\mathbf{i} \in I_v^p} (V_{\mathbf{i}}^p)_v \cup Z_v^p, \prec_v^p), I_v^p, \{(A_{\mathbf{i}}^p)_v\}_{\mathbf{i} \in I_v^p}, \{(F_{\mathbf{i}}^p)_v\}_{\mathbf{i} \in I_v^p}, \{(u_{\mathbf{i}}^p)_v\}_{\mathbf{i} \in I_v^p} \rangle$ is defined by:

1. $I_v^p = \{\bar{V}_i(\mathrm{v}_i) \mid \mathrm{v}_i \in (V_i)_v\}$, and for each $\mathbf{i} \in I_v^p$
2. $(V_{\mathbf{i}}^p)_v = \mathbf{i} \cap (V_i)_v$, $Z_v^p = Z_v$ and $\prec_v^p = \prec_v$
3. $(A_{\mathbf{i}}^p)_v(w, w') = (A_i)_v(w, w')$ with $w \in (V_{\mathbf{i}}^p)_v$ and $\mathbf{i} \subseteq V_i$
4. $(F_{\mathbf{i}}^p)_v = \{f \cap \mathbf{i} \mid f \in (F_i)_v \text{ and } f \cap \mathbf{i} \neq \emptyset\}$ with $\mathbf{i} \subseteq V_i$
5. $(u_{\mathbf{i}}^p)_v = (u_i)_v$ where $\mathbf{i} \subseteq V_i$

To find the strategic form of an EGU Γ, first transform it into its *p*-normal form, then take the strategic form of each game in Γ^p.

Definition 3.6 (Strategic Form for EGU)**.** Let $\Gamma = \{G_v\}_{v \in \mathcal{V}}$ be an EGU with G its master game. The *strategic form of* Γ is a collection of strategic games $sf(\Gamma) = \{g_{v'}\}_{v' \in \mathcal{V}'}$ defined as follows:

1. $\mathbf{i}_1 \dots \mathbf{i}_n \in \mathcal{V}'$ iff $\mathrm{v}_1 \dots \mathrm{v}_n \in \mathcal{V}$ and $\mathbf{i}_k$ $(1 \leq k \leq n)$ is the player at the viewpoint v_k in Γ^p.
2. Each $g_v = sf(G_v^p)$.

As discussed above, the definition of Γ^p ensures that the game assigned to each view v in $sf(\Gamma)$ is always unique. Therefore, the strategic form of an EGU is well-defined. Indeed let $v, u \in \mathcal{V}$ be distinct views and set $v = \mathrm{v}_1 \dots \mathrm{v}_n$ and $u = \mathrm{u}_1 \dots \mathrm{u}_n$ with $\mathrm{v}_k \in V_{\mathbf{i}_k}^p$ and $\mathrm{u}_k \in V_{\mathbf{j}_k}^p$ $(1 \leq k \leq n)$. If $\mathbf{i}_1 \dots \mathbf{i}_n = \mathbf{j}_1 \dots \mathbf{j}_n$, then v_k and u_k are either identical or in the same equivalence class under R_{V_i}. By the definition of *p*-normal form, $G_v^p = G_u^p$.

To show that $sf(\Gamma)$ is indeed a strategic game with unawareness, we need one proposition. Let G, G' be extensive games such that G' is a restriction of G. The following structural result is straightforward.

Proposition 1. $sf(G')$ is a restriction of $sf(G)$.

Given this proposition, it is easy to check that the conditions **CE1-5** on an EGU Γ in Definition 3.3 guarantee that $sf(\Gamma)$ satisfies the conditions **C1-4** in Definition 3.1. If $G_{v\hat{}\mathrm{v}}$ is a restriction of G_v, as required by **CE1-5**, the strategic form $red(G_{v\hat{}\mathrm{v}})$ (defined in Definition 2.5) is a restriction of $red(G_v)$, as required by **C1-4**. Therefore, we have:

Proposition 2. $sf(\Gamma)$ is a strategic game with unawareness.

We can similarly define the reduced strategic form for EGU.

Definition 3.7 (Reduced Strategic Form for EGU)**.** Let $\Gamma = \{G_v\}_{v \in \mathcal{V}}$ be an EGU with G its master game. The *reduced strategic form of* Γ is a collection of strategic games $red(\Gamma) = \{g_{v'}\}_{v' \in \mathcal{V}'}$ defined as follows:

1. $\mathbf{i}_1 \ldots \mathbf{i}_n \in \mathcal{V}'$ iff $\mathrm{v}_1 \ldots \mathrm{v}_n \in \mathcal{V}$ and $\mathbf{i}_k$ $(1 \leq k \leq n)$ is the player at the viewpoint v_k in Γ^p.

2. Each g_v is the restriction g^* of $red(G^p)$ such that $g^* \cong red(G^p_v)$.

(The reason that we can't simply take g_v to be $red(G^p_v)$ is that the operation *red* takes the equivalence classes of strategies in G^p_v, which may be different from those in G^p since G^p may contain more strategies than its restriction, G^p_v.)

A similar argument to the one above applies to $red(\Gamma)$. Thus it is straightforward to show:

Proposition 3. $red(\Gamma)$ is a strategic game with unawareness.

Next let us consider the notion that corresponds to p-reduced strategic form. As we saw in Section 2, the intuition behind the p-reduced strategic form is that it removes redundant players (i.e. players who share the same payoffs) from the reduced strategic form. For a game with unawareness, players may not be redundant, even if they share the same payoffs, since they may have different perspectives on the game. For example, there may be players $i, j \in I$ such that $u_i = u_j$ yet $A_i \neq A_j$. We can interpret such players in two ways. They may represent different epistemic states of the same player (say, his perspective on the game at different times), or they may represent players on the same "team" who share interests but have different states of knowledge about the game. In either case, the difference between such players is strategically significant, so we wish to keep both in the p-reduced strategic form. Therefore, in order to be considered redundant, players must share both payoffs *and* perspectives on the game.

We now make this idea precise for the definition of p-reduced strategic form on EGU. Let Γ be a p-normal EGU $\{G_v\}_{v \in \mathcal{V}}$. Let I be the set of players in the master game G. Define an equivalence relation R_I on I such that, for $i, j \in I$, $iR_I j$ iff

1. $u_i = u_j$

2. for all $v\hat{\ }\mathrm{v}_i\hat{\ }\tilde{v}, v\hat{\ }\mathrm{v}_j\hat{\ }\tilde{v} \in \mathcal{V}$ with $\mathrm{v}_i \in V_i$ and $\mathrm{v}_j \in V_j$,

$$G_{v\hat{\ }\mathrm{v}_i\hat{\ }\tilde{v}} = G_{v\hat{\ }\mathrm{v}_j\hat{\ }\tilde{v}}.$$

Item 2 is the formalization of the above preliminary discussion of perspectives. It guarantees that, if what any player (including i and j) perceives that i perceives and what that player perceives that j perceives are always the same, then it is redundant to consider i and j as distinct players.

Replace R_I in Definition 2.10 with its new statement and redefine red^p. We write $\bar{I}(i)$ for the equivalence relation under R_I that contains i. We can now define the p-reduced strategic form for EGU.

Definition 3.8 (p-Reduced Strategic Form)**.** Let $\Gamma = \{G_v\}_{v \in \mathcal{V}}$ be an EGU with G its master game. The *strategic form of* Γ is a collection of strategic games $red^p(\Gamma) = \{g_{v'}\}_{v' \in \mathcal{V}'}$ defined as follows:

1. $\bar{I}(\mathbf{i}_1) \ldots \bar{I}(\mathbf{i}_n) \in \mathcal{V}'$ iff $\mathrm{v}_1 \ldots \mathrm{v}_n \in \mathcal{V}$ and $\mathbf{i}_k$ $(1 \leq k \leq n)$ is the player at the viewpoint v_k in Γ^p.

2. Each g_v is the restriction g^* of $red^p(G^p)$ such that $g^* \cong red^p(G^p_v)$.

We need to check that the p-reduced strategic form is well-defined. Indeed let $v, u \in \mathcal{V}$ be distinct views and put $v = \mathrm{v}_1 \dots \mathrm{v}_n$ and $u = \mathrm{u}_1 \dots \mathrm{u}_n$ with $\mathrm{v}_k \in V_{i_k}$ and $\mathrm{u}_k \in V_{j_k}$ $(1 \leq k \leq n)$. If $\bar{I}(\mathbf{i}_1) \dots \bar{I}(\mathbf{i}_n) = \bar{I}(\mathbf{j}_1) \dots \bar{I}(\mathbf{j}_n)$, then (i) v_k, u_k are identical or in the same information set (by p-normality), or (ii) $\mathbf{i}_k R_I \mathbf{j}_k$ (by definition). In each case, we have $G_v = G_u$ by definition of R_I. Therefore, the above definition is well-defined.

Also, it is straightforward to check:

Proposition 4. $red^p(\Gamma)$ is a strategic game with unawareness.

Given the definition of p-reduced strategic form, we can now formulate the definition of strategic equivalence on EGU.

Definition 3.9 (Strategic Equivalence on EGU)**.** Γ_1 is equivalent to Γ_2, written as $\Gamma_1 \approx \Gamma_2$, if $red^p(\Gamma_1)$ is isomorphic to $red^p(\Gamma_2)$.

4 Transformations on EGU

Before introducing the transformations for games with unawareness, we need some supplementary definitions. Given an extensive game G, G^{v} is the subgame of G whose root is v. The following definition for $\mathrm{v}|_i\mathrm{u}$ ensures that two viewpoints, v and u, are the product of different actions by the player i.

Definition 4.1 (Inflation Suitability for i)**.** $\mathrm{v}|_i\mathrm{u}$ if there are $\mathrm{v}', \mathrm{u}' \in V_i$, v'', u'' such that $\mathrm{v}' \in f_i(\mathrm{u}')$, $\mathrm{v}'' \leq \mathrm{v}$, and $\mathrm{u}'' \leq \mathrm{u}$; $A_i(\mathrm{v}', \mathrm{v}'') \neq A_i(\mathrm{u}', \mathrm{u}'')$.

We can now define the transformations on EGU which preserve strategic equivalence. We will formulate transformations as equivalence relations on EGU, as Thompson did on EG. Below let $\Gamma = \{G_v\}_{v \in \mathcal{V}}$ and $\Gamma' = \{G'_{v'}\}_{v' \in \mathcal{V}'}$ be EGU's. Also let $G = \langle (W, \prec), I, \{A_i\}_{i \in I}, \{F_i\}_{i \in I}, \{u_i\}_{i \in I} \rangle$ be the master game of Γ.

Our first transformation takes an EGU into one that has the same p-normal form (as in Definition 3.5). If a player i has distinguishable viewpoints with different perspectives (in the sense of the definition of R_I above), this transformation "splits" player i and relabels the viewpoints with different perspectives as viewpoints of distinct players.

Definition 4.2 (Splitting of Players)**.** $\Gamma \sim_s \Gamma'$ if $\Gamma^p \cong \Gamma'$.

For the p-reduced strategic form of an EGU, we take the p-normal form of the EGU and transform extensive games in it into their p-strategic forms. Therefore, this transformation preserves strategic equivalence on EGU.

The second transformation allows us to consider distinct players i, j equivalent with respect to R_I as the same player.

Definition 4.3 (Coalescing of Players)**.** $\Gamma \sim_c \Gamma'$ if there is an i^* such that $\Gamma \cong \Gamma^* = \{G^*_v\}_{v \in \mathcal{V}}$, where each $G^*_v = \langle (\bigcup_{j \in I^*} V^*_j \cup Z^*, \prec^*), I^*, \{A^*_i\}_{i \in I^*}, \{F^*_i\}_{i \in I^*}, (u^*_i)_{i \in I^*} \rangle$ in Γ^* is defined by

1. $I^* = I_v - (\bar{I}_v(i^*) - \{i^*\})$
2. $V^*_{i^*} = \bigcup_{j \in \bar{I}(i^*)} (V_j)_v$ and $V^*_i = (V_i)_v$ for $i \notin \bar{I}(i^*)$

3. $Z^* = Z_v$ and $\prec^* = \prec_v$
4. $F^*_{i^*} = \bigcup_{j \in \bar{I}(i^*)} (F_j)_v$ and $F^*_i = (F_i)_v$ for $i \notin \bar{I}(i^*)$.
5. $A^*_{i^*} = \bigcup_{j \in \bar{I}(i^*)} (A_j)_v$ and $A^*_i = (A_i)_v$ for $i \notin \bar{I}(i^*)$.

For the p-reduced strategic form of an EGU, we take the equivalence class under R_I on the set of players I. Therefore, this transformation preserves strategic equivalence on EGU. Intuitively, coalescing of players and splitting of players together ensure that we can take any EGU into one with neither redundant players nor players whose perspective changes over the course of the game, i.e. one with the minimal number of distinct players.

The rest of the transformations are extensions of those in Thompson (1952). The essential idea here is that we transform the master game of Γ in accordance with Thompson's original transformation, then ensure that corresponding changes are made to all $G_v \in \Gamma$. Readers are referred again to Figure 1 for intuitive motivation.

Definition 4.4 (Inflation-Deflation)**.** $\Gamma \sim_1 \Gamma'$ if there are v, u $\in W$, and $i \in I$ such that for any v′ $\in f_i$(v) and u′ $\in f_i$(u), v′$|_i$u′ (in G), and $\Gamma' \cong \Gamma^* = \{G^*_v\}_{v \in \mathcal{V}}$, where each G^*_v in Γ^* is exactly the same as G_v except when both v and u are in W_v, in which case the information partition F^*_i in G^* is $((F_i)_v - \{(f_i)_v(\text{v}), (f_i)_v(\text{u})\}) \cup \{(f_i)_v(\text{v}) \cup (f_i)_v(\text{u})\}$.

Inflation-deflation adds and removes perfect recall. It preserves underlying strategic form because it does not change the set of strategies available to any player, merely the size of his information sets. Of course, in general, enlarging information sets will eliminate potential strategies, but inflation suitability (Definition 4.1) ensures that inflation-deflation can only be applied to information sets which are distinguished by an earlier move, thus ensuring that strategies remain distinct. Therefore the reduced strategic form is unchanged.

Definition 4.5 (Addition of a Superfluous Move)**.** $\Gamma \sim_2 \Gamma'$ if there are v, $\text{u}_1, \text{u}_2 \in W$ with $Succ(\text{v}) = \{\text{u}_1, \text{u}_2\}$ such that

1. for every $v \in \mathcal{V}$ such that u_1, $\text{u}_2 \in W_v$, there exist $\rho_v, \alpha_v, \upsilon_v, \phi_v, \iota_v$, as specified in Definition 3.4, such that
   - they isomorphically map $G^{\text{u}_1}_v$ to $G^{\text{u}_2}_v$
   - for any w $\geq \text{u}_1$, $\phi_v(\text{w}) \in (f_i)_v(\text{w})$
2. $\Gamma' \cong \Gamma^* = \{G_v\}_{v \in \mathcal{V}^*}$, where
   - $\mathcal{V}^* = \mathcal{V} - \{v | \exists x \in \{\{\text{v}\} \cup \{\text{w}|\text{u}_2 \preceq \text{w}\}\}(x \in v)\}$
   - each G^*_v is exactly the same as G_v except that the tree W_v in G is replaced with $W_v - (\{\text{v}\} \cup \{\text{w}|\text{u}_2 \preceq \text{w}\})$ and $\prec_v$ and other items in G_v are restricted to $W_v - (\{\text{v}\} \cup \{\text{w}|\text{u}_2 \preceq \text{w}\})$.

Addition of a superfluous move adds a move which has no effect on the strategic structure of the game (in this case, the move at node v). The irrelevance of this move is ensured by the bijection between u_1 and u_2; since the move produces identical, indistinguishable outcomes, it is irrelevant for decision making purposes. One potential worry in extending this transformation to games with unawareness is the role that node v has in views on the game. However, since the irrelevance of the move at v obtains in all restrictions of the master game, no such view can affect strategic decisions and the corresponding reduced strategic form is preserved.

Definition 4.6 (Coalescing of Moves)**.** $\Gamma \sim_3 \Gamma'$ if there are $\{v_1, \ldots, v_k\} = f_i(v_1)$ and $\{u_1, \ldots, u_k\} = f_i(u_1)$ such that

1. $u_m \in Succ(v_m)$ for $1 \leq m \leq k$
2. $A_i(v_m, u_m) = A_i(v_n, u_n)$ for $1 \leq m, n \leq k$.
3. $v\hat{}v_m\hat{}\tilde{v}, v\hat{}u_m\hat{}\tilde{v} \in \mathcal{V}$ implies $G_{v\hat{}v_m\hat{}\tilde{v}} = G_{v\hat{}u_m\hat{}\tilde{v}}$ for $1 \leq m \leq k$
4. $\Gamma' \cong \Gamma^* = \{G^*_v\}_{v \in \mathcal{V}^*}$, where Γ^* is defined by:
   - $\mathcal{V}^* = \mathcal{V} - \{v \mid \exists x \in \{u_1, \ldots, u_k\}(x \in v)\}$
   - each G^*_v is the same as G_v except that W_v is replaced with $W_v - \{u_1, \ldots u_k\}$ and $<_v$ and other items in G are restricted to $W_v - \{u_1, \ldots u_k\}$.

Coalescing of moves combines into a single decision point moves made by the same player in sequence. Again, the main worry in extending this transformation to games with unawareness is ensuring that the transformation only applies when the player's state of awareness does not change between the information sets. This restriction is taken care of by condition 3, ensuring that the reduced strategic form remains unchanged.

We now define the final transformation: *interchange of moves*. We say an extensive game G is *binary*, if for any viewpoint v in G, $Succ(v) \leq 2$. Similarly EGU $\Gamma = \{G_v\}_{v \in \mathcal{V}}$ is *binary*, if for all $v \in \mathcal{V}$, G_v is binary.

Definition 4.7 (IM Suitability)**.** Let $G = \langle (W, <), I, \{A_i\}_{i \in I}, \{F_i\}_{i \in I}, \{u_i\}_{i \in I} \rangle$ be a binary extensive game. G is *IM suitable* if there are distinct viewpoints, $v, u_1, u_2 \in W$ and $i, j \in I$ such that

1. $v \in V_i$ and $u_1, u_2 \in V_j$
2. $Succ(v) = \{u_1, u_2\}$
3. $u_2 \in f_j(u_1)$

If the master game of Γ is IM suitable, we would like to be able to apply Thompson's interchange of moves transformation. However, we need to guarantee our extension of this transformation preserves conditions **CE1–CE5**, in particular, that all $G_v \in \Gamma$ are still restrictions of the master game. In order to do this, we need to consider several cases which might obtain with respect to the viewpoints, v, u_1, u_2 in any G_v. We define the operation *IM* to produce an extensive game $IM(G_0)$ on restrictions G_0 of the extensive game G in the following way. Below we denote each item of $IM(G)$ by items in G superscripted with ^{IM}.

IM1 If none of v, u_1, u_2 are in G_0, $IM(G_0) = G_0$.

IM2 If all of v, u_1, u_2 are in G_0 and there are distinct viewpoints, w_1, w_2, x_1, x_2 in G_0 such that $Succ(u_1) = \{w_1, w_2\}$, $Succ(u_2) = \{x_1, x_2\}$, then:

(a) $G - G^v = IM(G) - (IM(G))^v$

(b) $(IM(G))^{w_k} = G^{w_k}$ and $IM(G)^{x_k} = G^{x_k}$ $(k \in \{1,2\})$

(c) $u_1, u_2 \in V_i^{IM}$ and $v \in V_j^{IM}$ in $IM(G)$

(d) $Succ^{IM}(u_1) = \{w_1, x_1\}$ and $Succ^{IM}(u_2) = \{w_2, x_2\}$

(e) $f_i^{IM}(u_1) = (f_i(v) - \{v\}) \cup \{u_1, u_2\}$ and $f_j^{IM}(v) = (f_j(u_1) - \{u_1, u_2\}) \cup \{v\}$

(f)
- $A_i^{IM}(u_1, w_1) = A_i^{IM}(u_2, x_1) = A_i(v, u_1)$
- $A_i^{IM}(u_1, w_2) = A_i^{IM}(u_2, x_2) = A_i(v, u_2)$
- $A_j^{IM}(v, u_1) = A_j(u_1, w_1) = A_j(u_2, x_1)$
- $A_j^{IM}(v, u_2) = A_j(u_1, w_2) = A_j(u_2, x_2)$

IM3 If all of v, u_1, u_2 are in G_0 and there are distinct viewpoints, w, x in G_0 such that $Succ(u_1) = \{w\}$, $Succ(u_2) = \{x\}$, then:

(a) $G - G^v = IM(G) - (IM(G))^v$

(b) $(IM(G))^w = G^w$ and $IM(G)^x = G^x$

(c) $u_1 \in V_i^{IM}$ and $v \in V_j^{IM}$

(d) $Succ^{IM}(u_1) = \{w, x\}$

(e) $f_i^{IM}(u_1) = (f_i(v) - \{v\}) \cup \{u_1\}$ and $f_j^{IM}(v) = (f_j(u_1) - \{u_1, u_2\}) \cup \{v\}$

(f)
- $A_j^{IM}(v, u_1) = A_j(u_1, w) = A_j(u_2, x)$
- $A_i^{IM}(u_1, w) = A_i(v, u_1)$
- $A_i^{IM}(u_1, x) = A_i(v, u_2)$

IM4 If only v, u_1 are in G_0 and there are distinct viewpoints, w_1, w_2 in G_0 such that $Succ(u_1) = \{w_1, w_2\}$, then:

(a) $G - G^v = IM(G) - (IM(G))^v$

(b) $(IM(G))^{w_k} = G^{w_k}$ $(k \in \{1,2\})$

(c) $u_1, u_2 \in V_i^{IM}$ and $v \in V_j^{IM}$ in $IM(G)$

(d) $Succ^{IM}(v) = \{u_1, u_2\}$, $Succ^{IM}(u_1) = \{w_1\}$, and $Succ^{IM}(u_2) = \{w_2\}$

(e) $f_i^{IM}(u_1) = (f_i(v) - \{v\}) \cup \{u_1, u_2\}$ and $f_j^{IM}(v) = (f_j(u_1) - \{u_1\}) \cup \{v\}$

(f)
- $A_j^{IM}(v, u_1) = A_j(u_1, w_1)$
- $A_j^{IM}(v, u_2) = A_j(u_1, w_2)$
- $A_i^{IM}(u_1, w_1) = A_i^{IM}(u_2, w_2) = A_i(v, u_1)$

IM5 If only v, u_2 are in G_0 and there are distinct viewpoints, x_1, x_2 in G_0 such that $Succ(u_2) = \{x_1, x_2\}$, then $IM(G_0)$ is defined just as in **IM4** except with u_1, u_2, w_1, w_2 replaced by u_2, u_1, x_1, x_2 respectively.

IM6 If only v, u_1 are in G_0 and there is a viewpoint w in G_0 such that $Succ(u_1) = \{w\}$, then:

(a) $G - G^v = IM(G) - (IM(G))^v$

(b) $(IM(G))^w = G^w$

(c) $u_1 \in V_i^{IM}$ and $v \in V_j^{IM}$

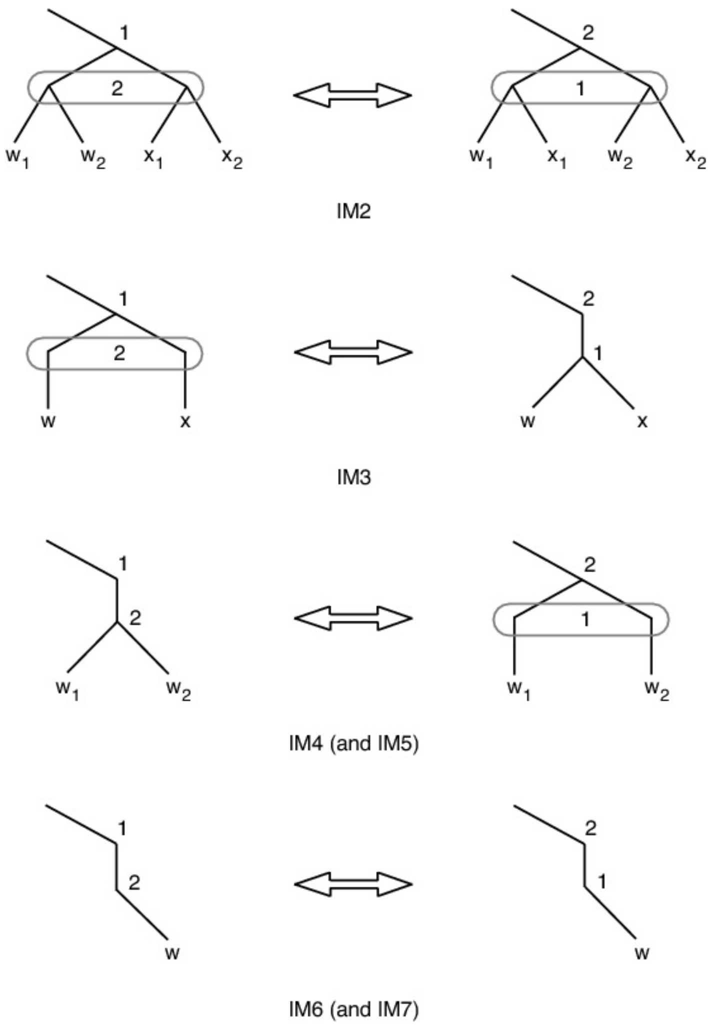


Figure 3: **IM2–IM7**

(d) $f_i^{IM}(\mathrm{u}_1) = (f_i(\mathrm{v}) - \{\mathrm{v}\}) \cup \{\mathrm{u}_1\}$ and $f_j^{IM}(\mathrm{v}) = (f_j(\mathrm{u}_1) - \{\mathrm{u}_1\}) \cup \{\mathrm{v}\}$

(e)
- $A_j^{IM}(\mathrm{v}, \mathrm{u}_1) = A_j(\mathrm{u}_1, \mathrm{w})$
- $A_i^{IM}(\mathrm{u}_1, \mathrm{w}) = A_i(\mathrm{v}, \mathrm{u}_1)$

IM7 If only v, u_2 are in G_0 and there is a viewpoint x in G_0 such that $Succ(\mathrm{u}_2) = \{\mathrm{x}\}$, then $IM(G_0)$ is defined just as in **IM6** except with u_1, w replaced with u_2, x respectively.

The operation *IM* "interchanges" the moves of players i and j in the region of G_0 corresponding to $\mathrm{v}, \mathrm{u}_1, \mathrm{u}_2 \in W$ (as in Definition 4.7), leaving the rest of the game tree unchanged. When u_1 and u_2 are each followed by two daughter nodes, it transforms the game as in Thompson's original interchange of moves. *IM* also interchanges the moves of players i and j when the specified part of the tree is restricted in G_0. Since we assume that G is binary, the cases provided in the above definition exhaust all possible restrictions of G affecting the interchange of nodes v, u_1, and u_2 (see Figure 3).

Definition 4.8 (Interchange of Moves)**.** $\Gamma \sim_4 \Gamma'$ if (i) Γ is binary, (ii) there are $\mathrm{v}, \mathrm{u}_1, \mathrm{u}_2 \in W$ and $i, j \in I$ in the master game G of Γ that satisfy *IM Suitability*, and (iii) $\Gamma' \cong \Gamma^* = \{G^*_{v^*}\}_{v^* \in \mathcal{V}^*}$, where $\mathcal{V}^*$ and each $G^*_{v^*}$ are defined as follows:

1. The master game of $\Gamma^* = G^*_\lambda$ is $IM(G)$.
2. $v^*\hat{\ }\mathrm{v}^* \in \mathcal{V}^*$ iff $v^* \in \mathcal{V}^*$ and v^* is in G_{v^*}
3. $G^*_{v^*} = IM(G_{v^*})$ for $v^* \in \mathcal{V}$
4. $G^*_{v^*\hat{\ }\mathrm{v}^*}$ is a restriction of $G^*_{v^*}$ that is isomorphic to $IM(G_{v_0})$ for $v^*\hat{\ }\mathrm{v}^* \notin \mathcal{V}$, where v_0 is in $\mathcal{V}$ and obtained by replacing some occurrences of u_1 (respectively, u_2) in $v^*\hat{\ }\mathrm{v}^*$ with u_2 (respectively, u_1).

A brief comment on point 4 is in order. It ensures that the node introduced in **IM4** (respectively, **IM5**) agrees in its awareness with the other node in its information set. We have stated the condition slightly imprecisely to suppress irrelevant bookkeeping details.

It is straightforward to check that the Γ^* constructed in Definition 4.8 is an EGU. Given *IM* and the fourth condition in this definition, restrictions G_v of G get transformed to restrictions of $IM(G)$.

Two remarks. First, interchange of moves may only be applied to a binary EGU. However, it is always possible to convert an EGU into a binary EGU through successive applications of coalescing of moves and addition of superfluous moves (as in Thompson's original result). Second, $\mathcal{V}^*$ is different from $\mathcal{V}$. If G_v ($v \in \mathcal{V}$) is of the form specified in **IM4** (or **IM5**), then G_v does not contain u_2 (respectively, u_1), but $IM(G_v)$ will contain u_2 (respectively, u_1). This means that $v\hat{\ }\mathrm{u}_2 \notin \mathcal{V}$ (since u_2 is not in G_v, by **CE1** in Definition 3.3), yet $v\hat{\ }\mathrm{u}_2 \in \mathcal{V}^*$. However, the corresponding reduced strategic form is not changed because it collapses all nodes within an information set to a single decision point.

Definition 4.9 (Transformability on EGU)**.** Γ_1 is transformable into Γ_2, written as $\Gamma_1 \sim \Gamma_2$, if there is a sequence of EGU's, $\Gamma^*_1, \ldots, \Gamma^*_n$ such that $\Gamma^*_1 = \Gamma_1$, Γ^*_n is isomorphic to Γ_2, and $\Gamma^*_i \sim_t \Gamma_{i+1}$ ($1 \leq i \leq n-1$) with $t \in \{s, c, 1, 2, 3, 4\}$ (Γ_i is the result of applying one of the rules defined above to Γ_{i+1}).

5 Equivalence and Transformability

Given the definitions for strategic equivalence and transformability on EGU, we now prove the following result on EGU analogous to Thompson's result on extensive games: *For every EGU* Γ_1, Γ_2, $\Gamma_1 \sim \Gamma_2$ *iff* $\Gamma_1 \approx \Gamma_2$.

As we discussed above, it is straightforward to see that the transformations defined in Definitions 4.2–4.6 and 4.8 preserve strategic equivalence. Readers are invited to verify this.

Theorem 3. $\Gamma_1 \sim \Gamma_2$ implies $\Gamma_1 \approx \Gamma_2$.

For the other direction, we need some lemmas. First note the following facts about the transformations defined above. An application of a transformation transforms the master game of a given EGU Γ in the same way as the corresponding rule for extensive games; furthermore, the games $G_v \in \Gamma$ are each transformed similarly. For instance, an application of coalescing of moves to Γ will coalesce moves in the master game G of Γ in exactly the same fashion as Thompson's coalescing of moves would on G as a standard extensive game. Likewise, if they appear, the same moves are coalesced in each $G_v \in \Gamma$. Therefore, the following is a consequence of Lemma 2:

Observation 1. Let Γ be an EGU and G its master game. There is an EGU Γ' with G' its master game such that $\Gamma \sim \Gamma'$ and $sf(G')$ is isomorphic to $red^p(G')$.

Next, note that the extensive game assigned to each view is a restriction of the master game G. Also a brief inspection of Definition 2.10 reveals that any restriction of an extensive game in canonical form is an extensive game in canonical form. Therefore, in the above proposition, if $sf(G')$ is isomorphic to $red^p(G')$, for any restriction G'_v in Γ', $sf(G'_v)$ is isomorphic to $red^p(G'_v)$. Therefore, we have the following result that is analogous to Lemma 2 on extensive games.

Lemma 4. For every EGU Γ, there is an EGU Γ' such that $\Gamma \sim \Gamma'$ and $sf(\Gamma')$ is isomorphic to $red^p(\Gamma')$.

Let us say an EGU Γ is in *canonical form* if $sf(\Gamma')$ is isomorphic to $red^p(\Gamma)$.

Finally, we need a result analogous to Lemma 3.

Lemma 5. Let Γ, Γ' be EGU's in canonical form. $red^p(\Gamma)$ is isomorphic to $red^p(\Gamma')$ iff there is an EGU Γ^* such that Γ is isomorphic to Γ^* with $\Gamma' \sim \Gamma^*$.

Proof. The right-to-left direction is clear by Theorem 3. For the other direction, by Lemma 4, we can assume without loss of generality that Γ, Γ' are in canonical form. Moreover, by splitting of players (Definition 4.2), we can transform Γ, Γ' into p-normal form. Therefore, we can assume that Γ and Γ' are both isomorphic to their p-normal forms.

Next, let G, G' be the master games of Γ, Γ' respectively. By Lemma 3, there is an extensive game G^* such that $G \cong G^*$ and $G' \sim G^*$. Thus there is a function ϕ as specified in Definition 3.4 between the set of viewpoints in G and the set of viewpoints in G^*. Also, given $G' \sim G^*$, there is an EGU Γ^* such that $\Gamma' \sim \Gamma^*$ and G^* is the master game of Γ^*. Set $\Gamma = \{G_v\}_{v \in \mathcal{V}}$ and $\Gamma^* = \{G^*_{v^*}\}_{v^* \in \mathcal{V}^*}$. Define a function Φ from $\mathcal{V}$ to $\mathcal{V}^*$ so that, for all $v = \mathrm{v}_1 \ldots \mathrm{v}_n \in \mathcal{V}$, $\Phi(v) = \phi(\mathrm{v}_1) \ldots \phi(\mathrm{v}_n)$.

Our goal is to show that Γ is isomorphic to Γ^*. Let $\mathcal{V}_k = \{v \mid v \in \mathcal{V} \text{ and } \mathsf{len}(v) = k\}$, where $\mathsf{len}(v)$ is the length of v. By Definition 3.4, it suffices to show that, for all $v \in \mathcal{V}$ with $\mathsf{len}(v) = k$,

1. $\{\Phi(w) \mid w \in \mathcal{V}_{\mathsf{len}(v)}\} = \mathcal{V}^*_k$, and
2. $G_v \cong G_{\Phi(v)}$ with G_v and $G_{\Phi(v)}$ in canonical form.

Since 1 implies Φ is one-to-one, and, from 2 and Definition 3.4, it follows that Γ is isomorphic to Γ^*.

We prove the claim by induction on the length of $v \in \mathcal{V}$. The base case is clear. For the inductive step, first note that the functions that isomorphically map G to G^* induce the functions that isomorphically map $red^p(G)$ to $red^p(G^*)$. Also, given Definition 2.2, if $red^p(\Gamma) \cong red^p(\Gamma^*)$, the functions induce the functions that isomorphically map $red^p(\Gamma)$ to $red^p(\Gamma^*)$. In particular, define a function $\rho : I \to I^*$ such that, for all v_i in G and $\mathrm{v}^*_{i^*}$ in G^* with $\phi(\mathrm{v}_i) = \mathrm{v}^*_{i^*}$, $\rho(i) = i^*$. Then ρ is a function as specified in Definition 3.2. Put $red^p(\Gamma) = \{g_u\}_{u \in \mathcal{U}}$ and $red^p(\Gamma^*) = \{g_{u^*}\}_{u^* \in \mathcal{U}^*}$. Now, by the inductive hypothesis, $\{\Phi(w) \mid w \in \mathcal{V}_{\mathsf{len}(v)}\} = \mathcal{V}^*_k$ for an arbitrary k. Let $v = v_1 \dots v_k \in \mathcal{V}$ with $\mathsf{len}(v) = k$ and $\Phi(v) = v^*$. Let $\mathbf{i}=i_1 \dots i_k$ be the sequence of players such that $\mathrm{v}_l \in V_{i_l}$ $(1 \le l \le k)$. Let $\mathbf{i}\hat{\ }i$ be an arbitrary extension such that $\mathbf{i}\hat{\ }i \in \mathcal{U}$. We need to show that $\{\phi(\mathrm{v}) \in V_i \mid v\hat{\ }\mathrm{v} \in \mathcal{V}\} = \{\mathrm{v}^* \in V^*_{\rho(i)} \mid v^*\hat{\ }\mathrm{v}^* \in \mathcal{V}^*\}$. This follows from the inductive hypothesis that $G_v \cong G_{\Phi(v)}$, which implies by **CE1**: if $v\hat{\ }\mathrm{v} \in \mathcal{V}$, then $v^*\hat{\ }\phi(\mathrm{v}) \in \mathcal{V}^*$; and if $v^*\hat{\ }\mathrm{v}^* \in \mathcal{V}^*$, then there must be $\mathrm{v} \in \mathcal{V}$ such that $\phi(\mathrm{v}) = \mathrm{v}^*$. Since i and v are arbitrary, this proves that $\{\Phi(w) \mid w \in \mathcal{V}_{\mathsf{len}(v)}\} = \mathcal{V}^*_{k+1}$.

For the other part of the claim, by inductive hypothesis, $G_v \cong G_{\Phi(v)}$ for an arbitrary v with $\mathsf{len}(v) = k$, where $G_v, G_{\Phi(v)}$ are in canonical form. Let $v\hat{\ }\mathrm{v} \in \mathcal{V}$. First, it is clear (by Definitions 2.5 and 2.10) that $G_{v\hat{\ }\mathrm{v}}$ and $G^*_{\Phi(v\hat{\ }\mathrm{v})}$ are both in canonical form, since $G_{v\hat{\ }\mathrm{v}}, G^*_{\Phi(v\hat{\ }\mathrm{v})}$ are restrictions of $G_v, G^*_{v^*}$ respectively (by **CE 2**, **3**, and **5**), which are already in canonical form. Therefore, $sf(G_{v\hat{\ }\mathrm{v}}) \cong red^p(G_{v\hat{\ }\mathrm{v}})$ and $sf(G_{\Phi(v\hat{\ }\mathrm{v})}) \cong red^p(G_{\Phi(v\hat{\ }\mathrm{v})})$. Next, by assumption, $red^p(\Gamma) \cong red^p(\Gamma^*)$. By the definition of Φ above, we have $red^p(G_{v\hat{\ }\mathrm{v}}) \cong red^p(G_{\Phi(v\hat{\ }\mathrm{v})})$. This gives us $sf(G_{v\hat{\ }\mathrm{v}}) \cong sf(G_{\Phi(v\hat{\ }\mathrm{v})})$. Also, by assumption, Γ and Γ^* are in p-normal form, and $G_{v\hat{\ }\mathrm{v}}, G_{\Phi(v\hat{\ }\mathrm{v})}$ are restrictions of G_v and $G_{\Phi(v)}$, respectively, which are isomorphic to each other. Therefore, it follows from Definition 2.5 that $G_{v\hat{\ }\mathrm{v}} \cong G_{\Phi(v\hat{\ }\mathrm{v})}$. □

Theorem 4. For every EGU Γ_1, Γ_2, $\Gamma_1 \sim \Gamma_2$ iff $\Gamma_1 \approx \Gamma_2$.

Proof. The desired result follows from Theorem 3 and Lemmas 4 and 5 by an argument analogous to that given for Theorem 1. □

6 Conclusion

We have extended the Thompson transformations to games with unawareness. Along the way, we identified a novel transformation, coalescing of players, based on the insight that players are differentiated by their payoffs and, therefore, players with identical payoffs are strategically equivalent. In the case of games with unawareness, we discovered that players are differentiated by both payoffs and awareness. Consequently, from a strategic standpoint, a single player should be analyzed as two distinct agents if he exhibits two distinct states of awareness at different stages of a temporally extended game. The fact that these two agents share payoffs, i.e. they constitute a team, will ensure their actions are strategically coordinated. The next stage in this project is the investigation of solution concepts: which solution concepts are preserved under each of these transformations and which are not?

Acknowledgements The authors would like to thank Eric Pacuit for extensive comments at early stages of this project. They would like to thank Johan van Benthem not only for his constructive comments, but also his moral and financial support at various stages of this work. Finally, they thank the participants in both the Stanford *Logical Dynamics Workshop* and the ILLC *Seminar on Logics for Dynamics of Information and Preferences* for stimulating comments and discussion at presentations of earlier versions of this material.

References

G. Bonanno. Set-theoretic equivalence of extensive-form games. *International Journal of Game Theory*, 20:429–447, 1992.

B. de Bruin. Game transformations and game equivalence. Technical Report X-1999-01, ILLC, Amsterdam, 1999.

S. Elmes and P. J. Reny. On the strategic equivalence of extensive form games. *Journal of Economic Theory*, 62(1):1–23, 1994.

Y. Feinberg. Games with unawareness. preprint (2009), available at `http://www.stanford.edu/ yossi/OpenGwithU.html`, 2009.

F. Klein. A comparative review of recent researches in geometry. *Bulletin of the New York Mathematical Society*, 2:215–249, 1893.

E. Kohlberg and J.-F. Mertens. On the strategic stability of equilibria. *Econometrica*, 54(5):1003–1037, 1986.

D. Lewis. *Convention*. Harvard UP, Cambridge, MA, 1969.

M. J. Osborne and A. Rubinstein. *A Course in Game Theory*. The MIT Press, Cambridge, MA, 1994.

F. B. Thompson. Equivalence of games in extensive form. Technical Report RM-759, RAND Corporation, Washington, D.C., 1952.

Logic of Information Flow on Communication Channels

Yanjing Wang, Floor Sietsma and Jan van Eijck

Centrum Wiskunde en Informatica,
`y.wang, f.sietsma, Jan.van.Eijck@cwi.nl`

Abstract

In this paper[1], we develop an epistemic logic to specify and reason about information flow on the underlying communication channels. By combining ideas from Dynamic Epistemic Logic (DEL) and Interpreted Systems (IS), our semantics offers a natural and neat way of modelling multi-agent communication scenarios with different assumptions about the observational power of agents. We relate our logic to the standard DEL and IS approaches and demonstrate its use by studying a telephone call communication scenario. [2]

1 Introduction

The 1999 'National Science Quiz' of *The Netherlands Organisation for Scientific Research (NWO)*[3] had the following question:

> *Six friends each have one piece of gossip. They start making phone calls. In every call they exchange all pieces of gossip that they know at that point. How many calls at least are needed to ensure that everyone knows all six pieces of gossip?*

To reason about the information flow in such a scenario, we want to take into account the following issues: the messages that the agents possess (e.g. secrets), the knowledge of the agents, the dynamics of the system in terms of information passing (e.g. telephone calls) and the underlying communication channels (e.g. the network of landlines). To incorporate specific designs for such issues, we first need to make a choice between two mainstream logical frameworks for multi-agent systems: *Interpreted Systems* and *Dynamic Epistemic Logic*.

[1] This paper is based on an extended abstract to appear in the proceedings of AAMAS10.
[2] The first author is supported by NWO project VEMPS 612.000.528.
[3] For a list of references about the problem c.f. Hurkens (2000).

Interpreted Systems (ISs), introduced by Parikh and Ramanujam (1985) and Fagin et al. (1995) independently, are mathematical structures that combine history-based temporal components of a system with epistemic ones (defined in terms of *local states* of the agents). ISs are convenient to model knowledge development based on the given temporal development of a system. In ISs the epistemic structure is generated from the temporal structure in a uniform way. However, the generation of temporal structures is not specified in the framework.

A different perspective on the dynamics of multi-agent systems is provided by Dynamic Epistemic Logic (DEL) (Gerbrandy and Groeneveld 1997, Baltag and Moss 2004). The main focus of DEL is not on the temporal structure of the system but on the epistemic impact of events as the agents perceive them. The development of a system through time is essentially generated by executing so-called *action models* on a static initial model, to generate an updated static model. The epistemic relations in the initial static model and in the action models are not generated uniformly as in IS. Instead, they are designed by hand. It is customary to start out from a static situation of universal ignorance, where the ignorance is supposed to be common knowledge[4].

In recent years, much has been said about the comparison of the two frameworks, based on the observation that certain temporal developments of the system in IS can be generated by sequences of DEL updates on static models (see, e.g., van Benthem et al. (2009), Hoshi and Yap (2009), Hoshi (2009)). In this paper, we will demonstrate further benefits of combining the two approaches by presenting a framework where epistemic relations are generated by matching local states and a history of observations as in ISs, while keeping the flexibility of explicit actions as in DEL approaches.

The puzzle of the telephone calls was briefly discussed in van Ditmarsch (2000, Ch. 6.6) within the original DEL framework. van Benthem (2002) raised the research question whether the communication network can be made explicit in DEL. An early proposal to fill in this line of research can be found in Roelofsen (2005). Communication channels in an IS framework made their appearance in Parikh and Ramanujam (2003). Recent work in (Pacuit and Parikh 2007, Apt et al. 2009) addresses the information passing on so-called *communication graphs* or *interaction structures*, where *"messages"* are either atomic propositions or Boolean combinations of atomic propositions. In Wang et al. (2009) a PDL-style DEL language is developed that allows explicit specification of protocols. The present paper attempts to blend the DEL and IS approaches to model communication along channels. More specifically, the contributions of this paper are:

- Combining insights from Dynamic Epistemic Logics and Interpreted Systems, we propose a logic $\mathcal{L}_\iota^{I,M}$ to specify and reason about the information flow over underlying communication channels. Unlike in previous work in Pacuit and Parikh (2007), Apt et al. (2009), Roelofsen (2005), we can *specify* the communication protocols in our language and deal with information flow in terms of both *messages* and higher-order formulas.
- The semantics of $\mathcal{L}_\iota^{I,M}$ is given on single-state models with respect to

[4]In a situation with n atomic propositions, this gives an initial model consisting of 2^n worlds, with universal accessibility relations for all agents.

different observational equivalence relations generated in IS-style, which are also studied and compared in this paper.

- The DEL-style actions in $\mathcal{L}_\iota^{I,M}$ allow us to model various communicative actions such as message passing and group announcements. In particular we define an external informing action, which essentially announces the protocol that agents are supposed to follow, thus making it common knowledge that the future behavior of the agents is constrained. It turns out to make a crucial difference whether epistemic protocols such as those discussed in van Ditmarsch et al. (2007) are assumed to be common knowledge among the agents carrying out the protocol or not (see also Wang et al. (2009)).
- Taking advantage of our semantics, we also propose a generic method of epistemic modeling where the initial model is simply the *real world* and all the initial assumptions are specified explicitly by means of formulas of $\mathcal{L}_\iota^{I,M}$. This significantly simplifies the modeling procedure. According to our semantics, the relevant possible states can be automatically constructed while evaluating the formulas. In particular, there is no need to specify the whole state space at the beginning.
- As a case study, we model telephone communications among agents. We show that it is impossible to obtain new common knowledge by telephone calls or voice mails but that we can get arbitrarily close to common knowledge if we not only can send messages but also make statements like "I know j got message m".

The paper is organized as follows. We introduce our logic $\mathcal{L}_\iota^{I,M}$ in Section 2. Section 3 relates our logic to the standard DEL and IS approaches. Section 4 introduces a modeling method and illustrates this method by a study of variations on the puzzle that was mentioned above. The final section concludes and lists future work.

2 Logic $\mathcal{L}_\iota^{I,M}$

2.1 Language

Let I be a finite set of agents, M a finite set of message terms and A a finite set of basic actions with internal structures given by an action map ι defined later. A communication network *net* is represented as a hypergraph of agents in I, namely a set of subsets of I as in Apt et al. (2009). For example if $net = \{\{1,2\},\{1,2,3\}\}$ then there is a private channel $\{1,2\}$ between agents 1 and 2 and there is a public channel used by all three agents.

The set $Prop_{I,A,M}$ of basic propositions is defined by

$$p ::= has_i m \mid com(G) \mid past(\bar{\alpha}) \mid future(\bar{\alpha})$$

with $i \in I$, $m \in M$, $G \subseteq I$ and $\bar{\alpha} = \alpha_0; \alpha_1; \ldots; \alpha_k \in A^*$.

$has_i m$ is intended to mean that i possesses the message m[5]; while $com(G)$ expresses that group G forms a channel in the network; $past(\bar{\alpha})$ says that the

[5] *has* is a commonly used predicate in the logic of security protocols to model declarative knowledge about messages c.f., e.g., Ramanujam and Suresh (2005).

sequence of actions $\bar{\alpha}$ just happened and $future(\bar{\alpha})$ means that $\bar{\alpha}$ can be executed according to the current protocol. The formulas of $\mathcal{L}_\iota^{I,M}$ are built from the set $Prop_{I,A,M}$ as follows:

$$\begin{array}{lll}\phi & ::= & \top \mid p \mid \neg\phi \mid \phi_1 \wedge \phi_2 \mid \langle\pi\rangle\phi \mid C_G\phi \\ \pi & ::= & \alpha \mid \varepsilon \mid \delta \mid \pi_1;\pi_2 \mid \pi_1 \cup \pi_2 \mid \pi^*\end{array}$$

with $p \in Prop_{I,A,M}$, $G \subseteq I$, $\alpha \in A$ and ε, δ as constants for empty sequence and deadlock respectively.

The intended meaning of the formulas is mostly as usual as in dynamic epistemic logics: $C_G\phi$ expresses "*the agents in group G commonly know* ϕ", $\langle\pi\rangle\phi$ expresses "*the protocol* π *can be executed, and at least one execution of* π *yields a state where* ϕ *holds*". Let Π be the set of all protocols π and $Form^{-\langle\pi\rangle}(\mathcal{L}_\iota^{I,M})$ be the set of all the $\mathcal{L}_\iota^{I,M}$ formulas without $\langle\pi\rangle$ modalities. Each $\alpha \in A$ has an internal structure given by $\iota : A \to \mathcal{P}(I) \times Form^{-\langle\pi\rangle}(\mathcal{L}_\iota^{I,M}) \times (\mathcal{P}(M))^{|I|} \times (\Pi \cup \{\#\})$. Thus $\iota(\alpha)$ is a tuple:

$$\langle G, \phi, M_0 \ldots M_{|I|}, \rho\rangle$$

Here we define $Obs(\iota(\alpha)) = G$ as the set of agents that can observe α; $Pre(\iota(\alpha)) = \phi$ is the precondition that should hold in order for α to be executable; $Pos(\iota(\alpha)) = \langle M_0 \ldots M_{|I|}, \rho\rangle$ (with $\rho \in \Pi \cup \{\#\}$) is the postcondition which lists the set of messages M_i that get delivered to i by action α for each i and the protocol ρ that the agents are going to follow after execution of α. If $\rho = \#$, then the agents should keep following the current protocol. If $\rho = \pi \in \Pi$ then they should change their protocol to π. In this paper we assume that the agents can always observe the actions that deliver messages to them: if $\iota(\alpha)$ has $M_j \neq \emptyset$ then $j \in Obs(\iota(\alpha))$. The converse does not hold since agents may also observe actions that do not deliver any messages to them.

Note that by excluding the preconditions of the form $\langle\pi\rangle\phi$, the interdependence of actions are limited but still useful, e.g., for action α, $future(\alpha)$ is allowed as a precondition meaning that α can be executed only when it was planned according to the current protocol.

As usual, we define $\bot$, $\phi \vee \psi$, $\phi \to \psi$, $\langle C_G\rangle\phi$ and $[\pi]\phi$ as the abbreviations of $\neg\top$, $\neg(\neg\phi \wedge \neg\psi)$, $\neg\phi \vee \psi$, $\neg C_G\neg\phi$ and $\neg\langle\pi\rangle\neg\phi$ respectively. Moreover, we use the following additional abbreviations:

$$\begin{array}{|rl|}\hline K_j\phi := & C_{\{j\}}\phi \\ has_i M' := & \bigwedge_{m\in M'} has_i m \\ dhas_G M' := & \bigwedge_{m\in M'} \bigvee_{j\in G} has_j m \\ com(net) := & \bigwedge_{G\in net} com(G) \wedge \bigwedge_{G\notin net} \neg com(G) \\ \pi^n := & \underbrace{\pi;\pi;\ldots;\pi}_{n} \\ \Sigma\Pi' := & \bigcup_{\pi\in\Pi'} \pi \text{ where } \Pi' \subset \Pi \text{ is finite.} \\ \hline\end{array}$$

where $K_j\phi$ means that agent j knows ϕ; $dhas_G M'$ says the messages from M' are distributed knowledge among agents in G; $com(net)$ specifies the communication channels in the network.

By having both *has* and K operator in the language, we can make the distinction between knowing about a message and knowing about its content.

$K_i has_j m \wedge \neg has_i m$ and $K_i has_j m \wedge has_i m$ can express the *de dicto* and *de re* reading of knowing a message m respectively. For example, let m be the hiding place of Bin Laden, then $K_{CIA} has_{Al-Qaeda} m \wedge \neg has_{CIA} m$ expresses that CIA knows that Al-Qaeda knows the hiding place, which is, however, a secret to CIA.

2.2 Semantics

In order to interpret basic propositions $Prop_{I,A,M}$ we let the finer structure of the basic propositions correspond with a finer structure in the states, replacing the traditional valuation in Kripke structures used in DEL-approaches:

Definition 2.1. Let the state space $S = \mathcal{P}(\mathcal{P}(I)) \times (\mathcal{P}(M))^{|I|} \times (A)^* \times (\mathcal{P}(M))^{|I|} \times \Pi$. A state $s \in S$ for $\mathcal{L}_\iota^{I,M}$ is thus a tuple:

$$\langle net, M_0, \ldots, M_{|I|}, \bar{\alpha}, M'_0, \ldots, M'_{|I|}, \pi \rangle$$

Here $IS(s,i) = M'_i$ is i's current set of messages (information set), $AM(s) = \bar{\alpha}$ is the action history, $CC(s) = net$ is the available communication network and $Prot(s) = \pi$ is the protocol that agents have to follow from this state. We let $AM_k(s) = \alpha_k$ in $\bar{\alpha}$ and $l(s) = |AM(s)|$ be the *length* of s. Note that each state also contains information of the initial distribution of the messages: $M_0, \ldots, M_{|I|}$. From s we can recover the initial state of the system before any actions were executed:

$$Init(s) = \langle net, M_0, \ldots, M_{|I|}, \epsilon, M_0, \ldots, M_{|I|}, (\Sigma A)^* \rangle.$$

The action history in the initial state should be empty, thus $AM(Init(s)) = \epsilon$. We also assume that all the actions are allowed initially, thus $Prot(Init(s)) = (\Sigma A)^*$.

Intuitively, each state represents a deterministic temporal development of the system with its constraint for the future actions. Note that the past is linear ($AM(s)$ is a single sequence of actions), while the future can be branching ($Prot(s)$ may allow several possible sequences of actions).

$has_i m$, $com(G)$ and $past(\bar{\alpha})$ can be interpreted in a straightforward way at a state s according to $IS(s,i)$, $AM(s)$ and $CC(s)$ respectively. To give the semantics for $future(\bar{\alpha})$ at a state s, we need to check whether $\bar{\alpha}$ *complies* with the current protocol $Prot(s)$ and compute the remaining protocol after the execution of $\bar{\alpha}$ in order to know what the new protocol is. For this, we recall the *input derivate* $\pi \backslash \alpha$ of the regular expression $\pi \in \Pi$ and the output function $o : \Pi \to \{\delta, \varepsilon\}$ (cf. Brzozowski (1964), Conway (1971)):

$$\begin{array}{ll}
\varepsilon\backslash\alpha = \delta\backslash\alpha = \beta\backslash\alpha = \delta \quad (\alpha \neq \beta) & \alpha\backslash\alpha = \varepsilon \\
(\pi;\pi')\backslash\alpha = (\pi\backslash\alpha);\pi' + o(\pi);(\pi'\backslash\alpha) & (\pi \cup \pi')\backslash\alpha = \pi\backslash\alpha \cup \pi'\backslash\alpha \\
(\pi)^*\backslash\alpha = \pi\backslash\alpha;(\pi)^* & o(\pi;\pi) = o(\pi);o(\pi') \\
o(\pi^*) = \varepsilon & o(\varepsilon) = \varepsilon \\
o(\delta) = o(\alpha) = \delta & o(\pi \cup \pi') = o(\pi) \cup o(\pi')
\end{array}$$

Let $\pi\backslash(\alpha_0;\alpha_1;\ldots;\alpha_n) = (\pi\backslash\alpha_0)\backslash\alpha_1\ldots\backslash\alpha_n$. Together with the axioms of Kleene algebra we can derive syntactically the remaining protocol after executing a sequence of basic actions. For example: $(\alpha \cup (\beta;\gamma))^*\backslash\beta = (\alpha\backslash\beta \cup (\beta;\gamma)\backslash\beta);(\alpha \cup \beta;\gamma)^* = (\delta \cup (\varepsilon;\gamma));(\alpha \cup \beta;\gamma)^* = \gamma;(\alpha \cup (\beta;\gamma))^*$. Note that in general we do not have $\bar{\beta};(\pi\backslash\bar{\beta}) = \pi$.

Let $L(\pi)$ be the language of the regular expressions defined as follows:

$$L(\delta) = \emptyset \qquad L(\varepsilon) = \{\epsilon\} \qquad L(\alpha) = \{\alpha\}$$
$$L(\pi;\pi') = \{\bar{\alpha};\bar{\beta} \mid \bar{\alpha} \in L(\pi), \bar{\beta} \in L(\pi')\}$$
$$L(\pi \cup \pi') = L(\pi) \cup L(\pi')$$
$$L(\pi^*) = \{\bar{\alpha}_1;\ldots;\bar{\alpha}_n \mid \bar{\alpha}_1,\ldots,\bar{\alpha}_n \in L(\pi)\}$$

From Conway (1971), we have:

Proposition 1. $L(\pi\backslash\bar{\alpha}) = \{\bar{\beta} \mid \bar{\alpha};\bar{\beta} \in L(\pi)\}$.

Similar to Cohen and Dam (2007), Apt et al. (2009), we give the truth value of complex $\mathcal{L}_\iota^{I,M}$ formula on *single* states instead of *pointed Kripke models*. The interpretation of epistemic formulas depends on the relation $\sim_i^x$ to be defined later.

Let $s = \langle net, M_0, \ldots, M_{|I|}, \bar{\beta}, M'_0, \ldots, M'_{|I|}, \pi\rangle$, we define:

$s \vDash has_i(m)$	$\Leftrightarrow$	$m \in IS(s,i)$
$s \vDash com(G)$	$\Leftrightarrow$	$G \in CC(s)$
$s \vDash past(\bar{\alpha})$	$\Leftrightarrow$	$\bar{\alpha}$ is a suffix of $AM(s)$
$s \vDash future(\bar{\alpha})$	$\Leftrightarrow$	$Prot(s)\backslash\bar{\alpha} \neq \delta$
$s \vDash \neg\phi$	$\Leftrightarrow$	$s \nvDash \phi$
$s \vDash \phi \wedge \psi$	$\Leftrightarrow$	$s \vDash \phi$ and $s \vDash \psi$
$s \vDash C_G\phi$	$\Leftrightarrow$	for all v, if $s \sim_G^x t$ then $t \vDash \phi$
$s \vDash \langle\pi\rangle\phi$	$\Leftrightarrow$	$\exists s' : s[\![\pi]\!]s'$ and $s' \vDash \phi$

where $\sim_G^x$ is the reflexive transitive closure of $\bigcup_{i\in G} \sim_i^x$. The protocols π function as *state changers*:

$s[\![\varepsilon]\!]s'$	$\Leftrightarrow$	$s = s'$
$s[\![\delta]\!]s'$	$\Leftrightarrow$	never
$s[\![\alpha]\!]s'$	$\Leftrightarrow$	$s \vDash Pre(\iota(\alpha))$ and $s' = s\vert_{Pos(\iota(\alpha))}$
$s[\![\pi_1;\pi_2]\!]s'$	$\Leftrightarrow$	$s[\![\pi_1]\!] \circ [\![\pi_2]\!]s'$
$s[\![\pi_1 \cup \pi_2]\!]s'$	$\Leftrightarrow$	$s[\![\pi_1]\!] \cup [\![\pi_2]\!]s'$
$s[\![(\pi_1)^*]\!]s'$	$\Leftrightarrow$	$s[\![\pi_1]\!]^*s'$

where $\circ, \cup$ and * at right-hand side express the usual composition, union and reflexive transitive closure on relations respectively. Given $Pos(\iota(\alpha)) = \langle N_0, \ldots, N_{|I|}, \rho\rangle$, $s|_{Pos(\iota(\alpha))}$ is the result of executing action α at s defined as:

$$s|_{Pos(\iota(\alpha))} = \langle net, M_0, \ldots, M_{|I|}, \bar{\beta};\alpha, M'_0 \cup N_0, \ldots, M'_{|I|} \cup N_{|I|}, f(\rho)\rangle$$

where $f(\rho) = \begin{cases} \pi\backslash\alpha \text{ if } \rho = \# \\ \pi' \quad \text{if } \rho = \pi' \end{cases}$.

Now we define $\sim_i^x$, the epistemic relation of an agent i between states. A state s is said to be *consistent* if $Init(s)[\![AM(s)]\!]s$. It is easy to see that for any s, $Init(s)$ is always consistent[6].

We define that $t \sim_i^x t'$ iff the following conditions are met:

[6]Note that we can actually omit the current information sets $IS(s,i)$ in the definition of a state, and compute it by applying the actions in $AM(s)$, thus only generate consistent states. We keep the current information sets there to simplify notations and make it more efficient to evaluate basic propositions according to the semantics.

consistency t and t' are consistent.

local initialization $IS(Init(t), i) = IS(Init(t'), i)$

local history $AM(t)|_i^x = AM(t')|_i^x$, where x is the *type of observational power* of agents.

Many definitions of $AM(t)|_i^x$ are possible, giving the agents different observational powers. Several reasonable definitions are:

1. $AM(t)|_i^{set} = \{\alpha \text{ appearing in } AM(t) \mid i \in Obs(\iota(\alpha))\}$ as in Apt et al. (2009).
2. $AM(t)|_i^{1st}$ is the subsequence of $AM(t)$ which only keeps the first occurrence of each $\alpha \in AM(t)|_i^{set}$ as in Baskar et al. (2007).
3. $AM(t)|_i^{asyn}$ is the subsequence of $AM(t)$ which only keeps all the occurrences of each $\alpha \in AM(t)|_i^{set}$, as in *asynchronous* systems (cf., e.g., Shilov and Garanina (2002)).
4. $AM(t)|_i^{\tau}$ is the sequence obtained by replacing each occurrence of $\alpha \notin AM(t)|_i^{set}$ in $AM(t)$ by τ, as in *synchronous* systems with prefect recall (cf., e.g., van der Meyden and Shilov (1999)).

It is clear from the above definition that $\sim_i^x$ is an equivalence relation and the following holds:

Proposition 2. $\sim_i^{\tau} \subseteq \sim_i^{asyn} \subseteq \sim_i^{1st} \subseteq \sim_i^{set}$.

We then call the semantics defined by $\sim_i^x$ the *x-semantics*, and denote the corresponding satisfaction relation as $\vDash^x$.

Recall that we require that the agents can always observe the actions that change his information set. Then we have:

Proposition 3. For any consistent state t: $t \sim_i^x t'$ implies $IS(t, i) = IS(t', i)$ where $x \in Sem = \{set, asyn, 1st, \tau\}$.

Proof. By Proposition 2, $t \sim_i^x t'$ implies $t \sim_i^{set} t'$ for all $x \in Sem$. Therefore we only need to prove the claim for $x = set$. Suppose $t \sim_i^{set} t'$ then by the definition of $\sim_i^{set}$, $IS(Init(t), i) = IS(Init(t'), i)$ and $AM(t)|_i^{set} = AM(t')|_i^{set}$. So at t and t' agent i initially had the same messages and has observed the same actions. Since agents can always observe the actions that change his information set then we know the same message passing actions relavent to i have happened for t and t'. Since the actions can only add messages to the information set and never delete messages from them, it doesn't matter how often or in which order those actions have been executed. Therefore the information sets of agent i in t and t' are identical. □

2.3 Communicative Actions

In this section, we will define some useful basic actions with their internal structures. To simplify the presentation, we abuse the notation of action names to stand for their internal structures as well, when the context is clear. Thus we let $Obs(\beta) = Obs(\iota(\beta))$ and similar for $Pre(\beta)$ and $Pos(\beta)$. Recall that the internal structure of an action β is a tuple $\langle F, \phi, N_0, \ldots, N_{|I|}, \rho\rangle$ such that $N_j = \emptyset$ for $j \notin Obs(\beta)$. We now list some basic actions with their internal structures defined in the table below:

β (communication by the agents):	*Obs* :	*Pre* : common part is: $com(Obs(\beta)) \wedge future(\beta) \wedge ...$	*Pos* $(j \in Obs(\beta))$:
$send^i_G(M')$	$G \cup \{i\}$	$has_i M'$	$N_j = M', \rho = \#$
$share_G(M')$	G	$dhas_G M'$	$N_j = M', \rho = \#$
$sendall^i_G(M')$	$G \cup \{i\}$	$has_i M' \wedge \bigwedge_{m \notin M'} \neg has_i m$	$N_j = M', \rho = \#$
$shareall_G(M')$	G	$dhas_G M' \wedge \bigwedge_{m \notin M'} \neg dhas_i m$	$N_j = M', \rho = \#$
$inform^i_G(\phi)$	$G \cup \{i\}$	$K_i \phi$	$N_j = \emptyset, \rho = \#$
β (external actions):	*Obs* :	*Pre* :	*Pos* :
$exinfo(\phi)$	I	ϕ	$\rho = \#$
$exprot(\pi')$	I	$\top$	$\rho = \pi'$

The first group of actions are communicative actions that are done by the agents. These actions must abide by the communication channels and the protocol, which is enforced by having $com(Obs(\beta)) \wedge future(\beta)$ in the precondition. $send^i_G(M')$ is the action that i sends the set of messages M' to the group G. Apart from respecting the channel and the protocol, the precondition $has_i M'$ enforces that agent i should possess any messages he wants to send. The postcondition of $send^i_G(M')$ expresses that the messages in M' get added to the message sets of the agents in G. $share_G(M')$ shares the messages from M' within the group G. A precondition is that the messages from M' are already distributed knowledge in the group. $sendall^i_G(M')$ differs from $send^i_G(M')$ in the extra precondition that M' should contain *all* the messages that i has. Similarly for $shareall_G(M')$. $inform^i_G(\phi)$ is the group announcement of an arbitrary formula ϕ within $G \cup \{i\}$. A precondition is that i should know ϕ is true before he can announce it.

The second group of actions are public announcements that do not respect the channels or the protocol. They model the external information that is given to the agents. $exinfo(\phi)$ models the public announcement of a formula ϕ. The only precondition of this announcement is that ϕ should hold. The postcondition is empty. Knowledge of ϕ among the agents is created by the fact that all agents can observe the action. Since all agents know the execution of this action would only be possible if ϕ would hold, all agents know that ϕ holds at the moment it is announced. $exprot(\pi')$ announces the protocol π' that the agents are supposed to follow in the future. Its postcondition changes the protocol to π' and knowledge of the protocol is created because all agents observe the announcement.

We can define more complex actions based on the above basic actions, as we will demonstrate in Section 4.

3 Comparison with IS and DEL

The results in this section relate our logic to IS and DEL approaches. Theorem 1 shows that by the semantics of $\mathcal{L}^{I,M}_{\iota}$, an interpreted system is generated implicitly from a single state. Together with Theorem 1, Proposition 4 demonstrates that compared to DEL, our approach is powerful and concise in modelling actions. Let us compare our approach to IS first. In the following we only consider consistent states.

Let the history of s be a sequence: $hist(s) = s_0 s_1 \ldots s_{l(s)}$ where $s_0 = Init(s)$, $s_{l(s)} = s$ and $s_k [\![\alpha_k]\!] s_{k+1}$ for any k such that $\alpha_k = AM_k(s)$. Clearly then $s_0 s_1 \ldots s_k =$

$hist(s_k)$ for any $k \leq l(s)$. Let $ExpT^x$ be the Interpreted System with action labels with respect to x-semantics $\{H, \rightarrow_\alpha, \{R_i \mid i \in I\}, V\}$, where:

- $H = \{hist(s) \mid s \text{ is consistent.}\}$
- $\langle s_0 \dots s_n \rangle \rightarrow_\alpha \langle s_0 \dots s_n s_{n+1} \rangle \Leftrightarrow s_n [\![\alpha]\!] s_{n+1}$.
- $\langle s_0 \dots s_n \rangle R_i \langle s'_0 \dots s'_m \rangle$ iff $s_n \sim_i^x s'_m$.
- $V(\langle s_0 \dots s_n \rangle)(p) = \top \Leftrightarrow s_n \vDash^x p$ where $p \in Prop_{I,A,M}$.

The language of $\mathcal{L}_\iota^{I,M}$ can be seen as a fragment of *Propositional Dynamic Logic* (PDL): $\mathcal{L}_{pdl}^I$ with basic action set $A \cup I$. Here C_G can be seen as $(\Sigma G)^*$. Let $\Vdash_{PDL}$ denote the usual semantics of $\mathcal{L}_{pdl}^I$, then it is not hard to see:

Theorem 1. For any formula $\phi \in \mathcal{L}_\iota^{I,M}$ and for each consistent $\mathcal{L}_\iota^{I,M}$-state s:

$$s \vDash^x \phi \Leftrightarrow ExpT^x, hist(s) \Vdash_{PDL} \phi.$$

This result shows that if we abstract away the inner structure of basic propositions and actions, then our logic can be seen as a PDL language interpreted on ISs that are generated in a particular way w.r.t some constraints. Note that this result does not imply the decidability of $\mathcal{L}_\iota^{I,M}$ since although PDL is decidable on general Kripke structures, we do not know yet whether it is decidable on the restricted class of generated models $ExpT^x$.

Now consider the DEL language $\mathcal{L}_{del}^I$:

$$\phi \quad ::= \quad \top \mid p \mid \neg\phi \mid \phi_1 \wedge \phi_2 \mid \langle \mathbb{A}, e \rangle \phi \mid C_G \phi$$

where p is in a set of basic propositions $Prop$, $G \subseteq I$ and $\mathbb{A}$ is an *action model* with e as its designated action. Action models are tuples of the form $(E, \{\asymp_i\}_{i \in I}, Pre, Pos)$ where $\asymp_i$ models agents i's observational power on events in E (e.g. $e_1 \asymp_i e_2$ means i is not sure which one of e_1 and e_2 happened); the precondition function $Pre : E \rightarrow \mathcal{L}_{del}^I$ describes when an event can happen and the postcondition $Pos : E \rightarrow (Prop \rightarrow \mathcal{L}_{del}^I)$ models the factual changes caused by the event by changing the truth values of basic propositions p to $Pos(e)(p)$ van Benthem et al. (2006). The semantics for epistemic formulas is as usual and

$$\mathbb{M}, s \Vdash_{DEL} \langle \mathbb{A}, e \rangle \phi \Leftrightarrow \mathbb{M} \otimes \mathbb{A}, (s, e) \vDash \phi$$

Where, given a static Kripke model $\mathbb{M} = (W, \{R_i\}_{i \in I}, V)$ and an action model $\mathbb{A} = (E, \{\asymp_i\}_{i \in I}, Pre, Pos)$, the updated model is $\mathbb{M} \otimes \mathbb{A} = (W', \{R'_i\}_{i \in I}, V')$ with:

$$\begin{array}{rcl} W' &=& \{\langle w, e \rangle \mid \mathbb{M}, w \Vdash Pre(e)\} \\ R'_i &=& \{(\langle w, e \rangle, \langle v, e' \rangle) \mid wR_iv \text{ and } e \asymp_i e'\} \\ V'(\langle w, e \rangle)(p) &=& \top \Leftrightarrow \mathbb{M}, w \Vdash Pos(e)(p) \end{array}$$

To facilitate a comparison, let us consider $\mathcal{L}_\iota^{I,M,-*}$, the star-free fragment of $\mathcal{L}_\iota^{I,M}$. Let $ExpK^x(s)$ be the Kripke model $\{W, \{R_i \mid i \in I\}, V\}$ obtained by the *expansion* of the state s according to $x-$semantics, with:

- $W = \{s' \mid s \sim_I^x s'\}$ where $\sim_I^x$ is the reflexive transitive closure of $\{\sim_i^x \mid i \in I\}$.
- $R_i = \sim_i^x|_{W \times W}$.
- $V(s)(p) = \top \Leftrightarrow s \vDash^x p$ where $p \in Prop_{I,A,M}$.

Note that although I, A, M are assumed to be finite, W in $ExpK^x(s)$ can still be infinite due to the fact that we record the past explicitly in the states and there may be infinitely many possible histories.

Based on $ExpK^x(s)$ it seems plausible to obtain a similar correspondence result as Theorem 1 for $\mathcal{L}_\iota^{I,M,-*}$and $\mathcal{L}_{del}^I$, since the basic actions in $\mathcal{L}_\iota^{I,M,-*}$ look like special cases of pointed action models in DEL. However, the result does not hold in general. To see this, we first recall a fact from van Benthem et al. (2009): If we see $\langle \mathbb{A}, e\rangle$ as a basic action modality when considering *PDL* semantics, then for any formula $\phi \in \mathcal{L}_{del}^I$:

$$\mathbb{M}, s \Vdash_{DEL} \phi \Leftrightarrow Forest(\mathbb{M}, \mathcal{A}), (s) \Vdash_{PDL} \phi \quad (\star)$$

where $\mathcal{A}$ is the set of action models and $Forest(\mathbb{M}, \mathcal{A})$ is the IS generated by executing all possible sequences of action models in $\mathcal{A}$ on $\mathbb{M}, s$[7]. We now show that the effects of actions in $\mathcal{L}_\iota^{I,M}$ cannot be simulated by action models.

Proposition 4. There is no translation of action models $T : A \to \mathcal{A}$ such that for all consistent $\mathcal{L}_\iota^{I,M}$-states s:

$$T(ExpT^x), hist(s) \leftrightarrows Forest(ExpK^x(s), \mathcal{A}), (s)$$

where $x \in \{set, 1st, asyn\}$, $T(ExpT^x)$ is the IS obtained from $ExpT^x$ by replacing each label of $\alpha \in A$ by $T(\alpha) \in \mathcal{A}$ and $\leftrightarrows$ is the bisimulation for transitions labled by $I \cup \mathcal{A}$.

Proof. van Benthem et al. (2009) shows that $Forest(ExpK^x(s), \mathcal{A})$ must satisfy the property of *Perfect Recall* meaning that if the agents can not distinguish two sequences of action $\bar{\alpha}; \alpha$ and $\bar{\beta}; \beta$ then they can not distinguish $\bar{\alpha}$ and $\bar{\beta}$. However, $ExpT^x$ clearly does not satisfy this property for $x \in \{set, 1st, asyn\}$ in general. For example, $send_j^i(M); \gamma \sim_j^x \gamma; send_j^i(M)$ where $x \in \{set, 1st, asyn\}$ and γ is some action j cannot observe, but $send_j^i(M) \nsim_j^x \gamma$. □

If we consider τ−semantics, then a correspondence result can be obtained. Let $T_{DEL} : \mathcal{L}_\iota^{I,M,-*} \to \mathcal{L}_{del}^I$ be defined as follows:

$$\begin{array}{rcl} T_{DEL}(\top) & = & \top \\ T_{DEL}(p) & = & p \\ T_{DEL}(\neg\phi) & = & \neg T_{DEL}(\phi) \\ T_{DEL}(\phi_1 \wedge \phi_2) & = & T_{DEL}(\phi_1) \wedge T_{DEL}(\phi_2) \\ T_{DEL}([\alpha]\phi) & = & [ExpA_\iota^\tau(\alpha)]T_{DEL}(\phi) \\ T_{DEL}([\pi_1 \cup \pi_2]\phi) & = & T_{DEL}([\pi_1]\phi) \wedge T_{DEL}([\pi_2]\phi) \\ T_{DEL}([\pi_1; \pi_2]\phi) & = & T_{DEL}([\pi_1][\pi_2]\phi) \end{array}$$

where $ExpA_\iota^\tau(\alpha)$ is the pointed action model $\{E, \{R_i \mid i \in I\}, V, e_\alpha\}$ obtained by the *saturation* of the action α according to τ−semantics:

- $E = \{e_\beta \mid \beta \in A\}$
- $e_\beta R_i e_{\beta'} \Leftrightarrow \iota(\beta) = \iota(\beta')$ or $i \notin Obs(\beta) \cup Obs(\beta')$.
- $Pre(e_\beta) = T_{DEL}(Pre(\beta))$.

[7]Due to the limit of space, readers are referred to van Benthem et al. (2009) for details.

- If $Pos(\beta) = \langle M_0, \ldots, M_I, x\rangle$ then:

$$Pos(e_\beta)(has_i m) = \begin{cases} \top & \text{if } m \in M_i \\ has_i m & \text{otherwise} \end{cases}$$

$$Pos(e_\beta)(com(G)) = com(G)$$

$$Pos(e_\beta)(past(\bar{\gamma}; \gamma)) = \begin{cases} past(\bar{\gamma}) & \text{if } \gamma = \beta \\ \bot & \text{otherwise} \end{cases}$$

$$Pos(e_\beta)(future(\bar{\gamma})) = \begin{cases} future(\beta; \bar{\gamma}) & \text{if } \rho \text{ in } Pos(\beta) \text{ is } \sharp \\ \top & \text{if } \rho \text{ in } Pos(\beta) \text{ is } \pi \text{ and } \pi\backslash\bar{\gamma} \neq \delta \\ \bot & \text{if } \rho \text{ in } Pos(\beta) \text{ is } \pi \text{ and } \pi\backslash\bar{\gamma} = \delta \end{cases}$$

Based on the above translation, the star-free fragment of $\mathcal{L}_\iota^{I,M}$ can be seen as a version of *DEL* on generated models:

Theorem 2. For any $\phi \in \mathcal{L}_\iota^{I,M,-*}$ and for any consistent $\mathcal{L}_\iota^{I,M}$-state s:

$$s \vDash^\tau \phi \Leftrightarrow ExpK^\tau(s), s \Vdash_{DEL} T_{DEL}(\phi).$$

However, without the internal structure of basic propositions and protocol constraints in action models, the translation to standard *DEL* relies on infinitely many atomic propositions and action models which change infinitely many atomic propositions.

4 Applications

4.1 Common Knowledge

Our framework gives an interesting perspective on common knowledge. We first focus on asynchronous semantics. It may not be surprising that we cannot reach common knowledge without public communication. We might think that achieving common knowledge becomes easier if we can publicly agree on a common protocol before the communication is limited to non-public communication. However, in the case of asynchronous semantics we still can not reach common knowledge, even if we can publicly agree on a protocol. Recall that we say an action α *respects the communication channel* if $Pre(\alpha) \vDash com(Obs(\alpha))$.

Theorem 3. For any state s with $I \notin CC(s)$, any protocol π containing only communications that respect the communication channels, any $\varphi \in \mathcal{L}_\iota^{I,M}$ and any sequence of actions $\bar{\alpha}$:

$$s \vDash^{asyn} \langle exprot(\pi)\rangle(\neg C_I\varphi \rightarrow \neg\langle\bar{\alpha}\rangle C_I\varphi)$$

Proof. Let $s[\![exprot(\pi)]\!]t$ and suppose $t \vDash^{asyn} \neg C_I\phi$. Towards a contradiction, let $\bar{\alpha}$ be the minimal sequence of actions such that $t \vDash^{asyn} \langle\bar{\alpha}\rangle\phi$. Let $\bar{\alpha} = \bar{\beta}; \alpha$, $t[\![\bar{\beta}]\!]u$ and $u[\![\alpha]\!]v$. Since $I \notin CC(s)$ and α respects the communication channel, $obs(\alpha) \neq I$ so there exists $j \notin Obs(\alpha)$. Then $AM(u)|_j^{asyn} = AM(v)|_j^{asyn}$ so $u \sim_j^{asyn} v$. Since $\bar{\alpha}$ was minimal, $u \nvDash^{asyn} C_I\varphi$. But then $u \vDash^{asyn} \neg K_j C_I\varphi$ so $v \vDash^{asyn} \neg K_j C_I\varphi$. So $v \nvDash^{asyn} C_I\varphi$. □

Essentially, even if the agents agree on a protocol beforehand, the agents that cannot observe the final action of the protocol will never know whether this final action has been executed and thus common knowledge is never established. This is because in the asynchronous semantics, there is no sense of time. If we could add some kind of clock and the agents would agree to do an action on every "tick", the agents would be able to establish common knowledge. This is exactly what we try to achieve with our τ-semantics. Here every agent observes a "tick" the moment some action is executed. This way, they can agree on a protocol *and* know when it is finished. We will show examples of how this can result in common knowledge in the next section on the telephone call scenario.

Here we will first investigate what happens in τ-semantics if we *cannot* publicly agree on a protocol beforehand. We will show that in this case we cannot reach common knowledge of basic formulas. We start out with a lemma stating that actions preserve the agent's relations.

Lemma 1. For any two states s and t and any action α, if $s \sim_i^\tau t$ and we have s', t' such that $s[\![\alpha]\!]s'$ and $t[\![\alpha]\!]t'$ then $s' \sim_i^\tau t'$.

Proof. Suppose $s \sim_i^\tau t$. Then $AM(s)|_i^\tau = AM(t)|_i^\tau$. Suppose $i \in Obs(\alpha)$. Then $AM(s')|_i^\tau = (AM(s)|_i^\tau;\alpha) = (AM(t)|_i^\tau;\alpha) = AM(t')|_i^\tau$. Suppose $i \notin Obs(\alpha)$. Then $AM(s')|_i^\tau = (AM(s)|_i^\tau;\tau) = (AM(t)|_i^\tau;\tau) = AM(t')|_i^\tau$. So $s' \sim_i^\tau t'$. □

This result may seem counter-intuitive, since for example a public announcement action may give the agents new information and thus destroy their epistemic relations. However, in our framework we model the new knowledge introduced by communicative actions by the fact that these actions would not be possible in states that do not satisfy the precondition of the action. In this lemma we assume that there are s', t' such that $s[\![\alpha]\!]s'$ and $t[\![\alpha]\!]t'$. This means that s and t both satisfy the preconditions of α, so essentially no knowledge that distinguishes s and t is introduced by α.

Now we define a fragment $\mathcal{L}_{bool}$ of our logic as follows:

$$\phi ::= has_i m \mid com(G) \mid \neg\phi \mid \phi_1 \wedge \phi_2$$

It is trivial to show that any action that does not change the agent's message sets or the protocol does not change the truth value of these basic formulas:

Lemma 2. Let α be an action that does not change the agent's message sets or the protocol. For any $\phi \in \mathcal{L}_{bool}$ and any state s: $s \vDash \phi \leftrightarrow \langle\alpha\rangle\phi$.

Combining the properties of the actions from the previous lemma, we call an action α_d^G to be a *dummy action* for a group of agents G if its internal structure has the precondition $com(G) \wedge future(\alpha_d^G)$, it does not change the message sets of the agents or the protocol and $Obs(\alpha_d^G) = G$. An example of dummy action is $inform_G^i(\top)$. We could see it as "talking about irrelevant things".

Theorem 4. Let A be a set of basic actions respecting the communication channels such that for any agent i there is a dummy action α_d^G such that $i \notin G$. Let s be a state such that $I \notin CC(s)$ and it is common knowledge at s that the protocol is $\pi = (\Sigma A)^*$ (any action in A is allowed). Then for any $\phi \in \mathcal{L}_{bool}$ and any sequence of actions $\bar{\alpha}$,

$$s \vDash^\tau \neg C_I\phi \to \neg\langle\bar{\alpha}\rangle C_I\phi$$

Proof. Suppose towards a contradiction that $s \models \neg C_I\phi$ and there is a minimal sequence $\bar{\alpha}$ such that $s \models^\tau \langle\bar{\alpha}\rangle C_I\phi$. Let $\bar{\alpha} = \bar{\beta}; \alpha$ and let $i \notin Obs(\alpha)$. Such i always exists since $I \notin CC(s)$. Let α_d^G be a dummy action such that $i \notin G$. Let $s[\![\bar{\beta}]\!]u$. Since $\bar{\alpha}$ is minimal, $u \models^\tau \neg C_I\phi$, so there is a $\sim_I$-path from u to a world t such that $t \not\models^\tau \phi$. Since it is common knowledge that any action in A is possible, then we can execute α_d^G at any world on the path from u to t. By lemma 1 α_d^G preserves the relations between states so there are states u', t' such that $u[\![\alpha_d^G]\!]u'$, $t[\![\alpha_d^G]\!]t'$ and $u' \sim_I t'$. Also, since $t \not\models^\tau \phi$ and by lemma 2, $t' \not\models^\tau \phi$. So $u' \not\models^\tau C_I\phi$. This means that if we would execute α_d^G in state u, then $C_I\phi$ would not hold.

Let $u[\![\alpha_d^G]\!]u'$ and $u[\![\alpha]\!]v$. Because $i \notin G$, i cannot see the difference between executing α_d^G and α: $AM(u')|_i^\tau = (AM(u)|_i^\tau; \tau) = AM(v)|_i^\tau$ so $u' \sim_i^\tau v$. But we just saw that $u' \not\models^\tau C_I\phi$, so then $v \not\models^\tau C_I\phi$. But this contradicts our assumption that $\bar{\beta}; \alpha$ induced common knowledge of ϕ. □

4.2 Telephone Calls

Before going to the specific scenario of the telephone calls, we propose the following general modeling method:

1. Select a finite set of suitable actions A with internal structures to model the communications in the scenario.
2. Design a single state as the *real world* to model the initial setting, i.e., $s = \langle net, \bar{M}_i, \epsilon, \bar{M}_i, (\Sigma A)^* \rangle$ where net models the communication network and $\bar{M}_i$ models *"who has what information"*.
3. Translate the informal assumptions of the scenario into formulas ϕ and protocols π in $\mathcal{L}_\iota^{I,M}$.
4. Use $exinfo(\phi)$ and $exprot(\pi)$ to make the assumptions and the protocol common knowledge.

We will demonstrate how we use this method to model the telephone call scenario. Let us first recall the scenario: in a group of people, each person has one secret. They can make private telephone calls among themselves in order to communicate these secrets. The original puzzle concerns the minimal number of telephone calls needed to ensure everyone gets to know all secrets. We start out by selecting a set of suitable actions A. We define:

$$call_j^i(M') := \bigcup_{M'' \subseteq M'} shareall_{\{i,j\}}(M'')$$
$$mail_j^i(M') := \bigcup_{M'' \subseteq M'} sendall_{\{j\}}^i(M'')$$

Here $call_j^i(M')$ is the call between agent i and j where they share all messages out of M' they possess[8]. Later on we will also be interested in what happens if the agents can only leave voicemail messages instead of making two-way calls. For this purpose we use $mail_j^i(M')$, where agent i sends all messages out of M' he possesses to agent j. The third kind of communication we are interested in will be when the agents can call each other and communicate any formula from the language instead of only their messages. This is modeled by $inform_j^i(\phi)$. Let

[8]Here M' encodes the *relevant context* e.g. messages that are *"about work"*.

$M_I = \{m_0, ..., m_{|I|}\}$ be the set of all secrets. For suitable finite sets of formulas Φ and protocols Π[9], we define

$$A = \bigcup_{\phi\in\Phi} exinfo(\phi) \cup \bigcup_{\pi\in\Pi} exprot(\pi) \cup \bigcup_{i,j\in I} call^i_j(M_I) \cup \bigcup_{i,j\in I} mail^i_j(M_I) \cup \bigcup_{i,j\in I,\phi\in\Phi} inform^i_j(\phi),$$

where we include $exinfo(\phi)$ and $exprot(\pi)$ because we need them to make the assumptions and the protocol of the scenario common knowledge.

Next, we define the communication network and the agent's message sets. Each agent has one secret so we define $M_i = \{m_i\}$. The agents can only communicate in pairs, so the communication network is $net^{tel}_I = \{\{i, j\} \mid i \neq j \in I\}$. Then the initial state is:

$$s^{tel}_I = \langle net^{tel}_I, \{m_0\} \dots \{m_{|I|}\}, \epsilon, \{m_0\} \dots \{m_{|I|}\}, (\Sigma A)^* \rangle$$

We are interested in situations with different communicative powers for the agents, which can be characterized by protocols that restrict the possible basic actions. We define $\pi_{call} := (\bigcup_{i,j\in I} call^i_j(M_I))^*$, $\pi_{mail} := (\bigcup_{i,j\in I} mail^i_j(M_I))^*$ as the protocols where the agents can only make telephone calls or send voicemails, respectively. We define $\pi_{\text{call, inform}} := (\bigcup_{i,j\in I} call^i_j(M_I) \cup \bigcup_{i,j\in I} mail^i_j(M_I))^*$.

As for the informal assumptions of the scenario, we assume it is common knowledge that every agent has one secret, and we assume the communication network is common knowledge. We use the following abbreviations:

$$\begin{aligned} \texttt{OneSecEach}_I &:= \textstyle\bigwedge_{i\in I}(has_i m_i \wedge \bigwedge_{j\neq i} \neg has_j m_i) \\ \texttt{TP} &:= exinfo(com(net^{tel}_I) \wedge \texttt{OneSecEach}_I) \\ \texttt{TP}_x &:= \texttt{TP}; exprot(\pi_x) \\ \texttt{HasAll}_I &:= \textstyle\bigwedge_{i\in I} has_i M_I \end{aligned}$$

$\texttt{OneSecEach}_I$ states that every agent has one secret known only to him. $\texttt{TP}_x$ is the action of announcing the assumptions of the scenario and protocol π_x. $\texttt{HasAll}_I$ expresses that every agent knows every secret, which is the goal we want to reach.

In order to reason about the number of calls the agents need to make to reach their goal, we use the following abbreviations:

$$\begin{aligned} \langle\rangle^{\leq n}\phi &:= \langle \textstyle\bigcup_{k\leq n} (\Sigma A')^k \rangle \phi \\ \langle\rangle^{min(n)}\phi &:= \langle\rangle^{\leq n}\phi \wedge \neg\langle\rangle^{\leq n-1}\phi \end{aligned}$$

where A' is the set of all actions in A that respect the channels, i.e., excluding $exprot$, $exinfo$ and other external actions.
$\langle\rangle^{\leq n}\phi$ expresses that we can reach a state where ϕ holds by sequentially executing at most n actions from A without external information or any changes in protocol. $\langle\rangle^{min(n)}\phi$ expresses that n is the minimal such number. The reason we exclude these actions is because we essentially want to know whether we can reach ϕ with the current protocol. The external actions do not abide by the protocol, so we should not consider them[10].

Then the following result states that we need exactly $2|I| - 4$ calls to make sure every agent knows all secrets:

[9] For example, the sets of formulas/protocols up to the length of certain large number.

[10] Note that $\langle\rangle^{\leq n}$ serves as a generalization of the *arbitrary announcement* that is added to DEL in Ågotnes et al. (2009).

Proposition 5. For any $x \in Sem$:

$$s_I^{tel} \vDash^x \langle \mathtt{TP}_{call} \rangle \langle \rangle^{min(2|I|-4)} \mathtt{HasAll}_I$$

A proof of this proposition is given in Hurkens (2000). The protocol given there is the following: pick a group of four agents 1 ... 4 and let 4 be their informant. Let all other agents call agent 4, then let the four agents communicate all their secrets within their group and let all other agents call agent 4 again. In our framework we can express this as follows: $call_5^4(M_I); ...; call_{|I|}^4(M_I); call_2^1(M_I); call_4^3(M_I); call_3^1(M_I); call_4^2(M_I); call_5^4(M_I); ...; call_{|I|}^4(M_I)$

Another interesting question arises when the agents cannot make direct telephone calls, but they can only leave voicemail messages. This means that any agent can tell the secrets he knows to another agent, but he cannot in the same call also learn the secrets the other agent knows. How many voicemail messages would we need in this case?

Intuitively we can use $mail_j^i(M_I)$; $mail_i^j(M_I)$ to mimic each $call_j^i(M_I)$, thus we have:

$$s_I^{tel} \vDash^x \langle \mathtt{TP}_{mail} \rangle \langle \rangle^{\leq 4|I|-8} \mathtt{HasAll}_I.$$

However, we can do much better:

Proposition 6. For any $x \in Sem$:

$$s_I^{tel} \vDash^x \langle \mathtt{TP}_{mail} \rangle \langle \rangle^{min(2|I|-2)} \mathtt{HasAll}_I$$

Proof. Consider the following protocol: $mail_2^1(M_I)$; $mail_3^2(M_I)$; ...; $mail_{|I|}^{|I|-1}(M_I)$; $mail_1^{|I|}(M_I)$; $mail_2^{|I|}(M_I)$; ...; $mail_{|I|-1}^{|I|}(M_I)$. Clearly, this results in all agents knowing all secrets. The length of this protocol is $2|I| - 2$. We claim this protocol is minimal. To see why this claim holds, first observe that there has to be one agent who is the first to learn all secrets. For this agent to exist all other agents will first have to make at least one call to reveal their secret to someone else. This is already $|I| - 1$ calls. The moment that agent learns all secrets, since he is the first, all other agents do not know all secrets. So each of them has to receive at least one more call in order to learn all secrets. This also takes $|I| - 1$ calls which brings the total number of calls to $2|I| - 2$. □

As we saw above, it is possible to make sure all agents know all secrets. However, in these results the secrets are not *common knowledge* yet, since the agents do not know that everyone knows all secrets. We will investigate whether we can establish common knowledge of $\mathtt{HasAll}_I$. If there are only three agents, this is possible by making telephone calls:

Proposition 7. If $|I| \leq 3$ then for some $n \in \mathbb{N}$:

$$s_I \vDash^\tau \langle \mathtt{TP}_{call} \rangle \langle \rangle^{\leq n} C_I \mathtt{HasAll}_I$$

Proof. For $|I| < 3$ the proof is trivial. Suppose $|I| = 3$, say $I = \{1, 2, 3\}$. A protocol that results in the desired property is $call_2^1(M_I); call_3^2(M_I); call_1^2(M_I)$. After execution of this protocol all agents know all secrets, and agent 2 knows this. Also, since agent 1 learned the secret of agent 3 from agent 2, he knows that agent 2 and 3 must have communicated after the last time he spoke to agent 2, so agent 3 must know the secret of agent 1. Regarding agent 3, he knows agent 2 has all secrets the moment he communicated with agent 2, and he observed a τ when agent 2 called agent 1 after that. Since there are only three agents agent 3 can deduce that agent 1 and 2 communicated so he knows agent 1 knows all secrets. Since all agents can reason about each others knowledge it is common knowledge that all agents have all secrets. □

We do not extend this result for the case with more than three agents. If there are more than three agents, agents that are not participating in the phone call will never know which of the other agents are calling, which makes it much harder to establish common knowledge. In the above results the communicative power of the agents is still fairly limited. They can only communicate their messages and they cannot talk about higher-order knowledge. An interesting question is whether the agents will be able to reach common knowledge if they can tell each other arbitrary formulas of the language, using the *inform* action. Interestingly, this reduces the possibilities to reach common knowledge since the dummy action $inform^i_G(\top)$ is allowed. The agents have no clue whether any information is transferred when they observe a τ action so they can never reach common knowledge, not even in the case that $|I| = 3$. This directly follows from Theorem 4.

Proposition 8. For any $n \in \mathbb{N}$, if $|I| > 2$ then:

$$s_I \not\models^\tau \langle \mathtt{TP}_{call,inform} \rangle \langle\rangle^{\leq n} C_I \mathtt{HasAll}_I$$

However, we can approach common knowledge arbitrarily close. For any finite sequence of agents $w = ij...k$ define:

$$K_w\varphi := K_iK_j...K_k\varphi$$

Proposition 9. For any finite sequence w of agents from I, there exists some $n \in \mathbb{N}$ such that:

$$s_I \models^\tau \langle \mathtt{TP}_{call,inform} \rangle \langle\rangle^{\leq n} K_w \mathtt{HasAll}_I$$

Proof. We will give a protocol that results in the desired property. First we execute the protocol given in the proof of Proposition 6. Note that after executing this protocol, agent $|I|$ knows that everyone knows all secrets. Let $w = a_1...a_n$. We execute $inform^{|I|}_{a_n}(\mathtt{HasAll}_I); inform^{|I|}_{a_{n-1}}(K_{a_n}\mathtt{HasAll}_I); ...; inform^{|I|}_{a_1}(K_{a_2}...K_{a_n}\mathtt{HasAll}_I)$ and clearly, after these actions the desired property will hold. □

Now imagine a situation where the agents are allowed to publicly announce beforehand a specific protocol they are going to follow which is more complex than just the set of actions they can choose from. Then, in our τ-semantics, it is possible to reach common knowledge:

Proposition 10. There is a protocol π of *call* actions such that

$$s_I \models^\tau \langle \mathtt{TP}; exprot(\pi) \rangle \langle\rangle^{\leq n} C_I \mathtt{HasAll}_I$$

Proof. Let π be the protocol given in the proof of proposition 5. Since each agent observes a τ at every communicative action, they can all count the number of communicative actions that have been executed and they all know when the protocol has been executed. So at that moment, it will be common knowledge that everyone has all secrets. □

This shows the use of the ability to communicate about the future protocol and not only about the past and present. There are many more situations where announcing the protocol is very important, for example in the puzzle of 100 prisoners and a light bulb Dehaye et al. (2003) or many situations in distributed computing.

However, when we use *asyn*-semantics, the agents cannot count the number of communicative actions happening, so they never know when the protocol has been executed. Because of this they can never reach common knowledge:

Proposition 11. There is no protocol π of *call* and *inform* actions such that

$$s_I \vDash^{asyn} \langle \mathtt{TP}; exprot(\pi)\rangle\langle\rangle^{\leq n} C_I \mathtt{HasAll}_I$$

Proof. Follows from Theorem 3. □

These results show the way we can use our framework to model a lot of different situations, often with surprising outcomes.

5 Conclusions and Future work

We developed an expressive dynamic epistemic logic tailored to specify and reason about the information flow over communication channels. We also proposed an intuitive lightweight modeling method for multi-agent communication scenarios. The logic and the modeling method were put to use in the telephone call example.

Our framework is very flexible in modeling different observational powers of agents and various communicative actions. For example, we can define the communicative action in Pacuit and Parikh (2007) : "i gets j's information without j noticing that" as $\alpha = download^i_j(M)$ with $Obs(\alpha) = i$, $Pre(\alpha) = com(\{i, j\}) \wedge has_j M$ and a suitable postcondition adding messages to i's information set[11]. Therefore our framework can facilitate the comparison among different approaches with different assumptions. The table below summarizes the setting of our framework compared to others:

Reference	Actions	Information flow	Obs. Power
Roelofsen (2005)	inform	propositions	$\equiv^\tau$
Pacuit and Parikh (2007)	download	Boolean atomic propositions	$\equiv^\tau$
Apt et al. (2009)	inform	positive atomic propositions	$\equiv^{set}$
Our work	by design	messages or formulas	by design

We end with a list of further issues to be explored:

Theoretical Issues Many theoretical issues are left for future work e.g. the model checking and satisfiability problem of (the fragments of) $\mathcal{L}_\iota^{I,M}$ w.r.t

[11]Pacuit and Parikh (2007) phrases such download action with propositions instead of messages.

different x-semantics and the expressivity of $\mathcal{L}_{\iota}^{I,M}$ compared to various fixed point logics. Another interesting issue is the logical characterization of the observational equivalences defined in our work.

Network In this work, we take the hypergraphs of Apt et al. (2009) as networks, thus assuming the communication channels to be symmetric. More constrained network definitions with asymmetric channels are also possible. Moreover, different social networks/organizations may have different properties, e.g. the network of a group of gossiping girls is usually connected and transtive[12] while the network of a secret society is usually not transitive due to a hierarchy and secrecy. Thus leaking a secret to your closest girl friend may cause it to be a shared knowledge among all the girls on the next day, but gossiping about your boss with the juniors under your supervision might be safe in a secret society.

Actions There are other useful actions that we did not cover in this paper. For example, we have assumed that message passing actions are always monotonic, but there are cases when deleting messages from memory or buffer is natural. Another assumption is that the agents either clearly observe an action or observe nothing at all. This excludes the modeling of actions which may give some agents partial observations e.g. BCC in email. Roelofsen (2005) also mentioned the possibility of changing the channels, e.g. deleting people from your Christmas card sending list if they did not reply to your card last year. Furthermore, the actions that change the protocol may also need to be constrained by the communication channels, as discussed in Moses et al. (1986). This will raise more interesting issues, e.g., whether different levels of knowledge of the protocol (weaker than common knowledge) suffice to facilitate the successful runs of certain class of protocols. Such actions could be handled within our framework with little adaption.

Protocol We use regular expressions without tests to specify sequential protocols. We leave out tests since the observation of a test is not clear, unless it is grouped with follow-up actions. It seems that this is expressive enough for many useful applications. In a more general setting, we would like to have tests and parallel composition in the protocol language and model the protocol by composing local protocols for each agent.

Knowledge Transfer Our framework paves a way to discuss message passing and knowledge transfer over communication channels at the same time. It may be applicable to a security setting where information flow should be controlled strictly complying certain knowledge requirements.

Acknowledgements We would like to thank Krzysztof Apt and Johan van Benthem for the detailed discussions and comments on earlier versions of this work, and also Alexandra Silva for pointing out the references to the derivatives of regular expressions. We also thank the anonymous referees of AAMAS10 for their insightful comments.

[12]In the sense that if girl A can call girl B and girl B can call girl C then A is in touch with C.

References

K. R. Apt, A. Witzel, and J. A. Zvesper. Common knowledge in interaction structures. In A. Heifetz, editor, *TARK*, pages 4–13, 2009.

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

A. Baskar, R. Ramanujam, and S. P. Suresh. Knowledge-based modelling of voting protocols. In *TARK '07: Proceedings of the 11th conference on Theoretical aspects of rationality and knowledge*, pages 62–71, New York, NY, USA, 2007. ACM.

J. A. Brzozowski. Derivatives of regular expressions. *J. ACM*, 11(4):481–494, October 1964. ISSN 0004-5411.

M. Cohen and M. Dam. A complete axiomatization of knowledge and cryptography. In *LICS*, pages 77–88. IEEE Computer Society, 2007.

J. H. Conway. *Regular Algebra and Finite Machines (Chapman and Hall mathematics series)*. Chapman and Hall, September 1971. ISBN 0412106205.

P. O. Dehaye, D. Ford, and H. Segerman. One hundred prisoners and a light bulb. *Mathematical Intelligencer*, 24(4):53–61, 2003.

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

J. Gerbrandy and W. Groeneveld. Reasoning about information change. *Journal of Logic, Language and Information*, 6(2):147–169, April 1997.

T. Hoshi. *Epistemic Dynamics and Protocol Information.* PhD thesis, Department of Philosophy, Stanford University, 2009.

T. Hoshi and A. Yap. Dynamic epistemic logic with branching temporal structures. *Synthese*, 169(2):259–281, July 2009.

C. A. J. Hurkens. Spreading gossip efficiently. *Nieuw Archief voor Wiskunde*, 5/1(2):208–210, 2000.

Y. Moses, D. Dolev, and J. Y. Halpern. Cheating husbands and other stories: A case study of knowledge, action, and communication. *Distributed Computing*, 1(3):167–176, September 1986. ISSN 0178-2770.

E. Pacuit and R. Parikh. Reasoning about communication graphs. In J. van Benthem, D. Gabbay, and B. Löwe, editors, *Interactive Logic — Proceedings of the 7th Augustus de Morgan Workshop*, Texts in Logic and Games, pages 135–157, Amsterdam, 2007.

R. Parikh and R. Ramanujam. A knowledge based semantics of messages. *Journal of Logic, Language and Information*, 12(4), 2003.

R. Parikh and R. Ramanujam. Distributed processes and the logic of knowledge. In *Proceedings of the Conference on Logic of Programs*, pages 256–268, London, UK, 1985. Springer-Verlag. ISBN 3-540-15648-8.

T. Ågotnes, P. Balbiani, H. van Ditmarsch, and P. Seban. Group announcement logic. *Journal of Applied Logic*, July 2009. ISSN 15708683.

R. Ramanujam and S. P. Suresh. Deciding knowledge properties of security protocols. In *Proc. Theoretical Aspects of Rationality and Knowledge*, pages 219–235. Morgan Kaufmann, 2005.

F. Roelofsen. Exploring logical perspectives on dis- tributed information and its dynamics. Master's thesis, University of Amsterdam, 2005.

N. V. Shilov and N. O. Garanina. Model checking knowledge and fixpoints. In Z. Ésik, A. Ingólfsdóttir, Z. Ésik, and A. Ingólfsdóttir, editors, *FICS*, volume NS-02-2 of *BRICS Notes Series*, pages 25–39. University of Aarhus, 2002.

J. van Benthem. 'one is a lonely number': on the logic of communication. In Z. Chatzidakis, P. Koepke, and W. Pohlers, editors, *Logic Colloquium '02*, pages 96–129, Wellesley MA, 2002. ASL & A.K. Peters.

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

J. van Benthem, J. Gerbrandy, T. Hoshi, and E. Pacuit. Merging frameworks for interaction. *Journal of Philosophical Logic*, 38(5):491–526, October 2009. ISSN 0022-3611.

R. van der Meyden and N. Shilov. Model checking knowledge and time in systems with perfect recall. In *Foundations of Software Technology and Theoretical Computer Science*, pages 432–445, 1999.

H. van Ditmarsch. *Knowledge games*. PhD thesis, 2000.

H. van Ditmarsch, W. van der Hoek, and B. Kooi. *Dynamic Epistemic Logic*. (Synthese Library). Springer, 1st edition, November 2007. ISBN 1402069081.

Y. Wang, L. Kuppusamy, and J. van Eijck. Verifying epistemic protocols under common knowledge. In *TARK '09: Proceedings of the 12th Conference on Theoretical Aspects of Rationality and Knowledge*, pages 257–266, New York, NY, USA, 2009. ACM. ISBN 978-1-60558-560-4.