Logics of Informational Attitudes and Informative Actions

Eric Pacuit

July 15, 2010

1 Introduction

There is an extensive literature focused on using logical methods to reason about communities of agents engaged in some form of social interaction. Much of the work builds upon existing logical frameworks developed by philosophers and computer scientists incorporating insights and ideas from philosophy (especially epistemology and philosophy of action), game theory, decision theory and social choice theory. The result is a web of logical systems each addressing different aspects of rational agency and social interaction. This paper focuses one one aspect of this broad area: logical systems designed to model the agents' *informational attitudes* (eg., knowledge, belief, certainty) in social interactive situations. This includes notions of *group* knowledge and *informative action*. Indeed, a key challenge for the logician is to account for the many dynamic processes that govern the agents' (social) interactions over time. Inference, observation and communication are all examples of such processes that are the focus of current logics of informational update and belief revision (see, for example van Benthem, 2010a; van Ditmarsch et al., 2007; Parikh and Ramanujam, 2003)[1]. This paper will introduce these epistemic and doxastic logics as models of "rational interaction" and provide pointers to some current literature.

The point of departure for modern epistemic and doxastic logic is Jaakko Hintikka's seminal text *Knowledge and Belief: An Introduction to the Logic of the Two Notions* (1962)[2]. In fact, Hintikka was not the first to recognize that discourse about knowledge and belief could be the subject of a logical analysis. Indeed, Hintikka cites G.H. Von Wright's *An Essay in Modal Logic* (1951) as the starting point for his logical analysis. A comprehensive history of epistemic and

---

[1] Of course, one may argue that (logical) *inference* is the central topic of *any* logic. What we have in mind here is reasoning *about* agents that make inferences.

[2] This important book has recently been re-issued and extended with some of Hintikka's latest papers on epistemic logic (Hintikka, 2005).

doxastic logic is beyond the scope of this paper; however, the interested reader can consult the following three sources for relevant historical details:

1. Paul Gochet and Pascal Gribomont's article in the *Handbook of the History of Logic* (2006) has an extensive discussion of the main highlights in the technical development of epsitemic logic;

2. Robert Goldblatt's article in the *Handbook of the History of Logic* (2006) has a nearly complete history of the mathematical development of modal logic in the 20th century; and

3. Vincent Hendricks and John Symons (2006, Section 2) describe some key developments in modal logic that led to Hintikka's book.

While Hintikka's project sparked some discussion among mainstream epistemologists (especially regarding the "KK Principle": does knowing something imply that one knows that one knows it?[3]), much of the work on epistemic and doxastic logic was taken over by Game Theorists (Aumann, 1999a) and Computer Scientists (Fagin et al., 1995) in the 1990s. Recently, focus is shifting back to Philosophy with a growing interest in "bridging the gap between formal and mainstream epistemology": witness the collection of articles in (Hendricks, 2006) and the book *Mainstream and Formal Epistemology* by Vincent Hendricks (2005).

Thus, the field of Epistemic Logic has developed into an interdisciplinary area no longer immersed *only* in the traditional questions of mainstream epistemology. Much recent work focuses on explicating epistemic issues in, for example, game theory (Brandenburger, 2007) and economics (Samuelson, 2004), computer security (Halpern and Pucella, 2003; Ramanujam and Suresh, 2005), distributed systems (Halpern and Moses, 1983), and *social software* (Parikh, 2002)[4]. The situation is nicely summarized in a recent article by Robert Stalnaker who suggests that a logical analysis can

> "...bring out contrasting features of some alternative conceptions of knowledge, conceptions that may not provide plausible analyses of knowledge generally, but that may provide interesting models of knowledge that are appropriate for particular applications, and that may illuminate in an idealized way, one or another of the dimensions of the complex epistemological terrain." (Stalnaker, 2006, pg. 170)

---

[3]Timothy Williamson (2000, Chapter 5) has a well-known and persuasive argument against this principle (cf. Egré and Bonnay, 2009, for a discussion of interesting issues for epistemic logic deriving from Williamson's argument).

[4]See also Parikh's contribution to this volume.

In this survey, the modeling of informational attitudes of a group of (rational) agents engaged in some form of social interaction (eg. having a conversation or playing a card game) takes center stage.

Many logical systems today focus on (individual and group) informational attitudes often with a special focus on how the agents' information changes over time. Sometimes differences between "competing" logical systems are technical in nature reflecting different conventions used by different research communities. And so, with a certain amount of technical work, such frameworks are seen to be equivalent up to model transformations (cf. Halpern, 1999; Lomuscio and Ryan, 1997; Pacuit, 2007; van Benthem et al., 2009). Other differences point to key conceptual issues about rational interaction.

The main objective of this paper is to not only to introduce important logical frameworks but also help the reader navigate the extensive literature on (dynamic) epistemic and doxastic logic. Needless to say, we will not be able to do justice to all of this extensive literature. This would require a textbook presentation. Fortunately, there are a number of excellent textbooks on this material (Fagin et al., 1995; van Ditmarsch et al., 2007; van Benthem, 2010a). The article will be self-contained, though familiarity with basic concepts in modal logic may be helpful[5].

2 Informational Attitudes

Contemporary epistemology provides us with a rich typology of informational attitudes. There are numerous notions of knowledge around: the pre-Gettier "justified true belief" view, reliability accounts (Goldman, 1976), counterfactual accounts (Nozick, 1981), and *active* vs. *passive* knowledge (Stalnaker, 1999, pg. 299), to name just a few (cf. Sosa et al., 2008, for a survey). Similarly, beliefs come in many forms: graded or flat-out (Harman, 1986), conditional and lexicographic (Brandenburger, 2007), safe and strong (Baltag and Smets, 2006b). On top of all this, beliefs seem to be just one example in a large variety of "acceptance-like" attitudes (Shah and Velleman, 2005). In this paper, we concentrate on a general distinction between attitudes of *hard* and *soft* information (van Benthem, 2005; Baltag and Smets, 2006a) without taking a stance on which of these attitudes, if any, should be seen as primary, either for epistemology in general or for specific applications.

Hard information, and its companion attitude, is information that is *veridical* and *not revisable*. This notion is intended to capture what the agents are fully and correctly certain of in a given social situation. So, if an agent has hard

---

[5]See (van Benthem, 2010b) for a modern textbook introduction to modal logic and (Blackburn et al., 2002) for an overview of some of the more advanced topics.

information that some fact φ is true, then φ really is true. In absence of better terminology and following common usage in the literature, we use the term *knowledge* to describe this very strong type of informational attitude. However, we make no claim as to whether this notion captures one of the many notions of knowledge just mentioned (in fact, it probably does not) and simply note that "hard information" shares *some* of the characteristics that have been attributed to knowledge in the epistemological literature such as veridicality. *Soft information* is, roughly speaking, anything that is not "hard": it is not necessarily veridical and/or highly revisable in the presence of new information. As such, it comes much closer to *beliefs* or more generally attitudes that can be described as "regarding something as true" (Schwitzgebel, 2008).

Thus, we identify *revisability* as a key distinguishing feature. Typically, discussions of epistemic logic focus instead on the epistemic capabilities of the *agents* such as *introspection* or *logical omniscience*. For example, it is typically assumed that if an agent has the (hard or soft) information that φ is true, then this fact is fully transparent (to the agent). In order keep the presentation manageable, we do not go into details about these interesting issues (cf. Fagin et al., 1995, for extensive discussions).

Before going into details, a few comments about the general approach to modeling are in order. The formal models introduced below can be broadly described as "possible worlds models" familiar in much of the philosophical logic literature. These models assume and underlying set of *states of nature* describing the (ground) facts about the situation being modeled that do not depend on the agents' uncertainties. Typically, these facts are represented by sentences in some propositional (or first-order) language. Each agent is assumed to entertain a number of *possibilities*, called *possible worlds* or simply *(epistemic) states*. These "possibilities" are intended to represent "the current state of the world". So each possibility is associated with a *unique* state of nature (i.e., there is a function from possible worlds to sets of sentences "true" at that world, but this function need not be 1-1 or even onto). Crucial for the epistemic-logic analysis is the assumption that there may be *different* possible worlds associated with the same state of nature. Such possible worlds are important for representing higher-order information (eg., information about the other agents' information). One final common feature is that the agents' informational attitudes are directed towards *propositions*, also called *events* in the game-theory literature, represented as sets of possible worlds. These basic modeling choices are not uncontroversial, but such issues are beyond the scope of this paper[6] and so we opt for mathematical precision in favor of philosophical carefulness.

---

[6]The interested reader can consult (Parikh, 2008) for a discussion.

2.1 Models of Hard Information

Let Agt be a non-empty set of agents and At a (countable or finite) set of atomic sentences. Elements $p \in \mathsf{At}$ are intended to describe ground facts, for example, "it is raining" or "the red card is on the table", in the situation being modeled. A non-empty set W, called *possible worlds* or *states*, are intended to represent the different ways the situation being modeled may evolve. Rather than *directly* representing the agents' *hard information*, the models given below describe the "implicit consequences" of this information in terms of "*epistemic indistinguishability relations*"[7]. The idea is that each agent has some "hard information" about the situation being modeled and agents cannot distinguish between states that agree on this information. In basic epistemic models, this "epistemic indistinguishability" is represented by *equivalence relations* on W:

Definition 2.1 (Epistemic Model) An **epistemic model** (based on the set of agents Agt and set of atomic propositions At) is a tuple $\langle W, \{\sim_i\}_{i \in \mathsf{Agt}}, V \rangle$ where W is a non-empty set; for each $i \in \mathsf{Agt}$, $\sim_i \subseteq W \times W$ is reflexive, transitive and symmetric; and $V : \mathsf{At} \to \wp(W)$ is a valuation function. ◁

A simple propositional modal language will be used to describe properties of these structures. Formally, let $\mathcal{L}_{EL}$ be the (smallest) set of sentences generated by the following grammar:

$$\varphi \;:=\; p \mid \neg\varphi \mid \varphi \wedge \varphi \mid K_i\varphi$$

where $p \in \mathsf{At}$ and $i \in \mathsf{Agt}$. The additional propositional connectives $(\to, \leftrightarrow, \vee)$ are defined as usual and the dual of K_i, denoted L_i, is defined as follows: $L_i\varphi := \neg K_i \neg \varphi$. The intended interpretation of $K_i\varphi$ is "according to agent i's current (hard) information, φ is true" (following standard notation we can also say "agent i knows that φ is true"). Given a story or situation we are interested in modeling, each state $w \in W$ of an epistemic model $\mathcal{M} = \langle W, \{\sim_i\}_{i \in \mathsf{Agt}}, V \rangle$ represents a possible scenario which can be described in the formal language given above: if $\varphi \in \mathcal{L}_{EL}$, $\mathcal{M} = \langle W, \{\sim_i\}_{i \in \mathsf{Agt}}, V \rangle$ and $w \in W$, we write $\mathcal{M}, w \models \varphi$ if φ is a correct description of some aspect of the situation represented by w. This can be made precise as follows:

---

[7]The phrasing "epistemic indistinguishable", although common in the epistemic logic literature, is misleading since, as a relation, "indistinguishability" is *not* transitive. However, we typically assume that epistemic indistinguishability is an equivalence relation. A standard example is: a cup of coffee with n grains of sugar is indistinguishable from a cup with $n+1$ grains; however, transitivity would imply that a cup with 0 grains of sugar is indistinguishable from a cup with 1000 grains of sugar.

Definition 2.2 (Truth) Let $\mathcal{M} = \langle W, \{\sim_i\}_{i\in\mathsf{Agt}}, V\rangle$ be an epistemic model. For each $w \in W$, φ **is true at state** w, denoted $\mathcal{M}, w \models \varphi$, is defined by induction on the structure of φ:

- $\mathcal{M}, w \models p$ 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 $\mathcal{M}, w \models \psi$

- $\mathcal{M}, w \models K_i\varphi$ iff for all $v \in W$, if $w \sim_i v$ then $\mathcal{M}, v \models \varphi$
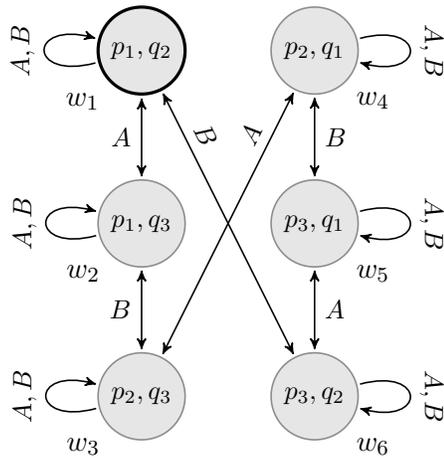
We say φ is **satisfiable** if there is an epistemic model $\mathcal{M} = \langle W, \{\sim_i\}_{i\in\mathsf{Agt}}, V\rangle$ and state $w \in W$ such that $\mathcal{M}, w \models \varphi$; and φ is **valid in** $\mathcal{M}$, denoted $\mathcal{M} \models \varphi$ if $\mathcal{M}, w \models \varphi$ for all $w \in W$. ◁

Given the definition of the dual of K_i, it is easy to see that

$$\mathcal{M}, w \models L_i\varphi \text{ iff there is a } v \in W \text{ such that } \mathcal{M}, v \models \varphi.$$

Thus an interpretation of $L_i\varphi$ is "φ is consistent with agent i's current (hard) information". The following example will illustrate the above definitions.

Suppose there are two agents, Ann (A) and Bob (B), and three cards labeled with the numbers 1, 2 and 3. Consider the following scenario: Ann is dealt one of the cards, Bob is given one of the cards and the third card is put face down on a table. What are the relevant possible worlds for this scenario? The answer to this question depends, in part, on the level of detail in the description of the situation being modeled. For example, relevant details may include whether Ann is holding the card in her right hand or left hand, the color of the cards or whether it is raining outside. The level of detail is fixed by the choice of atomic propositions. For example, suppose that $\mathsf{At} = \{p_1, p_2, p_3, q_1, q_2, q_3\}$ where p_i is intended to mean that "Ann has card i" and q_i is intended to mean "Bob has card i". Since each agent is given precisely one of the three possible cards, there are 6 relevant possible worlds, $W = \{w_1, w_2, w_3, w_4, w_5, w_6\}$, one for each way the cards could be distributed. What about the agents' information? Some of the aspects about the situation being modeled can be classified as "informative" for the agents. For example, since the third card is placed face down on the table, neither agent "knows" the number written on the other agent's card. The complete epistemic state of the agents is described by the epistemic model pictured below (in this picture, an i-labeled arrow from state w to state v means $w \sim_i v$ and each state is labeled with the atomic propositions true at that state):

A,B p_1, q_2 p_2, q_1 A,B
w_1 w_4

A   B   A   B

A,B p_1, q_3 p_3, q_1 A,B
w_2 w_5

B   A

A,B p_2, q_3 p_3, q_2 A,B
w_3 w_6

- $\mathcal{M}, w_1 \models K_A p_1 \wedge K_A \neg q_1$

- $\mathcal{M}, w_1 \models K_A(q_2 \vee q_3)$

- $\mathcal{M}, w_1 \models K_A(K_B p_2 \vee K_B \neg p_2)$

The reader is invited to check that the formulas to the right are indeed true at state w_1 according to Definition 2.2. The intuitive interpretation of these formulas describe (part of) the hard information that Ann has in the above situation. For example, Ann knows that she has card 1 (i.e., it is assumed that Ann is looking at her card); Ann knows that Bob does not have card 1 (because, for example, Ann has background knowledge that there are only three cards with no duplicates); and Ann knows that Bob either has card 2 or card 3 and she knows that Bob knows *whether* he has card 2 (this can also be derived from her background knowledge).

Notice that the set of states that Ann considers possible at w_1 is $\{w_1, w_2\}$. This set is the truth set of the formula p_1 (i.e., $\{x \mid \mathcal{M}, x \models p_1\} = \{w_1, w_2\}$); and so, we can say that *all Ann knows about the situation is that she has card 1*. The other propositions that Ann knows are (non-monotonic) *consequences* of this proposition (given her background knowledge about the situation). This suggests that it may be useful to include an operator "for all agent i knows". In fact, this notion was introduced by Hector Levesque (1990) and, although the logical analysis turned out to be a challenge (cf. Humberstone, 1987; Halpern and Lakemeyer, 1995, 2001; Engelfriet and Venema, 1998), has proven useful in the epistemic analysis of certain solution concepts (Halpern and Pass, 2009).

The above epistemic models are intended to represent the agents' *hard information* about the situation being modeled. In fact, we can be much more precise about the sense in which these models "represent" the agents' hard information by using standard techniques from the mathematical theory of modal logic (Blackburn et al., 2002). In particular, *modal correspondence theory* rigorously relates properties of the the relation in an epistemic model with modal

formulas (cf. Blackburn et al., 2002, Chapter 3)[8]. The following table lists some key formulas in the language $\mathcal{L}_{EL}$ with their corresponding (first-order) property and the relevant underlying assumption.

Assumption	Formula	Property
Logical Omniscience	$K_i(\varphi \to \psi) \to (K_i\varphi \to K_i\psi)$	—
Veridical	$K_i\varphi \to \varphi$	Reflexive
Positive Introspection	$K_i\varphi \to K_iK_i\varphi$	Transitive
Negative Introspection	$\neg K_i\varphi \to K_i\neg K_i\varphi$	Euclidean

Viewed as a description, even an idealized one, of *knowledge*, the above properties have raised many criticisms. While the logical omniscience assumption (which is valid on all models regardless of the properties of the accessibility relation) generated the most extensive criticisms (Stalnaker, 1991) and responses (cf. Fagin et al., 1995, Chapter 9), the two introspection principles have also been the object of intense discussion (cf. Williamson, 2000; Egré and Bonnay, 2009)[9]. These discussions are fundamental for the theory of knowledge and its formalization, but here we choose to bracket them, and instead take epistemic models for what they are: models of hard information, in the sense introduced above.

2.2 Varieties of Soft Information

A small modification of the above epistemic models allows us to model a softer informational attitude. Indeed, by simply replacing the assumption of reflexivity of the relation $\sim_i$ with seriality (for each state w there is a state v such that $w \sim_i v$), but keeping the other aspects of the model the same, we can capture

---

[8]To be more precise, the key notion here is *frame definability*: a frame is a pair $\langle W, R \rangle$ where W is a nonempty set and R a relation on W. A modal formula is valid on a frame if it is valid in every model (cf. Definition 2.1) based on that frame. It can be shown that some modal formulas have first-order *correspondents* P where for any frame $\langle W, R \rangle$, the relation R has property P iff φ is valid on $\langle W, R \rangle$. A highlight of this theory is *Sahlqvist's Theorem* which provides and algorithm for finding first-order correspondents for certain modal formulas. See (Blackburn et al., 2002, Sections 3.5 - 3.7) for an extended discussion.

[9]In fact, Hintikka explicitly rejects negative introspection: "The consequences of this principle, however, are obviously wrong. By its means (together with certain intuitively acceptable principles) we could, for example, show that the following sentence is self sustaining $p \to K_iL_ip$" (Hintikka, 1962, pg. 54). Hintikka regards this last formula as counter-intuitive since it means that if it possible that an agent knows some fact p then that fact must be true. However, it seems plausible that an agent can justifiably believe that she knows some fact p but p is in fact false. Other authors have pointed out difficulties with this principle in modal systems with both knowledge and belief modalities: see, in particular, (Stalnaker, 2006) and (Shoham and Leyton-Brown, 2009, Section 13.7).

what epistemic logicians have called "*beliefs*". Formally, a **doxastic model** is a tuple $\langle W, \{R_i\}_{i\in\mathsf{Agt}}, V \rangle$ where W is a nonempty set of states, R_i is a transitive, Euclidean and serial relation on W and V is a valuation function (cf. Definition 2.1). Truth is defined precisely as in Definition 2.2, replacing $\sim_i$ with R_i. This notion of belief is very close to the above hard informational attitude and, in fact, shares all the properties of K_i listed above except *Veracity* (this is replaced with a weaker assumption that agents are "consistent" and so cannot believe contradictions). This points to a logical analysis of both informational attitudes with various "bridge principles" relating knowledge and belief (such as knowing something implies believing it or if an agent believes φ then the agent knows that he believes it). However, we do not discuss this line of research (see, for example, Halpern, 1996; Stalnaker, 2006) here since these models are not our preferred ways of representing the agents' soft information.

A key aspect of beliefs which is not yet represented in the above models is that they are *revisable* in the presence of new information. While there is an extensive literature on the theory of belief revision (see the article by Booth and Meyer in this collection for a discussion), the focus here is how to extend the above models with a representation of softer, revisable informational attitudes. The standard approach is to include a *plausibility ordering* for each agent: a preorder (reflexive and transitive) denoted $\preceq_i \subseteq W \times W$. If $w \preceq_i v$ we say "player i considers v at least as plausible as w." For $X \subseteq W$, let

$$Min_{\preceq_i}(X) = \{v \in W \mid v \preceq_i w \text{ for all } w \in X \}$$

denote the set of minimal elements of X according to $\preceq_i$. Thus while the $\sim_i$ partitions the set of possible worlds according to the hard information the agents are assumed to have about the situation, the plausibility ordering $\preceq_i$ represents which of the possible worlds the agent considers more likely (i.e., it represents the players soft information). Models representing both the agents' hard and soft information have been used not only by logicians (van Benthem, 2004; Baltag and Smets, 2006b) but also by game theorists (Board, 2004) and computer scientists (Boutilier, 1992; Lamarre and Shoham, 1994):

Definition 2.3 (Epistemic-Doxastic Models) Suppose Agt is a set of agents and At a set of atomic propositions, an **epistemic doxastic model** is a tuple $\langle W, \{\sim_i\}_{i\in\mathsf{Agt}}, \{\preceq_i\}_{i\in\mathsf{Agt}}, V \rangle$ where $\langle W, \{\sim_i\}_{i\in\mathsf{Agt}}, V \rangle$ is an epistemic model and for each $i \in \mathsf{Agt}$, $\preceq_i$ is a well-founded[10], reflexive and transitive relation on W satisfying the following properties, for all $w, v \in W$

1. *plausibility implies possibility*: if $w \preceq_i v$ then $w \sim_i v$.

---

[10]Well-foundedness is only needed to ensure that for any set X, $Min_{\preceq_i}(X)$ is nonempty. This is important only when W is infinite.

2. *locally-connected*: if $w \sim_i v$ then either $w \preceq_i v$ or $v \preceq_i w$. ◁

Remark 2.4 *Note that if $w \not\sim_i v$ then, since $\sim_i$ is symmetric, we also have $v \not\sim_i w$, and so by property 1, $w \not\preceq_i v$ and $v \not\preceq_i w$. Thus, we have the following equivalence: $w \sim_i v$ iff $w \preceq_i v$ or $v \preceq_i w$.*

Let $[w]_i$ be the equivalence class of w under $\sim_i$. Then local connectedness implies that $\preceq_i$ totally orders $[w]_i$ and well-foundedness implies that $Min_{\preceq_i}([w]_i)$ is nonempty. This richer model allows us to formally define a variety of (soft) informational attitudes. We first need some additional notation: the plausibility relation $\preceq_i$ can be lifted to subsets of W as follows[11]

$$X \preceq_i Y \text{ iff } x \preceq_i y \text{ for all } x \in X \text{ and } y \in Y$$

Suppose $\mathcal{M} = \langle W, \{\sim_i\}_{i \in \mathsf{Agt}}, \{\preceq_i\}_{i \in \mathsf{Agt}}, V \rangle$ is an epistemic-doxastic model with $w \in W$, consider the following extensions to the language $\mathcal{L}_{EL}$

- *Belief*: $\mathcal{M}, w \models B_i\varphi$ iff for all $v \in Min_{\preceq_i}([w]_i)$, $\mathcal{M}, v \models \varphi$.
  This is the usual notion of belief which satisfies the standard properties discussed above (eg., positive and negative introspection).

- *Safe Belief*: $\mathcal{M}, w \models \square_i\varphi$ iff for all v, if $v \preceq_i w$ then $\mathcal{M}, v \models \varphi$.
  Thus, φ is safely believed if φ is true in *all* states the agent considers more plausible. This stronger notion of belief has also been called *certainty* by some authors (cf. Shoham and Leyton-Brown, 2009, Section 13.7).

- *Strong Belief*: $\mathcal{M}, w \models B_i^s\varphi$ iff there is a v such that $w \sim_i v$ and $\mathcal{M}, v \models \varphi$ and $\{x \mid \mathcal{M}, x \models \varphi\} \cap [w]_i \preceq_i \{x \mid \mathcal{M}, x \models \neg\varphi\} \cap [w]_i$.
  So φ is strongly believed provided it is epistemically possible and agent i considers *any* state satisfying φ more plausible than *any* state satisfying $\neg\varphi$. This notion has also been studied by Stalnaker (1994) and Battigalli and Siniscalchi (2002).

The logic of these notions has been extensively studied by Alexandru Baltag and Sonja Smets in a series of articles (2006a; 2008a; 2006b; 2009). We conclude this section with a few remarks about the relationship between these different notions. For example, it is not hard to see that if agent i knows that φ then i (safely, strongly) believes that φ. However, much more can be said about the logical relationship between these different notions (cf. Baltag and Smets, 2009).

As noted above, a crucial feature of these informational attitudes is that they are *defeasible* in light of new evidence. In fact, we can characterize these

---

[11]This is only one of many possible choices here, but it is the most natural in this setting (cf., Liu, 2008, Chapter 4).

attitudes in terms of the type of evidence which can prompt the agent to adjust her beliefs. To make this precise, we introduce the notion of a *conditional belief*: suppose $\mathcal{M} = \langle W, \{\sim_i\}_{i \in \mathsf{Agt}}, \{\preceq_i\}_{i \in \mathsf{Agt}}, V \rangle$ is an epistemic-doxastic and φ and ψ are formulas, then we say i *believes* φ *given* ψ, denoted $B_i^{\psi} \varphi$, provided

$$\mathcal{M}, w \models B_i^{\psi} \varphi \text{ iff for all } v \in Min_{\preceq_i}(\llbracket \psi \rrbracket_{\mathcal{M}} \cap [w]_i), \mathcal{M}, v \models \varphi$$

where $\llbracket \varphi \rrbracket_{\mathcal{M}} = \{w \mid \mathcal{M}, w \models \varphi\}$ is the *truth set* of φ. So, 'B_i^{ψ}' encodes what agent i will believe upon receiving (possibly misleading) evidence that ψ is *true*. Two observations are immediate. First of all, we can now define belief $B_i \varphi$ as $B_i^{\top} \varphi$ (belief in φ given a tautology). Second, unlike beliefs, conditional beliefs may be inconsistent (i.e., $B^{\psi} \perp$ may be true at some state). In such a case, agent i cannot (on pain of inconsistency) revise by ψ, but this will only happen if the agent has hard information that ψ is false. Indeed, $K \neg \varphi$ is logically equivalent to $B_i^{\varphi} \perp$ (over the class of epistemic-doxastic models). This suggests the following (dynamic) characterization of an agents' hard information as unrevisable beleifs:

$$\mathcal{M}, w \models K_i \varphi \text{ iff } \mathcal{M}, w \models B_i^{\psi} \varphi \text{ for all } \psi$$

Safe belief and strong belief can be similarly characterized by restricting the admissible evidence:

- $\mathcal{M}, w \models \Box_i \varphi$ iff $\mathcal{M}, w \models B_i^{\psi} \varphi$ for all ψ with $\mathcal{M}, w \models \psi$.
  That is, i safely beliefs φ iff i continues to believe φ given any true formula.

- $\mathcal{M}, w \models B_i^s \varphi$ iff $\mathcal{M}, w \models B_i \varphi$ and $\mathcal{M}, w \models B_i^{\psi} \varphi$ for all ψ with $\mathcal{M}, w \models \neg K_i(\psi \rightarrow \neg \varphi)$.
  That is, agent i strongly believes φ iff i believes φ and continues to believe φ given any evidence (truthful or not) that is not known to contradict φ.

Baltag and Smets (2009) provide an elegant logical characterization of the above notions by adding the safe belief modality ($\Box_i$) to the epistemic language $\mathcal{L}_{EL}$ (denote the new language $\mathcal{L}_{EDL}$). First of all, note that conditional belief (and hence belief) and strong belief are *definable* in this language:

- $B_i^{\varphi} \psi := L_i \varphi \rightarrow L_i(\varphi \wedge \Box_i(\varphi \rightarrow \psi))$

- $B_i^s \varphi := B_i \varphi \wedge K_i(\varphi \rightarrow \Box_i \varphi)$

All that remains is to characterize properties of an epistemic-doxastic model (Definition 2.3). As discussed in the previous Section, K_i satisfies logical omnsicience, veracity and both positive and negative introspection. Safe belief, $\Box_i$, shares all of these properties except negative introspection. Modal correspondence theory can again be used to characterize the remaining properties:

11

- Knowledge implies safe belief: $K_i \varphi \rightarrow \Box_i \varphi$
  (*Definition 2.3, property 1*)

- Locally connected: $K_i(\varphi \vee \Box \psi) \wedge K_i(\psi \vee \Box \varphi) \rightarrow K_i \varphi \vee K_i \psi$
  (*Definition 2.3, property 2*)

Remark 2.5 *The above models use a "crisp" notion of uncertainty, i.e., for each agent and state w, any other state $v \in W$ is either is or is not possible/more plausible than w. However, there is an extensive body of literature developing* graded, *or* quantitative, *models of uncertainty (Halpern, 2003). For instance, in the Game Theory literature it is standard to represent the players' beliefs by probabilities (Aumann, 1999b; Harsanyi, 1967). The idea here is to use probability distributions in place of the above plausibility orderings. Formally, a epistemic-proabalistic model is a tuple $\mathcal{M} = \langle W, \{\sim_i\}_{i \in \mathsf{Agt}}, \{P_i\}_{i \in \mathsf{Agt}}, V \rangle$ where $\langle W, \{\sim_i\}_{i \in \mathsf{Agt}}, V \rangle$ is an epistemic model and $P_i : W \rightarrow \Delta(W)$ ($\Delta(W) = \{p : W \rightarrow [0,1] \mid p$ is a probability measure $\}$) assigns to each state a probability measure over W. Write p_i^w for the i's probability measure at state w. We make two natural assumptions (cf. Definition 2.3):*

1. *For all $v \in W$, if $p_i^w(v) > 0$ then $p_i^w = p_i^v$ (i.e., if i assigns a non-zero probability to state v at state w then the agent uses the same probability measure at both states)*

2. *For all v, if $w \not\sim_i v$ then $p_i^w(v) = 0$ (i.e., assign nonzero probability only to the states in i's (hard) information set, cf. Definition 2.3 item 1).*

Many different formal languages have been used to describe these rich structures. Examples range from '$\Box_i \varphi$' with the intended meaning "φ is more probable than $\neg\varphi$ for agent i" (Herzig, 2003) to more expressive languages containing operators of the form $B_i^q \varphi$ (with q a rational number) and interpreted as follows:

$$\mathcal{M}, w \models B_i^q(\varphi) \text{ iff } p_i^w(\{v \mid \mathcal{M}, v \models \varphi\}) \geq q.$$

These models have also been the subject of sophisticated logical analyses (Fagin et al., 1990; Fagin and Halpern, 1994; Heifetz and Mongin, 2001) complementing the logical frameworks introduced in this paper (Baltag and Smets, 2007).

2.3 Group Attitudes

Suppose there are two friends Ann and Bob on a bus separated by a crowd. Before the bus comes to the next stop a mutual friend from outside the bus yells

"get off at the next stop to get a drink?". Say Ann is standing near the front door and Bob near the back door. When the bus comes to a stop, will they get off? Of course, this depends, in part, on Ann and Bob's preferences. Suppose that both Ann and Bob want to have a drink with their mutual friend, but *only if both are there for the drink*. So Ann will only get off the bus if she "knows" (justifiable believes) that Bob will also get off (similarly for Bob). But this does not seem to be enough (after all, she needs some assurance that Bob is thinking along the same lines). In particular, she needs to "know" (justifiably believe) that Bob "knows" (justifiably believes) that she is going to get off at the next stop. Is this state of knowledge sufficient for Ann and Bob to coordinate their actions? Lewis (1969) and Clark and Marshall (1981) argue that a condition of *common knowledge* is necessary for such coordinated actions. In fact, a seminal result by Halpern and Moses (1983) shows that, without synchronized clocks, such coordinated action is impossible. Chwe (2001) has a number of examples that point out the everyday importance of the notion of common knowledge.

Both the game theory community and the epistemic logic community have extensively studied formal models of common knowledge and belief. Barwise (1988) highlights three main approaches to formalize common knowledge: (i) the iterated view, (ii) the fixed-point view and (iii) the shared situation view. Here we will focus only on the first two approaches (cf. van Benthem and Sarenac, 2004, for a rigorous comparison between (i) and (ii)). Vanderschraaf and Sillari (2009) provide an extensive discussion of the literature (see also Fagin et al., 1995, for a general discussion).

Consider the statement "everyone in group G knows φ". If there are only finitely many agents, this can be easily defined in the epistemic language $\mathcal{L}_{EL}$:

$$E_G\varphi \quad := \quad \bigwedge_{i \in G} K_i\varphi$$

where $G \subseteq \mathsf{Agt}$. Following Lewis (1969)[12], the intended interpretation of "it is common knowledge in G that φ" (denoted $C_G\varphi$) is the infinite conjunction:

$$\varphi \wedge E_G\varphi \wedge E_G E_G\varphi \wedge E_G E_G E_G\varphi \wedge \cdots$$

However, this involves an *infinite* conjunction, so cannot be a formula in the language of epistemic logic. This suggests that common knowledge is not definable in the language of multi-agent epistemic logic[13]. Thus we need to add a new symbol to the language $C_G\varphi$ whose intended interpretation is "it is common

---

[12]Although see (Cubitt and Sugden, 2003) for an alternative reconstruction of Lewis' notion of common knowledge.

[13]In fact, one can prove this using standard methods in modal logic.

knowledge in the group G that φ". Let $\mathcal{L}_{EL}^C$ be the smallest set generated by the following grammar:

$$p \mid \neg\varphi \mid \varphi \wedge \varphi \mid K_i\varphi \mid C_G\varphi$$

with $p \in \mathsf{At}$ and $G \subseteq \mathsf{Agt}$.

Before giving semantics to $C_G\varphi$, we consider $E_G E_G E_G \varphi$. This formula says that "everyone from group G knows that everyone from group G knows that everyone from group G knows that φ". When will this be true at a state w in an epistemic model? First some notation: a **path on length n for** G in an epistemic model is a sequence of states $(w_0, w_2, \ldots, w_n)$ where for each $l = 0, \ldots, n-1$, we have $w_l \sim_i w_{l+1}$ for some $i \in G$ (for example $w_0 \sim_1 w_1 \sim_2 w_2 \sim_1 w_3$ is a path of length 3 for $\{1, 2\}$). Thus, $E_G E_G E_G \varphi$ is true at state w iff every path of length 3 for G starting at w leads to a state where φ is true. This suggests the following definition:

Definition 2.6 (Interpretation of C) Let $\mathcal{M} = \langle W, \{\sim_i\}_{i \in \mathcal{A}}, V \rangle$ be an epistemic model and $w \in W$. The truth of formulas of the form $C\varphi$ is:

$$\mathcal{M}, w \models C_G\varphi \quad \text{iff} \quad \text{for all } v \in W, \text{ if } wR_G^*v \text{ then } \mathcal{M}, v \models \varphi$$

where $R_G^* := (\bigcup_{i \in G} \sim_i)^*$ is the reflexive transitive closure of $\bigcup_{i \in G} \sim_i$. ◁

Sometimes it is useful to work with the following equivalent characterization:

$\mathcal{M}, w \models C_G\varphi$ iff every finite path for G from w ends with a state satisfying φ.

The logical analysis is much more complicated in languages with a common knowledge operator; however, the following two axioms can be said to characterize[14] common knowledge:

- Fixed-Point: $C_G\varphi \rightarrow E_G C_G\varphi$

- Induction: $\varphi \wedge C_G(\varphi \rightarrow E_G\varphi) \rightarrow C_G\varphi$

The first formula captures the "self-evident" nature of common knowledge: if some fact is common knowledge in some group G then everyone in G not only knows the fact but also that it is common knowledge. Aumann (1999a) uses this as an alternative characterization of common knowledge:

---

[14]Techniques similar to the previously mentioned *correspondence theory* can be applied here to make this precise: see (van Benthem, 2006) for a discussion.

Suppose you are told "Ann and Bob are going together,"' and respond "sure, that's common knowledge." What you mean is not only that everyone knows this, but also that the announcement is pointless, occasions no surprise, reveals nothing new; in effect, that the situation after the announcement does not differ from that before. ...the event "Ann and Bob are going together" — call it E — is common knowledge if and only if some event — call it F — happened that entails E and also entails all players' knowing F (like all players met Ann and Bob at an intimate party). (Aumann, 1999a, pg. 271)

Remark 2.7 *In this section we have focused only on the notion of common knowledge (eg., hard information). What about notions of common (safe, strong) belief? The general approach outlined above also works for these informational attitudes: for example, suppose $wR_i^B v$ iff $v \in Min_{\preceq_i}([w]_i)$ and define R_G^B to be the transitive closure of $\cup_{i \in G} R_i^B$. Of course, this does raise interesting technical and conceptual issues, but these are beyond the scope of this paper (cf. Bonanno, 1996; Lismont and Mongin, 1994, 2003).*

While it is true that coordinated actions do happen, the analysis of many social situations suggests that other "levels of knowledge", short of the above infinite-common knowledge level are also relevant. Such levels can arise in certain pragmatic situations:

Example 2.8 Suppose that Ann would like Bob to attend her talk; however, she only wants Bob to attend if he is interested in the subject of her talk, not because he is just being polite. There is a very simple procedure to solve Ann's problem: Have a (trusted) friend tell Bob the time and subject of her talk.

Taking a cue from computer science, perhaps we can *prove* that this simple procedure correctly solves Ann's problem. However, it is not so clear how to define a correct solution to Ann's problem. If Bob is actually present during Ann's talk, can we conclude that Ann's procedure succeeded? Not really. Bob may have figured out that Ann wanted him to attend, and so is there only out of politeness. Thus for Ann's procedure to succeed, she must achieve a certain "level of knowledge" (cf. Parikh, 2003) between her and Bob. Besides both Ann and Bob knowing about the talk and Ann knowing that Bob knows, we have

Bob *does not know* that Ann knows about the talk.

This last point is important, since, if Bob knows that Ann knows that he knows about the talk, he may feel social pressure to attend[15]. Thus, the procedure *to*

---

[15] Of course, this is not meant to be a complete analysis of "social politeness".

have a friend tell Bob about the talk, but not reveal that it is at Ann's suggestion, will satisfy all the conditions. Telling Bob directly will satisfy the first three, but not the essential last condition.

We conclude this section by briefly discussing another notion of "group knowledge": *distributed knowledge*. Intuitively, φ is distributed knowledge among a group of agents if φ would be known if all the agents in the group put all their information together. Formally, given an epistemic model (beliefs do not play a role here) $\mathcal{M} = \langle W, \{\sim_i\}_{i \in \mathsf{Agt}}, V \rangle$, let $R_G^D = \cap_{i \in G} \sim_i$, then define

$$\mathcal{M}, w \models D_G\varphi \text{ iff for all } v \in W, \text{ if } wR_G^Dv \text{ then } \mathcal{M}, v \models \varphi.$$

Note that $D_G\varphi$ is *not* simply equivalent to $\bigwedge_{i \in G} K_i\varphi$ (the reader is invited to prove this well-known fact). Indeed, the logical analysis has raised a number of interesting technical and conceptual issues (cf. Halpern and Moses, 1983; Gerbrandy, 1999a; van der Hoek et al., 1999; Roelofsen, 2007; van Benthem, 2010a).

3 Informative Actions

The logical models and languages introduced in the previous Section provide *static* descriptions of the situation being modeled. However, many of the situations we are interested in concern agents interacting over time, and this *dynamics* also calls for a logical analysis. Indeed, an important challenge for the logician is to account for the many dynamic processes that govern the agents' social interactions. Inference, observation and communication are all examples of such processes that are the focus of current logics of informational update and belief revision (see, for example, van Benthem, 2010a; van Ditmarsch et al., 2007; Parikh and Ramanujam, 2003)[16]. In this Section, we discuss some key issues that appear when shifting from a static to a dynamic perspective.

The main issue is how to incorporate *new* information into an epistemic-doxastic model. At a fixed moment in time the agents are in some *epistemic state* (which may be described by an epistemic(-doxastic) model). The question is how does (the model of) this epistemic state change during the course of some social interaction? The first step towards answering this question is identifying (and formally describing) the *informative* events that shape a particular social interaction. Typical examples include showing one's hand in a card game, make a public or private announcement or sending an email message. However, this step is not always straightforward since the information conveyed by a particular event may depend on many factors which need to be specified. Even the *absence*

---

[16]Of course, one may argue that (logical) *inference* is the central topic of *any* logic. What we have in mind here is reasoning *about* agents that make inferences.

of an event can trigger a change in an agent's informational state: Recall the famous observation of Sherlock Holmes in *Silver Blaze*: "Is there any point to which you would wish to draw my attention?" "To the curious incident of the dog in the night-time." "The dog did nothing in the night-time." "That was the curious incident," remarked Sherlock Holmes.

Current dynamic epistemic(-doxastic) logics focus on three key issues:

1. The agents' *observational* powers. Agents may perceive the same event differently and this can be described in terms of what agents do or do not observe. Examples range from *public announcements* where everyone witnesses the same event to private communications between two or more agents with the other agents not even being aware that an event took place.

2. The *type* of change triggered by the event. Agents may differ in precisely how they incorporate new information into their epistemic states. These differences are based, in part, on the agents' perception of the *source* of the information. For example, an agent may consider a particular source of information *infallible* (not allowing for the possibility that the source is mistaken) or merely *trustworthy* (accepting the information as reliable though allowing for the possibility of a mistake).

3. The underlying *protocol* specifying which events (observations, messages, actions) are available (or permitted) at any given moment. This is intended to represent the rules or conventions that govern many of our social interactions. For example, in a conversation, it is typically not polite to "blurt everything out at the beginning", as we must speak in small chunks. Other natural conversational protocol rules include "do not repeat yourself", "let others speak in turn", and "be honest". Imposing such rules *restricts* the legitimate sequences of possible statements or events.

A comprehensive theory of rational interaction focuses on the sometimes subtle interplay between these three aspects (cf. van Benthem, 2010a).

The most basic type of informational change is a so-called *public announcement* (Plaza, 1989; Gerbrandy, 1999b). This is the event where some proposition φ (in the language of $\mathcal{L}_{EL}$) is made *publicly* available. That is, it is completely open and all agents not only observe the event but also observe everyone else observing the event, and so on *ad infinitum* (cf. item 1 above). Furthermore, all agents treat the source as *infallible* (cf. item 2 above). Thus the effect of such an event on an epistemic(-doxastic) model should be clear: *remove* all states that do not satisfy φ. Formally,

Definition 3.1 (Public Announcement) Suppose $\mathcal{M} = \langle W, \{\sim_i\}_{i \in \mathsf{Agt}}, \{\preceq_i\}_{i \in \mathsf{Agt}}, V \rangle$ is an epistemic-doxastic model and φ is a formula (in $\mathcal{L}_{EDL}$). The

model updated by the **public announcement of** φ is the structure $\mathcal{M}^\varphi = \langle W^\varphi, \{\sim_i^\varphi\}_{i\in\mathsf{Agt}}, \{\preceq_i^\varphi\}_{i\in\mathsf{Agt}}, V^\varphi \rangle$ where $W^\varphi = \{w \in W \mid \mathcal{M}, w \models \varphi\}$, for each $i \in \mathsf{Agt}$, $\sim_i^\varphi = \sim_i \cap W^\varphi \times W^\varphi$, $\preceq_i^\varphi = \preceq_i \cap W^\varphi \times W^\varphi$, and for all atomic proposition p, $V^\varphi(p) = V(p) \cap W^\varphi$. $\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\triangleleft$

It is not hard to see that if $\mathcal{M}$ is an epistemic-doxastic model then so is $\mathcal{M}^\varphi$. So, the models $\mathcal{M}$ and $\mathcal{M}^\varphi$ describe two different moments in time with $\mathcal{M}$ describing the current or initial information state of the agents and $\mathcal{M}^\varphi$ the information state *after* the information that φ is true has been incorporated in $\mathcal{M}$. This temporal dimension needs to also be represented in our logical language: let $\mathcal{L}_{PAL}$ extend $\mathcal{L}_{EDL}$ with expressions of the form $[\varphi]\psi$ with $\varphi \in \mathcal{L}_{EDL}$. The intended interpretation of $[\varphi]\psi$ is "ψ is true after the public announcement of φ" and truth is defined as $\mathcal{M}, w \models [\varphi]\psi$ iff if $\mathcal{M}, w \models \varphi$ then $\mathcal{M}^\varphi, w \models \psi$.

For the moment, focus only on the agents' hard information and consider the formula $\neg K_i \psi \wedge [\varphi] K_i \psi$: this says that "agent i (currently) does not know ψ but after the announcement of φ, agent i knows ψ". So, the language of $\mathcal{L}_{PAL}$ describes what is true both before and after the announcement. A fundamental insight is that there is a strong logical relationship between what is true before and after an announcement in the form of so-called *reduction axioms*:

$$
\begin{array}{lcl}
[\varphi]p & \leftrightarrow & \varphi \to p, \quad \text{where } p \in \mathsf{At} \\
[\varphi]\neg\psi & \leftrightarrow & \varphi \to \neg[\varphi]\psi \\
[\varphi](\psi \wedge \chi) & \leftrightarrow & [\varphi]\psi \wedge [\varphi]\chi \\
[\varphi][\psi]\chi & \leftrightarrow & [\varphi \wedge [\varphi]\psi]\chi \\
[\varphi]K_i\varphi & \leftrightarrow & \varphi \to K_i(\varphi \to [\varphi]\psi)
\end{array}
$$

These are reduction axioms in the sense that going from left to right either the number of announcement operators is reduced or the complexity of the formulas within the scope of announcement operators is reduced. These reductions axioms provide an insightful syntactic analysis of announcements which complements the semantic analysis. In a sense, the reduction axioms describe the effect of an announcement in terms of what is true before the announcement. By relating pre- and postconditions for each logical operator, the reduction axioms completely characterize the announcement operator.

The above reductions axioms also illustrate the mixture of factual and *procedural* truth that drives conversations or processes of observation (cf. item 3 above). To be more explicit about this point, consider the formula $\langle\varphi\rangle\top$ (with $\langle\varphi\rangle\psi = \neg[\varphi]\neg\psi$ the dual of $[\varphi]$) which means "φ is *announceable*". It is not hard to see that $\langle\varphi\rangle\top \leftrightarrow \varphi$ is derivable using standard modal reasoning and the above reduction axioms. The left-to-right direction represents a semantic fact

about public announcements (only true facts can be announced), but the right-to-left direction represents specific *procedural information*: every true formula is available for announcement. But this is only one of many different protocols and different assumptions about the protocol is reflected in a logical analysis. Consider the following variations of the knowledge reduction axiom (cf. van Benthem et al., 2009, Section 4):

1. $\langle\varphi\rangle K_i\psi \leftrightarrow \varphi \wedge K_i\langle\varphi\rangle\psi$

2. $\langle\varphi\rangle K_i\psi \leftrightarrow \langle\varphi\rangle\top \wedge K_i(\varphi \rightarrow \langle\varphi\rangle\psi)$

3. $\langle\varphi\rangle K_i\psi \leftrightarrow \langle\varphi\rangle\top \wedge K_i(\langle\varphi\rangle\top \rightarrow \langle\varphi\rangle\psi)$

Each of these axioms represent a different assumption about the underlying protocol and how that affects the agents' knowledge. The first is the above reduction axiom (in the dual form) and assumes a specific protocol (which is common knowledge) where all true formulas are always available for announcement. The second (weaker) axiom is valid when there is a fixed protocol that is common knowledge. Finally, the third adds a requirement that the agents must know which formulas are currently available for announcement. Of course, the above three formulas are all *equivalent* given our definition of truth in an epistemic(-doxastic) model (Definition 2.2) and public announcement (Definition 3.1). In order to see a difference, the *protocol information* must be explicitly represented in the model (cf. Section 3.1 and van Benthem et al., 2009).

We end this introductory Section with a few comments about the effect of a public announcement on the agents' soft information. In particular, it is natural to wonder about the precise relationship between $B_i^\varphi\psi$ and $[\varphi]B_i\psi$. *Prima Facie*, the two statements seem to express the same thing; and, in fact, they are equivalent provided ψ is a *ground formula* (i.e., does not contain any modal operators). However, consider state w_1 in the following epistemic-doxastic model:
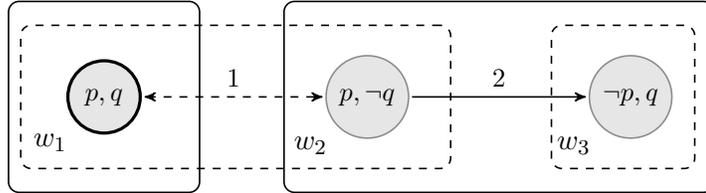
In this model, the solid lines represent agent 2's hard and soft information (the box is 2's hard information $\sim_2$ and the arrow represent 2's soft information $\preceq_2$) while the dashed lines represent 1's hard and soft information. (Reflexive arrows are not drawn to keep down the clutter in the picture.) Note that at state w_1, agent 2 *knows* p and q (eg., $w_1 \models K_2(p \wedge q)$), and agent 1 believes p but not q

($w_1 \models B_1 p \wedge \neg B_1 q$). Now, although agent 1 does not *know* that agent 2 knows p, agent 1 does believe that agent 2 believes q ($w_1 \models B_1 B_2 q$). Furthermore, agent 1 maintains this belief *conditional on* p: $w_1 \models B_1^p B_2 q$. However, public announcing the true fact p, removes state w_3 and so we have $w_1 \models [p]\neg B_1 B_2 q$. Thus a belief in ψ conditional on φ is *not* the same as a belief in ψ *after* the public announcement of φ. This point is worth reiterating: the reader is invited to check that $B_i^p(p \wedge \neg K_i p)$ is satisfiable but $[!p]B_i(p \wedge \neg K_i p)$ is not satisfiable. The situation is nicely summarized as follows: "$B_i^\psi \varphi$ says that if agent i would learn φ then she would come to believe that ψ was the case (before the learning)...$[!\varphi]B_i\psi$ says that after learning φ, agent i would come to believe that ψ is the case (in the worlds after the learning)." (Baltag and Smets, 2008b, pg. 2). While a public announcement increases the agents' knowledge about the state of the world by reducing the total number of possibilities, it also reveals inaccuracies agents may have about the *other* agents' information. The example above is also interesting because the announcement of a *true* fact misleads agent 1 by forcing her to drop her belief that agent 2 believes q (cf. van Benthem, 2010a, pg. 182). Nonetheless, we do have a reduction axiom for conditional beliefs:

$$[\varphi]B_i^\psi \chi \leftrightarrow (\varphi \rightarrow B_i^{\varphi \wedge [\varphi]\psi}[\varphi]\chi)$$

What about languages that include group knowledge operators (note that $w_1 \models [p]C_{\{1,2\}}p$)? The situation is much more complex in languages with common knowledge/belief operators. Baltag et al. (1998) proved that the extension of $\mathcal{L}_{EL}$ with common knowledge and public announcement operators is strictly more expressive than with common knowledge alone. Therefore a reduction axiom for formulas of the form $[\varphi]C_G\psi$ does not exist. Nonetheless, a reduction axiom-style analysis is still possible, though the details are beyond the scope of this paper (see van Benthem et al., 2006).

3.1   Two Models of Informational Dynamics

Many different logical systems today describe the dynamics of information over time in a social situation. However, two main approaches can be singled out. The first is exemplified by *epistemic temporal logic* (ETL, Fagin et al., 1995; Parikh and Ramanujam, 1985) which uses linear or branching time models with added epistemic structure induced by the agents' different capabilities for observing events. These models provide a "grand stage" where histories of some social interaction unfold constrained by a *protocol* (cf., item 3. in the previous Section). The other approach is exemplified by *dynamic epistemic logic* (DEL, Gerbrandy, 1999b; Baltag et al., 1998; van Ditmarsch et al., 2007) which describes social interactions in terms of epistemic **event models** (which may occur inside modalities of the language). Similar to the way epistemic models are used to

capture the (hard) information the agents' have about a *fixed* social situation, an **event model** describes the agents' information about which actual events are currently taking place (cf. item 1 in the previous Section). The temporal evolution of the situation is then computed from some initial epistemic model through a process of successive "product updates". In this Section, we demonstrate each approach by formalizing Example 2.8.

Epistemic Temporal Logic. Fix a finite set of agents $\mathcal{A}$ and a (possibly infinite) set of events[17] Σ. A **history** is a finite sequence of events[18] from Σ. We write Σ^* for the set of histories built from elements of Σ. For a history h, we write he for the history h followed by the event e. Given $h, h' \in \Sigma^*$, we write $h \preceq h'$ if h is a prefix of h', and $h \prec_e h'$ if $h' = he$ for some event e.

For example, consider the social interaction described in Example 2.8. There are three participants: Ann (A), Bob (B) and Ann's friend (call him Charles (C)). What are the relevant primitive events? To keep things simple, assume that Ann's talk is either at 2PM or 3PM and initially none of the agents know this. Say, that Ann receives a message stating that her talk is at 2PM (denote this event — Ann receiving a private message saying that her talk is at 2PM — by e_A^{2PM}). Now, after Ann receives the message that the talk is at 2PM, she proceeds to tell her trusted friend Charles that the talk is at 2PM (and that she wants him to inform Bob of the time of the talk without acknowledging that the information can from her — call this event e_C^A), then Charles tells Bob this information (call this event e_B^C). Thus, the history

$$e_A^{2PM} \ e_C^A \ e_B^C$$

represents the sequence of events where "Ann receives a (private) message stating that the talk is at 2PM, Ann tells Charles the talk is at 2PM, then Charles tells Bob the talk is at 2PM". Of course, there are other events that are also relevant to this situation. For one thing, Ann could have received a message stating that her talk is at 3PM (denote this event by e_A^{3PM}). This will be important to capture Bob's uncertainty about whether Ann knows that he knows about the

---

[17]There is a large literature addressing the many subtleties surrounding the very notion of an *event* and when one event *causes* another event (see, for example, Cartwright, 2007). However, for this paper we take the notion of event as primitive. What is needed is that if an event takes place at some time t, then the fact that the event took place can be observed by a relevant set of agents at t. Compare this with the notion of an event from probability theory. If we assume that at each clock tick a coin is flipped exactly once, then "the coin landed heads" is a possible event. However, "the coin landed head more than tails" would not be an event, since it cannot be observed at any one moment. As we will see, the second statement will be considered a *property* of histories, or sequences of events.

[18]To be precise, elements of Σ should, perhaps, be thought of as event *types* whereas elements of a history are event *tokens*.

talk. Furthermore, Charles may learn about the time of the talk independently of Ann (denote these two events by e_C^{2PM}, e_C^{3PM}). So, for example, the history

$$e_A^{2PM} \; e_C^{2PM} \; e_B^C$$

represents the situation where Charles independently learns about the time of the talk and informs Bob.

There are a number of simplifying assumptions that we adopt in this section. They are not crucial for the analysis of Example 2.8, but do simplify the some of the formal details. Since, histories are sequences of (discrete) events, we assume the existence of a global discrete clock (whether the agents have access to this clock is another issue that will be discussed shortly). The length of the history then represents the amount of time that has passed. Note that this implies that we are assuming a finite past with a possibly infinite future. Furthermore, we assume that at each clock tick, or moment, *some* event takes place (which need not be an event that any agent directly observes). Thus, we can include an event e_t (for 'clock tick') which can represent that "Charles does *not* tell Bob that the talk is at 2PM." So the history

$$e_A^{2PM} \; e_C^A \; e_t$$

describes the sequence of events where, after learning about the time of the talk, Ann informs Charles, but Charles does *not* go on to tell Bob that the talk is at 2PM. Once a set of events Σ is fixed, the temporal evolution and moment-by-moment uncertainty of the agents can be described.

Definition 3.2 (ETL Models) Let Σ be a set of events and At a set of atomic propositions. A **protocol** is a set $\mathsf{H} \subseteq \Sigma^*$ closed under non-empty prefixes. An **ETL model** is a tuple $\langle \Sigma, \mathsf{H}, \{\sim_i\}_{i \in \mathcal{A}}, V \rangle$ with H a protocol, for each $i \in \mathcal{A}$, an equivalence relation $\sim_i$ on H and V a valuation function ($V : \mathsf{At} \to 2^{\mathsf{H}}$). ◁

An ETL model describes how the agents' *hard* information evolves over time in some social situation. The protocol describes (among other things) the temporal structure, with h' such that $h \prec_e h'$ representing the point in time after e has happened in h. The relations $\sim_i$ represent the uncertainty of the agents about how the current history has evolved. Thus, $h \sim_i h'$ means that from agent i's point of view, the history h' looks the same as the history h.
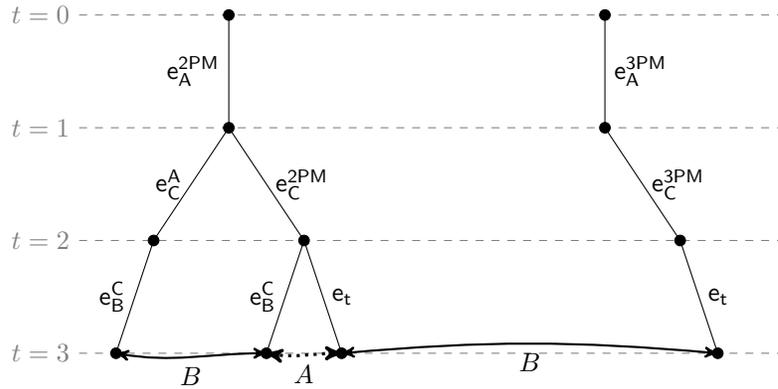
A protocol in an ETL model captures not only the temporal structure of the social situation but also assumptions about the nature of the participants. Typically, a protocol does not include all *possible* ways a social situation could evolve. This allows us to account for the for the *motivation* of the agents. For example in Example 2.8, the history

$$e_A^{3PM} \; e_C^A \; e_B^C$$

describes the sequence of events where Ann learns the talk is at 3PM but tells Charles (who goes on to inform Bob) that the talk is at *2PM*. Of course, given that Ann *wants* Bob to attend her talk, this should not be part of (Ann's) protocol. Similarly, since we assume Charles is trustworthy, we should not include any histories where e_t follows the event e_C^A. Taking into account these underlying assumptions about the motivations (eg. Ann wants Bob to attend the talk) and dispositions (eg. Charles tells the truth and lives up to his promises) of the agents we can drop a number of histories from the protocol shown above. Note that we keep the history

$$e_A^{2PM} \; e_C^{2PM} \; e_t$$

in the protocol, since if Charles learns independently about the time of the talk, then he is under no obligation to inform Bob. In the picture below, we also add some of the uncertainty relations for Ann and Bob (to keep the picture simple, we do not draw the full ETL model). The solid line represents Bob's uncertainty while the dashed line represents Ann's uncertainty. The main assumption is that Bob can only observe the event (e_B^C). So, for example, the histories $h = e_A^{2PM} \; e_C^A \; e_B^C$ and $h' = e_A^{2PM} \; e_C^{2PM} \; e_B^C$ look the same to Bob (i.e., $h \sim_B h'$).[19]

Assumptions about the underlying protocol in an ETL model corresponds to "fixing the playground" where the agents will interact. As we have seen, the protocol not only describes the temporal structure of the situation being modeled, but also any *causal* relationships between events (eg., sending a message must always proceed receiving that message) plus the motivations and dispositions of the participants (eg., liars send messages that they *know* — or believe — to be false). Thus the "knowledge" of agent i at a history h in some ETL model is derived from both i's observational powers (via the $\sim_i$ relation) and i's information about the (fixed) protocol.

---

[19]Again we do not include any reflexive arrows in the picture in order to keep things simple.

We give the bare necessities to facilitate a comparison between ETL and DEL. Different modal languages describe ETL models (see, for example, Hodkinson and Reynolds, ming; Fagin et al., 1995), with 'branching' or 'linear' variants. Let At be a countable set of atomic propositions. The language $\mathcal{L}_{ETL}$ extends the epistemic language $\mathcal{L}_{EL}$ with "event" modalities:

$$p \mid \neg\varphi \mid \varphi \wedge \psi \mid K_i\varphi \mid \langle e \rangle\varphi$$

where $i \in \mathcal{A}$, $e \in \Sigma$ and $p \in$ At. The boolean connectives $(\vee, \rightarrow, \leftrightarrow)$ and the dual modal operators $(L_i, [e])$ are defined as usual. The intended interpretation of '$\langle e \rangle\varphi$' is "after event e (does) take place, φ is true." Formulas are interpreted at histories: Let $\mathcal{H} = \langle \Sigma, \mathsf{H}, \{\sim_i\}_{i \in \mathcal{A}}, V \rangle$ be an ETL model, φ a formula and $h \in \mathsf{H}$, we define $\mathcal{H}, h \models \varphi$ inductively as follows (we only give the modal definitions)

1. $\mathcal{H}, h \models K_i\varphi$ iff for each $h' \in \mathsf{H}$, if $h \sim_i h'$ then $\mathcal{H}, h' \models \varphi$

2. $\mathcal{H}, h \models \langle e \rangle\varphi$ iff there exists $h' \in \mathsf{H}$ such that $h \prec_e h'$ and $\mathcal{H}, h' \models \varphi$

Natural extensions of the $\mathcal{L}_{ETL}$ include group operators (cf. Section **??**) and more expressive temporal operators (e.g., arbitrary future or past modalities).

Dynamic Epistemic Logic. An alternative account of interactive dynamics was elaborated by Gerbrandy (1999b); Baltag et al. (1998); van Benthem (2002); van Benthem et al. (2006) and others. From an initial epistemic model, temporal structure evolves as explicitly triggered by complex informative events.

Returning to our running example (Example 2.8), initially we assume that none of the agents knows the time of Ann's talk. Let P be the atomic proposition "Ann's talk is at 2PM." Whereas an ETL model describes the agents' information at all moments, **event models** are used to build new epistemic models as needed.

Definition 3.3 (Event Model) An **event model** is a tuple $\langle S, \{\longrightarrow_i\}_{i \in \mathcal{A}}, \mathsf{pre} \rangle$, where S is a nonempty set of **primitive events**, for each $i \in \mathcal{A}$, $\longrightarrow_i \subseteq S \times S$ and $\mathsf{pre} : S \to \mathcal{L}_{EL}$ is the **pre-condition function**.   ◁

Given two primitive events e and f, the intuitive meaning of $e \longrightarrow_i f$ is "if event e takes place then agent i *thinks* it is event f" Event models then describe an "epistemic event". In Example 2.8 the first event is Ann receiving a private message that the talk is at 2PM. This can be described by a simple event model with two primitive events e (with precondition P) and f (with precondition $\top$: f is the "skip" event),
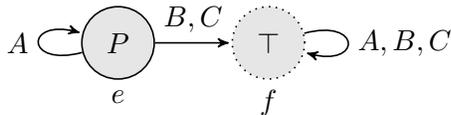
Thus, initially Ann observes the actual event e (and so, learning that P is true) while Bob an Charles observe a skip event (and so, their information does not change). What is the effect of this event on the initial situation (where no one knows the time of the talk)? Intuitively, it is not hard to see that after this event, Ann knows that P while Bob and Charles are still ignorant of P *and the fact that Ann knows P*. That is, incorporating this event into the initial epistemic model should yield (for simplicity we only draw Ann and Bob's uncertainty relations):
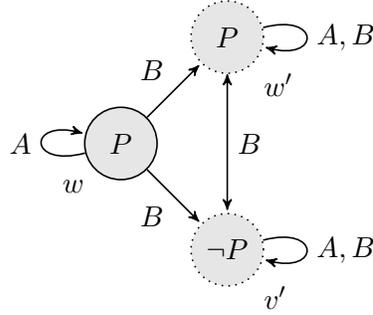

The following definition gives a general procedure for constructing a new epistemic model from a given epistemic model and an event model.

Definition 3.4 (Product Update) The **product update** $\mathcal{M} \otimes \mathcal{E}$ of an epistemic model $\mathcal{M} = \langle W, \{R_i\}_{i \in \mathcal{A}}, V \rangle$ and event model $\mathcal{E} = \langle S, \{\longrightarrow_i\}_{i \in \mathcal{A}}, \mathsf{pre} \rangle$ is the epistemic model $\langle W', R'_i, V' \rangle$ with

1. $W' = \{(w, e) \mid w \in W, e \in S \text{ and } \mathcal{M}, w \models \mathsf{pre}(e)\}$,

2. $(w, e)R'_i(w', e')$ iff wR_iw' in $\mathcal{M}$ and $e \longrightarrow_i e'$ in $\mathcal{E}$, and

3. For all $P \in \mathsf{At}$, $(s, e) \in V'(P)$ iff $s \in V(P)$ ◁

We illustrate this construction using our running example. The main event in Example 2.8 is "Charles telling Bob (without Ann present) that Ann's talk is at 2PM". This can be described using the following event model (again only the Ann and Bob relations will be drawn): Ann is aware of the actual event taking place while Bob thinks the event is a private message to himself.

25
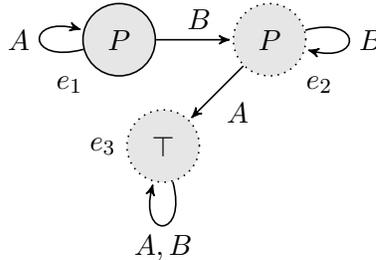
As in the previous section, there are implicit assumptions here about the motivations and dispositions of the agents. Thus, even though Ann is not present during the actual event[20], she *trusts* that Charles will honestly tell Bob that the talk is at 2PM (without revealing he received the information from her). This explains why in the above event model, $e_1 \longrightarrow_A e_1$. Starting from a slightly modified epistemic model from the one given above (where Bob now knows that Ann knows *whether* the talk is at 2PM), using Definition 3.4, we can *calculate* the effect of the above event model as follows:
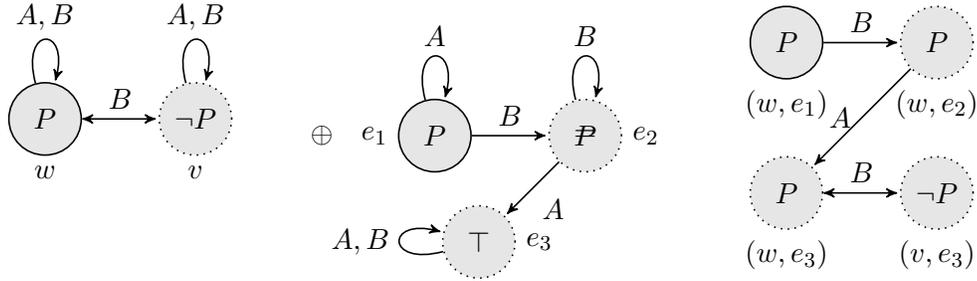
Again, for simplicity, not all the reflexive arrows are drawn.

Finally, a few comments about syntactic issues. The language $\mathcal{L}_{DEL}$ extends $\mathcal{L}_{EL}$ with operators $\langle \mathcal{E}, e \rangle$ for each pair of event models $\mathcal{E}$ and event e in the domain of $\mathcal{E}$. Truth is defined as usual: We only give the typical DEL modalities:

$$\mathcal{M}, w \models \langle \mathcal{E}, e \rangle \varphi \text{ iff } \mathcal{M}, w \models \mathsf{pre}(e) \text{ and } \mathcal{M} \otimes \mathcal{E}, (w, e) \models \varphi$$

Remark 3.5 *We conclude by noting that the public announcement of the previous Section is a special case of Definition 3.3. Given a formula $\varphi \in \mathcal{L}_{EL}$, the public announcement is the event model $\mathcal{E}_\varphi = \langle \{e\}, \{\longrightarrow_i\}_{i \in \mathcal{A}}, \mathsf{pre} \rangle$ where for each $i \in \mathcal{A}$, $e \longrightarrow_i e$ and $\mathsf{pre}(e) = \varphi$. As the reader is invited to verify, the product update of an epistemic model $\mathcal{M}$ with a public announcement event $\mathcal{E}_\varphi$ ($\mathcal{M} \otimes \mathcal{E}_\varphi$) is (isomorphic) to the model $\mathcal{M}^\varphi$ of Definition 3.1.*

3.2 Varieties of Informational Change

The dynamic models discussed in the previous Section focus on the agents' observational powers and procedural information. The assumption is that precisely how an agent incorporates new information depends on only two factors: what the agent has observed and the underlying protocol (which is typically assumed to be common knowledge). To what degree the agent *trusts* the source of the information is not taken into account (cf. item 2 from Section 3). In this section,

---

[20]Of course, we must assume that she knows precisely *when* Charles will meet with Bob.

we show how to extend our logical analysis this aspect of model this aspect of the agents' extend our logical analysis with this information. We only have the space here for some introductory remarks: see (van Benthem, 2010a, Chapter 7) and (Baltag and Smets, 2009) for more extensive discussions.

The general problem we focus on in the Section is how to incorporate the evidence that φ is true into an epistemic-doxastic model $\mathcal{M}$. The approach taken so far is *eliminate* all worlds inconsistent with (each agent's observation of) the evidence that φ is true. (This may reveal *more* than φ is true given an underlying protocol). However, not all sources of evidence are 100% reliable opening the door to the possibility that later evidence may contradict earlier evidence. Consider the situation for agent i's point-of-view: Abstractly, the problem is how to define a *new ordering* over i's (hard) information cell given i's current soft information (represented as a total ordering over the set of states that i considers possible) and the incoming information represented as the *truth set* of some formula φ:
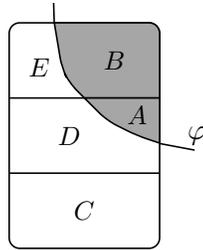
Rather than *removing* the states inconsistent with φ (in the above case, this would be the states in the set $C \cup D \cup E$), the goal is to *rearrange* the states in such a way that φ is believed. In the above example, this means that at least the set A should become the new minimal set. But there is a variety ways to fill in the rest of the order with each way corresponding to a different "policy" the agent takes towards the incoming information (Rott, 2006). We only have space here to discuss two of these policies (both have been widely discussed in the literature, see for example, van Benthem, 2010a, Chapter 7). The first captures the situation where the agent only tentatively accepts the incoming information φ by making the best φ the new minimal set and keeping the rest of the ordering the same. Before formally defining the policy we need some notation: given an epistemic-doxastic model $\mathcal{M}$, let $best_i(\varphi, w) = Min_{\preceq_i}([w]_i \cap \{x \mid \mathcal{M}, x \models \varphi\})$ denote the best φ worlds at state w.

Definition 3.6 (Conservative Upgrade) Given an epistemic-doxastic model $\mathcal{M} = \langle W, \{\sim_i\}_{i \in \mathsf{Agt}}, \{\preceq_i\}_{i \in \mathsf{Agt}}, V \rangle$ be an epistemic-doxastic model and a formula φ, the *conservative* upgrade of $\mathcal{M}$ with φ is the model $\mathcal{M}^{\uparrow\varphi} = \langle W^{\uparrow\varphi}, \{\sim_i^{\uparrow\varphi}$

$\}_{i\in\mathsf{Agt}}, \{\preceq_i^{\uparrow\varphi}\}_{i\in\mathsf{Agt}}, V^{\uparrow\varphi}\rangle$ with $W^{\uparrow\varphi} = W$, for each i, $\sim_i^{\uparrow\varphi}=\sim_i$, $V^{\uparrow\varphi} = V$ and for all $i \in \mathsf{Agt}$ and $w \in W^{\uparrow\varphi}$ we have:

1. If $v \in best_i(\varphi, w)$ then $v \preceq_i^{\uparrow\varphi} x$ for all $x \in [w]_i$, and

2. for all $x, y \in [w]_i - best_i(\varphi, w)$, $x \preceq_i^{\uparrow\varphi} y$ iff $x \preceq_i y$. $\quad\triangleleft$

In the above picture a conservative upgrade with φ results in the new ordering $A \preceq_i C \preceq_i D \preceq_i B \cup E$. A logical analysis of this type of information change includes formulas of the form $[\uparrow_i\varphi]\psi$ intended to mean "after i's conservative upgrade of φ, ψ is true" and interpreted as follows: $\mathcal{M}, w \models [\uparrow_i\varphi]\psi$ iff $\mathcal{M}^{\uparrow_i\varphi}, w \models \psi$. We also have reduction axioms for conditional beliefs:

$$[\uparrow\varphi]B^\psi\chi \leftrightarrow (B^\varphi\neg[\uparrow\varphi]\psi \wedge B^{[\uparrow\varphi]\psi}[\uparrow\varphi]\chi) \vee (\neg B^\varphi\neg[\uparrow\varphi]\psi \wedge B^{\varphi\wedge[\uparrow\varphi]\psi}[\uparrow\varphi]\chi)$$

(We leave out the i subscripts to make the formula easier to read). The reader is invited to check the validity of the above axiom. The second policy we introduce here models a more "radical" change to the agent's plausibility ordering: *all* φ worlds are moved ahead of all other worlds. Thus, rather than focusing on only the best φ worlds, the agent shifts *all* φ worlds consistent with i's current information: let $[\![\varphi]\!]_i^w = \{x \mid \mathcal{M}, x \models \varphi\} \cap [w]_i$ denote this set of φ worlds:

Definition 3.7 (Radical Upgrade) Given an epistemic-doxastic model $\mathcal{M} = \langle W, \{\sim_i\}_{i\in\mathsf{Agt}}, \{\preceq_i\}_{i\in\mathsf{Agt}}, V\rangle$ be an epistemic-doxastic model and a formula φ, the *conservative* upgrade of $\mathcal{M}$ with φ is the model $\mathcal{M}^{\Uparrow\varphi} = \langle W^{\Uparrow\varphi}, \{\sim_i^{\Uparrow\varphi}\}_{i\in\mathsf{Agt}}, \{\preceq_i^{\Uparrow\varphi}\}_{i\in\mathsf{Agt}}, V^{\Uparrow\varphi}\rangle$ with $W^{\Uparrow\varphi} = W$, for each i, $\sim_i^{\Uparrow\varphi}=\sim_i$, $V^{\Uparrow\varphi} = V$ and for all $i \in \mathsf{Agt}$ and $w \in W^{\Uparrow\varphi}$ we have:

1. for all $x \in [\![\varphi]\!]_i^w$ and $y \in [\![\neg\varphi]\!]_i^w$, set $x \preceq_i^{\Uparrow\varphi} y$,

2. for all $x, y \in [\![\varphi]\!]_i^w$, set $x \preceq_i^{\Uparrow\varphi} y$ iff $x \preceq_i y$, and

3. for all $x, y \in [\![\neg\varphi]\!]_i^w$, set $x \preceq_i^{\Uparrow\varphi} y$ iff $x \preceq_i y$. $\quad\triangleleft$

In the above picture a conservative upgrade with φ results in the new ordering $A \preceq_i B \preceq_i C \preceq_i D \preceq_i E$. A logical analysis of this type of information change includes formulas of the form $[\Uparrow_i\varphi]\psi$ intended to mean "after i's radical upgrade of φ, ψ is true" and interpreted as follows: $\mathcal{M}, w \models [\Uparrow_i\varphi]\psi$ iff $\mathcal{M}^{\Uparrow_i\varphi}, w \models \psi$. As the reader is invited to check, the conservative upgrade is a special case of this radical upgrade: the conservative upgrade of φ at w is the radical upgrade of $best_i(\varphi, w)$. In fact, both of these operations can be seen as instances of a more general *lexicographic update* (cf. van Benthem, 2010a, Chapter 7). In fact, the above reduction axiom for conservative upgrade can be *derived* from

the following reduction axiom for radical upgrade: (again, we leave out the i subscripts to make the formula easier to read)

$$[\Uparrow\varphi]B^\psi\chi \leftrightarrow (L(\varphi \wedge [\Uparrow\varphi]\psi) \wedge B^{\varphi \wedge [\Uparrow\varphi]\psi}[\Uparrow\varphi]\chi) \vee (\neg L(\varphi \wedge [\Uparrow\varphi]\psi) \wedge B^{[\Uparrow\varphi]\psi}[\Uparrow\varphi]\chi)$$

4   Conclusions

Agents are faced with many diverse tasks as they interact with the environment and one another. At certain moments, they must *react* to their (perhaps surprising) observations while at other moments they must be *proactive* and choose to perform a specific action. One central underlying assumption is that "rational" agents obtain what they want via the implementation of (successful) *plans* (cf. Bratman, 1987). And this implementation often requires, among other things, representation of various informational attitudes of the other agents involved in the social interaction. In social situations there are many (sometimes competing) *sources* for these attitudes: for example, the type of "communicatory event" (public announcement, private announcement), the disposition of the other participants (liars, truth-tellers) and other implicit assumptions about procedural information (reducing the number of possible histories). This naturally leads to different notions of "knowledge" and "belief" that drive social interaction.

An overarching theme in is this paper is that during a social interaction, the agents' "knowledge" and "beliefs" both influence *and* are shaped by the *social* events. The following example (taken from Pacuit et al., 2006) illustrates this point. Suppose that Uma is a physician whose neighbour Sam is ill and consider the following cases

Case 1: . Uma does not know and has not been informed. Uma has no obligation (as yet) to treat her neighbour.

Case 2: The neighbour's daughter Ann comes to Uma's house and tells her. Now Uma does have an obligation to treat Sam, or perhaps call in an ambulance or a specialist.

In both of these cases, the issue of an obligation arises. This obligation is circumstantial in the sense that in other situations, the obligation might not apply. If Sam is ill, Uma needs to know that he is ill, and the nature of the illness, but not where Sam went to school. Thus an agent's obligations are often dependent on what the agent knows, and indeed one cannot reasonably be expected to respond to a problem if one is not aware of its existence. This, in turn, creates a secondary obligation on Ann to inform Uma that her father is ill.

Based on the logical framework discussed in Section 3.1 and (Horty, 2001), Pacuit et al. (2006) develop a logical framework that formalizes the reasoning of

Uma and Ann in the above examples. It is argued that this reasoning is shaped by the assumption that Uma and Ann's preferences are aligned (i.e., both want Sam to get better). For example, Ann will not be under any obligation to tell Uma that her father is ill, if Ann justifiably believes that Uma would not treat her father even if she knew of his illness. Thus, in order for Ann to *know* that she has an obligation to tell Uma about her father's illness, Ann must *know* that "Uma will, in fact, treat her father (in a reasonable amount of time) upon learning of his illness". More formally, in all the histories that Ann currently considers possible, the event where her father is treated for his illness is always preceded by the event where she tells Uma about his illness. That is, the histories where Uma learns of Sam's illness but does not treat him are not part of the protocol. Similar reasoning is needed for Uma to derive that she has an obligation to treat Sam. Obviously, if Uma has a good reason to believe that Ann always lies about her father being ill, then she is under no obligation to treat Sam. See (Pacuit et al., 2006) for a formal treatment of these examples.

This paper surveyed a number of logical systems that model the reasoning and dynamic processes that govern many of our social interactions. This is a well-developed area attempting to balance sophisticated logical analysis with philosophical insight. Furthermore, the logical systems discussed in this paper have been successfully used to sharpen the analysis of key epistemic issues in a variety of disciplines. However, they represent only one component of a logical analysis of *rational interaction*. Indeed, as the above example illustrates, a comprehensive account of rational interaction cannot always be isolated from other aspects of rational agency and social interaction (such as the agents' motivational attitudes or social obligations).

References

Aumann, R. (1999a). Interactive epistemology I: Knowledge. *International Journal of Game Theory 28*, 263–300.

Aumann, R. (1999b). Interactive epistemology II: Probability. *International Journal of Game Theory 28*, 301 – 314.

Baltag, A., L. Moss, and S. Solecki (1998). The logic of common knowledge, public announcements and private suspicions. In I. Gilboa (Ed.), *Proceedings of the 7th Conference on Theoretical Aspects of Rationality and Knowledge (TARK 98)*, pp. 43 – 56.

Baltag, A. and S. Smets (2006a). Conditioanl doxastic models: A qualitative approach to dynamic belief revision. In *Electornic notes in theoretical computer science*, Volume 165, pp. 5 – 21. Springer.

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

Baltag, A. and S. Smets (2007). From conditional probability to the logic of doxastic actions. In *TARK '07: Proceedings of the 11th conference on Theoretical aspects of rationality and knowledge*, pp. 52–61. ACM.

Baltag, A. and S. Smets (2008a). The logic of conditional doxastic actions. In R. van Rooij and K. Apt (Eds.), *New Perspectives on Games and Interaction*. Texts in Logic and Games, Amsterdam University Press.

Baltag, A. and S. Smets (2008b). A qualitative theory of dynamic interactive belief revision. In G. Bonanno, W. van der Hoek, and M. Wooldridge (Eds.), *Logic and the Foundation of Game and Decision Theory (LOFT7)*, Volume 3 of *Texts in Logic and Games*, pp. 13–60. Amsterdam University Press.

Baltag, A. and S. Smets (2009). ESSLLI 2009 course: Dynamic logics for interactive belief revision. Slides available online at `http://alexandru.tiddlyspot.com/#%5B%5BESSLLI09%20COURSE%5D%5D`.

Barwise, J. (1988). Three views of common knowledge. In *TARK '88: Proceedings of the 2nd conference on Theoretical aspects of reasoning about knowledge*, San Francisco, CA, USA, pp. 365–379. Morgan Kaufmann Publishers Inc.

Battigalli, P. and M. Siniscalchi (2002). Strong belief and forward induction reasoning. *Journal of Economic Theory 105*, 356 – 391.

Blackburn, P., M. de Rijke, and Y. Venema (2002). *Modal Logic*. Campbridge University Press.

Board, O. (2004). Dynamic interactive epistemology. *Games and Economic Behavior 49*, 49 – 80.

Bonanno, G. (1996). On the logic of common belief. *Mathematical Logical Quarterly 42*, 305 – 311.

Boutilier, C. (1992). *Conditional Logics for Default Reasoning and Belief Revision*. Ph. D. thesis, University of Toronto.

Brandenburger, A. (2007). The power of paradox: some recent developments in interactive epistemology. *International Journal of Game Theory 35*, 465–492.

Bratman, M. (1987). *Intention, Plans and Practical Reason*. London: Harvard University Press.

Cartwright, N. (2007). *Hunting Causes and Using Them: Approaches in Philosophy and Economics*. Cambridge University Press.

Chwe, M. S.-Y. (2001). *Rational Ritual*. Princeton University Press.

Clark, H. and C. R. Marshall (1981). Definite reference and mutual knowledge. In Joshi, Webber, and Sag (Eds.), *Elements of Discourse Understanding*. Cambridge University Press.

Cubitt, R. P. and R. Sugden (2003). Common Knowledge, Salience and Convention: A Reconstruction of David Lewis' Game Theory. *Economics and Philosophy 19*(2), 175–210.

Egré, P. and D. Bonnay (2009). Inexact knowledge with introspection. *Journal of Philosophical Logic 38*(2), 179 – 228.

Engelfriet, J. and Y. Venema (1998). A modal logic of information change. In *Proceedings of TARK 1998*, pp. 125–131.

Fagin, R. and J. Halpern (1994). Reasoning about knowledge and probability. *Journal of the ACM 41*(2), 340 – 367.

Fagin, R., J. Halpern, and N. Megiddo (1990). A logic for reasoning about probabilities. *Information and Computation 87*(1), 78 – 128.

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

Gabbay, D. and J. Woods (Eds.) (2006). *The Handbook of the History of Logic: Logic and the Modalities in the Twentieth Century*, Volume 7. Elsevier.

Gerbrandy, J. (1999a). *Bisimulations on Planet Kripke*. Ph. D. thesis, University of Amsterdam.

Gerbrandy, J. (1999b). *Bisimulations on Planet Kripke*. Ph. D. thesis, Institute for Logic, Language and Computation (DS-1999-01).

Gochet, P. and P. Gribomont (2006). Epistemic logic. In *Gabbay and Woods (2006)*. Elsevier.

Goldblatt, R. (2006). Mathematical modal logic: A view of its evolution. In *Gabbay and Woods (2006)*. Elsevier.

Goldman, A. (1976). What is justified belief? In G. Pappas (Ed.), *Justification and Kowledge*. D. Reidel.

Halpern, J. (1996). Should knowledge entail belief? *Journal of Philosophical Logic 25*(5), 483 – 494.

Halpern, J. (1999). Set-theoretic completeness for epistemic and conditional logic. *Annals of Mathematics and Artificial Intelligence 26*, 1 – 27.

Halpern, J. (2003). *Reasoning about Uncertainty*. The MIT Press.

Halpern, J. and G. Lakemeyer (1995). Levesque's axiomatization of only knowing is incomplete. *Artificial Intelligence 74*(2), 381 – 387.

Halpern, J. and G. Lakemeyer (2001). Multi-agent only knowing. *Journal of Logic and Computation 11*, 41 – 70.

Halpern, J. and Y. Moses (1983). Knowledge and common knowledge in a distributed environment. *ACM-PODC*, 50 – 61.

Halpern, J. and R. Pass (2009). A logical characterization of iterated admissibility. In *TARK '09: Proceedings of the 12th Conference on Theoretical Aspects of Rationality and Knowledge*, New York, NY, USA, pp. 146–155. ACM.

Halpern, J. and R. Pucella (2003). Modeling adversaries in a logic for security protocol analysis. In *Formal Aspects of Security*.

Harman, G. (1986). *Change in View*. MIT Press.

Harsanyi, J. C. (1967). Games with incompletet information played by bayesian players parts I-III. *Management Sciences 14*.

Heifetz, A. and P. Mongin (2001). Probability logic for type spaces. *Games and Economic Behavior 35*, 31 – 53.

Hendricks, V. (2005). *Mainstream and Formal Epistemology*. Cambridge University Press.

Hendricks, V. (2006). Editor, special issue: "8 bridges between formal and mainstream epistemology". *Philosophical Studies 128*(1), 1–227.

Hendricks, V. and J. Symons (2006). Where's the bridge? epistemology and epistemic logic. *Philosophical Studies 128*, 137 – 167.

Herzig, A. (2003). Modal probability, belief, and actions. *Fundam. Inf. 57*(2-4), 323–344.

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

Hintikka, J. (2005). *Knowledge and Belief: An Introduction to the Logic of the Two Notions (with an Introduction by V. Hendricks and J. Symons).* King's College Publications.

Hodkinson, I. and M. Reynolds (forthcoming). Temporal logic. In *Handbook of Modal Logic.*

Horty, J. (2001). *Agency and Deontic Logic.* Oxford University Press.

Humberstone, L. (1987). The modal logic of 'all and only'. *Notre Dame Journal of Formal Logic 28*, 177 – 188.

Lamarre, P. and Y. Shoham (1994). Knowledge, certainty, belief and conditionalisation. In *Proceedings of the International Conference on Knoweldge Representation and Reasoning*, pp. 415 – 424.

Levesque, H. J. (1990). All I know: a study in autoepistemic logic. *Artificial Intelligence 42*(3), 263 – 309.

Lewis, D. (1969). *Convention.* Harvard University Press.

Lismont, L. and P. Mongin (1994). On the logic of common belief and common knowledge. *Theory and Decision 37*(1), 75 – 106.

Lismont, L. and P. Mongin (2003). Strong Completeness Theorems for Weak Logics of Common Belief. *Journal of Philosophical Logic 32*(2), 115 – 137.

Liu, F. (2008). *Changing for the better: Preference dynamics and agent diversity.* Ph. D. thesis, Institute for logic, lanugage and computation (ILLC).

Lomuscio, A. and M. Ryan (1997). On the relation between interpreted systems and kripke models. In *Proceedings of the AI97 Workshop on Theoretical and Practical Foundation of Intelligent Agents and Agent-Oriented Systems*, Volume LNCS 1441.

Nozick, R. (1981). Knowledge and skepticism. In *Philosophical Investigations*, pp. 172 – 185. The MIT Press.

Pacuit, E. (2007). Some comments on history based structures. *Journal of Applied Logic Applied Logic 5*(4), 613–624.

Pacuit, E., R. Parikh, and E. Cogan (2006). The logic of knowledge based obligation. *Knowledge, Rationality and Action: A Subjournal of Synthese 149*(2), 311 – 341.

Parikh, R. (2002, September). Social software. *Synthese 132*, 187–211.

Parikh, R. (2003). Levels of knowledge, games, and group action. *Research in Economics 57*, 267 — 281.

Parikh, R. (2008). Sentences belief and logical omniscience or what does deduction tell us? *Review of Symbolic Logic 1*(4), 514 – 529.

Parikh, R. and R. Ramanujam (1985). Distributed processes and the logic of knowledge. In *Logic of Programs*, Volume 193 of *Lecture Notes in Computer Science*, pp. 256 – 268. Springer.

Parikh, R. and R. Ramanujam (2003). A knowledge based semantics of messages. *Journal of Logic, Language and Information 12*, 453 – 467.

Plaza, J. (1989). Logics of public communications. In M. L. Emrich, M. S. Pfeifer, M. Hadzikadic, and Z. Ras (Eds.), *Proceedings, 4th International Symposium on Methodologies for Intelligent Systems*, pp. 201–216 (republished as (Plaza, 2007)).

Plaza, J. (2007). Logics of public communications. *Synthese: Knowledge, Rationality, and Action 158*(2), 165 – 179.

Ramanujam, R. and S. Suresh (2005). Deciding knowledge properties of security protocols. In *Proceedings of Theoretical Aspects of Rationality and Knowledge*, pp. 219–235.

Roelofsen, F. (2007). Distributed knowledge. *Journal of Applied Non-Classical Logics 17*(2), 255 – 273.

Rott, H. (2006). Shifting priorities: Simple representations for 27 itereated theory change operators. In H. Lagerlund, S. Lindström, and R. Sliwinski (Eds.), *Modality Matters: Twenty-Five Essays in Honour of Krister Segerberg*, Volume 53 of *Uppsala Philosophical Studies*, pp. 359 – 384.

Samuelson, L. (2004). Modeling knowledge in economic analysis. *Journal of Economic Literature 57*, 367 – 403.

Schwitzgebel, E. (2008). Belief. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Fall 2008 ed.).

Shah, N. and J. Velleman (2005). Doxastic deliberation. *The Philosophical Review 114*(4), 497 – 534.

Shoham, Y. and K. Leyton-Brown (2009). *Multiagent Systems: Algorithmic, Game-Theoretic, and Logical Foundations*. Cambridge University Press.

Sosa, E., J. Kim, J. Fantl, and M. McGrath (Eds.) (2008). *Epistemology: An Anthology.* Wily-Blackwell.

Stalnaker, R. (1991). The problem of logical omniscience I. *Synthese 89*, 425 – 440.

Stalnaker, R. (1994). On the evaluation of solution concepts. *Theory and Decision 37* (42).

Stalnaker, R. (1999). Extensive and strategic forms: Games and models for games. *Research in Economics 53*, 293 – 319.

Stalnaker, R. (2006). On logics of knowledge and belief. *Philosophical Studies 128*, 169 – 199.

van Benthem, J. (2002). 'One is a lonely number': on the logic of communication. In Z. Chatzidakis, P. Koepke, and W. Pohlers (Eds.), *Logic Colloquium '02*, pp. 96 – 129. ASL and A. K. Peters. Available at http://staff.science.uva.nl/~johan/Muenster.pdf.

van Benthem, J. (2004). Dynamic logic for belief revision. *Journal of Applied Non-Classical Logics 14* (2), 129 – 155.

van Benthem, J. (2005). Rational animals: What is 'KRA'? invited lecture Malaga ESSLLI Summer School 2006.

van Benthem, J. (2006). Modal frame correspondences and fixed-points. *Studia Logica 83* (1-3), 133–155.

van Benthem, J. (2010a). *Logical Dynamics of Information and Interaction.* Cambridge University Press.

van Benthem, J. (2010b). *Modal Logic for Open Minds.* CSLI Publications.

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

van Benthem, J. and D. Sarenac (2004). The geometry of knowledge. In *Aspects of Universal Logic,*, Volume 17, pp. 1–31.

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

van der Hoek, W., B. van Linder, and J.-J. Meyer (1999). Group knowledge is not alwasy distributed (neither is it always implicity). *Mathematical Social Sciences 38*, 215 – 240.

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

Vanderschraaf, P. and G. Sillari (2009). Common knowledge. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Spring 2009 ed.).

von Wright, G. H. (1951). *An Essay in Modal Logic.* North-Holland, Amsterdam.

Williamson, T. (2000). *Knowledge and its Limits.* Oxford University Press.