Dynamic Epistemic Logic
Part I: Modeling Knowledge and Belief

Eric Pacuit

April 3, 2012

1 Introduction

The first to propose a (modal) logic of knowledge and belief was Jaako Hintikka in his seminal book *Knowledge and Belief: An Introduction to the Logic of the Two Notions*, published in 1962. However, the general study of formal semantics for knowledge and belief (and their logic) really began to flourish in the 1990s with fundamental contributions from computer scientists (Fagin et al., 1995; Meyer and van der Hoek, 1995) and game theorists (Aumann, 1999; Bonanno and Battigalli, 1999). As a result, the field of Epistemic Logic developed into an interdisciplinary area focused on explicating epistemic issues in, for example, game theory (Brandenburger, 2007), economics (Samuelson, 2004), computer security (Halpern and Pucella, 2003; Ramanujam and Suresh, 2005), distributed and multiagent systems (Halpern and Moses, 1990; van der Hoek and Wooldridge, 2003), and the social sciences (Parikh, 2002; Gintis, 2009). Nonetheless, the field has not loss touch with its philosophical roots: See (Holliday, 2012; Egré, 2011; Stalnaker, 2006; van Benthem, 2006; Sorensen, 2002) for logical analyses aimed at mainstream epistemology.

Inspired, in part, by issues in these different "application" areas, a rich repertoire of epistemic and doxastic attitudes have been identified and analyzed in the epistemic logic literature. The challenge for a logician is not to argue that one particular account of belief or knowledge is *primary*, but, rather, to explore the logical space of definitions and identify interesting relationships between the different notions. I do not have the space for a comprehensive overview of the many different logics of knowledge and belief. Instead, I will confine the discussion to three logical frameworks needed for the survey of *dynamic* logics of knowledge an belief found in part 2.

The formal models introduced in this article can be broadly described as "possible worlds models," familiar in much of the philosophical logic literature.

Let $\mathcal{A}$ be a finite set of agents and At a (finite or countable) set of atomic sentences. Elements $p \in \mathsf{At}$ are intended to describe *basic properties* of the situation being modeled, such as "it is raining" or "the red card is on the table". Setting aside any conceptual difficulties surrounding the use of "possible worlds"[1], a non-empty set W of *states*, or *possible worlds*, will be used to represent different possible *scenarios* or *states of affairs*. A **valuation function** associates with each atomic proposition p a set of states where p is true: $V : \mathsf{At} \to \wp(W)$.

In this article, I will introduce different logical systems that have been used to reason about the knowledge and beliefs of a group of agents. My focus is on the underlying intuitions and a number of illustrative examples rather than the general logical theory of these systems (eg., issues of decidability, complexity, completeness and definability). Although I will include pointers to this more technical material throughout this article, interested readers are encouraged to consult two recent textbooks on modal logic (Blackburn et al., 2002; van Benthem, 2010b) for a systematic presentation.

2 Knowledge

The model in this section is based on a very simple (and ubiquitous) idea: An agent is *informed* that φ is true when φ is true throughout the agent's current range of possibilities. In an epistemic model, the agents' "range of possibilities" are described in terms of *epistemic accessibility relations* R_i on the set of states W (i.e., $R_i \subseteq W \times W$).

Definition 2.1 (Epistemic Model) Let $\mathcal{A}$ be a finite set of agents and At a (finite or countable) set of atomic propositions. An **epistemic model** is a tuple $\langle W, \{R_i\}_{i\in\mathcal{A}}, V\rangle$ where $W \neq \emptyset$, for each $i \in \mathcal{A}$, $R_i \subseteq W \times W$ is a relation, and $V : \mathsf{At} \to \wp(W)$ is a valuation function. ◁

Epistemic models are often used to describe what the agents *know* about the situation being modeled. A simple propositional modal language is often used to make this precise: Let $\mathcal{L}_K$ 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 $p \in \mathsf{At}$ (the set of atomic propositions). The additional propositional connectives $(\to, \leftrightarrow, \vee)$ are defined as usual and the dual of K_i, denoted L_i, is

[1] Note that there is nothing "metaphysical" attached to the term "possible world" here. Nonetheless, the basic modeling choices are not without controversy. The "conceptual difficulties" are related to issues raised by Jon Barwise and John Perry in their development of *situation semantics* (Barwise and Perry, 1983) and issues surrounding the underlying assumption of logical omniscience (Stalnaker, 1991; Parikh, 2005).

defined as follows: $L_i\varphi := \neg K_i \neg\varphi$. Following the standard usage in the epistemic logic and game theory literature, the intended interpretation of $K_i\varphi$ is "agent i knows that φ". An alternative interpretation (which is more natural in many situations) is "agent i is informed that φ is true".

Each state of an epistemic model represents a possible scenario which can be described in the formal language given above: If $\varphi \in \mathcal{L}_K$, I write $\mathcal{M}, w \models \varphi$ provided φ is a correct description of some aspect of the situation represented by w. This can be made precise as follows:

Definition 2.2 (Truth for $\mathcal{L}_K$) Let $\mathcal{M} = \langle W, \{R_i\}_{i\in\mathcal{A}}, V\rangle$ 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 wR_iv then $\mathcal{M}, v \models \varphi$ $\triangleleft$

It is important to recognize that $K_i\varphi$ *describes* a set of states in which agent i "knows" that φ (or agent i is "informed" that φ). It does not *explain* why or how agent i came to the conclusion that φ is true. Indeed, there are different ways to understand exactly how these models represent the agents' *knowledge* (and later, beliefs). The crucial interpretative step is to explain what it means for a state v to be accessible for agent i from state w.

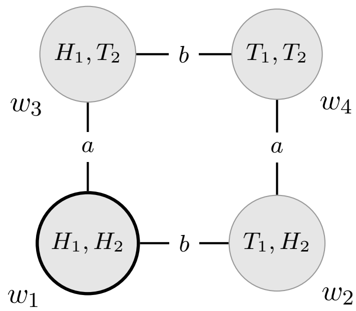
The first, and most neutral, interpretation of wR_iv is *everything that agent i knows in state w is true in state v*. Under this interpretation, the agents' knowledge is not *defined* in terms of more primitive notions. Instead, an epistemic model represents the "implicit consequences" of what the agents are assumed to *know* in the situation being modeled.

A second use of epistemic models is to formalize a substantive theory of knowledge. In this case, the agents' knowledge is *defined* in terms of more primitive concepts (which are built into the definition of the accessibility relation). For instance, suppose that wR_iv means that *agent i has the same experiences and memories in both w and v* (Lewis, 1996). Another example is wR_iv means that *at state w, agent i cannot* rule-out *state v (according to i's current observations and any other evidence she has obtained).* In both cases, what the agents *know* is defined in terms of more primitive notions (in terms of "sameness of experience and memory" or a "ruling-out" operation).

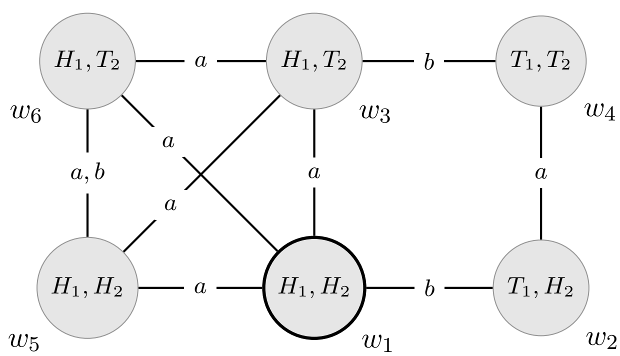
Finally, epistemic models have been used to make precise informal notions of knowledge found in both the computer science and game theory literature. For example, consider a system of processors executing some distributed program. When designing such a distributed program, it is natural to use statements of the form "if processor i *knows* that the message was delivered, then ... " The notion of knowledge being used here can be given a concrete interpretation as follows: A *run*[2] is a complete description of what happens in the system over time. A state, or possible world, is a pair (r, t) where r is a run and t a moment in time. At each possible world, the processors are in a *local state*, which can be given a concrete definition in terms of the values of the variables that the processor has access to. Then, a state (r, t) is accessible from another state (r', t') for processor i provided i is in the same local state in (r, t) and (r', t'). A similar approach can be used to analyze game-theoretic situations. For example, in a poker game, the states are the different distribution of cards. Then, a state v is accessible for player i from state w provided i has the same *information* in both states (eg., i was dealt the same cards in both states). In both cases, the goal is to develop a useful model of knowledge based on a concrete definition of possible worlds and the epistemic accessibility relation rather than a general analysis of "knowledge".

The issues raised above are important conceptual and methodological considerations, and they help us understand the scope of an epistemic analysis using the logical system introduced in this section. However, the general distinctions should not be overstated as they tend to fade when analyzing specific examples. Consider the following running example: Suppose that there are two coins each sitting in different drawers and two agents, Ann (a) and Bob (b). We are interested in describing what Ann and Bob know about the two coins. Of course, there are many facts about the coins that we may want to represent (eg., the exact position of the coins in the drawer, what type of coins are in the drawers: are they nickels? are they quarters?, etc.), but to keep things simple consider four atomic propositions: for $i = 1, 2$, let H_i denote "the coin in drawer i is lying heads up" and T_i denote "the coin in drawer i is lying tails up". Suppose that Ann looked at the coin in drawer 1 and Bob looked at the coin in drawer 2. Given what Ann and Bob have observed, we can represent their knowledge about the coins in the diagram below: Suppose that both coins are lying heads up (so w_1 is the actual state). We draw an edge labeled with a (b) between two states if a (b) cannot distinguish them (based on their observations).

[2]Also called a *history* in (Parikh and Ramanujam, 2003).



The reader is invited to verify that $K_aH_1 \wedge K_bH_2$ is true at state w_1, as expected. However, much more is true. In particular, while Ann does not know which way the coin is facing the the second drawer, she does know that Bob knows whether the coin is laying heads up or tails up (i.e., $K_a(K_bH_2 \vee K_bT_2)$ is true at w_1). Thus, in an epistemic model also describe the agents' *higher-order knowledge*: the information about what the other agents know about each other. The implicit assumption underlying the above model is that the agents correctly observe the face of the coin when they look in the drawer *and* they take it for granted that the other agent correctly perceived the coin. This is a substantive assumption about what the agents know about each other and can be dropped by adding states to the model:



Now, at state w_1 in the above model, Ann considers it possible that Bob does not know the position of the coin in the second drawer (so, $K_a(K_bH_2 \vee K_bT_2)$ is not true at w_1). It is not hard to see that one always finds substantive assumptions in finite structures. Are there models that make no, or at least as few as possible, substantive assumptions? This question has been extensively discussed in the literature on the epistemic foundations of game theory—see the discussion in (Samuelson, 2004) and the discussion and references in (Roy and Pacuit, 2011).

In the above example, the agents' accessibility relation satisfied a number of additional properties. In particular, the relations are reflexive (for each w,

wR_iw), transitive (for each w, v, s if wR_iv and vR_is then wR_is) and Euclidean (for each w, v, s if wR_iv and wR_is then vR_is) (such relations are called *equivalence relations*). Assuming that the agents' accessibility relations are equivalence relations (in the remainder of this article and in part 2 we use the notation $\sim_i$ instead of R_i when the accessibility relations are equivalence relations) is a substantive assumption about the nature of the informational attitude being represented in these models. In fact, by using standard techniques from the mathematical theory of modal logic, I can be much more precise about what properties of knowledge are being assumed. In particular, *modal correspondence theory* rigorously relates properties of the relation in an epistemic model with modal formulas (cf. Blackburn et al., 2002, Chapter 3)[3]. The following table lists some key formulas in the language $\mathcal{L}_K$ 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 drawn 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 (Halpern and Pucella, 2011), the two introspection principles have also been the subject of intense discussion (cf. Williamson, 2000; Egré and Bonnay, 2009). These discussions are fundamental to the theory of knowledge and its formalization, but here I choose to bracket them, and, instead, take epistemic models for what they are: precise descriptions of what the (modeler takes the) agents (to) know about the situation being modeled.

[3] 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 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 an algorithm for finding first-order correspondents for modal formulas with a certain syntactic shape. See (Blackburn et al., 2002, Sections 3.5 - 3.7) for an extended discussion.

3 Adding Belief

Philosophers and psychologists alike have repeatedly pointed out that there is much more to intelligent behavior than simply keeping track of the implicit consequences of correct observations. A characteristic feature of a rational agent are her *beliefs*, and the ability to correct them when they turn out to be wrong. Many different formal representations of beliefs have been proposed; however, I do not have the space to discuss this line of research here — see (Huber, 2011) and (Halpern, 2003) for overviews.

A simple modification of the above epistemic models allows us to represent both the agents' knowledge and "beliefs": An epistemic-doxastic model is a tuple $\langle W, \{R_i^K\}_{i\in\mathcal{A}}, \{R_i^B\}_{i\in\mathcal{A}}, V\rangle$ where both R_i^K and R_i^B are relations on W. Truth of a belief operator $B_i\varphi$ is defined precisely as in Definition 2.2, replacing R_i with R_i^B. 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 here[4] since these models are not our preferred ways of representing the agents' beliefs (see, for example, Halpern, 1996; Stalnaker, 2006).

3.1 Models of Belief via Plausibility

A key aspect of beliefs not yet represented in the models sketched above is that they are *revisable* in the presence of new information. While there is an extensive literature on the theory of belief revision, starting with (Alchourrón et al., 1985), truly logical models of the dynamics of beliefs have only been developed recently. One appealing approach is to endow epistemic ranges with a *plausibility ordering* for each agent: a pre-order (reflexive and transitive) $w \preceq_i v$ that says "agent i considers world w at least as plausible as v." As a convenient notation, for $X \subseteq W$, we set $Min_{\preceq_i}(X) = \{v \in X \mid v \preceq_i w \text{ for all } w \in X\}$, the set of minimal elements of X according to $\preceq_i$. This is the subset of X that agent i considers the "most plausible". Thus, while the $\sim_i$ partitions the set of possible worlds according to i's "hard information", the plausibility ordering $\preceq_i$ represents i's "soft information" about which of the possible worlds agent i

[4] A key issue here is that assuming B_i and K_i both satisfy positive and negative introspection, K_i satisfies veracity, B_i is consistent ($\neg B_i(\varphi \wedge \neg\varphi)$), knowledge implies belief ($K_i\varphi \to B_i\varphi$) and belief implies believing you know it ($B_i\varphi \to B_iK_i\varphi$) leads to contradiction. Suppose that p is something agent i certain of, but is false: i.e., $\neg p \wedge B_ip$ is true. Then, since K_i satisfies the veracity axiom, $\neg K_ip$ must be true. By negative introspection, this implies $K_i\neg K_ip$ is true. Since knowledge implies belief, we have $B_i\neg K_ip$ is true. Since B_ip is true and belief implies believing it is known, we have B_iK_ip is true. But this means $B_i(K_ip \wedge \neg K_ip)$ is true, which contradicts the assumption that B_i is consistent. See (Halpern, 1996) for an analysis.

considers more "plausible". Models representing both knowledge and beliefs have been extensively studied by logicians (van Benthem, 2007; van Ditmarsch, 2005; Baltag and Smets, 2006), game theorists (Board, 2004), and computer scientists (Boutilier, 1992; Lamarre and Shoham, 1994):

Definition 3.1 (Epistemic-Plausibility Models) Let $\mathcal{A}$ be a finite set of agents and At a (finite or countable) set of atomic propositions. An **epistemic-plausibility model** is a tuple $\mathcal{M} = \langle W, \{\sim_i\}_{i\in\mathcal{A}}, \{\preceq_i\}_{i\in\mathcal{A}}, V\rangle$ where $\langle W, \{\sim_i\}_{i\in\mathcal{A}}, V\rangle$ is an epistemic model and, for each $i \in \mathcal{A}$, $\preceq_i$ is a well-founded[5] reflexive and transitive relation on W satisfying, for all $w, v \in W$:

1. *plausibility implies possibility*: if $w \preceq_i v$ then $w \sim_i v$.
2. *locally-connected*: if $w \sim_i v$ then either $w \preceq_i v$ or $v \preceq_i w$. $\triangleleft$

Remark 3.2 *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 $w \succeq_i v$ or $v \succeq_i w$. In what follows, unless otherwise stated, I will assume that $\sim_i$ is* defined *as $w \sim_i v$ iff $w \preceq_i v$ or $v \preceq_i w$.*

In order to reason about these structures, extend the basic epistemic language $\mathcal{L}_K$ with a conditional belief operator: Let $\mathcal{L}_{KB}$ be the (smallest) set of sentences generated by the following grammar:

$$\varphi \ := \ p \mid \neg\varphi \mid \varphi \wedge \psi \mid B_i^{\varphi}\psi \mid K_i\varphi$$

where $p \in \mathsf{At}$ (the set of atomic propositions) and $i \in \mathcal{A}$. The same conventions apply as above with the additional convention that we write $B_i\varphi$ for $B_i^{\top}\varphi$.

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 $Min_{\succeq_i}([w]_i \cap X)$ is nonempty if $[w]_i \cap X \neq \emptyset$.

Definition 3.3 (Truth for $\mathcal{L}_{KB}$) Suppose that $\mathcal{M} = \langle W, \{\sim_i\}_{i\in\mathcal{A}}, \{\preceq_i\}_{i\in\mathcal{A}}, V\rangle$ is an epistemic-plausibility model. The definition of truth for formulas from $\mathcal{L}_K$ is given in Definition 2.2. The conditional belief operator is defined as follows:

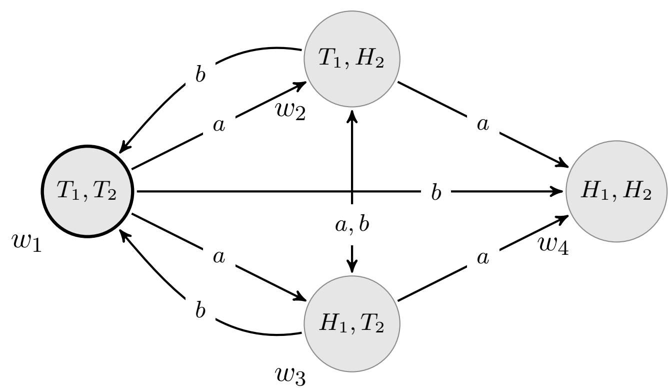
- $\mathcal{M}, w \models B_i^{\varphi}\psi$ iff for all $v \in Min_{\preceq_i}([w]_i \cap [\![\varphi]\!]_{\mathcal{M}})$, $\mathcal{M}, v \models \varphi$ $\triangleleft$

Thus, ψ is believed conditional on φ, if i's most plausible φ-worlds (i.e., the states satisfying φ that i has not ruled out and considers most plausible) all satisfy ψ. Then, the definition of plain belief (which is defined to be $B^{\top}$) is:

[5] 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 – and there are ways around this in current logics. Moreover, the condition of connectedness can also be lifted, but we use it here for convenience.

$$\mathcal{M}, w \models B_i\varphi \text{ iff for each } v \in Min_{\preceq_i}([w]_i),\ \mathcal{M}, v \models \varphi$$

Recall the example of Ann and Bob and the two coins in separate drawers. The following epistemic-plausibility model describes a possible configuration of beliefs before the agents observe their respective coins: I draw an arrow from v to w if $w \preceq v$ (to keep the clutter down, I do not include all arrows. The remaining arrows can be inferred by transitivity and reflexivity).



Suppose that both coins are laying tails up, so w_1 is the actual state. Following the convention from Remark 3.2, we have $[w_1]_a = [w_1]_b = \{w_1, w_2, w_3, w_4\}$, and so, neither Ann nor Bob knows this fact. Furthermore, both Ann and Bob believe that both coins are lying heads up (i.e., $w_1 \models B_a(H_1 \wedge H_2) \wedge B_b(H_1 \wedge H_2)$ since $Min_{\preceq_a}([w_1]_a) = Min_{\preceq_b}([w_1]_b) = \{w_4\}$). However, Ann and Bob do have different *conditional* beliefs. Ann believes that the position of the coins in the two drawers are independent; and so, she believes that H_2 is true even under the supposition that T_1 is true (and vice versa for the other coin: $w_1 \models B_a^{T_1} H_2 \wedge B_a^{T_2} H_1$). On the other hand, Bob believes that the coins are somehow correlated; and so, under the supposition that T_1 is true, Bob believes that the coin in the second drawer must also be laying tails up (and vice versa for the other coin: $w_1 \models B_b^{T_1} T_2 \wedge B_b^{T_2} T_1$).

So, conditional beliefs describe an agent's *disposition* to change her beliefs in the presence of (perhaps surprising) evidence.[6] This is reflected in the logic of conditional belief. An immediate observation is that $B_i(\varphi \to \psi)$ does not imply $B_i^{\varphi}\psi$.[7] [8] Another, more fundamental, observation is that conditioning

[6]See (Leitgeb, 2007) for an extensive discussion about a number of philosophical issues surrounding how to interpret conditional beliefs *qua* belief.

[7]Suppose that there are two states w and v with p true only at w and q true only at v. Then $p \to q$ is true at v and q is false at w. Suppose that $v \preceq_i w$. Then, $B_i(p \to q)$ is true at w, but $B^p q$ is not true at w.

[8]This example shows that conditional belief does not reduce to belief in a *material conditional*. This is, of course, not surprising. A much more interesting question is about the

is nonmonotonic in the sense that $B_i^\varphi \alpha \rightarrow B_i^{\varphi\wedge\psi}\alpha$ is not valid. Nonetheless, weaker forms of monotonicity do hold. These and other key logical principles of conditional belief include:

Success:	$B_i^\varphi \varphi$		
Knowledge entails belief	$K_i\varphi \rightarrow B_i^\psi \varphi$		
Full introspection:	$B_i^\varphi \psi \rightarrow K_i B_i^\varphi \psi$	and	$\neg B_i^\varphi \psi \rightarrow K_i \neg B_i^\varphi \psi$
Cautious Monotonicity:	$(B_i^\varphi \alpha \wedge B_i^\varphi \beta) \rightarrow B_i^{\varphi\wedge\beta}\alpha$		
Rational Monotonicity:	$(B_i^\varphi \alpha \wedge \neg B_i^\varphi \neg\beta) \rightarrow B_i^{\varphi\wedge\beta}\alpha$		

The reader is invited to check that all of the above formulas are valid in any epistemic plausibility model. The full introspection principles are valid since we assume that each agent's plausibility ordering is uniform. That is, each agent i has a single plausibility measure over the set of all possible worlds which totally orders each of i's information cells[9]. The intuition is that all the worlds that an agent i has ruled out at state w (based on her observations and/or background knowledge) are considered the least plausible overall at state w. A more general semantics is needed if one wants to drop the assumption of full introspection and that knowledge entails belief. The key idea is to define the agent's plausibility ordering as a *ternary* relation where $x \preceq_i^w y$ means i considers x at least as plausible as y in state w. So, agents may have different plausibility orderings at different states. The logic of these more general structures has been studied by Oliver Board (2004) (see also an earlier paper by John Burgess, 1981).

The other principles highlight the close connection with the AGM postulates of rational belief revision (Alchourrón et al., 1985). [10] While success and cautious monotonicity are often taken to be constitutive of believing something *under the supposition that φ is true*, rational monotonicity has generated quite a bit of discussion. The most well-known criticism comes from Robert Stalnaker (1994a) who suggests the following counterexample: Suppose that Ann initially believes that the composer Verdi is Italian $I(v)$ and Bizet and Satie are French $(F(b) \wedge F(s))$. Conditioning on the fact that Verdi and Bizet are compatriots

relationship between a conditional belief and a belief in an (*indicative*) *conditional*. There is a broad literature on this issue stemming from Ramsey's famous footnote 'If two people are arguing 'If p, then q?' and are both in doubt as to p, they are adding p hypothetically to their stock of knowledge and arguing on that basis about q..." (Ramsey, 1929) and Gärdenfors' impossibility theorem (Gärdenfors, 1986). A good entrée into the field is (Leitgeb, 2010a).

[9] In general, each agent's plausibility ordering will not be a total relation as states in different information cells cannot be compared by the plausibility ordering.

[10] This close connection should not come as a surprise, since epistemic plausibility models are essentially the sphere models of Grove (1988).

$(C(v,b))$,[11] Ann still believes that Satie is French ($B_a^{C(v,b)}F(s)$ is true since supposing that Verdi and Bizet are compatriots does not conflict with the belief that Satie is French). Under the supposition that Verdi and Bizet are compatriots, Ann thinks it is (doxastically) possible that Verdi and Satie are compatriots ($\neg B^{C(v,b)}\neg C(v,s)$ is true since $C(v,b)$ is consistent with all three being French). Rational monotonicity gives us that Ann believes Satie is French under the suppositions that $C(v,b) \wedge C(v,s)$ (i.e., $B^{C(v,b)\wedge C(v,s)}F(s)$ must be true). However, supposing that $C(v,b) \wedge C(v,s)$ is true implies all three composers are compatriots, and this could be because they are all Italian. Much has been written in response to this counterexample. However, a detailed analysis of this literature would take us too far away from the main objective of this section: a concise introduction to the main formal models of knowledge and belief.

I conclude this short section by discussing two additional notions of belief. To that end, we need some additional notation: The plausibility relation $\preceq_i$ can be lifted to subsets of W as follows: [12]

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

Exend the language $\mathcal{L}_{KB}$ with two belief operators: B_i^r ("robust belief") and B_i^s ("strong belief"). Let $\mathcal{L}_{KB}^+$ be the resulting language. Semantics for this language is given by adding the following clauses to Definition 3.3.

- *Robust Belief*[13]: $\mathcal{M}, w \models B_i^r\varphi$ iff for all v if $v \preceq_i w$ then $\mathcal{M}, v \models \varphi$. Thus, φ is robustly believed if φ is true in *all* states the agent considers more plausible than the current state. This stronger notion of belief has also been called *certainty* by some authors (Shoham and Leyton-Brown (2009), Section 13.7).

- *Strong Belief*: $\mathcal{M}, w \models B_i^s\varphi$ iff there is a v with $w \sim_i v$ and $\mathcal{M}, v \models \varphi$ and $\{x \mid \mathcal{M}, x \models \varphi\} \cap [w]_i \preceq_i \{x \mid \mathcal{M}, x \models \neg\varphi\} \cap [w]_i$, where $[w]_i$ is the equivalence class of w under $\sim_i$. This notion has been studied by (Stalnaker, 1994b; Battigalli and Siniscalchi, 2002).

The following example illustrates the logical relationships between the various notions of belief and knowledge we have discussed. Consider the following plausibility model with four states for a single agent (since there is only one agent, I do not use subscripts):

[11] Note that in this example, $I(v)$, $F(b)$ and $F(s)$ are all atomic propositions and $C(i,j)$ is defined to be $(I(i) \wedge I(j)) \vee (F(i) \wedge F(j))$.

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

[13] In the dynamic doxastic logic literature, this notion is often called *safe belief*. However, this terminology conflicts with Williamson's notion of "safety" (Williamson, 2000).

Note that the set of minimal states is $\{v_2\}$, so if $v_2 \in V(p)$, then the agent believes p (Bp is true at all states). Suppose that w is the "actual world" and consider the following truth assignments of an atomic proposition p.

- $V(p) = \{v_0, w, v_2\}$. Then $\mathcal{M}, w \models B^r p$, but $\mathcal{M}, w \not\models B^s p$, so robust belief need not imply strong belief.
- $V(p) = \{v_2\}$. Then $\mathcal{M}, w \models B^s p$, but $\mathcal{M}, w \not\models B^r p$, so strong belief need not imply robust belief.
- $V(p) = \{v_0, v_2, w, v_2\}$. Then $\mathcal{M}, w \models Kp \wedge B^s p \wedge B^r p \wedge Bp$ (in fact, it is easy to see that knowledge implies belief, robust belief and strong belief).

Note that, unlike beliefs, conditional beliefs may be inconsistent (i.e., $B_i^\varphi \bot$ may be true at some state). In such a case, agent i cannot (on pain of inconsistency) revise by φ, but this will happen only if the agent has hard information that ψ is false. Indeed, $K_i \neg\varphi$ is logically equivalent to $B_i^\varphi \bot$ (over the class of epistemic plausibility models). This suggests the following (dynamic) characterization of an agent's hard information as unrevisable beliefs:

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

Robust and strong belief can be similarly characterized by restricting the set of formulas that an agent can condition on:

- $\mathcal{M}, w \models B_i^r \varphi$ iff $\mathcal{M}, w \models B_i^\psi \varphi$ for all ψ with $\mathcal{M}, w \models \psi$: That is, the agent robustly believes φ iff she continues to believe φ given any true formula.
- $\mathcal{M}, w \models B_i^s \varphi$ iff $\mathcal{M}, w \models B\varphi$ 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, the agent strongly believes φ iff she believes φ and continues to believe φ given any evidence (truthful or not) that is not known to contradict φ.

Finally, there is an elegant axiomatization of epistemic plausibility models in a modal language containing knowledge operators and robust belief operators using the following characterizations of conditional and strong belief (Baltag and Smets, 2009):

- $B_i^\varphi \psi \ := \ L_i\varphi \rightarrow L_i(\varphi \wedge [\preceq_i](\varphi \rightarrow \psi))$
- $B_i^s \varphi \ := L_i\varphi \wedge K_i(\varphi \rightarrow [\preceq_i]\varphi)$

As discussed above, each K_i satisfies logical omniscience, veracity and both positive and negative introspection. Robust belief, B_i^r, shares all of these properties except negative introspection. Modal correspondence theory can again be used to characterize the remaining properties:

Knowledge entails robust belief:	$K_i\varphi \rightarrow B_i^r\varphi$
Local connectedness:	$(K_i(\varphi \vee B_i^r\psi) \wedge K_i(\psi \vee B_i^r\varphi)) \rightarrow K_i\varphi \vee K_i\psi$

3.2 Models of Belief via Probability

The logical frameworks introduced in the previous section represent an agent's "all-out" or "full" beliefs. However, there is an large body of literature within (formal) epistemology and game theory that works with a *quantitative* conception of belief. Graded beliefs have also been subjected to sophisticated logical analyses (see, for example, Fagin et al., 1990; Fagin and Halpern, 1994; Heifetz and Mongin, 2001; Zhou, 2009; Goldblatt, 2008).

The dominant approach to formalizing graded beliefs is (subjective) probability theory. A **probability measure** on a set W is a function assigning a positive real number to (some) subsets of W such that $\pi(W) = 1$ and for disjoint subsets $E, F \subseteq W$ $(E \cap F = \emptyset)$ $\pi(E \cup F) = \pi(E) + \pi(F)$. For simplicity, I assume in this section that W is finite. Then, the definition of a probability measure can be simplified: a probability measure on a finite set W is a function $\pi : W \rightarrow [0, 1]$ such that for each $E \subseteq W$, $\pi(E) = \sum_{w \in E} \pi(w)$ and $\pi(W) = 1$. Nothing that follows hinges on the assumption that W is finite, but if W is infinite, then there are a number of important mathematical details that add some complexity to the forthcoming definitions.[14] **Conditional probability** is defined in the usual way: $\pi(E \mid F) = \frac{\pi(E \cap F)}{\pi(F)}$ if $\pi(F) > 0$ ($\pi(E \mid F)$ is undefined when $\pi(F) = 0$).

The model we study in this section is very close to the epistemic plausibility of Definition 3.1 with probability measures in place of plausibility orderings:

Definition 3.4 (Epistemic-Probability Model) Suppose that $\mathcal{A}$ is a finite set of agents, At is a (finite or countable) set of atomic propositions and W is a finite set of states. An **epistemic probability model** is a tuple $\mathcal{M} = \langle W, \{\sim_i\}_{i \in \mathcal{A}}, \{\pi_i\}_{i \in \mathcal{A}}, V \rangle$ where $\langle W, \{\sim_i\}_{i \in \mathcal{A}}, V \rangle$ is an epistemic model and, for each $i \in \mathcal{A}$, π_i is a probability measure on W. We also assume that each π_i is *weakly regular*[15] in the sense that for each $w \in W$, $\pi_i([w]_i) > 0$. $\triangleleft$

[14] See (Halpern, 2003, Chapter 1) and (Billingsley, 1995) for details.

[15] A probability measure is **regular** provided $\pi(E) > 0$ for each (measurable) subset E.

The probability measures π_i represent agent i's prior beliefs about the likelihood of each element of W. Agents then receive private information, represented by the equivalence relations $\sim_i$, and update their initial beliefs with that information. A variety of modal languages have been proposed to reason about graded beliefs. In this section, we focus on a very simple language containing a knowledge modality $K_i\varphi$ ("i is informed that φ is true") and $B_i^q\varphi$ ("i believes φ is true to degree at least q", or "i's degree of belief in φ is at least q") where q is a rational number. More formally, let $\mathcal{L}_{KB}^{prob}$ be the smallest set of formulas generated by the following grammar:

$$p \mid \neg\varphi \mid \varphi \wedge \psi \mid B_i^q\varphi \mid K_i\varphi$$

where $q \in \mathbb{Q}$ (the set of rational numbers), $i \in \mathcal{A}$ and $p \in \mathsf{At}$.

Definition 3.5 (Truth for $\mathcal{L}_{KB}^{prob}$) Suppose that $\mathcal{M} = \langle W, \{\sim_i\}_{i\in\mathcal{A}}, \{\pi_i\}_{i\in\mathcal{A}}, V\rangle$ is an epistemic-probaiblity model. The definition of truth for formulas from $\mathcal{L}_K$ is given in Definition 2.2. The belief operator is defined as follows:

- $\mathcal{M}, w \models B_i^q\varphi$ iff $\pi_i([\![\varphi]\!]_{\mathcal{M}} \mid [w]_i) \geq q$

where $[\![\varphi]\!]_{\mathcal{M}} = \{w \mid \mathcal{M}, w \models \varphi\}$. ◁

Note that since we assume for each $w \in W$, $\pi_i([w]_i) > 0$, the above definition is always well-defined. The logical framework presented here has much in common with the qualitative version presented in the previous section. In particular, knowledge entails belief ($K_i\varphi \to B_i^q\varphi$) and full introspection ($B_i^q\varphi \to K_iB_i^q\varphi$ and $\neg B_i^q\varphi \to K_i\neg B_i^q\varphi$) are both valid. As before, full introspection can be dropped by assuming agents have different probability measures at different states. [16]

The graded notion of belief has much in common with its qualitative version. In particular, B_i^q satisfies both positive and negative introspection (these both follow from full introspection and that fact that knowledge entails belief) and logical omniscience (if $\varphi \to \psi$ is valid then so is $B_i^q\varphi \to B_i^q\psi$). There is also an analogue to the success axiom (although we cannot state it in our language since we do not have conditional belief operators in $\mathcal{L}_{KB}^{prob}$):

$$\pi_i([\![\varphi]\!]_{\mathcal{M}} \mid [\![B_i^q\varphi]\!]_{\mathcal{M}}) \geq q$$

It is a simple (and instructive!) exercise to verify that the above property is true in any epistemic-probability model. In addition to principles describing what the agents know and believe about their own beliefs, there are principles that ensure the different B_i^q operators fit together in the right way:

[16] Formally, $P_i : W \to \Delta(W)$ where $\Delta(W)$ is the class of probability measures on W.

- $B_i^0\varphi$
- $B_i^1\top$
- $B_i^q(\varphi \wedge \psi) \wedge B_i^p(\varphi \wedge \neg\psi) \to B_i^{q+p}\varphi, \quad q+p \leq 1$
- $\neg B_i^q(\varphi \wedge \psi) \wedge \neg B_i^p(\varphi \wedge \neg\psi) \to \neg B_i^{q+p}\varphi, \quad q+p \leq 1$
- $B_i^q\varphi \to \neg B_i^p\neg\varphi, \quad q+p > 1$

A distinctive feature of the complete logic of epistemic-probability models is the following inference rule reflecting the completeness (in the topological sense) of the real numbers:

Archimedian Rule: If $\psi \to B_i^p\varphi$ is valid for each $p < q$, then $\psi \to B_i^q\varphi$ is valid.

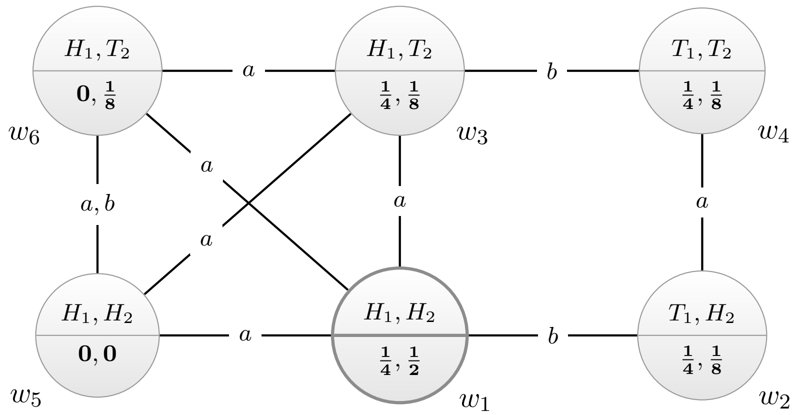
There is also a probabilistic analogue of robust belief (recall the definition from Section 3.1). The key idea is that φ is robustly believed above a threshold q provided the probability i assigns to φ conditional on any true event (i.e., any subset of W containing the current state) is at least q.

Definition 3.6 (Graded Robust Belief) Agent i **q-robustly believes** φ, denoted $B_i^{r:q}\varphi$, is defined as follows: Suppose that $\mathcal{M} = \langle W, \{\sim_i\}_{i\in\mathcal{A}}, \{\pi_i\}_{i\in\mathcal{A}}, V\rangle$ is an epistemic-probability model. [17]

- $\mathcal{M}, w \models B_i^{r:q}\varphi$ iff $\min_{w\in X} \pi(\llbracket\varphi\rrbracket_{\mathcal{M}} \mid X \cap [w]_i) \geq q$ $\triangleleft$

I conclude this section with some brief comments about probability zero events and the relationship between knowledge ($K_i\varphi$) and belief with probability one ($B_i^1\varphi$). Note that it is possible that in an epistemic-probability model there are states w and v with $w \sim_i v$ and $\pi_i(v) = 0$. In such a case, agent i is sure that state v is note that actual state, but does not consider v *impossible* (i.e., i cannot "rule out" v according to her information). In particular, this means that $B_i^1\varphi \to K_i\varphi$ is not valid. The following example illustrates these issues: The numbers in the lower half of the circle indicate Ann and Bob's initial probability for that state (Ann's probability is on the left and Bob's is on the right, so, for example, $\pi_a(w_1) = \frac{1}{4}$ and $\pi_b(w_1) = \frac{1}{2}$).

[17] If W is infinite, then the definition should be $\mathcal{M}, w \models B_i^{r:q}\varphi$ iff $\inf_{w\in X} \pi(\llbracket\varphi\rrbracket_{\mathcal{M}} \mid X \cap [w]_i) \geq q$, where inf is the *infimum* (i.e., the greatest lower bound).



The following observation illustrates the notions introduced above:

- Ann does not know the direction the coin is laying in the second drawer and believes that H_2 and T_2 are equally likely: $\mathcal{M}, w_1 \models \neg K_a H_2 \wedge \neg K_a T_2 \wedge B_a^{\frac{1}{2}} H_2 \wedge B_a^{\frac{1}{2}} T_2$
- Bob does not know the direction the coin is laying in the first drawer, but believes it is more likely laying heads up: $\mathcal{M}, w_1 \models \neg K_b H_1 \wedge \neg K_b T_1 \wedge B_b^{\frac{4}{5}} H_1 \wedge B_b^{\frac{1}{5}} T_1$
- Ann does not know that Bob knows whether H_2, but she is certain that he knows whether H_2 (in the sense that she assigns probability one to him knowing whether): $\mathcal{M}, w_1 \models \neg K_a(K_b H_2 \vee K_b T_2) \wedge B_a^1(K_b H_2 \vee K_b T_2)$

Of course, epistemic-probability models provide a more fine-grained representation of the agents' beliefs than their qualitative counterparts. However, the relationship between the two models is more subtle and touches on many issues beyond the scope of this article. [18]

[18] It is tempting to identify "more plausible" with "more probable". There are a number of reasons why this is not a good idea. First of all, if we define $w \preceq_i v$ iff $\pi_i(w) \geq \pi_i(v)$, then it is easy to construct examples where $B_i\varphi$ is true (using Definition 3.3), but $B_i^1\varphi$ is not true (using Definition 3.5). This is a problem if we equate full belief ($B_i\varphi$) and belief with probability one ($B_i^1\varphi$). (It is also easy to find examples where full belief cannot be identified with probability above a threshold.) The difficulty here boils down to a deep foundational problem: Can a rational agent's full beliefs be *defined* in terms of her graded beliefs, and/or vice versa? A second, more conceptual, observation is that the two models represent different aspects of the agents' states of belief. To illustrate the difference, suppose that Ann is flipping a biased coin. It may be much more likely to land heads, but landing heads and tails are both *plausible* outcomes whereas landing on it side is not a plausible outcome. So, the plausibility ordering describes the agents' all-out judgements about the priority between the states, which

3.3 Group Notions

Both game theorists and logicians have extensively discussed different notions of knowledge and belief for a group, such as common knowledge and belief. These notions have played a fundamental role in the analysis of distributed algorithms (Halpern and Moses, 1990) and social interactions (Chwe, 2001). In this section, I introduce and formally define various group informational attitudes. I can only scratch the surface of the extensive literature discussing the numerous logical and philosophical issues that arise here (see Vanderschraaf and Sillari (2009) for an in-depth discussion of this literature[19]).

Consider the statement "everyone in group G knows that φ". With finitely many agents, this can be easily defined in the epistemic language $\mathcal{L}_{KB}$:

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

where $G \subseteq \mathcal{A}$. The first nontrivial informational attitude for a group that we study is *common knowledge*. If φ is common knowledge for the group G, then not only does everyone in the group know that φ is true, but this fact is completely transparent to all members of the group. There are different ways to make precise what it means for something to be "completely transparent" to a group of agents.[20] The approach I follow here is to iterate the everyone knows operator:

$$\varphi \wedge K_G\varphi \wedge K_GK_G\varphi \wedge K_GK_GK_G\varphi \wedge \cdots$$

The above formula is an *infinite* conjunction, and so is not a formula in our epistemic language $\mathcal{L}_{KB}$ (by definition, there can be at most finitely many conjunctions in any formula). In fact, using standard modal logic techniques, one can show that there is no formula of $\mathcal{L}_{KB}$ that is logically equivalent to the above infinite conjunction. Thus, we must extend our basic epistemic language with a modal operator $C_G\varphi$ with the intended meaning "φ is common knowledge among the group G". The idea is to *define* $C_G\varphi$ to be true precisely when φ is true, everyone in G knows that φ is true, everyone in G knows that everyone in

is not directly represented in a probability measure.

A good overview of the main philosophical issues is found in (Christensen, 2007). See for discussions that are particularly relevant See (Hawthorne and Bovens, 1999; Leitgeb, 2010b; Arló-Costa and Pedersen, 2011) for discussions that are particularly relevant to the logical frameworks introduced in this paper.

[19]The textbooks Fagin et al. (1995) and van Benthem (2010a) also provide illuminating discussions of key logical issues issues.

[20]Barwise (1988) discusses three main approaches: (i) the iterated view, (ii) the fixed-point view and (iii) the shared situation view. In this paper, I focus only on the first two approaches.

G knows that φ is true, and so on *ad infinitum*.[21]

Before giving the details of this definition, consider $K_G K_G K_G \varphi$. 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 of length n for G** in an epistemic model is a sequence of states $(w_0, w_1, \dots, w_n)$ where for each $l = 0, \dots, n-1$, 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, $K_G K_G K_G \varphi$ 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 3.7 (Interpretation of C_G) Let $\mathcal{M} = \langle W, \{\sim_i\}_{i \in \mathcal{A}}, V\rangle$ be an epistemic model[22] and $w \in W$. The truth of formulas of the form $C_G\varphi$ is:

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

where $R_G^C := (\bigcup_{i\in G} \sim_i)^*$ is the reflexive transitive closure[23] of $\bigcup_{i \in G} \sim_i$. ◁

It is well-known that for any relations R on W, if wR^*v then there is a finite R-path starting at w ending in v. Thus, we have $\mathcal{M}, w \models C_G\varphi$ iff every finite path for G from w ends with a state satisfying φ. The logical analysis is more complicated in languages with a common knowledge operator; however, the following two axioms can be said to characterize[24] common knowledge:

Fixed-Point	$C_G\varphi \rightarrow K_G C_G \varphi$
Induction	$(\varphi \wedge C_G(\varphi \rightarrow K_G\varphi)) \rightarrow C_G\varphi$

The first formula captures the "self-evident" nature of common knowledge: if φ is common knowledge then everyone in the group knows this. This suggests the following alternative characterization of common knowledge: Call a subset

[21] It is worth pointing out that this is not what David Lewis had in mind when he first formalized common knowledge Lewis (1973). For Lewis, the infinite conjunction is a necessary but not a sufficient condition for common knowledge. See Cubitt and Sugden (2003) for an illuminating discussion and a reconstruction of Lewis' notion of common knowledge. Nonetheless, following Aumann (1976), the definition given in this section has become standard in the game theory and epistemic logic literature.

[22] The same definition will of course hold for epistemic-plausibility and epistemic-probability models.

[23] The reflexive transitive closure of a relation R is the smallest relations R^* containing R that is reflexive and transitive.

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

$X \subseteq W$ i-**closed** provided for all $w \in W$, if $w \in X$ and $w \sim_i v$ then $v \in X$. If X is i-closed then it is "self-evident" for agent i in the sense that if X obtains (i.e., the current state is in X) then i knows that X obtains (formally, we have $X \subseteq \{w \mid [w]_i \subseteq X\}$). Then,

Fact 3.8 *Suppose that $\mathcal{M}$ is an epistemic (-plausibility/-probability) model. Then, $\mathcal{M}, w \models C_G\varphi$ iff there is a set $X \subseteq W$ that is i-closed for all $i \in G$ and $X \subseteq [\![\varphi]\!]_{\mathcal{M}}$.*

It is sometimes convenient to use the above characterization common knowledge as the definition of C_G. On epistemic (-plausibility/-probability) models, the two definitions are mathematically equivalent (see van Benthem and Sarenac, 2004, for a more general semantics where the two definitions are no longer equivalent.).

The approach to defining common knowledge outlined above can be viewed as a recipe for defining common (safe/strong) belief. For example, suppose wR_i^Bv iff $v \in Min_{\preceq_i}([w]_i)$ and define R_G^B to be the transitive closure[25] of $\cup_{i \in G} R_i^B$. Then, **common belief of** φ, denoted $C_G^B\varphi$, is defined in the usual way:

$$\mathcal{M}, w \models C_G^B\varphi \text{ iff for each } v \in W, \text{ if } wR_G^Bv \text{ then } \mathcal{M}, v \models \varphi.$$

While common belief also validates the fixed-point and induction axiom, its logic does differ from logic of common knowledge. The most salient difference is that common knowledge satisfies negative introspection ($\neg C_G\varphi \to C_G\neg C_G\varphi$ is valid) while common belief does not ($\neg C_G^B\varphi \to C_G^B\neg C_G^B\varphi$ is not valid). See (Bonanno, 1996; Lismont and Mongin, 1994, 2003) for more information about the logic of common belief.

A probabilistic variant of common belief was introduced by Monderer and Samet (1989). It is convenient to give the definition in terms of "self-evident sets".

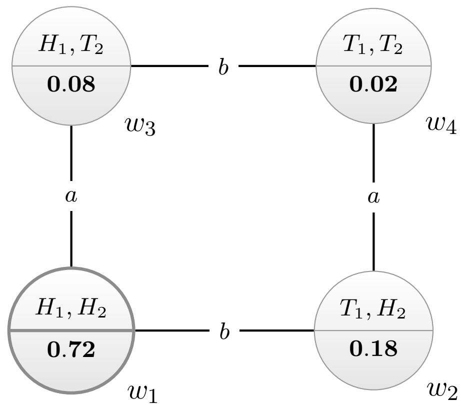
Definition 3.9 (Common q-belief) Call a set $X \subseteq W$ an **evident q-belief for** i when $X \subseteq \{w \mid \pi_i(X \mid [w]) \geq q\}$. Then, **common q-belief for a group** G, denoted $C_G^q\varphi$, is defined as follows: Let $\mathcal{M} = \langle W, \{\sim_i\}_{i \in \mathcal{A}}, \{\pi\}_{i \in \mathcal{A}}, V\rangle$ be an epistemic-probability model.

- $\mathcal{M}, w \models C_G^q\varphi$ iff there is a set X such that $w \in X$, X is an evident q-belief for all $i \in G$ and $X \subseteq [\![B_i^q\varphi]\!]_{\mathcal{M}}$ for all $i \in G$. $\triangleleft$

The following examples illustrates the above definition. Suppose that after observing their respective coins, there is an announcement over a loudspeaker that both coins are laying heads up ($H_1 \wedge H_2$). Assume that Ann was listening closely

[25] Since beliefs need not be factive, we do not force R_G^B to be reflexive.

but the announcement was not perfectly clear, and so the the probability that she heard correctly is 0.9. Bob was not paying as close attention and so the probability that he heard correctly is 0.8. Given that these probabilities are commonly known by Ann and Bob, the initial probability of state w_1 (where $H_1 \wedge H_2$ is true) is $0.9 \times 0.8 = 0.72$. Similar calculations for the remaining states gives us the following epistemic-probability model (note that both agents have the same initial probability measure, so I only record one number for each states):



Then, $w_1 \models B_a^{0.9}(H_1 \wedge H_2) \wedge B_b^{0.8}(H_1 \wedge H_2)$. This means $X = \{w_1\}$ is an evident 0.8-belief for both Ann and Bob and $X \subseteq [\![B_a^{0.8}(H_1 \wedge H_2)]\!]_{\mathcal{M}} = \{w_1, w_3\} = [\![B_b^{0.8}(H_1 \wedge H_2)]\!]_{\mathcal{M}}$. Hence, $w_1 \models C_{a,b}^{0.8}(H_1 \wedge H_2)$. That is, while it is not common knowledge that the coins are both laying heads up, it is common 0.8-belief that they are both laying heads up.

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) $\mathcal{M} = \langle W, \{\sim_i\}_{i \in \mathcal{A}}, V\rangle$, let $R_G^D = \cap_{i \in G} \sim_i$, then define

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

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 (see Fagin et al., 1995; van Benthem, 2010a; Gerbrandy, 1999; Baltag and Smets, 2010; Roelofsen, 2007).

References

Alchourrón, C. E., P. Gärdenfors, and D. Makinson (1985). On the logic of theory change: Partial meet contraction and revision functions. *Journal of Symbolic Logic 50*, 510 – 530.

Arló-Costa, H. and A. P. Pedersen (2011). Belef and probability: A general theory of probability cores. *Journal of Approximate Reasoning*, forthcoming.

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

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

Baltag, A. and S. Smets (2006). 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, pp. 5 – 21.

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`.

Baltag, A. and S. Smets (2010). Correlated knowledge, an epistemic-logic view on quantum entanglement. *International Journal of Theoretical Physics 49*(12), 3005 – 3021.

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.

Barwise, J. and J. Perry (1983). *Situations and Attitudes*. The MIT Press.

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

Billingsley, P. (1995). *Probability and Measure*. Wiley-Interscience.

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.

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

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.

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

Christensen, D. (2007). *Putting Logic in Its Place: Formal Constraints on Rational Belief*. Oxford University Press.

Chwe, M. S.-Y. (2001). *Rational Ritual*. Princeton 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. (2011). Epistemic logic. In L. Horsten and R. Pettigrew (Eds.), *The Continuum Companion to Philosophical Logic*. Continuum.

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

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.

Gärdenfors, P. (1986). Belief revision and the ramsey test. *Philosophical Review 95*, 81 – 93.

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

Gintis, H. (2009). *The Bounds of Reason: Game Theory and the Unification of the Behavioral Sciences*. Princeton University Press.

Goldblatt, R. (2008). Deductive systems for coalgebras over measurable spaces. *Journal of Logic and Computation*, doi: 10.1093/logcom/exn092.

Grove, A. (1988). Two modellings for theory of change. *Journal of Philosophical Logic 17*, 157 – 179.

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

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

Halpern, J. and Y. Moses (1990). Knowledge and common knowledge in a distributed environment. *Journal of the ACM 37*(3), 549 – 587.

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

Halpern, J. and R. Pucella (2011). Dealing with logical omniscience: Expressiveness and pragmatics. *Artificial Intelligence 175*(1), 220 – 235.

Hawthorne, J. and L. Bovens (1999). The preface, the lottery, and the logic of beleif. *Mind 108*, 241 – 264.

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

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

Holliday, W. H. (2012). Epistemic logic and epistemology. In V. Hendricks and S. O. Hansson (Eds.), *Handbook of Formal Philosophy.* Springer.

Huber, F. (2011). Formal representations of belief. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Winter 2011 ed.).

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.

Leitgeb, H. (2007). Beliefs in conditionals vs. conditional beliefs. *Topoi 26*(1), 115 – 132.

Leitgeb, H. (2010a). On on the ramsey test without triviality. . *Notre Dame Journal of Formal Logic 51*(1), 21 – 54.

Leitgeb, H. (2010b). Reducing belief simpliciter to degrees of belief. Technical report, Bristol. Manuscript.

Lewis, D. (1973). *Counterfactuals*. Oxford: Blackwell Publishers.

Lewis, D. (1996). Elusive knowledge. *Australasian Journal of Philosophy 74*(4), 549 – 567.

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).

Meyer, J.-J. and W. van der Hoek (1995). *Epistemic Logic for AI and Computer Science*. Cambridge University Press.

Monderer, D. and D. Samet (1989). Approximating common knowledge with common beliefs. *Games and Economic Behavior 1*(2), 170 – 190.

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

Parikh, R. (2005). WHAT do we know, and what do WE know? In R. van der Meyden (Ed.), *Proceedings of TARK 2005*. National University of Singapore.

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

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

Ramsey, F. P. (1929). General propositions and causality. In R. B. Braithwaite (Ed.), *The Foundations of Mathematics and other Logical Essays*, Chapter Philosophical Papers, ed. D.H. Mellor, Cambridge:, 1990, pp 145-163. Kegan Paul, Trench & Trubner.

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

Roy, O. and E. Pacuit (2011). Substantive assumptions in interaction: A logical perspective. *Synthese*.

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

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

Sorensen, R. (2002). Formal problems about knowledge. In P. K. Moser (Ed.), *Oxford Handbook of Epistemology*.

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

Stalnaker, R. (1994a). Nonmonotonic consequence relations. *Fundamenta Informaticae 21*, 7 – 21.

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

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

van Benthem, J. (2006). Epistemic logic and epistemology: The state of their affairs. *Philosophical Studies 128*, 49 – 76.

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

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

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

van der Hoek, W. and M. Wooldridge (2003). Cooperation, knowledge, and time: Alternating-time temporal epistemic logic and its applications. *Studia Logica 75*(1), 125 – 157.

van Ditmarsch, H. (2005). Prolegomena to dynamic logic for belief revision. *Synthese: Knowledge, Rationality, and Action 147*, 229 – 275.

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

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

Zhou, C. (2009). A complete deductive system for probability logic. *Journal of Logic and Computation 19*(6), 1427 – 1454.