Chapter 6
Procedural Information and the Dynamics of Belief

Eric Pacuit

6.1 Introduction

The point of departure for modern epistemic and doxastic logic is Jaakko Hintikka's seminal book *Knowledge and Belief: An Introduction to the Logic of the Two Notions* (Hintikka 1962).[1] 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?),[2] much of the work on epistemic and doxastic logic was taken over by game theorists (Aumann 1999) and computer scientists (Fagin et al. 1995) in the 1990s.[3] See Bonanno and Battigalli (1999) and Brandenburger (2007) for a survey of epistemic issues that arise in game theory and Fagin et al. (1995) for applications of epistemic logic in computer science.

This focus on different areas of "application" has pushed the analysis beyond the basic epistemic logic of Hintikka (1962) and Aumann (1999) (representing an agent's "hard" information) to "softer" informational attitudes that may be revised. Recent work by epistemic logicians has identified and analyzed a rich repertoire

[1] This important book has recently been reissued and extended with some of Hintikka's latest papers on epistemic logic (Hintikka 2005).

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

[3] Recently, focus has shifted back to Philosophy, with a growing interest in "bridging the gap between formal and mainstream epistemology". Witness the collection of articles (Hendricks 2006) and the book *Mainstream and Formal Epistemology* by Vincent Hendricks (2006).

E. Pacuit (✉)
Tilburg Institute for Logic and Philosophy of Science, Tilburg University,
Warandelaan 2, 5037 AB Tilburg, The Netherlands
e-mail: epacuit@umd.edu

M. Rebuschi et al. (eds.), *Interdisciplinary Works in Logic, Epistemology, Psychology and Linguistics*, Logic, Argumentation & Reasoning 3,
DOI 10.1007/978-3-319-03044-9_6,

of *informational attitudes*. Examples that have been subjected to a logical analysis include different flavors of belief, such as "strong" and "safe" belief (van Benthem 2007; Baltag and Smets 2006); "syntactic" notions, such as awareness (Halpern and Rego 2009) and "explicit knowledge" (Ågotnes and Alechina 2007); variants of "knowing how", such as the "constructive" knowledge" of Jamroga and Ågotnes (2007); and, of course, the many different representations of *graded beliefs* found in Artificial Intelligence and Decision and Game Theory (see Halpern (2005), and references therein). The goal of a logical analysis is to see how these different notions of knowledge and belief fit together.

In this paper, I am not interested in these *static* logics of informational attitudes per se. Rather, my focus is on the dynamic operations that change these informational attitudes during a social interaction or rational inquiry. Current *dynamic* logics of belief revision and information update focus on two key aspects of informative actions:

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 no other agents 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).

One of the goals of this paper is to introduce the key ideas and main definitions that form the foundations of these dynamic logics of interaction and inquiry.

Many of the recent developments in this area have been driven by analyzing *concrete* examples. These range from toy examples, such as the infamous muddy children puzzle, to philosophical quandaries, such as Fitch's Paradox, to everyday examples of social interaction. Different logical systems are then judged, in part, on how well they conform to the analyst's intuitions about the relevant set of examples. But this raises an important methodological issue: Implicit assumptions about what the actors know and believe about the situation being modeled often guide the analyst's intuitions. In many cases, it is crucial to make these underlying assumptions explicit.

The general point is that *how* an agent comes to know or believe that some proposition p is true is as important (or, perhaps, more important) than the fact that the agent knows or believes that p is the case (cf. the discussion in van Benthem (2009, Sect. 2.5)). One lesson to take away is that during a social interaction, the agents' "knowledge" and "beliefs" are both influenced by *and* shaped by the *social* events. The following example taken from Pacuit et al. (2006) illustrates this point.

Suppose that Uma is a physician whose neighbor Sam is ill, and consider the following cases:

Case 1: Uma does not know and has not been informed that Sam is ill. Uma has no obligation (as yet) to treat her neighbor.

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

These simple examples highlight the observation that an agent's obligation often depends 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. But these obligations depend on certain (implicit) information that Uma and Ann have about each other. For example, Ann is not under any obligation to tell Uma that her father is ill if she 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". Furthermore, 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.

Two "types" of information play a role in the above discussion. The first, which might be called "meta-information" (cf. the discussion in Stalnaker (2009)) is information about how "trusted" or "reliable" the sources of the information are. This is particularly important when analyzing how an agent's beliefs change over an extended period of time. For example, rather than taking a stream of contradictory incoming evidence (i.e., the agent receives the information that p, then the information that q, then the information that $\neg p$, then the information that $\neg q$) at face value (and performing the suggested belief revisions), a rational agent may consider the stream itself as evidence that the source is not reliable.[4]

There is much more to say about logical models of trust and reliability, but, in this paper, I am interested in a second "type" of information: **procedural information**. This is information about the underlying *protocol* specifying which events (observations, messages, actions) are available (or permitted) at any given moment. Procedural information 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, but, rather, to 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.

[4]Cf. the very interesting discussion of *higher-order evidence* in the (formal) epistemology literature (Christensen 2010).

A *protocol* describes what the agents "can" or "cannot" do (say, observe) in a social interactive situation or rational inquiry. This leads to *substantive* assumptions about the formal model, such as which actions (observations, messages, utterances) are available (permitted) at any given moment. These assumptions can be roughly categorized according to the different uses of "can":

1. To describe physical, temporal or historical possibilities: A typical example is the assumption that an agent *cannot* receive a message unless another agent sent it earlier. Such assumptions limit the options available to the agents at any given moment.
2. To describe the agents' abilities, or skills: The options available to an agent at any given moment are defined not only by what is "physically possible", but also by the agent's *capacity* to perform various actions. For example, "Ann *can* throw a bulls-eye" typically means that Ann has the ability to (repeatedly) throw a bulls-eye.
3. To describe compliance to some type of norm: The social or conversational[5] norms at play in the interactive situation being modeled (i.e., the "rules of the game") impose further constraints on options available to each agent.

So, a protocol encodes not only which options are *feasible*, but also what is *permissible* for the agents to do or say. Of course, an interesting and important component of a logical analysis of rational agency is to disambiguate these different meanings of "can" (I do not discuss these issues here, see John F. Horty (2001), Dag Elgesem (1997) and Charles B. Cross (1986) for discussions).

A typical assumption is that there is a fixed, global protocol that all the agents have (explicitly or implicitly) agreed to follow (and this is commonly known). This raises an important question: *In what sense do the agents know the protocol?* Formally, the protocol describes which states or histories are "in the model", so the *proposition* expressing that "the protocol is being followed" is the set of *all* elements in the model (i.e., the set W of all possible worlds in the model). Thus, in terms of the agents' *propositional knowledge*, "knowing the protocol" amounts to knowing that "the set of possible states is W", but this just means that the agent knows that "$\top$". Nonetheless, "knowing the protocol" has important practical and pragmatic ramifications on the agents' information.[6] First, the protocol explicitly limits the observations, messages and/or actions available (or permitted) to the agent. Second, the protocol affects how the agents interpret their observations (Parikh and Ramanujam 2003).

This is an exploratory paper focused on ideas and concepts rather than on concrete results. I focus only on dynamic logics of knowledge and belief for a single agent. This is not because I do not find the many-agent situation interesting or important. Quite the opposite: I focus on a single agent only to simplify the exposition and technical details. Section 6.2 is a general introduction to the many

[5] See Parikh and Ramanujam (2003), Sect. 6, for a discussion of Gricean norms in this context.

[6] See Pacuit and Simon (2011), and references therein, for a logic to reasoning about what agents know about a protocol, or plan, that they are executing.

different flavors of dynamic epistemic and doxastic logics for non-specialists. Section 6.3 is an extended discussion of the role that procedural information plays in dynamic logics of belief revision. Finally, I offer some conclusions in Sect. 6.4.

6.2 A Primer on Logics of Informational Change

In this section, I introduce the key logical frameworks that incorporate how a (rational) agent's information changes in response to new information or evidence. This is a well-developed area attempting to balance sophisticated logical analysis with philosophical insight. Of course, I will not be able to do justice to the entire literature here, see van Benthem (2011) and references therein for a broad overview.

6.2.1 Static Models of Hard and Soft Information

The formal models introduced below can be broadly described as "possible worlds models", familiar in much of the philosophical logic literature. Setting aside any conceptual difficulties surrounding the use of these models, the structures I study in this paper are instances of a relational model:

Definition 6.2.1 (Relational Model). Let At be a (finite) set of atomic sentences. A **relational model** (based on At) is a tuple $\langle W, R, V\rangle$ where W is a finite set whose elements are called *possible worlds* or *states*; $R \subseteq W \times W$ is a relation; and $V : \mathsf{At} \to \wp(W)$ is a valuation function mapping atomic propositions to sets of states. □

Elements $p \in \mathsf{At}$ are intended to describe ground facts about the situation being modeled, such as "it is raining" or "the red card is on the table". A nonempty set W is intended to represent the different possible "scenarios" (elements of W are called possible worlds or states). The valuation function V associates with every ground fact the set of situations where that fact holds. Finally, the agent's informational attitude is defined in terms of the relation R. Different properties of R give rise to different types of attitudes. There are two types of attitudes that are important for this paper.

The first is the attitude that is associated with the agent's *hard* information. For lack of a better term (and following standard usage), I call this the agent's *knowledge*. In this case, I assume that R is an equivalence relation (i.e., reflexive, transitive and symmetric) and write '$\sim$' for R. Rather than *directly* representing the agent's *hard information*, the relation $\sim$ describes the "implicit consequences" of this information in terms of an "*epistemic indistinguishability relations*".[7] The idea

[7]The phrasing "epistemic indistinguishability", although common in the epistemic logic literature, is misleading since, as a relation, "indistinguishability" is *not* transitive. A standard example is: A

is that each agent has some "hard information" about the situation being modeled, and agents cannot distinguish between states that agree on this information. I call structures $\langle W, \sim, V\rangle$ an **epistemic model**.

A simple propositional modal language is often used to describe the agent's knowledge at states in an epistemic model. Formally, let $\mathcal{L}_{EL}$ be the (smallest) set of sentences generated by the following grammar:

$$\varphi \ := \ p \mid \neg\varphi \mid \varphi \wedge \varphi \mid K\varphi$$

where $p \in \mathsf{At}$ (the set of atomic propositions). The additional propositional connectives ($\rightarrow, \leftrightarrow, \vee$) are defined as usual and the dual of K, denoted L, is defined as follows: $L\varphi \ := \neg K \neg\varphi$. The intended interpretation of $K\varphi$ is "according to the agent's current (hard) information, φ is true" (again, I can also say that "the agent knows that φ is true"). Given a story or situation we are interested in modeling, each state $w \in W$ of an epistemic model $\mathcal{M} = \langle W, \sim, V\rangle$ represents a possible scenario which can be described in the formal language given above: If $\varphi \in \mathcal{L}_{EL}$, 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 6.2.2 (Truth). Let $\mathcal{M} = \langle W, \sim, V\rangle$ 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 $\mathcal{M}, w \models \psi$
- $\mathcal{M}, w \models K\varphi$ iff for all $v \in W$, if $w \sim v$ then $\mathcal{M}, v \models \varphi$ $\lhd$

The above epistemic models are intended to represent the agent's *hard information* about the situation being modeled. In fact, by using standard techniques from the mathematical theory of modal logic, I can be much more precise about the sense in which these models "represent" the agent's hard information. In particular, *modal correspondence theory* (see Blackburn et al. (2002, Chap. 3)) rigorously relates properties of the relation in an epistemic model with modal formulas (cf. Blackburn et al. 2002, Chap. 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.

These properties have generated much discussion among philosophers, computer scientists and logicians. While the logical omniscience assumption (which is valid

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 1,000 grains of sugar. In this context, two states are "epistemicly indistinguishable" for an agent if the agent has the "same information" in both states. This is indeed an equivalence relation.

[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 based on that frame. It can be shown that some modal formulas have first-order

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

on all models regardless of the properties of the accessibility relation) has generated the most extensive criticisms (cf. Stalnaker 1991) and responses (cf. Fagin et al. 1995, Chap. 9) the two introspection principles have also been the object 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 take epistemic models for what they are: formal models of hard information, in the sense introduced above.

The theory of belief *revision* started with the seminal paper by Alchourrón et al. (1985). In this paper, I focus on logical models of belief revision. The standard approach is to use a relational model where the relation is a *connected preorder* (reflexive and transitive). Such orders are typically called *plausibility orderings* and are denoted '$\preceq$'. While $\sim$ partitions the set of possible worlds according to the agent's hard information, the ordering $\preceq$ represents the possible worlds that the agent considers more plausible (i.e., it represents the agent's soft information). A **plausibility model** is a relational structure $\mathcal{M} = \langle W, \preceq, V\rangle$. David Lewis (1973) first used these structures as a semantics for *conditionals* (Grove 1988). These structures 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).

The richer models allows us to define a variety of (soft) informational attitudes. I first need some additional notation. For $X \subseteq W$, let

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

denote the set of minimal elements of X according to $\preceq$. This set is interpreted as the set of worlds the agent considers most plausible.[9] Also, the plausibility relation $\preceq$ can be *lifted* to subsets of W as follows[10]

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

correspondents P where for any frame $\langle W, R\rangle$, the relation R has property P iff φ is valid on $\langle W, R\rangle$.

[9]It is a convention in this literature that going down according to $\preceq$ corresponds to being *more* plausible. This is just a convention which can be easily changed.

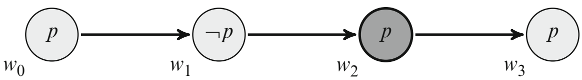
[10]This is only one of many possible choices here, but it is the most natural in this setting (cf. Liu 2008, Chap. 4).

Suppose that $\mathcal{M} = \langle W, \preceq, V \rangle$ is a plausibility model with $w \in W$, and consider the following modalities:

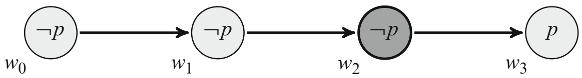
- *Belief*: $\mathcal{M}, w \models B\varphi$ iff for all $v \in Min_{\preceq}(W)$, $\mathcal{M}, v \models \varphi$.
  This is the usual notion of belief that satisfies the standard properties discussed above (e.g., positive and negative introspection).
- *Robust Belief*: $\mathcal{M}, w \models \Box\varphi$ iff for all v, if $v \preceq w$ then $\mathcal{M}, v \models \varphi$.
  Thus, φ is robustly believed if φ is true in *all* states that the agent considers more plausible. This stronger notion of belief has also been called *certainty* by some authors (cf. Shoham and Leyton-Brown 2009, Sect. 13.7).
- *Strong Belief*: $\mathcal{M}, w \models B^s\varphi$ iff there is a $v \in W$ such that $\mathcal{M}, v \models \varphi$ and $\{x \mid \mathcal{M}, x \models \varphi\} \preceq \{x \mid \mathcal{M}, x \models \neg\varphi\}$.
  So, φ is strongly believed provided it is epistemically possible and the agent considers *any* state satisfying φ more plausible than *any* state satisfying $\neg\varphi$. This notion has also been studied in Stalnaker (1994) and Battigalli and Siniscalchi (2002).
- *Knowledge*: $\mathcal{M}, w \models K\varphi$ iff for all $v \in W$, $\mathcal{M}, v \models \varphi$.
  Knowledge is interpreted as a universal modality here. The intuition is that the agent's plausibility ordering ranges over the states that the agent has not ruled out according to her hard information.

The logic of these notions has been extensively studied by Alexandru Baltag and Sonja Smets in a series of articles (Baltag and Smets 2006, 2008a, 2009). The following example illustrates the relationship between these different notions.

Example 6.2.3 (Relationships between the different notions of belief). It is not hard to see that if an agent knows p (Kp is true) then the agent believes p according to all the definitions above (i.e., $Kp \to (Bp \wedge \Box p \wedge B^s p)$ is valid). Furhtermore, both strong belief and robust belief in p implies the agent believes p. What about the relationship between strong belief and robust belief? These two notions of belief are logically independent. Consider the following plausibility model where $w_2 \models \Box p \wedge \neg B^s p$. I draw an arrow from v to w if $w \preceq v$ (to keep the notation down, I do not include all arrows. The remaining arrows can be inferred by transitivity).



To see that strong belief need not imply robust belief, consider the following variant of the above plausibility model where $w_2 \models B^s p \wedge \neg\Box p$:



□

As noted above, a crucial feature of these informational attitudes is that they are *defeasible* in light of new evidence. In fact, these attitudes can be characterized in terms of how an agent would change her beliefs in response to certain types of evidence. The notion of *conditional belief* is needed to make this idea precise. Suppose that $\mathcal{M} = \langle W, \preceq, V \rangle$ is a plausibility model and φ and ψ are formulas; then, we say that *the agent believes φ given ψ* (or *believes φ conditional on ψ*), denoted $B^{\psi\varphi}$, provided

$$\mathcal{M}, w \models B^{\psi\varphi} \text{ iff for all } v \in Min_{\preceq}(\llbracket\psi\rrbracket_{\mathcal{M}}), \mathcal{M}, v \models \varphi$$

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

$$\mathcal{M}, w \models K\varphi \text{ iff } \mathcal{M}, w \models B^{\psi}\varphi \text{ for all } \psi.$$

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

- $\mathcal{M}, w \models \Box\varphi$ iff $\mathcal{M}, w \models B^{\psi}\varphi$ for all ψ with $\mathcal{M}, w \models \psi$.
  That is, the agent safely believes φ iff she continues to believe φ given any true formula.
- $\mathcal{M}, w \models B^{s}\varphi$ iff $\mathcal{M}, w \models B\varphi$ and $\mathcal{M}, w \models B^{\psi}\varphi$ for all ψ with $\mathcal{M}, w \models \neg K(\psi \rightarrow \neg\varphi)$.
  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.

Baltag and Smets (2009) provide an elegant logical characterization of the above notions. First of all, note that conditional belief (and, hence, belief) and strong belief are *definable* in this language:

- $B^{\varphi\psi} := L\varphi \rightarrow L(\varphi \wedge \Box(\varphi \rightarrow \psi))$
- $B^{s}\varphi := B\varphi \wedge K(\varphi \rightarrow \Box\varphi)$

Thus, we can consider a modal language containing a universal modality (which I have called knowledge) and the usual modality for the plausibility ordering (which I have called robust belief). As discussed above, K satisfies logical omniscience, veracity and both positive and negative introspection. Safe belief, $\Box$, shares all of these properties except negative introspection. Modal correspondence theory can again be used to characterize the remaining properties:

- Knowledge implies safe belief: $K\varphi \rightarrow \Box\varphi$
- Connectedness: $K(\varphi \vee \Box\psi) \wedge K(\psi \vee \Box\varphi) \rightarrow K\varphi \vee K\psi$

6.2.2 *Dynamics of Beliefs*

The central issue here is how to incorporate *new* information into an epistemic or plausibility model. At a fixed moment in time, the agents are in some *epistemic state* (which may be described by an epistemic or plausibility model). The question is: How does (the model of) this epistemic state change during the course of some social interaction?

The most basic type of informational change is a so-called *public announcement* (Plaza 1989; Gerbrandy 1999). 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. the first aspect of informative actions discussed in the introduction). Furthermore, all agents treat the source as *infallible* (cf. the second aspect of informative actions discussed in the introduction). Thus, the effect of such an event on an epistemic or plausibility model should be clear: *Remove* all states that do not satisfy φ. Formally:

Definition 6.2.4 (Public Announcement). Suppose that $\mathcal{M} = \langle W, R, V\rangle$ is a relational model and φ is a formula (in the language of epistemic logic or conditional beliefs). The model updated by the **public announcement of** φ is the structure $\mathcal{M}^{\varphi} = \langle W^{\varphi}, R^{\varphi}, V^{\varphi}\rangle$ where $W^{\varphi} = \{w \in W \mid \mathcal{M}, w \models \varphi\}$, $R^{\varphi} = R \cap W^{\varphi} \times W^{\varphi}$, and for all atomic propositions p, $V^{\varphi}(p) = V(p) \cap W^{\varphi}$. ◻

It is not hard to see that if $\mathcal{M}$ is a relational model (i.e., an epistemic or plausibility model), then so is $\mathcal{M}^{\varphi}$. 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 agent and $\mathcal{M}^{\varphi}$ the information state *after* the information that φ is true has been incorporated in $\mathcal{M}$. This temporal dimension can also be represented in the logical language with modalities of the form $[!\varphi]\psi$. 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, let us focus on epistemic models and consider the formula $\neg K\psi \wedge [!\varphi]K\psi$: This says that "the agent (currently) does not know ψ, but after the announcement of φ, the agent knows ψ". So, languages with these announcement modalities can describe 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 *recursion axioms*:

$$\begin{array}{lll} [!\varphi]p & \leftrightarrow & \varphi \rightarrow p, \text{ where } p \in \mathsf{At} \\ [!\varphi]\neg\psi & \leftrightarrow & \varphi \rightarrow \neg[!\varphi]\psi \\ [!\varphi](\psi\wedge\chi) & \leftrightarrow & [!\varphi]\psi\wedge[!\varphi]\chi \\ [!\varphi]K\varphi & \leftrightarrow & \varphi \rightarrow K(\varphi \rightarrow [!\varphi]\psi) \end{array}$$

These recursion axioms can be used to show that the announcement modalities do not add any expressive power to the standard epistemic modal language (without common knowledge).[11] More than that, these recursion axioms provide an insightful syntactic analysis of announcements that complements the semantic analysis: The recursion axioms describe the effect of an announcement in terms of what is true before the announcement.

Now, what is the effect of a public announcement on the agents' soft information? I will start by clarifying the relationship between conditional belief $B^{\varphi}\psi$ and beliefs after a public announcement $[!\varphi]B\psi$. *Prima facie*, the two statements seem to express the same thing; and, in fact, they are equivalent provided that ψ is a *true ground formula* (i.e., does not contain any modal operators). However, the formulas are not equivalent in general: The reader is invited to check that $B^{p}(p \wedge \neg Kp)$ is satisfiable, but $[!p]B(p\wedge\neg Kp)$ is not satisfiable. The situation is nicely summarized as follows: "$B^{\psi}\varphi$ says that if the agent would learn φ, then she would come to believe that ψ was the case (before the learning)… $[!\varphi]B\psi$ says that after learning φ, the agent would come to believe that ψ is the case (in the worlds after the learning)" (Baltag and Smets 2008b, p. 2). Thus, the conditional beliefs *encode* how the agent's beliefs will change in the presence of new information. In particular, conditional beliefs are crucial for a recursion axiom analysis. Note that the above recursion axiom for knowledge is not valid when replacing K with B on plausibility models. We do, however, have the following recursion axioms (valid on the class of plausibility models):

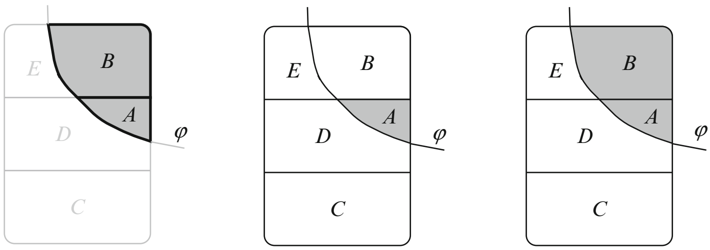
$$\begin{array}{lll} [!\varphi]B\psi & \leftrightarrow & B^{\varphi}[!\varphi]\psi \\ [!\varphi]B_i^{\psi}\chi & \leftrightarrow & (\varphi \rightarrow B_i^{\varphi\wedge[!\varphi]\psi}[!\varphi]\chi) \end{array}$$

There are also recursion axioms for robust and strong belief, but I do not discuss them here (see van Benthem (2011) for a discussion).

[11] This is not true for multiagent languages with a common knowledge operator. Nonetheless, a recursion axiom-style analysis is still possible, though the details are beyond the scope of this paper, see van Benthem et al. (2006).

A public announcement is only one type of informative action. It is an action where the agent is certain about *what* is being observed and treats the incoming information as infallible. Other types of informative actions can be defined by varying these these two aspects. In order to model situations where the agent is *misinformed* or *uncertain* about what she is observing, there must be a way to describe this uncertainty. Based on the logical framework introduced in Baltag et al. (1998), the key idea is to model such a complex epistemic event as a relational structure. I will not discuss this approach here (consult van Ditmarsch et al. (2007) for an overview of this approach). In this paper, I am primarily interested in informative actions where the source is trusted, but not necessarily treated as *infallible*.

As is well known from the belief revision literature, there are many ways to transform a plausibility model given some new information (Rott 2006). I do not have the space to survey this entire literature here (see van Benthem (2011) and Baltag and Smets (2009) for modern introductions). Instead, I will sketch some key ideas. The pictures below illustrate different ways that a plausibility model can incorporate φ.



$!(\varphi) : A \prec B$ $\quad\quad$ $\uparrow(\varphi): A \prec C \prec D \prec B \cup E$ $\quad\quad$ $\Uparrow(\varphi): A \prec B \prec C \prec D \prec E$

The general approach is to define a way of *transforming* a plausibility model given a formula φ. The operation on the left is the *public announcement* operation discussed above. For the other transformations, while the players do *trust* the source of φ, they do not treat the source as infallible. Perhaps the most ubiquitous policy is *conservative upgrade* ($\uparrow \varphi$), which lets the agent only tentatively accept the incoming information φ by making the best φ-worlds the new minimal set and keeping the old plausibility ordering the same on all other worlds. The operation on the right, *radical upgrade* ($\Uparrow \varphi$), is stronger, moving *all* φ worlds before all the $\neg\varphi$ worlds and otherwise keeping the plausibility ordering the same. I will make use of conservative upgrade in the next section, so I state the formal definition below:

Definition 6.2.5 (Conservative Upgrade). Given a plausibility model $\mathcal{M} = \langle W, \preceq, V\rangle$ and a formula φ, the *radical* upgrade of $\mathcal{M}$ with φ is the model $\mathcal{M}^{\uparrow\varphi} = \langle W^{\uparrow\varphi}, \preceq^{\uparrow\varphi}, V^{\uparrow\varphi}\rangle$ with $W^{\uparrow\varphi} = W$, $V^{\uparrow\varphi} = V$ and $\preceq^{\uparrow\varphi}$ is the smallest relation satisfying:

1. For all $x \in Min_{\preceq}(\llbracket\varphi\rrbracket_{\mathcal{M}})$ and $y \notin Min_{\preceq}(\llbracket\varphi\rrbracket_{\mathcal{M}})$, $x \prec^{\uparrow\varphi} y$;
2. For all $x, y \in Min_{\preceq}(\llbracket\varphi\rrbracket_{\mathcal{M}})$, $x \preceq^{\uparrow\varphi} y$; and
3. For all $x, y \notin Min_{\preceq}(\llbracket\varphi\rrbracket_{\mathcal{M}})$, $x \preceq^{\uparrow\varphi} y$ iff $x \preceq y$. ◁

These dynamic operations satisfy a number of interesting logical principles (van Benthem 2011; Baltag and Smets 2009), but a full discussion is beyond the scope of this paper.

6.3 Making the Protocol Explicit

A number of authors have forcefully argued that the underlying protocol (i.e., the procedural information) is an important component of any analysis of (social) interactive situations and should be explicitly represented in a formal model (cf. Fagin et al. 1995; van Benthem et al. 2009; Parikh and Ramanujam 2003; Hoshi 2009; Wang 2010). Indeed, much of the work over the past 20 years using epistemic logic to reason about distributed algorithms has provided interesting case studies highlighting the interplay between "protocol analysis" and epistemic reasoning (an important example here is the seminal paper by Halpern and Moses (1990) on the "generals problem").

The first observation is that the recursion axioms from Sect. 6.2.2 already illustrate the mixture of factual and *procedural* truth that drives conversations or processes of observation. 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 reduction axiom for knowledge (van Benthem et al. 2009, Sect. 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 represents a different assumption about the underlying protocol and how it affects the agent's knowledge. The first is the above recursion axiom (in 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 model (Definition 6.2.2) and public announcement (Definition 6.2.4). In order to see a difference, the *protocol information* must be explicitly represented in the model (see van Benthem et al. (2009) for a fuller discussion).

6.3.1 Protocol Information in Dynamic Logics of Belief Revision

The problem of *iterated revision* has been extensively studied (Boutilier 1996; Darwiche and Pearl 1997; Nayak et al. 2003; Stalnaker 2009), and although there are many proposals, there still remain a number of conceptual problems (see Stalnaker (2009) for a discussion). In this section, I focus on one such issue.

The main problem is this: Suppose that the agent receives a sequence of consistent formulas and uses, for example, radical upgrade to adjust her plausibility orderings. Since the information is consistent, no matter what the order in which she incorporates the information, she will always end up with the same beliefs. However, the different orders can lead to very different *conditional* beliefs, and this, in turn, means that there could be drastic differences in the result of incorporating information that contradicts one of the previous pieces of information.

Consider an example that has been extensively discussed in the literature. Suppose that you are in the forest and happen to see a strange-looking animal. You consult your animal guidebook and find a picture that seems to match the animal you see. The guidebook says that the animal is a type of bird, so that is what you conclude: The animal before you is a bird. After looking more closely, you also notice that the animal is red. So, you also update your beliefs with that fact. Now, suppose that an expert (whom you trust) happens to walk by and tells you that the animal is, in fact, not a bird. After incorporating this information into your beliefs (using conservative upgrade), you will no longer believe that the bird is red. Below is the sequence of upgrades (let b denote the proposition "the animal is a bird", $\overline{b}$ the negation of b, r is the proposition "the animal is red" and $\overline{r}$ the negation of r).

Note that in the last model, $\mathcal{M}_3$, the agent does not believe that the bird is red. The problem is that there does not seem to be any justification for why the agent drops her belief that the bird is red. There has been much discussion of this problem in the literature on iterated belief revision. Note that using radical upgrade, the agent would still believe the bird is red in $\mathcal{M}_3$ (as the reader is invited to check). My goal here is not to argue for or against one particular solution to this puzzle (see, for example Nayak et al. (2003, Sect. 5.1)). Rather, I want to highlight some general points about the underlying protocol specifying the order in which propositions are incorporated into the agent's epistemic state. In particular, the following sequence of updates is not problematic:

Of course, if we update the third model $\mathcal{M}_2$ with $\uparrow\overline{r}$, then the agent will drop her belief that b is true, which is equally problematic. This discussion highlights the importance of "procedural information" when reasoning about how an agent's beliefs change over time.

I conclude this section by introducing a logical framework that can reason about an agent's beliefs, and how her beliefs change in response to an explicit protocol describing which formulas (and types of updates) are available to her.

I start by being more precise about the definition of a protocol. A **tree** is a pair $\langle T, \rightarrowtail\rangle$ where T is a (finite) set of moments and $\rightarrowtail \subseteq T \times T$ satisfies the following properties:

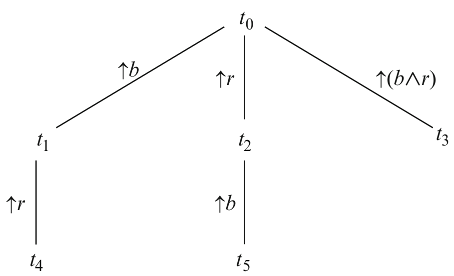
- For each $t_1, t_2, t_3 \in T$, if $t_1 \rightarrowtail t_2$ and $t_3 \rightarrowtail t_2$ then $t_1 = t_3$, and
- If $(t_1, \ldots, t_n)$ is a sequence in T with $t_i \rightarrowtail t_{i+1}$ for each $i = 1, \ldots, n-1$, then $t_n \neq t_1$.

If $t_1 \rightarrowtail t_2$, we say t_2 is an immediate successor of t_1. A path p in T starting at node t is a sequence $(t_1, \ldots, t_n)$ where $t_1 = t$, for each $i = 1, \ldots, n-1$, $t_i \rightarrowtail t_{i+1}$. We say a path $p = (t_1, \ldots, t_n)$ is maximal if t_n does not have any immediate successors.

A *protocol* describes the different ways in which an agent can incorporate available information into her beliefs. Formally, a protocol is a labeled tree where the edges are labeled with specific types of belief transformations.

Definition 6.3.1 (Protocol). A **protocol** for a language $\mathcal{L}$ and set of model transformations X is a tuple $\langle T, \rightarrowtail, l\rangle$ where $\langle T, \rightarrowtail\rangle$ is a tree and l assigns to each edge (i.e., pair (t, t') where t' is an immediate successor of t) a symbol $\tau(\varphi)$ where $\tau \in X$ is a model transformation and $\varphi \in \mathcal{L}$ is a formula. □

Let $\mathcal{P} = \langle T, \rightarrowtail, l\rangle$ be a protocol and $\mathcal{M} = \langle W, \preceq, V\rangle$ an initial plausibility model. The plausibility model at instant $t \in T$ is defined as follows by iteratively updating $\mathcal{M}$ according to the (unique) path in T leading to node t. Rather than giving a formal definition, I discuss an example. Consider the following protocol:



If $\mathcal{M}$ is the initial model in Fig. 6.1 (i.e., $\mathcal{M}_0$), then $\mathcal{M}_{t_4}$ is the model $\mathcal{M}_2$ in Fig. 6.1 and $\mathcal{M}_{t_5}$ is the model $\mathcal{M}_2$ in Fig. 6.2. We are interested in pairs $(\mathcal{M}_t, \mathcal{P})$ where t is a node in $\mathcal{P}$, and $\mathcal{M}_t$ is the model generated from an initial model $\mathcal{M}$ as described above.

The above protocol represents the different ways in which the agent from the previous example can go about incorporating the information that the animal she is looking at is a red bird. Why would a rational agent prefer one path over another in a given protocol? One answer might be that this is part of the description

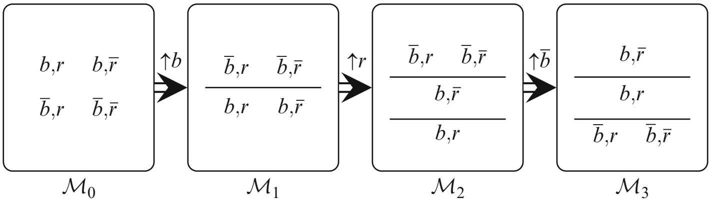


Fig. 6.1 A conservative upgrade sequence

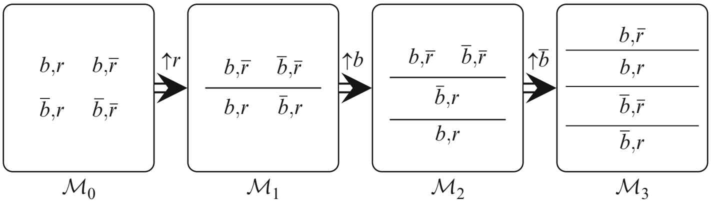


Fig. 6.2 Another conservative upgrade sequence

of the problem (i.e., that Ann first received the information that b and *then* the information that r). But this means that the agent has (implicitly or explicitly) agreed to conform to this specific protocol (a tree with a single branch with the labels $\uparrow b$ and $\uparrow r$), *not* to the protocol displayed above. The branching structure in a protocol represents situations where the agent has not (yet) committed to a particular way of incorporating the received evidence. Now, some beliefs might be *robust* in the sense that every (maximal) path in the protocol leads to a model where the agent has that belief. In the above protocol, all maximal paths lead to models (namely models $\mathcal{M}_{t_3}$, $\mathcal{M}_{t_4}$, and $\mathcal{M}_{t_5}$) where the agent believes that the animal is a red bird.

Of course, the situation becomes more interesting when the agent receives information that contradicts evidence found on some or all of the paths in the current protocol. This is the case when she receives the information that the animal is not a bird (denoted by $\overline{b}$). Rather than asking how the agent should incorporate this information into her current beliefs, we should ask how she should incorporate this information into her current protocol. One response would be to add $\uparrow \overline{b}$ at the end of all paths in the protocol. But other operations make sense. For example, a more cautious response would add an edge labeled by $\uparrow\overline{b}$ only to the node t_5. This analysis raises the following question: What are the natural operations on protocols and rational principles that these operations should conform to?

There are many temporal extensions of our basic doxastic language that one can use to reason about these structures (see Bonanno (2007) and Bonanno (2012); Dégremont (2010) for some examples). A complete account of these different logical systems will be left for future work. Here is one example: Include an operator '$\Diamond$' that quantifies over maximal paths in the protocol. Suppose that $\mathcal{M}$ is an initial plausibility model, $\mathcal{P}$ is a protocol, w is a state in $\mathcal{M}$ and t a moment in $\mathcal{P}$. Interpret formulas at pairs $(\mathcal{M}_t, \mathcal{P}, w)$ where $\mathcal{M}_t$ is defined as above (assuming the initial model is $\mathcal{M}$). The definition of the different informative attitudes (e.g., conditional beliefs) is as it is in Sect. 6.2.1. Here, I give only the definition of the new temporal operator:

- $\mathcal{M}_t, \mathcal{P}, w \models \Diamond\varphi$ provided that there exists a maximal path $p = (t, t_1, \ldots, t_n)$ such that $\mathcal{M}_{t_n}, \mathcal{P}_0, w \models \varphi$, where $\mathcal{P}_0$ is a single node protocol.

Thus, $\Diamond\varphi$ not only "moves time forward", but also "resets" the protocol.[12] Let $\Box$ be the dual of $\Diamond$ (i.e., $\Box\varphi$ is $\neg\Diamond\neg\varphi$). Then, $\Box\varphi$ means that φ is true after every way of updating beliefs consistent with the current protocol. But then we need some way to build up a protocol. One proposal is to reinterpret the dynamic modalities $[\Uparrow\varphi]$ as operations that change the protocol:

- $\mathcal{M}_t, \mathcal{P}, w \models [\uparrow\varphi]\psi$ iff $\mathcal{M}_t, \mathcal{P}^{\uparrow\varphi}, w \models \psi$, where $\mathcal{P}^{\uparrow\varphi}$ is the protocol that incorporates φ.

To make things concrete, suppose that $\mathcal{P}^{\uparrow\varphi}$ is the protocol that adds edges labeled by $\uparrow\varphi$ at *all* of the leave nodes in $\mathcal{P}$. This language can then express precisely what is puzzling about the example discussed in this section:

$$\Box Br \wedge [\uparrow\overline{b}]\neg\Box Br$$

The belief that the animal is red is *robust* in the given protocol, but after incorporating a proposition that is "irrelevant" to r (i.e., $\overline{b}$), this belief is no longer robust. This formula is true given the above protocol and the initial model where all four possible states are equally plausible.

These are only some initial ideas, but they illustrate the richness of the proposed framework. A complete logical analysis will be left for future work.

6.4 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 (informative) action. In interactive and learning 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 (are the sources of information *trustworthy*?);

[12] Of course, one could drop this assumption and assume that the protocol remains fixed. I do not pursue this line of inquiry here.

and other implicit assumptions about procedural information (reducing the number of possible observations). A key aspect of any formal model of a (social) interactive situation or situation of rational inquiry is the way it accounts for the

> …information about how I learn some of the things I learn, about the sources of my information, or about what I believe about what I believe and don't believe. If the story we tell in an example makes certain information about any of these things relevant, then it needs to be included in a proper model of the story, if it is to play the right role in the evaluation of the abstract principles of the model (Stalnaker 2009, p. 203).

I had two goals in this paper. First and foremost, I surveyed recent dynamic logics of belief revision (see van Benthem (2011) for full coverage of this topic). My second goal was to discuss why it is important to make explicit the underlying assumptions about the procedural information available to the agents in the situation being modeled. I also sketched some initial ideas of a logic for reasoning about this procedural information. There are a number of papers that explore the ideas touched on in this paper in much more detail. The interested readers is invited to consult Hoshi (2009), Wang (2010), van Ditmarsch et al. (2011), Rodenhäuser (2011) and Pacuit and Simon (2011) for more information.

References

T. Ågotnes, N. Alechina, The dynamics of syntactic knowledge. J. Log. Comput. **17**(1), 83–116 (2007)

C.E. Alchourrón, P. Gärdenfors, D. Makinson, On the logic of theory change: partial meet contraction and revision functions. J. Symb. Log. **50**, 510–530 (1985)

R. Aumann, Interactive epistemology I: knowledge. Int. J. Game Theory **28**, 263–300 (1999)

A. Baltag, S. Smets, Conditional doxastic models: a qualitative approach to dynamic belief revision, in *Proceedings of WOLLIC 2006, Electronic Notes in Theoretical Computer Science*, Stanford University, eds. by G. Mints, R. de Queiroz, 2006, Vol. 165, pp. 5–21

A. Baltag, S. Smets, The logic of conditional doxastic actions, in *New Perspectives on Games and Interaction*, eds. by R. van Rooij, K. Apt. Texts in Logic and Games (Amsterdam University Press, Amsterdam, 2008a), pp. 9–31

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

A. Baltag, S. Smets, ESSLLI 2009 course: dynamic logics for interactive belief revision (2009), Slides available online at: http://alexandru.tiddlyspot.com/#[[ESSLLI%2709%20Slides]]

A. Baltag, L. Moss, S. Solecki, The logic of common knowledge, public announcements and private suspicions, in *Proceedings of the 7th Conference on Theoretical Aspects of Rationality and Knowledge (TARK 98)*, Evanston, ed. by I. Gilboa, 1998, Morgan Kaufmann Publishers Inc. San Francisco, CA, USA, pp. 43–56

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

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

O. Board, Dynamic interactive epistemology. Games Econ. Behav. **49**, 49–80 (2004)

G. Bonanno, Axiomatic characterization of AGM belief revision in a temporal logic. Artif. Intell. **171**, 144–160 (2007)
G. Bonanno, Belief change in branching time: AGM-consistency and iterated revision. J. Philos. Log. **41**(1), 201–236 (2012). Manuscript
G. Bonanno, P. Battigalli, Recent results on belief, knowledge and the epistemic foundations of game theory. Res. Econ. **53**(2), 149–225 (1999)
C. Boutilier, Conditional Logics for Default Reasoning and Belief Revision. PhD thesis, University of Toronto, 1992
C. Boutilier, Iterated revision and minimal revision of conditional beliefs. J. Philos. Log. **25**, 262–304 (1996)
A. Brandenburger, The power of paradox: some recent developments in interactive epistemology. Int. J. Game Theory **35**, 465–492 (2007)
D. Christensen, Higher-order evidence. Philos. Phenomenol. Res. **81**(1), 185–215 (2010)
C.B. Cross, 'Can' and the logic of ability. Philos. Stud. **50**(1), 53–64 (1986)
A. Darwiche, J. Pearl, On the logic of iterated belief revision. Artif. Intell. **89**, 1–29 (1997)
C. Dégremont, The Temporal Mind. Observations on the Logic of Belief Change in Interactive Systems. PhD thesis, Institute for Logic, Language and Computation. (DS-2010-03, 2010)
P. Egré, B. Bonnay, Inexact knowledge with introspection. J. Philos. Log. **38**(2), 179–228 (2009)
D. Elgesem, The modal logic of agency. Nord. J. Philos. Log. **2**, 1–46 (1997)
R. Fagin, J. Halpern, Y. Moses, M. Vardi, *Reasoning about Knowledge* (MIT, Cambridge, 1995)
J. Gerbrandy, Bisimulations on Planet Kripke. PhD thesis, University of Amsterdam, 1999
A. Grove, Two modellings for theory change. J. Philos. Log. **17**, 157–170 (1988)
J. Halpern, *Reasoning About Uncertainty* (MIT, Cambridge, 2005)
J. Halpern, Y. Moses, Knowledge and common knowledge in a distributed environment. J. ACM **37**(3), 549–587 (1990)
J. Halpern, L. Rego, Reasoning about knowledge of unawareness revisited, in *Proceedings of Theoretical Aspects of Rationality and Knowledge (TARK'09)*, Stanford, ed. by A. Heifetz, 2009
V. Hendricks, Editor, special issue: "8 bridges between formal and mainstream epistemology". Philos. Stud. **128**(1), 1–227, 2006
V. Hendricks, *Mainstream and Formal Epistemology* (Cambridge University Press, Cambridge, 2006)
J. Hintikka, *Knowledge and Belief: An Introduction to the Logic of the Two Notions* (Cornell University Press, Ithaca, 1962)
J. Hintikka, *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, London, 2005)
J. Horty, *Agency and Deontic Logic* (Oxford University Press, Oxford/New York, 2001)
T. Hoshi, Epistemic Dynamics and Protocol Information. PhD thesis, Stanford University, 2009
W. Jamroga, T. Ågotnes, Constructive knowledge: what agents can achieve under imperfect information. J. Appl. Non-class. Log. **17**(4), 423–475 (2007)
P. Lamarre, Y. Shoham, Knowledge, certainty, belief and conditionalisation, in *Proceedings of the International Conference on Knowledge Representation and Reasoning*, Bonn, 1994, pp. 415–424
D. Lewis, *Counterfactuals* (Blackwell, Oxford, 1973)
F. Liu, Changing for the better: preference dynamics and agent diversity. PhD thesis, Institute for Logic, Language and Computation (ILLC), 2008
A. Nayak, M. Pagnucco, P. Peppas, Dynamic belief revision operators. Artif. Intell. **146**, 193–228 (2003)
E. Pacuit, S. Simon, Reasoning with protocols under imperfect information. Rev. Symb. Log. **4**, 412–444 (2011)
E. Pacuit, R. Parikh, E. Cogan, The logic of knowledge based obligation. Synthese **149**(2), 311–341 (2006)

R. Parikh, R. Ramanujam, A knowledge based semantics of messages. J. Log Lang. Inf. **12**, 453–467 (2003)

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

J. Plaza, Logics of public communications. Synthese **158**(2), 165–179 (2007)

B. Rodenhäuser, A logic for extensional protocols, J. Appl. Non-class. Log. **21**(3–4), 477–502 (2011)

H. Rott, Shifting priorities: simple representations for 27 itereated theory change operators, in *Modality Matters: Twenty-Five Essays in Honour of Krister Segerberg*, eds. by H. Lagerlund, S. Lindström, R. Sliwinski, Vol. 53. Uppsala Philosophical Studies, 2006, pp. 359–384

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

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

R. Stalnaker, On the evaluation of solution concepts. Theory Decis. **37**(42), 49–73 (1994)

R. Stalnaker, Iterated belief revision. Erkentnis **70**, 189–209 (2009)

J. van Benthem, Dynamic logic for belief revision. J. Appl. Non-class. Log. **17**(2), 129–155 (2007)

J. van Benthem, The information in intuitionistic logic. Synthese **167**, 329–348 (2009)

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

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

J. van Benthem, J. Gerbrandy, T. Hoshi, E. Pacuit, Merging frameworks for interaction. J. Philos. Log. **38**(5), 491–526 (2009)

H. van Ditmarsch, Prolegomena to dynamic logic for belief revision. Synth. Knowl. Ration. Action **147**, 229–275 (2005)

H. van Ditmarsch, W. van der Hoek, B.P. Kooi, *Dynamic Epistemic Logic*, Synthese Library (Springer, Dordrecht, 2007)

H. van Ditmarsch, S. Ghosh, R. Verbrugge, Y. Wang, Hidden protocols, in *Proceedings of the 13th Conference on Theoretical Aspects of Rationality and Knowledge (TARK)*, Groningen, ed. by K. Apt (ACM Digital Library, 2011) http://dl.acm.org/citation.cfm?id=645876

Y. Wang, Epistemic Modelling and Protocol Dynamics. PhD thesis, CWI, 2010

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