Principles of Knowledge, Belief and Conditional Belief

Guillaume Aucher

Principles of knowledge, belief and conditional belief

Guillaume Aucher

1 Introduction

Elucidating the nature of the relationship between knowledge and belief is an old issue in epistemology dating back at least to Plato. Two approaches to addressing this problem stand out from the rest. The first consists in providing a *definition* of knowledge, in terms of belief, that would somehow pin down the essential ingredient binding knowledge to belief. The second consists in providing a complete characterization of this relationship in terms of logical *principles* relating these two notions. The accomplishement of either of these two objectives would certainly contribute to solving this problem.

The success of the first approach is hindered by the so-called 'Gettier problem'. Until recently, the view that knowledge could be defined in terms of belief as 'justified true belief' was endorsed by most philosophers. This notion of justification, or "right to be sure" as Ayer called it (Ayer, 1956), was therefore the key element relating knowledge to belief, even though Ayer admitted that determining the general conditions needed to "have the right to be sure" would be too complicated, if at all posssible. Gettier's seminal three page paper presents two counterexamples which shatters this classical analysis (Gettier, 1963).[1] Following this publication, a large number of other definitions of knowledge were proposed, analyzed and refined in order to determine the additional clause needed to define knowedge in terms of belief. However, no consensus came out of this undertaking and the exact nature of the relationship between knowledge and belief remains to this day elusive (Lycan, 2006).

The second approach is related to the method employed by epistemic logicians such as Hintikka or Lenzen to better understand and "explicate" the notions of knowledge and

G. Aucher
Université de Rennes 1 – INRIA, Campus de Beaulieu, 35042 Rennes, France
E-mail: guillaume.aucher@irisa.fr

[1] One of these two examples is the following. Suppose that Smith has strong evidence that 'Jones owns a Ford' (1) (for instance, Jones has owned a Ford ever since Smith has known him). Then, because of (1) and by propositional logic, Smith is also justified in believing that 'Jones owns a Ford *or* his friend Brown is in Barcelona' (2), even if Smith has no clue where Brown is at the moment. However it turns out that Jones does not own a Ford and that by pure coincidence Brown is actually in Barcelona. Then, (a) (2) is true, (b) Smith believes (2), and (c) Smith is justified in believing (2). So Smith has a true and justified belief in (2). Intuitively, however, one could not say that Smith knows (2).

belief. In his seminal book (Hintikka, 1962), Hintikka examines the validity of various principles with the help of a logical model based on the Kripke semantics of modal logic. This publication sparked numerous discussions about the inherent properties of these epistemic notions, and a large spectrum of informational attitudes were explored ((Lenzen, 1978) provides a good overview of that period). Many axioms, viewed as reasoning principles, were proposed and discussed, especially interaction axioms relating the notions of knowledge and belief. This quest for reasoning principles somehow vanished in the 1980's when epistemic logic was taken over by computer scientists to address other problems related to various applications. In the early 1990's, however, new interaction axioms relating knowledge and *conditional* belief were elicited by some researchers in artificial intelligence (Moses and Shoham, 1993; Lamarre and Shoham, 1994).

To better grasp the relationship binding knowledge to belief, we review and examine in this paper the validity of the different axioms (and inference rules) relating knowledge to belief which have been proposed in the epistemic logic literature. In doing so, we are bound to encounter many of the problems that epistemic logic has had to face in its relatively short (modern) history. This paper is therefore more an exposition than a research paper. However, we will also contribute to this area by providing conditions under which the notion of belief can be formally defined in terms of knowledge, and vice versa. We will also prove that certain convoluted axioms dealing only with the notion of knowledge can be derived from understandable interaction axioms relating knowledge and *conditional* belief.

This paper is organized as follows. In Section 2, we will briefly describe the role that epistemic logic has played in the development of computer science (and to a lesser extent in philosophy). We will also set the modeling assumptions that will be used in the rest of the paper. Then, in Section 3, we will delve into our subject matter and review the most common epistemic principles (i.e. principles pertaining to the notion of knowledge) and doxastic principles (i.e. principles pertaining to the notion of belief) occurring in the epistemic logic literature. In Section 4, we will review the interaction principles relating knowledge to belief on the one hand, and relating knowledge to conditional belief on the other hand. In Section 5, we provide the logical apparatus needed to formalize our approach. In Section 6, we will investigate formally under which conditions knowledge can be defined in terms of belief, and vice versa. Finally, in Section 7, we will show that certain convoluted axioms for knowledge can be derived from simpler axioms of interaction between knowledge, belief and conditional belief.

2 Prolegomena

2.1 Epistemic logic in philosophy and computer science

Following the publication of (Hintikka, 1962), philosophers and logicians tried to formulate explicit principles governing expressions of the form "a knows that φ" (subsequently formalized as $K\varphi$) or "a believes that φ" (subsequently formalized as $B\varphi$), where a is a human agent and φ is a proposition. In other words, philosophers sought to determine 'the' logic of knowledge and belief. This quest was grounded in the observation that our intuitions of these epistemic notions comply to some systematic reasoning properties, and was driven by the attempt to better *understand* and *elucidate* them. Lenzen indeed claims,

following Hintikka, that the task of epistemic logic consists "1) in *explicating* the epistemic notions, and 2) in examining the validity of the diverse principles of epistemic logic *given such an explication*" (Lenzen, 1978, p. 15). As we shall see in the rest of this paper, assessing whether a given principle holds true or not does raise our own awareness of these epistemic notions and reveals to us some of their essential properties.

For many computer scientists, reaching such an understanding (via this kind of conceptual analysis) is not as central as for philosophers, partly because the agents considered in the "applications" of computer science are typically assumed to be non-human. Voorbraak even claims that his notion of objective knowledge "applies to *any* agent which is capable of processing information [and] may very well be a device like a thermostat or a television-receiver" (Voorbraak, 1993, p. 55). In computer science, epistemic logic is often viewed as a *formal tool* used to represent uncertainty in different kinds of settings.[2] From this perspective, a specific set of axioms and inference rules for knowledge and belief will apply to a specific applied context. Originally, this interest taken in epistemic logic by computer scientists stemmed from their observation that the notion of knowledge plays a central role in the informal reasoning used especially in the design of distributed protocols. So, in a sense, the logical analysis of epistemic notions carried out by logicians and philosophers provided computer scientists with formal models. These models were 'used' and developped further to address particular issues such as the problem of reaching an agreement in order to coordinate actions in distributed systems (Halpern and Moses, 1990) or the problem of diagnosing electric circuits (Friedman and Halpern, 1997), or even problems of computer security (Deschene and Wang, 2010). As a result of this shift, the computability properties of various epistemic logics were investigated systematically (Halpern and Moses, 1992) and other epistemic notions involving multiple agents were introduced in epistemic logic (Fagin et al., 1995), such as the notion of *distributed knowledge* and *common knowledge* (originally studied by the philosopher Lewis (Lewis, 1969)).

One should note that there is also a discrepancy between the analyses of the notion of knowledge in epistemic logic and in (mainstream) epistemology. As Castañeda already commented soon after the publication of (Hintikka, 1962), "Hintikka's 'K' ('B') does not seem to correspond to any of the senses of 'know' ('believe') that have been employed or discussed by philosophers. But Hintikka's systems are an excellent source from which we may eventually arrive at a formalization of the philosophers' senses of 'know' and 'believe' " (Castañeda, 1964, p. 133). As it turns out, some recent publications bear witness to a revival of the ties between epistemic logic and (mainstream) epistemology (Hendricks and Symons, 2006; Hendricks, 2005).

2.2 Modeling assumptions

If we want to *define* knowledge in terms of belief and give a complete and accurate account of this notion, then we should not limit our analyses to knowledge and belief only. Indeed, other related notions will inevitably play a role, such as the notions of justification,

[2] (Fagin et al., 1995) and (Meyer and van der Hoek, 1995) are the standard textbooks in computer science dealing with epistemic logic. Also see the survey (Gochet and Gribomont, 2006) for a more interdisciplinary approach and (Halpern, 2003) for a broader account of the different formalisms dealing with the representation of uncertainty.

(un)awareness, or even epistemic surprise.[3] In that respect, note that other related mental states such as goal, desire and intention are also necessary if we want to develop logics for rational agents (such as a computer program, a software or a machine) that need to act on their environment so as to reach certain goals (possibly in cooperation with other agents).[4] Nevertheless, if we are only interested in elucidating the nature of the relationship binding knowledge to belief, then it is possible to abstract away from these related notions and identify *principles* relating knowledge and belief only.

This said, we have to be a bit more explicit and accurate about the kind of principles we are interested in and also about the modeling assumptions we adopt. Firstly, these principles have to be interpreted as analytically true relations between the notions of knowledge and belief. This means that we do not take into account the pragmatic conditions of their utterance. Therefore, the fact that I cannot reasonably utter the so-called Moore sentence 'proposition p holds but I do not believe that p', or the fact that from the mere utterance of 'I know that p', the listener can only infer that I believe that p, will not be explained. For an account of these pragmatic issues, the interested reader can consult (Lenzen, 2004). Consequently, we depart from the approach developped in (Hintikka, 1962), because Hintikka studies what he calls epistemic "statements". According to Hintikka, "a statement is the act of *uttering, writing*, or otherwise *expressing* a declarative sentence. A sentence is the form of words which is uttered or written when a statement is made" (Hintikka, 1962, p. 6) (my emphasis). On the other hand, our choice of assumptions is supported by Lenzen's claim that "one may elaborate the meaning of epistemic expressions in a way that is largely independent of [...] the pragmatic conditions of utterability" (Lenzen, 2004, p. 17). Secondly, throughout this paper and as it is usually implicitly assumed in epistemic logic, we shall follow a perfect external approach. This means that the epistemic state of the agent under consideration is modeled from the point of view of an external modeler who has perfect and complete access to and knowledge of this state.[5] Therefore, the principles pertain to an agent other than the modeler who states them. Finally, as is often the case in epistemic logic, we shall be interested only in *propositional knowledge*, that is knowledge *that* something holds, in contrast to non-propositional knowledge, that is, knowledge *of* something (such as the knowledge of an acquaintance or a piece of music), and in contrast to knowledge of how to do something.[6]

3 Epistemic and doxastic principles

In this section, we will briefly review the most common principles of the logics of knowledge and belief that occur in the literature (spelled out in the form of axioms and inference

[3] The notion of justification is dealt with in the field of justification logic (Artemov and Fitting, 2011). Logical models of (un)awareness have been proposed in economics (Heifetz et al., 2006) and artificial intelligence (Fagin and Halpern, 1987) with a recent proposal in (Halpern and Rêgo, 2009). Some models for the notion of epistemic surprise can be found in (Aucher, 2007) and (Lorini and Castelfranchi, 2007).

[4] There are a number of logical frameworks that deal with rational agency: Cohen and Levesque's theory of intention (Cohen and Levesque, 1990), Rao and Georgeff's BDI architecture (Georgeff and Rao, 1991) (Rao and Georgeff, 1991), Meyer et. al.'s KARO architecture (van Linder et al., 1998) (Meyer et al., 2001), Wooldridge's BDI logic LORA (Wooldridge, 2000) and Broersen et. al's BOID architecture (Broersen et al., 2001)

[5] See (Aucher, 2010) for more details on the perfect external approach and its connection with the other modeling approaches, namely the internal and the imperfect external approaches.

[6] (Gochet, 2007) reviews the various attempts to formalize the notion of *knowing how* in artificial intelligence and logic.

rules). They have all been commented on and discussed extensively in the philosophical literature, and the interested reader can consult (Lenzen, 1978) for more details. That said, there is currently no real consensus in favour of any proposed set of epistemic principles, even among computer scientists.

3.1 Epistemic principles

Any normal modal logic contains the axiom K and the inference rule Nec. Hintikka's epistemic logic is no exception:

$$(K(\psi \to \varphi) \wedge K\psi) \to K\varphi \qquad \text{(K)}$$

$$\text{If } \varphi \text{ then } K\varphi \qquad \text{(Nec)}$$

Axiom K and rule Nec have been attacked ever since the beginning of epistemic logic. They state that the agent knows all tautologies (Nec) and knows all the logical consequences of her knowledge (K). This can indeed be considered as a non-realistic assumption as far as human agents are concerned, but it is also a problem in numerous applications of epistemic logic. In the context of computer security, we may want, for example, to reason about computationally bounded adversaries to determine whether or not they can factor a large composite number (Halpern and Pucella, 2002). It is not possible, however, to perform such reasoning if we assume that the adversary's knowledge complies to axiom K and inference rule Nec.[7] This problem, named the "logical omniscience problem", turns out to be one of the main problems in epistemic logic, and numerous and various proposals have been made over the years in order to solve it. It undermines not only the notion of knowledge but also the notion of belief (because, as we shall see, this notion also complies with the principles K and Nec). In this context, the notion of awareness plays an important role and it is also relevant to distinguish between *implicit* knowedge/belief and *explicit* knowledge/belief. An agent's implicit knowledge includes the logical consequences of her explicit knowledge (Levesque, 1984). We refer the interested reader to (Fagin et al., 1995, Chap. 9), (Gochet and Gribomont, 2006, p. 157-168) or (Halpern and Pucella, 2011) for more details on the logical omniscience problem.

Hintikka further claims in (Hintikka, 1962) that the logic of knowledge is S4, which is obtained by adding to K and Nec the axioms T and 4:

$$K\varphi \to \varphi \qquad \text{(T)}$$

$$K\varphi \to KK\varphi \qquad \text{(4)}$$

These axioms state that if the agent knows a proposition, then this proposition is true (axiom T for Truth), and if the agent knows a proposition, then she knows that she knows it (axiom 4, also known as the "KK-principle"or "KK-thesis"). Axiom T is often considered to be the hallmark of knowledge and has not been subjected to any serious attack. In epistemology, axiom 4 tends to be accepted by internalists, but not by externalists (Hemp, 2006) (also see (Lenzen, 1978, Chap. 4)). A persuasive argument against this axiom has been propounded by Williamson in (Williamson, 2000, Chap. 5) for the case of *inexact knowledge*, that is, knowledge that obeys a *margin for error principle*. The knowledge that

[7] See (Deschene and Wang, 2010) for a survey of approaches to computer security issues which use epistemic logic.

one gains by looking at a distant tree in order to know its height is an example of inexact knowledge. A solution to Williamson's luminosity paradox is proposed in (Bonnay and Egré, 2008) by resorting to a particular semantics for modal logic called "centered semantics", which validates axiom 4 without requiring the accessibility relation to be transitive. Axiom 4 is nevertheless widely accepted by computer scientists (but also by many philosophers, including Plato, Aristotle, Saint Augustine, Spinoza and Shopenhauer, as Hintikka recalls in (Hintikka, 1962)).

A more controversial axiom for the logic of knowledge is axiom 5:

$$\neg K\varphi \rightarrow K\neg K\varphi \tag{5}$$

This axiom states that if the agent does not know a proposition, then she knows that she does not know it. This addition of 5 to S4 yields the logic S5. Most philosophers (including Hintikka) have attacked this axiom, since numerous examples from everyday life seem to invalidate it. For example, assume that a university professor believes (is certain) that one of her colleague's seminars is on Thursday (formally Bp). She is actually wrong because it is on Tuesday ($\neg p$). Therefore, she does not know that her colleague's seminar is on Tuesday ($\neg Kp$). If we assume that axiom 5 is valid then we should conclude that she knows that she does not know that her colleague's seminar is on Tuesday ($K\neg Kp$) (and therefore she also believes that she does not know it: $B\neg Kp$). This is obviously counterintuitive. More generally, axiom 5 is invalidated when the agent has mistaken beliefs which can be due for example to misperceptions, lies or other forms of deception.[8] As it turns out, this axiom is often used by computer scientists because it fits very well with the assumptions they have to make in most of the applied contexts they deal with.

Finally, we examine an axiom which has not drawn much attention in epistemic logic. This axiom plays, however, a central role in the logic of the notion of 'being informed' which has recently been introduced in (Floridi, 2006).

$$\varphi \rightarrow K\neg K\neg\varphi \tag{B}$$

Axiom B states that if φ is true, then the agent knows that she considers it possible that φ is true. In other words, it cannot be the case that the agent considers it possible that she knows a false proposition (that is, $\neg(\neg\varphi \wedge \neg K\neg K\varphi)$). As pointed out in (Floridi, 2006), the validity of this axiom embeds a 'closed world assumption' similar to the assumption underlying the validity of axiom 5. As a matter of fact, adding axiom B to the logic S4 yields the logic S5. To be more precise, the sets $\{\mathsf{T},\mathsf{B},4\}$ and $\{\mathsf{T},5\}$ are logically equivalent. Therefore, if we assume that axioms T and 4 are valid, then axiom B falls prey to the same attack as the one presented in the previous paragraph, since in that case we can derive axiom 5. We may wonder if a similar argument against axiom B holds in the logic KTB, that is, if we drop axiom 4. Wheeler argues that it is indeed the case (Wheeler, 2012).[9]

The logic KTB (also known as B or Br or Brouwer's system) has been propounded by Floridi as the logic of the notion of 'being informed'. One of the main differences between the logic of this notion and the standard logic of knowledge is the absence of introspection

[8] (Sakama et al., 2010) and (van Ditmarsch et al., 2011) provide two independent logical accounts of the notion of lying and other kinds of deception using epistemic logic (resp. dynamic epistemic logic).

[9] Wheeler's argument against axiom B is based on two theorems derivable in the logic KTB. One of them is the following: $K(\varphi \rightarrow K\psi) \rightarrow (\neg K\neg\varphi \rightarrow \psi)$. If φ stands for 'the agent sees some smoke' and ψ stands for 'there is fire', then the consequent of this theorem states that if the agent considers it possible that he sees some smoke (without necessarily being sure of it), then there is fire. This conclusion is obviously counterintuitive.

(which is characterized by axiom 4). Floridi claims that his results "pave the way [...] to the possibility of a non-psychologistic, non-mentalistic and non-anthropomorphic approach to epistemology, which can easily be applied to artificial or synthetic agents such as computers, robots, webbots, companies, and organizations" (Floridi, 2006, p. 456). In that respect, his notion of 'being informed' is similar to Voorbraak's notion of objective knowledge, since, as we already mentioned in Section 2.1, objective knowledge "applies to *any* agent which is capable of processing information [and] may very well be a device like a thermostat or a television-receiver" (Voorbraak, 1993, p. 55). The claim that the notion of 'being informed' is an independent cognitive state which cannot be reduced to knowledge or belief has been attacked recently by Wheeler (Wheeler, 2012). His attack is based on the argument against axiom B sketched in Footnote 9 (where the notion of knowledge is replaced with the notion of being informed).

3.2 Doxastic principles

We have to be careful with the notion of belief, since the term 'belief' refers to different meanings: my belief that it will rain tomorrow is intuitively different from my belief that the Fermat-Wilson theorem is correct. This intuitive semantic difference that anyone can perceive stems from the fact that the doxastic strength of these two beliefs are not on the same 'scale'.

3.2.1 Weak and strong belief

Lenzen argues in (Lenzen, 1978) that there are two different kinds of belief, which he calls *weak* and *strong* belief (or *conviction*). We will now explain (succintly) the difference between these two types of belief.[10]

Weak belief. Assume that the agent conjectures an arithmetical theorem φ from a series of examples and particular cases she has examined. The more examples the agent will have checked, the more she will 'believe' that this theorem holds true. We can naturally give a probabilistic semantics to this notion of belief and define a corresponding belief operator as follows:

$$B^r_w\varphi \triangleq Prob(\varphi) > r$$

where r is a real number ranging over the interval [0.5;1[. It is read as 'the agent believes, at least to the degree r, that φ'. The formula $Prob(\varphi)$ represents the subjective probability the agent assigns to the likelihood of φ; the bigger r is, the more the agent 'believes' in φ. It turns out that the reasoning principles validated by this notion of belief do not depend on the value of r. In particular, the principle $(B^r_w\varphi \wedge B^r_w\psi) \rightarrow B^r_w(\varphi \wedge \psi)$ is *not* valid. For $r = 0.5$, this notion of belief is called *weak belief* in (Lenzen, 1978); we denote it here as $B_w\varphi$ and it stands for 'the agent *w*eakly believes φ' or 'the agent thinks φ more probable than not'. Note that this modal operator is studied from a logical point of view in (Herzig, 2003). Instead of resorting to probability to represent this continuum of degrees of belief,

[10] A relatively more detailed analysis distinguishing weak from strong belief is also presented in (Shoham and Leyton-Brown, 2009, p. 414-415). Also see (Lenzen, 1978).

we could also define a graded belief modality $B^n_w\varphi$, standing for 'the agent *w*eakly believes with degree at most n that φ', where n is a natural number.[11] A semantics for this modality based on Ordinal Conditional Functions (OCF) as introduced in (Spohn, 1988a) is proposed in (Aucher, 2004; Laverny and Lang, 2005; van Ditmarsch, 2005). However, the intended interpretation of OCF in these papers deviates from Spohn's intended interpretation, resulting in a definition of the graded belief modality which confuses the notions of weak and strong belief. As it turns out, the principle $(B^n\varphi \wedge B^n\psi) \rightarrow B^n(\varphi \wedge \psi)$ *is* valid with this OCF-based semantics, unlike with probabilistic semantics.

Strong belief. Now, if the agent comes up with a proof of this arithmetical theorem that she has checked several times, she will still 'believe' in this theorem, but this time with a different strength. Her belief will be a conviction, a certainty:

$$B\varphi \triangleq Prob(\varphi) = 1.$$

That said, her certainty might still be erroneous if there is a mistake in the proof that she did not notice. We will denote this second type of belief with the formula $B\varphi$ and read it as 'the agent strongly believes (is certain) that φ'.[12] Unlike weak belief (defined over a probabilistic semantics), strong belief validates the following axiom:

$$(B\varphi \wedge B\psi) \rightarrow B(\varphi \wedge \psi). \qquad (\mathsf{K'})$$

Note also that, according to probability theory, strong belief entails weak belief:

$$B\varphi \rightarrow B_w\varphi \qquad (\mathsf{BB_w1})$$

and that

$$B_w\varphi \wedge B\psi \rightarrow B_w(\varphi \wedge \psi). \qquad (\mathsf{BB_w2})$$

This notion of strong belief is also sometimes called *plain* belief (Spohn, 1988b) or *acceptance* (Gärdenfors, 1988).

Remark 1 The notions of weak and strong belief are often confused in the literature. This may lead to apparent paradoxes such as the lottery paradox (Kyburg, 1961). Weak and strong beliefs are indeed intertwined in the formulation of this paradox. Once these two notions are clearly identified and separated, the paradox vanishes. As Lenzen writes, "Consider a fair lottery with n tickets, only one of which is the winning ticket. For each ticket j, the chance of j being the winning ticket then is $\frac{1}{n}$. Thus, any individual a whose subjective expectation accords with the objective probabilities will have to presume [...] that j is not the winning ticket, $B_w\neg p_1 \wedge \ldots \wedge B_w\neg p_n$. But since a knows that one ticket will win, he *a fortiori* believes (strongly) that one ticket will win, $B(p_1 \vee \ldots \vee p_n)$. Hence his set of believings is neither consistent nor deductively closed." (Lenzen, 1978, p. 38)

[11] One should not confuse these graded belief modalities with the graded modalities $M_n\varphi$ found in (Fine, 1972; de Rijke, 2000; van der Hoek and Meyer, 1992). Indeed, the intended interpretation of $M_n\varphi$ is 'there are more than n accessible worlds that verify φ'.

[12] The modal operators of weak and strong belief are denoted "$B\varphi$" and "$C\varphi$" respectively in (Lenzen, 1978).

Conditional belief. The description of the agent's doxastic state can be enriched if we also consider what the agent *would* believe if she was confronted with new evidence about the current situation. This has led Lamarre and Shoham in (Lamarre and Shoham, 1994) to define two operators of conditional belief, $B^{\psi}\varphi$ and $B^{\psi}_{w}\varphi$.[13] The semantics of these operators of conditional belief is based on the semantics of default statements.

A default statement $\psi \supset \varphi$ can be read in various ways: 'if ψ holds, then typically φ holds' or 'if ψ, then by default φ'. The authors of (Friedman and Halpern, 1997) and (Lamarre and Shoham, 1994) interpret a default statement $\psi \supset \varphi$ as a conditional belief statement: 'the agent believes φ, given assumption ψ' or more precisely 'if ψ were announced to the agent, she would believe that φ held (before the announcement)'. Given this intended interpretation, the notion of strong belief $B\varphi$ (resp. weak belief) corresponds in this richer setting to the formula $B^{\top}\varphi$ (resp. $B^{\top}_{w}\varphi$). This epistemic interpretation of a default statement, and hence also of its underlying logical semantics, is meaningful. It is grounded in the relations set up in (Makinson and Gärdenfors, 1989) between AGM theory of belief change (Alchourrón et al., 1985; Gärdenfors, 1988) and default logic. This epistemic interpretation is also supported by the fact that the famous *Ramsey test* basically defines belief revision in terms of default logic. Indeed, the idea of the Ramsey test is that an agent should believe φ after learning ψ if and only if he currently believes that φ would be true if ψ were true (i.e. $\psi \supset \varphi$).

This notion of conditional belief gives rise in turn to a derived doxastic notion called "safe belief" by Baltag and Smets (Baltag and Smets, 2006, 2008a,b). A safe belief in φ is expressed by the formula $B^{\neg\varphi}\bot$. This notion corresponds intuitively to a belief which cannot be defeated by any assumption. It is therefore very close to the definition of knowledge as undefeated true belief proposed by Lehrer and Paxton in (Lehrer and Paxson, 1969), the only difference being that their notion of knowledge cannot be defeated by any *true* assumption. Originally introduced for technical reasons by Boutilier (Boutilier, 1994) to deal with defeasible reasoning, this operator of safe belief has been reintroduced recently in the context of *dynamic* epistemic logic (van Ditmarsch et al., 2007; van Benthem, 2011) together with the notions of "hard" and "soft" information, in order to deal with belief revision (unlike "hard" information, "soft" information is revisable).[14]

Remark 2 If we added dynamics to our framework, as in (Baltag and Smets, 2006, 2008a,b), then we would also have formulas of the form $[\psi!]B\varphi$, whose reading would be '*after* the announcement of ψ, the agent believes φ'. This reading is different from the (extended) reading of our formulas $B^{\psi}\varphi$: 'if ψ were announced to the agent, she would believe that φ held *before* the announcement'. The latter operator is a revision of the agent's beliefs about the state of the world as it was *before* the announcement, and the former is a revision of the state of the world as it is *after* the announcement. Note, however, that this important distinction between *static* belief revision and *dynamic* belief revision collapses in the case of propositional formulas ψ, which most interests us here.

For the rest of the paper, we will be interested only in the notion of strong belief (certainty) and its conditional version. We will show that convoluted axioms for knowledge such as .3 and .3.2, which can hardly be expressed in terms of intuitive interaction axioms

[13] These two operators are respectively denoted "$C^{\psi}\varphi$" and "$B^{\psi}\varphi$" in (Lamarre and Shoham, 1994).

[14] For more details, see (van Benthem, 2007; Baltag and Smets, 2006, 2008a,b; van Benthem, 2011) and also (Pacuit, 2012) in this book.

dealing with strong beliefs only, can be expressed in terms of interaction axioms dealing with *conditional* beliefs, which are easier to grasp.

3.2.2 Principles of strong belief

The logic of (strong) belief is less controversial than the logic of knowledge. It is usually considered to be KD45, which is obtained by adding to the axiom K and inference rule Nec (where the knowledge operator is replaced by the belief operator) to the following axioms D, 4 and 5:

$$B\varphi \rightarrow \neg B\neg\varphi \tag{D}$$
$$B\varphi \rightarrow BB\varphi \tag{4}$$
$$\neg B\varphi \rightarrow B\neg B\varphi \tag{5}$$

Axioms 4 and 5 state that the agent has positive and negative introspection over her own beliefs. Some objections have been raised against Axiom 4 (see (Lenzen, 1978, Chap. 4) for details). Axiom D states that the agent's beliefs are consistent. In combination with axiom K (where the knowledge operator is replaced by a belief operator), axiom D is in fact equivalent to a simpler axiom D' which conveys, maybe more explicitly, the fact that the agent's beliefs cannot be inconsistent ($B\bot$):

$$\neg B\bot \tag{D'}$$

In all the theories of rational agency developed in artificial intelligence (and in particular in the papers cited in Footnote 4), the logic of belief is KD45. Note that all these agent theories follow the perfect external approach. This is at odds with their intention to implement their theories in machines. In that respect, an internal approach seems to be more appropriate since, in this context, the agent needs to reason from its own internal point of view. For the internal approach, the logic of belief is S5, as proved in (Aucher, 2010) and (Arlo-Costa, 1999) (for the notion of *full belief*).[15]

3.2.3 Principles of conditional belief

The axioms and inference rules of an axiomatic system called system P form the core of any axiomatic system that deal with non-monotonic reasoning. A generalized version of this system (taken from (Friedman and Halpern, 1997)), which allows us to express boolean combinations of default statements is reproduced below (we omit modus ponens and all the substitution instances of propositional tautologies). We recall that $B^{\psi}\varphi$ reads as 'the agent (strongly) believes φ, given assumption ψ' or more precisely 'if ψ were announced to the agent, she would believe that φ held (before the announcement)'. We

[15] In both philosophy and computer science, there is formalization of the internal point of view. Perhaps one of the dominant formalisms for this is auto-epistemic logic (R.C.Moore, 1984, 1995). In philosophy, there are models of full belief like the one offered by Levi, (Levi, 1997) which is also related to ideas in auto-epistemic logic. See (Aucher, 2010) for more details on the internal approach and its connection to the other modeling approaches, namely the imperfect and the perfect external approaches.

leave the reader to find out the natural intuitions underlying these axioms and inference rules.

$$B^{\psi}\psi \tag{C1}$$
$$(B^{\psi}\varphi_1 \wedge B^{\psi}\varphi_2) \rightarrow B^{\psi}(\varphi_1 \wedge \varphi_2) \tag{C2}$$
$$(B^{\psi_1}\varphi \wedge B^{\psi_2}\varphi) \rightarrow B^{\psi_1 \vee \psi_2}\varphi \tag{C3}$$
$$(B^{\psi}\varphi \wedge B^{\psi}\chi) \rightarrow B^{\psi \wedge \varphi}\chi \tag{C4}$$
$$\text{If } \psi \leftrightarrow \psi' \text{ then } B^{\psi}\varphi \leftrightarrow B^{\psi'}\varphi \tag{RC1}$$
$$\text{If } \varphi \rightarrow \varphi' \text{ then } B^{\psi}\varphi \rightarrow B^{\psi}\varphi' \tag{RC2}$$

Note that axiom C2 is an indication that this notion of conditional belief is a generalization of the notion of *strong* belief rather than *weak* belief, since, as we have already noted, $(B\varphi \wedge B\psi) \rightarrow B(\varphi \wedge \psi)$ holds, but $(B_w\varphi \wedge B_w\psi) \rightarrow B_w(\varphi \wedge \psi)$ does not hold in general (at least for the probabilistic semantics of weak belief).

4 Principles of interaction

In this section, we will set out the interaction axioms which have been proposed and discussed in the epistemic logic literature and which connect the notions of belief or conditional belief with the notion of knowledge. We will start by reviewing interaction axioms that deal with strong belief, and then we will consider interaction axioms that deal with *conditional* belief. Note that a classification of certain interaction principles has been proposed in (van der Hoek, 1993).[16]

4.1 Principles of interaction with strong belief

The following interaction axioms are suggested by Hintikka (Hintikka, 1962) and are often encountered in the literature:

$$K\varphi \rightarrow B\varphi \tag{KB1}$$
$$B\varphi \rightarrow KB\varphi \tag{KB2}$$

Axiom KB1 is a cornerstone of epistemic logic. Just as axiom T, it follows from the classical analysis of knowledge of Plato presented in the Theaetethus. It turns out that axiom KB1 is rejected in Voorbraak's logic of objective knowledge, because his notion of knowledge does not necessarily apply to humans, but rather applies in general to any information-processing device. It is adopted by Halpern in (Halpern, 1996), but only for propositional formulas φ. Axiom KB2 highlights the fact that the agent has "privileged access" to his doxastic state. If we assume, moreover, that the axioms D, 4, 5 for belief

[16] The classification is as follows. If X, Y, Z are epistemic operators, $X\varphi \rightarrow YZ\varphi$ are called *positive introspection formulas*, $\neg X\varphi \rightarrow Y\neg Z\varphi$ are called *negative introspection formulas*, $XY\varphi \rightarrow Z\varphi$ are called *positive extraspection formulas*, $X\neg Y \rightarrow \neg Z\varphi$ are called *negative extraspection formulas*, and $X(Y\varphi \rightarrow \varphi)$ are called *trust formulas*.

hold, then we can derive the following principle (because in that case $\neg B\varphi \leftrightarrow B\neg B\varphi$ is valid):

$$\neg B\varphi \rightarrow K\neg B\varphi \tag{KB2'}$$

Axiom KB3 below confirms that our notion of belief does correspond to a notion of conviction or certainty. This axiom entails the weaker axiom $B\varphi \rightarrow B_w K\varphi$ (also discussed in (Lenzen, 1978)).

$$B\varphi \rightarrow BK\varphi \tag{KB3}$$

The underlying intuition of KB3 is that "*to the agent*, the facts of which he is certain appear to be knowledge" (Lamarre and Shoham, 1994, p. 415) (my emphasis). This informal analysis of the notion of strong belief is formally confirmed by the fact that the axiom $B(B\varphi \rightarrow \varphi)$ is valid in the KD45 logic of belief, and also by the fact that the axiom $B\varphi \rightarrow \varphi$, which is a key axiom of the notion of knowledge, is an axiom of the *internal* version of epistemic logic (Aucher, 2010).

Lenzen also introduces, in (Lenzen, 1979), the following interaction axiom:

$$\hat{B}K\varphi \rightarrow B\varphi \tag{KB3'}$$

This can be equivalently rewritten as $\hat{B}\varphi \rightarrow B\hat{K}\varphi$, where $\hat{B}\varphi$ and $\hat{K}\varphi$ are abbreviations of $\neg B\neg\varphi$ and $\neg K\neg\varphi$ respectively. In this form, this states that, if φ is compatible with everything the agent *believes*, then the agent actually believes that it is compatible with everything she *knows* that φ.

Remark 3 It is difficult to make sense intuitively of the distinction between $\hat{K}\varphi$ and $\hat{B}\varphi$, since they both refer to what the agent considers possible. Hintikka proposes in (Hintikka, 1962, p. 3), the following reading: the formula $\hat{K}\varphi$ should be read as "it is possible, for all that the agent knows, that φ" or "it is compatible with everything the agent knows that φ"; and the formula $\hat{B}\varphi$ should be read as "it is compatible with everything the agent believes that φ". In view of our modeling assumptions, we can add that the former possibility is ascribed externally by the modeler given her knowledge of the epistemic state of the agent, whereas the latter possibility can be determined internally by the agent herself.

Another interaction axiom also introduced by Lenzen (Lenzen, 1978) defines belief in terms of knowledge:

$$B\varphi \leftrightarrow \hat{K}K\varphi \tag{KB4}$$

Although this definition might seem a bit mysterious at first sight, it actually makes perfect sense, as explained in (Lenzen, 1978). Indeed, the left to right direction $B\varphi \rightarrow \hat{K}K\varphi$ can be rewritten $K\neg K\varphi \rightarrow \neg B\varphi$, that is, $\neg(K\neg K\varphi \wedge B\varphi)$. This first implication states that the agent cannot, at the same time, know that she does not know a proposition and be certain of this very proposition. The right to left direction $\hat{K}K\varphi \rightarrow B\varphi$ can be rewritten $\hat{B}\neg\varphi \rightarrow K\neg K\varphi$. This second implication states that, if the agent considers it possible that φ might be false, then she knows that she does not know φ.

Finally, the last interaction axiom we will consider is in fact a definition of knowledge in terms of belief:

$$K\varphi \leftrightarrow (\varphi \wedge B\varphi) \tag{KB5}$$

It simply states that knowledge is defined as true belief. This definition of knowledge in terms of belief lacks the notion of justification addressed in the field of justification logic (Artemov and Fitting, 2011). This definition has also been attacked by philosophers since, according to it, the agent's knowledge could simply be due to some "epistemic luck". Roughly speaking, this means that the agent could believe a proposition which turns out *by chance* to be true, although this belief cannot qualify as knowledge if one considers the whole epistemic context. An explanation of this notion of "epistemic luck" in logical terms is proposed in (Halpern et al., 2009a) (but also see (Prichard, 2004)).

The collapse of knowledge and belief. In any logic of knowledge and belief, if we adopt axiom 5 for the notion of knowledge, axiom D for the notion of belief and KB1 as the only interaction axiom, then we end up with counterintuitive properties. First, as noted by Voorbraak, we can derive the theorem $BK\varphi \rightarrow K\varphi$.[17] This theorem states that "one cannot believe to know a false proposition" (Voorbraak, 1993, p. 8). As it turns out, these axioms are adopted in the first logical framework combining modalities of knowledge and belief (Kraus and Lehmann, 1986). Moreover, if we add the axiom KB3, we can also prove that $B\varphi \rightarrow K\varphi$. This theorem collapses the distinction between the notions of knowledge and belief.

A systematic approach has been proposed by van der Hoek to avoid this collapse (van der Hoek, 1993). He showed, thanks to correspondence theory, that any multi-modal logic with both knowledge and belief modalities that includes the set of axioms $\{\mathsf{D},\mathsf{5},\mathsf{KB1},\mathsf{KB3}\}$ entails the theorem $B\varphi \rightarrow K\varphi$. He also showed, however, that for each proper subset of $\{\mathsf{D},\mathsf{5},\mathsf{KB1},\mathsf{KB3}\}$, counter-models can be built which show that none of those sets of axioms entail the collapse of the distinction between knowledge and belief. So we have to drop one principle in $\{\mathsf{D},\mathsf{5},\mathsf{KB1},\mathsf{KB3}\}$. Axioms D and KB3 are hardly controversial given our understanding of the notion of strong belief. In this case we have to drop either KB1 or 5. Voorbraak proposes to drop axiom KB1. His notion of knowledge, which he calls *objective knowledge*, is therefore unusual in so far as it does not require the agent to be aware of its belief state. But, as we have said, he clearly warns that this notion applies to any information-processing device, and not necessarily just to humans. Note that Floridi has similar reservations against axiom KB1 Floridi (2006), since his notion of *being informed* shares similar features with Voorbraak's notion of *objective knowledge*. Halpern also proposes in (Halpern, 1996) to drop axiom KB1 and to restrict to propositional formulas. This restriction looks a bit ad hoc at first sight. Dropping axiom 5 seems to be the most reasonable choice in light of the discussion about this axiom in Section 3.1.

By dropping 5, we then only have to investigate the logics between S4 and S5 as possible candidates for a logic of knowledge (S5 excluded), as Lenzen did in (Lenzen, 1979).

[17] Here is the proof:

1	$K\varphi \rightarrow B\varphi$	Axiom KB1
2	$K\neg K\varphi \rightarrow B\neg K\varphi$	KB1 : $\neg K\varphi/\varphi$
3	$B\varphi \rightarrow \neg B\neg\varphi$	Axiom D
4	$B\neg\varphi \rightarrow \neg B\varphi$	3, contraposition
5	$B\neg K\varphi \rightarrow \neg BK\varphi$	4 : $K\varphi/\varphi$
6	$\neg K\varphi \rightarrow K\neg K\varphi$	Axiom 5
7	$\neg K\varphi \rightarrow B\neg K\varphi$	6,2, Modus Ponens
8	$\neg K\varphi \rightarrow \neg BK\varphi$	7,5, Modus Ponens
9	$BK\varphi \rightarrow K\varphi$	8, contraposition.

4.2 Principles of interaction with conditional belief

The following axioms $\mathsf{KB1}^{\psi}$,$\mathsf{KB2}^{\psi}$ and $\mathsf{KB3}^{\psi}$ are natural conditional versions of the axioms KB1, KB2, KB3: if ψ is replaced by $\top$ in these three axioms correspond to the axioms KB1, KB2, KB3. Axioms $\mathsf{KB1}^{\psi}$ and $\mathsf{KB2}^{\psi}$ are first introduced in (Moses and Shoham, 1993) and are also adopted in (Friedman and Halpern, 1997). Axiom $\mathsf{KB3}^{\psi}$ is actually introduced in (Lamarre and Shoham, 1994) in the form $B^{\psi}\varphi \rightarrow B^{\psi}_{w}K(\psi \rightarrow \varphi)$.

$$K\varphi \rightarrow B^{\psi}\varphi \qquad (\mathsf{KB1}^{\psi})$$
$$B^{\psi}\varphi \rightarrow KB^{\psi}\varphi \qquad (\mathsf{KB2}^{\psi})$$
$$B^{\psi}\varphi \rightarrow B^{\psi}K(\psi \rightarrow \varphi) \qquad (\mathsf{KB3}^{\psi})$$

Axiom $\mathsf{KB1}^{\psi}$ states that, if the agent knows that φ, then she also believes that φ, and so on under any assumption ψ. Note that $\mathsf{KB1}^{\psi}$ entails the weaker principle $K\varphi \rightarrow (\psi \rightarrow B^{\psi}\varphi)$, which is tightly connected to the Lehrer and Paxton's definition of knowledge as undefeated true belief (Lehrer and Paxson, 1969). Indeed, this derived principle states that if the agent knows that φ (formally $K\varphi$), then her belief in φ cannot be defeated by any *true* information ψ (formally $\psi \rightarrow B^{\psi}\varphi$). Note that this very principle entails an even weaker variant of $\mathsf{KB1}^{\psi}$ introduced in (Moses and Shoham, 1993), namely $K\varphi \rightarrow (B^{\psi}\varphi \vee K\neg\psi)$, i.e. $K\varphi \rightarrow \big(\hat{K}\psi \rightarrow B^{\psi}\varphi\big)$. Axiom $\mathsf{KB2}^{\psi}$ is a straightforward generalization of KB2. As for $\mathsf{KB3}^{\psi}$ it states that, if the agent believes φ under the assumption that ψ, then, given this very assumption ψ, she also believes that she knows φ conditional on ψ.

The axioms $\mathsf{KB4}^{\psi}$ and $\mathsf{KB5}^{\psi}$ below are also introduced in (Lamarre and Shoham, 1994):

$$\neg B^{\psi}\varphi \rightarrow K(\hat{K}\psi \rightarrow \neg B^{\psi}\varphi) \qquad (\mathsf{KB4}^{\psi})$$
$$\hat{K}\psi \rightarrow \neg B^{\psi}\bot \qquad (\mathsf{KB5}^{\psi})$$

Axiom $\mathsf{KB4}^{\psi}$ is a conditional version of axiom KB2'. It is introduced in (Lamarre and Shoham, 1994) in the form $\neg B^{\psi}\varphi \rightarrow K(K\neg\psi \vee \neg B^{\psi}\varphi)$. Another possible conditional version of KB2' could have been $\neg B^{\psi}\varphi \rightarrow K\neg B^{\psi}\varphi$, and this axiom is indeed adopted in (Moses and Shoham, 1993). However, "this simpler axiom ignores the possibility of assumptions which are known to be false, and is valid only for the case of $\psi = \top$" (Lamarre and Shoham, 1994, p. 420).

Axiom $\mathsf{KB5}^{\psi}$ states that, if ψ is compatible with everything the agents knows, then her beliefs given this assumption cannot be inconsistent. In particular, if ψ holds then the agent's doxastic state given this assumption cannot be inconsistent: $\psi \rightarrow \neg B^{\psi}\bot$ (because $\psi \rightarrow \hat{K}\psi$ is valid according to axiom T). Axiom $\mathsf{KB5}^{\psi}$ is introduced in (Lamarre and Shoham, 1994) in the equivalent form $\hat{K}\psi \rightarrow (B^{\psi}\varphi \rightarrow \neg B^{\psi}\neg\varphi)$. Together with $\mathsf{KB1}^{\psi}$ and system P, it entails that knowledge is definable in terms of conditional belief. This definition of knowledge actually coincides with the notion of "safe belief" introduced in (Baltag and Smets, 2008b).

$$K\varphi \triangleq B^{\neg\varphi}\bot \qquad (\text{Def K})$$

Conversely, some definitions of conditional belief in terms of knowledge have been proposed in the literature. In (Moses and Shoham, 1993), the following three definitions

are introduced:

$$B^{\psi}\varphi \triangleq K(\psi \rightarrow \varphi) \qquad \text{(Def1 CB)}$$

$$B^{\psi}\varphi \triangleq K(\psi \rightarrow \varphi) \wedge (K\neg\psi \rightarrow K\varphi) \qquad \text{(Def2 CB)}$$

$$B^{\psi}\varphi \triangleq K(\psi \rightarrow \varphi) \wedge \neg K\neg\psi \qquad \text{(Def3 CB)}$$

The third definition entails the second definition, which itself entails the first definition. However, as one can easily check, none of these three definitions avoids the collapse of the notions of knowledge and belief. Indeed, if we replace ψ with $\top$ in these three definitions, we obtain that $B^{\top}\varphi \leftrightarrow K\varphi$ holds. Hence, if the operator $B^{\top}\varphi$ (i.e. the operator $B\varphi$) is interpreted as a strong belief operator, then these definitions are untenable.

In the spirit of these three definitions, we propose the following weaker interaction axiom which does not collapse the distinction between knowledge and belief:

$$B\neg\psi \rightarrow (B^{\psi}\varphi \rightarrow K(\psi \rightarrow \varphi)) \qquad (\mathsf{KB6}^{\psi})$$

If we assume, moreover, that $\mathsf{KB1}^{\psi}$ holds, then this axiom $\mathsf{KB6}^{\psi}$ entails that if the agent (strongly) believes that ψ does not hold, then her beliefs given ψ coincide with her knowledge given ψ, i.e. $B\neg\psi \rightarrow (B^{\psi}\varphi \leftrightarrow K(\psi \rightarrow \varphi))$. Indeed, given $\mathsf{KB1}^{\psi}$, one can prove that $K(\psi \rightarrow \varphi) \rightarrow B^{\psi}\varphi$ holds.

Finally, we note that the inference rules (RC1) and (RC2) of system P are translated in (Lamarre and Shoham, 1994) by the following two interaction axioms. The intuitive meaning of these axioms is clear.

$$K(\psi \leftrightarrow \psi') \rightarrow (B^{\psi}\varphi \leftrightarrow B^{\psi'}\varphi)$$
$$K(\varphi \rightarrow \varphi') \rightarrow (B^{\psi}\varphi \rightarrow B^{\psi}\varphi')$$

5 Logical formalization

In this section, we will see the standard formal semantics of knowledge, (strong) belief and (strong) conditional belief. For examples and applications of these semantics in computer science, the interested reader can consult (Fagin et al., 1995) or (Meyer and van der Hoek, 1995). We will also introduce the convoluted axioms .2, .3, .3.2 and .4 (together with the class of frame they define), and we will formally define what a (modal) logic is.

5.1 A semantics of knowledge and strong belief

In the rest of this paper, Φ is a set of propositional letters. We define the epistemic-doxastic language $\mathscr{L}_{KB}$ as follows:

$$\mathscr{L}_{KB}: \quad \varphi \ ::= \ p \ \mid \ \neg\varphi \ \mid \ \varphi \wedge \varphi \ \mid \ B\varphi \ \mid \ K\varphi$$

where p ranges over Φ. The propositional language $\mathscr{L}_0$ is the language $\mathscr{L}_{KB}$ without the knowledge and belief operators K and B. The language $\mathscr{L}_K$ is the language $\mathscr{L}_{KB}$ without the belief operator B, and the language $\mathscr{L}_B$ is the language $\mathscr{L}_{KB}$ without the knowledge operator K. The formula $B\varphi$ reads as 'the agent believes φ' and $K\varphi$ reads as 'the agent

knows φ'. Their dual operators $\hat{B}\varphi$ and $\hat{K}\varphi$ are abbreviations of $\neg B\neg\varphi$ and $\neg K\neg\varphi$ respectively.

In epistemic logic, a semantics of the modal operators of belief (B) and knowledge (K) is often provided by means of a Kripke semantics. The first logical framework combining these two operators with a Kripke semantics is proposed in (Kraus and Lehmann, 1986).

Epistemic-doxastic model. An *epistemic-doxastic model* $\mathscr{M}$ is a multi-modal Kripke model $\mathscr{M} = (W, R_B, R_K, V)$ where W is a non-empty set of possible worlds, $R_K, R_B \in 2^{W\times W}$ are binary relations over W called *accessibility relations*, and $V : \Phi \to 2^W$ is a mapping called a *valuation* assigning to each propositional letter p of Φ a subset of W. An *epistemic-doxastic frame* $\mathscr{F}$ is an epistemic-doxastic model without valuation. We often denote $R_K(w) = \{v \in W \mid wR_Kv\}$ and $R_B(w) = \{v \in W \mid wR_Bv\}$.

Let $\varphi \in \mathscr{L}_{KB}$, let $\mathscr{M}$ be an epistemic-doxastic model and let $w \in \mathscr{M}$. The satisfaction relation $\mathscr{M}, w \models \varphi$ is defined inductively as follows:

$$\begin{array}{lll}
\mathscr{M}, w \models p & \text{iff} & w \in V(p) \\
\mathscr{M}, w \models \varphi \wedge \varphi' & \text{iff} & \mathscr{M}, w \models \varphi \text{ and } \mathscr{M}, w \models \varphi' \\
\mathscr{M}, w \models \neg\varphi & \text{iff} & \text{not } \mathscr{M}, w \models \varphi \\
\mathscr{M}, w \models B\varphi & \text{iff} & \text{for all } v \in R_B(w), \mathscr{M}, v \models \varphi \\
\mathscr{M}, w \models K\varphi & \text{iff} & \text{for all } v \in R_K(w), \mathscr{M}, v \models \varphi.
\end{array}$$

We denote $[\![\varphi]\!]_{\mathscr{M}} = \{w \in \mathscr{M} \mid \mathscr{M}, w \models \varphi\}$. We abusively write $w \in \mathscr{M}$ for $w \in W$. If Γ is a set of formulas of $\mathscr{L}_{KB}$, then we write $\mathscr{M} \models \Gamma$ when for all $\varphi \in \Gamma$ and all $w \in \mathscr{M}$, it holds that $\mathscr{M}, w \models \varphi$. Likewise, if $\mathscr{F} = (W, R_B, R_K)$ is an epistemic-doxastic frame, then we abusively write $w \in \mathscr{F}$ for $w \in W$. If Γ is a set of formulas of $\mathscr{L}_{KB}$, then we write $\mathscr{F} \models \Gamma$ when for all $\varphi \in \Gamma$ and all valuation V, $(\mathscr{F}, V) \models \varphi$, and we say that Γ is *valid in* $\mathscr{F}$.

5.2 A semantics of knowledge and conditional belief

Taking up the work of (Friedman and Halpern, 2001), we define the syntax of the language $\mathscr{L}_{KB^\psi}$ inductively as follows:

$$\mathscr{L}_{KB^\psi}: \quad \varphi \ ::= \ p \mid \neg\varphi \mid \varphi \wedge \varphi \mid B^\varphi\varphi \mid K\varphi$$

where p ranges over Φ. The symbol $\top$ is an abbreviation for $p \vee \neg p$, and $B\varphi$ is an abbreviation for $B^\top\varphi$. The language $\mathscr{L}_K$ is $\mathscr{L}_{KB^\psi}$ without the belief operator B^ψ, and the language $\mathscr{L}_{B^\psi}$ is $\mathscr{L}_{KB^\psi}$ without the knowledge operator K.

Numerous semantics have been proposed for default statements, such as preferential structures (Kraus et al., 1990), ε-semantics (Adams, 1975), possibilistic structures (Dubois and Prade, 1991), and κ-ranking (Spohn, 1988a,b). They all have in common that they validate the axiomatic system P originally introduced in (Kraus et al., 1990). A slightly different version of this system is reproduced in Section 3.2.3. This remarkable fact is explained in (Friedman and Halpern, 2001), where a general framework based on *plausibility measures* is proposed. As proved in that paper, plausibility measures generalize all these semantics. We can nevertheless mention that other logical formalisms dealing with

conditional beliefs are proposed in the economics literature (Board, 2004). These other formalisms have been taken up in the field of *dynamic* epistemic logic (Baltag and Smets, 2006, 2008a,b).

We adopt the general framework of plausibility measures to provide a semantics for $\mathscr{L}_{KB^\psi}$. Plausibility spaces and epistemic-plausibility spaces are introduced respectively in (Friedman and Halpern, 1997) and (Friedman and Halpern, 2001). Because these structures will play a role only in the proofs of the subsequent theorems, their definitions and the truth conditions of the language $\mathscr{L}_{KB^\psi}$ are postponed until the appendix, together with the proofs of all the theorems and propositions in this paper.

5.3 Logics of knowledge, belief and conditional belief

A *(modal) logic* L for a modal language $\mathscr{L}$ is a set of formulas of $\mathscr{L}$ that contains all propositional tautologies and is closed under modus ponens (that is, if $\varphi \in \mathsf{L}$ and $\varphi \to \psi \in \mathsf{L}$, then $\psi \in \mathsf{L}$) and uniform substitution (that is, if φ belongs to L then so do all of its substitution instances (Blackburn et al., 2001, Def. 1.18)). A modal logic is usually defined by a set of axioms and inference rules. A formula belongs to the modal logic if it can be derived by successively applying (some of) the inference rules to (some of) the axioms. We are interested here in *normal modal logics*. These modal logics contain the formulas $(B(\varphi \to \psi) \wedge B\varphi) \to B\psi$ and $(K(\varphi \to \psi) \wedge K\varphi) \to K\psi$ (i.e. axiom K), and the inference rules of belief and knowledge necessitation: from $\varphi \in \mathsf{L}$, infer $B\varphi \in \mathsf{L}$, and from $\varphi \in \mathsf{L}$, infer $K\varphi \in \mathsf{L}$ (i.e. inference rule Nec). A modal logic *generated* by a set of axioms Γ is the smallest normal modal logic containing the formulas Γ.

Below, we give a list of properties of the accessibility relations R_B and R_K that will be used in the rest of the paper. We also give, below each property, the axiom which *defines* the class of epistemic-doxastic frames that fulfill this property (see (Blackburn et al., 2001, Def. 3.2) for a definition of the notion of *definability*). We choose, without any particular reason, to use the knowledge modality to write these conditions.

The logic $(\mathsf{KD45})_B$ is the smallest normal modal logic for $\mathscr{L}_B$ generated by the set of axioms $\{\mathsf{D}, \mathsf{4}, \mathsf{5}\}$. The logic $(\mathsf{P})_{B^\psi}$ is the smallest logic for $\mathscr{L}_{B^\psi}$ containing the axioms $\mathsf{C1}$-$\mathsf{C4}$ and inference rules $\mathsf{RC1}$-$\mathsf{RC2}$ from Section 3.2.3.[18] For any $\mathsf{x} \in \{.2, .3, .3.2, .4\}$, the logic $(\mathsf{S4.x})_K$ is the smallest normal modal logic for $\mathscr{L}_K$ generated by the set of axioms $\{\mathsf{T}, \mathsf{4}, \mathsf{x}\}$. We have the following relationship between these logics:

$$(\mathsf{S4})_K \subset (\mathsf{S4.2})_K \subset (\mathsf{S4.3})_K \subset (\mathsf{S4.3.2})_K \subset (\mathsf{S4.4})_K \subset (\mathsf{S5})_K$$

If L and L' are two sets of formulas (possibly logics), we denote by $\mathsf{L}+\mathsf{L}'$ the smallest normal modal logic containing L and L'. Note that $\mathsf{L}+\mathsf{L}'$ may be different from $\mathsf{L}\cup\mathsf{L}'$ in general, because $\mathsf{L}\cup\mathsf{L}'$ may not be closed under modus ponens or uniform substitution.

[18] Note that the axiom $(B^\psi(\varphi \to \varphi') \wedge B^\psi\varphi) \to B^\psi\varphi'$ and the inference rule from φ infer $B^\psi\varphi$ are both derivable in $(\mathsf{P})_{B^\psi}$. Therefore, $(\mathsf{P})_{B^\psi}$ is also a *normal* modal logic.

serial: D:	$R_K(w) \neq \emptyset$ $K\varphi \rightarrow \hat{K}\varphi$
transitive: 4:	If $w' \in R_K(w)$ and $w'' \in R_K(w')$, then $w'' \in R_K(w)$ $K\varphi \rightarrow KK\varphi$
Euclidean: 5:	If $w' \in R_K(w)$ and $w'' \in R_K(w)$, then $w' \in R_K(w'')$ $\neg K\varphi \rightarrow K\neg K\varphi$
reflexive: T:	$w \in R_K(w)$ $K\varphi \rightarrow \varphi$
symetric: B:	If $w' \in R_K(w)$, then $w \in R_K(w')$ $\varphi \rightarrow K\neg K\neg\varphi$
confluent: .2:	If $w' \in R_K(w)$ and $w'' \in R_K(w)$, then there is v such that $v \in R_K(w')$ and $v \in R_K(w'')$ $\hat{K}K\varphi \rightarrow K\hat{K}\varphi$
weakly connected: .3:	If $w' \in R_K(w)$ and $w'' \in R_K(w)$, then $w' = w''$ or $w' \in R_K(w'')$ or $w'' \in R_K(w')$ $\hat{K}\varphi \wedge \hat{K}\psi \rightarrow \hat{K}(\varphi \wedge \psi) \vee \hat{K}(\psi \wedge \hat{K}\varphi) \vee \hat{K}(\varphi \wedge \hat{K}\psi)$
semi-Euclidean: .3.2:	If $w'' \in R_K(w)$ and $w \notin R_K(w'')$ and $w' \in R_K(w)$, then $w'' \in R_K(w')$ $(\hat{K}\varphi \wedge \hat{K}K\psi) \rightarrow K(\hat{K}\varphi \vee \psi)$
R1: .4:	If $w'' \in R_K(w)$ and $w \neq w''$ and $w' \in R_K(w)$, then $w'' \in R_K(w')$ $(\varphi \wedge \hat{K}K\varphi) \rightarrow K\varphi$

Fig. 1 List of properties of the accessibility relations R_B and R_K and corresponding axioms.

6 Defining knowledge in terms of belief and vice versa

The definability of modalities in terms of other modalities is studied from a theoretical point of view in (Halpern et al., 2009b). This study is subsequently applied to epistemic logic in (Halpern et al., 2009a). Three notions of definability emerge from this work: explicit definability, implicit definability and reducibility. It has been proven that, for modal logic, explicit definability coincides with the conjunction of implicit definability and reducibility (unlike first-order logic, where the notion of explicit definability coincides with implicit definability only). In this paper, we are interested only in the notion of *explicit* definability, which is also used by Lenzen in (Lenzen, 1979). Here is its formal definition:

Definition 1 (Halpern et al., 2009a) Let L be a (modal) logic for $\mathscr{L}_{KB}$ (resp. $\mathscr{L}_{KB^\psi}$).

- We say that K is *explicitly defined in* L *by the definition* $Kp \triangleq \delta$, where $\delta \in \mathscr{L}_B$ (resp. $\delta \in \mathscr{L}_{B^\psi}$), if $Kp \leftrightarrow \delta \in \mathsf{L}$.
- We say that B (resp. B^ψ) is *explicitly defined in* L *by the definition* $Bp \triangleq \delta$, where $\delta \in \mathscr{L}_K$, if $Bp \leftrightarrow \delta \in \mathsf{L}$ (resp. $B^\psi p \leftrightarrow \delta \in \mathsf{L}$).

Obviously, putting together an epistemic logic and a doxastic logic, for example $(\mathsf{S4})_K + (\mathsf{KD45})_B$, does not yield a genuine epistemic-doxastic logic since the two notions will not interact. We need to add interaction axioms. In (Halpern et al., 2009a), only the interaction axioms $\mathsf{KB1}$ and $\mathsf{KB2}$ suggested by Hintikka (Hintikka, 1962) are considered. In this section, we will also add the interaction axiom $\mathsf{KB3}$, suggested by Lenzen (Lenzen, 1978), since this axiom is characteristic of the notion of strong belief, as we explained in Section 4.1.

6.1 Defining belief in terms of knowledge

We will address the problem of defining belief in terms of knowledge from a syntactic perspective and from a semantic perspective.

6.1.1 Syntactic perspective

Lenzen is the first to note that the belief modality can be defined in terms of knowledge if we adopt $\{\mathsf{KB1}, \mathsf{KB2}, \mathsf{KB3}\}$ as interaction axioms:

Theorem 1 *(Lenzen, 1979) The belief modality B is explicitly defined in the logic $\mathsf{L} = (\mathsf{S4})_K + (\mathsf{KD45})_B + \{\mathsf{KB1}, \mathsf{KB2}, \mathsf{KB3}\}$ by the following definition:*

$$B\varphi \triangleq \hat{K}K\varphi \qquad \text{(Def B)}$$

Consequently, the belief modality B is also defined by $\mathsf{Def}\, B$ *in any logic containing* L.

As a consequence of this theorem, the belief modality is also explicitly defined by $B\varphi \triangleq \hat{K}K\varphi$ in the logics $(\mathsf{S4.x})_K + (\mathsf{KD45})_B + \{\mathsf{KB2}, \mathsf{KB1}, \mathsf{KB3}\}$, where x ranges over $\{.2, .3, .3.2, .4\}$. This result is in contrast with Theorem 4.8 in (Halpern et al., 2009a), from which it follows that the belief modality can*not* be explicitly defined in the logic $(\mathsf{S4.x})_K + (\mathsf{KD45})_B + \{\mathsf{KB1}, \mathsf{KB2}\}$, and so on for any $\mathsf{x} \in \{.2, .3, .3.2, .4\}$. We see here that the increase in expressivity due to the addition of the interaction axiom $\mathsf{KB3}$ plays an important role in bridging the gap between belief and knowledge. Note that the definition $\mathsf{Def}\ \mathsf{B}$ of belief in terms of knowledge corresponds to the interaction axiom $\mathsf{KB4}$, which has already been discussed in Section 4.1.

6.1.2 Semantic perspective

Given that Theorem 1 shows that the belief modality B can be defined in terms of the knowledge modality K, we would expect that the belief accessibility relation R_B could also be 'defined' in terms of the knowledge accessibility relation R_K in any frame that validates $\mathsf{L} = (\mathsf{S4})_K + (\mathsf{KD45})_B + \{\mathsf{KB1}, \mathsf{KB2}, \mathsf{KB3}\}$. The following result, already pointed out in (Stalnaker, 2006) (without proof), shows that this is indeed the case:

Theorem 2 *Let $\mathscr{F}$ be a frame such that $\mathscr{F} \models (S4)_K + (KD45)_B + \{KB1, KB2, KB3\}$. Then, for all $w, v \in \mathscr{F}$, it holds that*

$$wR_Bv \quad \textit{iff} \quad \textit{for all } u \in \mathscr{F},\ wR_Ku \textit{ implies } uR_Kv \qquad (\mathsf{Def\ R_B})$$

Note that if we are in a world w such that wR_Bw, then the accessibility relation for knowledge R_K is Euclidean at w and axiom $\mathsf{5}$ holds at w. But according to our analysis in Section 4.1, this also entails that the notions of knowledge and belief collapse into one another (the proof in Footnote 17 can be adapted to this particular setting). Therefore, in the logic $\mathsf{L} = (\mathsf{S4})_K + (\mathsf{KD45})_B + \{\mathsf{KB1}, \mathsf{KB2}, \mathsf{KB3}\}$, the following principle holds:

If *all* the agent's beliefs hold true,
then her beliefs are actually all knowledge.

If it turns out that the agent has a single erroneous belief, then the conclusion of this principle obviously does not hold anymore. This principle is intuitively correct and can be explained informally by the following reasoning. If all my beliefs are correct (true), then the justification of any specific belief φ is also 'correct', since this very justification is based on my own beliefs. Therefore, any specific belief φ is justified and this justification is in a certain sense 'correct'. Consequently, all my beliefs φ turn out in fact to be knowledge.

Note that this principle holds in any logic that extends L. In particular, all the logics considered in the rest of this paper validate this reasonable principle.

6.2 Defining knowledge in terms of belief

We will address the problem of defining knowledge in terms of belief from a syntactic perspective and from a semantic perspective.

6.2.1 Syntactic perspective

Defining knowledge in terms of belief depends on the logic of knowledge that we deal with. As the following proposition shows, knowledge can be defined in terms of belief if the logic of knowledge is $\mathsf{S4.4}$, but not if the logic of knowledge is $\mathsf{S4}$ and $\mathsf{S4.x}$, where x ranges over $\{\mathsf{.2}, \mathsf{.3}, \mathsf{.3.2}\}$.

Theorem 3 – *The knowledge modality K is explicitly defined in the logic $(S4.4)_K + (KD45)_B + \{KB1, KB2, KB3\}$ by the following definition:*

$$K\varphi \triangleq \varphi \wedge B\varphi \qquad (\mathsf{Def\ K})$$

– *The knowledge modality K cannot be explicitly defined in the logics $(S4.x)_K + (KD45)_B + \{KB1, KB2, KB3\}$ for any $x \in \{.2, .3, .3.2\}$.*

This result can be contrasted with Theorem 4.1 in (Halpern et al., 2009a), from which it follows that the knowledge modality can*not* be explicitly defined in the logic $(\mathsf{S4.4})_K + (\mathsf{KD45})_B + \{\mathsf{KB1},\mathsf{KB2}\}$. We see once again that the increase in expressivity due to the addition of the interaction axiom KB3 plays an important role in bridging the gap between belief and knowledge.

6.2.2 Semantic perspective

As a semantic counterpart to Theorem 3, the knowledge accessibility relation K cannot be 'defined' in a frame that validates the logic $\mathsf{L} = (\mathsf{S4})_K + (\mathsf{KD45})_B + \{\mathsf{KB1},\mathsf{KB2},\mathsf{KB3}\}$. Therefore, there are, in principle, several possible ways to 'extend' the belief accessibility relation R_B to a knowledge accessibility relation R_K. Indeed, each interaction axiom *defines* a class of epistemic-doxastic frames (Blackburn et al., 2001, Def. 3.2). This imposes some constraints on the knowledge accessibility relation R_K, though without determining it completely. We are now going to present these constraints.

The interaction axiom KB1 defines the class of epistemic-doxastic frames $\mathscr{F}$ such that for all $w,v \in \mathscr{F}$,

$$\text{If } wR_Bv \text{ then } wR_Kv \tag{1}$$

The addition of the interaction axioms KB2 and KB3 to KB1 defines the class F of epistemic-doxastic frames $\mathscr{F}$ such that for all $w,v \in \mathscr{F}$,

$$\text{If } wR_Bw \text{ then } (wR_Kv \text{ iff } wR_Bv) \tag{2}$$

So, we still have to specify the worlds accessible by R_K for the worlds w such that it is *not* the case that wR_Bw. Indeed, if wR_Bw, then it holds that $R_K(w) = R_B(w)$ according to Equation 2.

In (Stalnaker, 2006), Stalnaker introduces four possible extensions of the belief accessibility relation R_B to a knowledge accessibility relations R_K. These four possible extensions turn out to correspond to our four logics of knowledge: S4.2, S4.3, S4.3.2 and S4.4.

1. The first extension consists in the reflexive closure of the accessibility relation R_B. This is the minimal extension possible and it yields the objectionable definition of knowledge as true belief, whose logic is S4.4.[19]
2. The second extension consists in defining wR_Kv as $((wR_Bw$ and $wR_Bv)$ or (not $wR_Bw))$. This is the maximal extension possible and it yields the logic S4.3.2.[20]
3. The third extension consists in defining knowledge as true belief which cannot be defeated by any true fact. In other words, a fact is known if and only if it is true and it will still be believed after any possible truthful announcement.[21] This yields the logic

[19] That is, S4 plus .4: $(\varphi \wedge \hat{K}K\psi) \rightarrow K(\varphi \vee \psi)$; see Section 5.3.

[20] That is, S4 plus .3.2: $(\hat{K}\varphi \wedge \hat{K}K\psi) \rightarrow K(\hat{K}\varphi \vee \psi)$; see Section 5.3.

[21] For this definition to be consistent, we have to add another constraints that Stalnaker does not mention: in this definition, knowledge should only deal with propositional facts belonging to the propositional language $\mathscr{L}_0$. Indeed, assume that the agent believes non-p (formally $B\neg p$). Then clearly the agent knows that she believes non-p by KB2 (formally $KB\neg p$). However, assume that p is actually true. If we apply this definition of knowledge, then, if she learnt that p (which is true), she should still believe that she believes non-p (formally $BB\neg p$), so she should still believe non-p (formally $B\neg p$), which is of course counterintuitive. This restriction on propositional knowledge does not produce a loss of generality because we assume that the agent knows everything about her own beliefs and disbeliefs.

S4.3.[22] Lehrer and Paxson proposed to add this last condition to the classical notion of knowledge as justified true belief in order to cope with the 'Gettier Problem' (Lehrer and Paxson, 1969).

4. The last extension consists in weakening the condition of the third extension. Stalnaker indeed argues in addition that this definition of knowledge as undefeated true belief should not be a sufficient and necessary condition for knowledge, but rather only a sufficient one. This contention gives the last possible extension of the accessibility relation for belief to an accessibility relation for knowledge.

Note that Rott also investigates systematically, but with the help of a 'sphere' semantics, how a number of epistemological accounts of the notion of knowledge (including Nozick's account) convert belief into knowledge (Rott, 2004). Like us, he does so not by considering the notion of justification, but by resorting to other properties such as the stability of beliefs, the sensitivity to truth or the strength of belief and of epistemic position.

7 A derivation of axioms .2, .3, .3.2, .4 from interaction axioms

In this section, we show that the convoluted axioms for knowledge .2, .3, .3.2 and .4 can be derived from understandable interaction axioms if we consider the logic $(\mathsf{S4})_K$ for the notion of knowledge and the logic $(\mathsf{KD45})_B$ (or $(\mathsf{P})_{B^\psi}$) for the notion of belief (or conditional belief).

7.1 Derivation of axiom .2

Theorem 1 can be equivalently formulated as $(\mathsf{S4.x})_K + (\mathsf{KD45})_B + \{\mathsf{KB1}, \mathsf{KB2}, \mathsf{KB3}\} = (\mathsf{S4.x})_K + (\mathsf{KD45})_B + \{\mathsf{KB1}, \mathsf{KB2}, \mathsf{KB3}\} + \{B\varphi \leftrightarrow \hat{K}K\varphi\}$. Note, however, that Lenzen proved, in (Lenzen, 1979), an even stronger result, which is the following:[23]

$$(\mathsf{S4})_K + (\mathsf{KD45})_B + \{\mathsf{KB2}, \mathsf{KB1}, \mathsf{KB3}\} = (\mathsf{S4.2})_K + \{B\varphi \leftrightarrow \hat{K}K\varphi\}$$

This proposition states not only that the belief modality is definable in terms of knowledge, but also that axiom .2 is derivable from the interaction axioms $\{\mathsf{KB2}, \mathsf{KB1}, \mathsf{KB3}\}$ in the logic $(\mathsf{S4})_K + (\mathsf{KD45})_B$, that is:

$$.2 \in (\mathsf{S4})_K + (\mathsf{KD45})_B + \{\mathsf{KB1}, \mathsf{KB2}, \mathsf{KB3}\} \qquad (.2)$$

S4.2 is the logic of knowledge propounded by Lenzen and Stalnaker. It is also the logic of the notion of *justified knowledge* studied by Voorbraak in (Voorbraak, 1993).

[22] That is, S4 plus .3: $\hat{K}\varphi \wedge \hat{K}\psi \rightarrow \hat{K}(\varphi \wedge \hat{K}\psi) \vee \hat{K}(\varphi \wedge \psi) \vee \hat{K}(\psi \wedge \hat{K}\varphi)$; see Section 5.3.

[23] Lenzen uses axiom $\mathsf{KB3}'$ instead of KB3, but one can easily show that the replacement does not invalidate the proposition.

7.2 Derivation of axiom .3

Lenzen does not provide an intuitive characterization of axiom .3 in terms of interaction axioms. In fact, I believe that such a characterization is not possible if we consider the language $\mathscr{L}_B$ only, and that we need to consider a more expressive language. It turns out that $\mathscr{L}_{KB^\psi}$ is sufficiently expressive to derive .3:

$$.3 \in (\mathsf{S4})_K + (\mathsf{P})_{B^\psi} + \{\mathsf{KB1}^\psi, \mathsf{KB5}^\psi, \mathsf{KB4}^\psi\} \tag{.3}$$

We can recall that $\mathsf{KB1}^\psi$ stands for $K\varphi \to B^\psi\varphi$, $\mathsf{KB5}^\psi$ for $\hat{K}\psi \to \neg B^\psi\bot$ and $\mathsf{KB4}^\psi$ for $\neg B^\psi\varphi \to K(\hat{K}\psi \to \neg B^\psi\varphi)$.

The logic S4.3 is propounded as the logic of knowledge by van der Hoek (van der Hoek, 1993).

7.3 Derivation of axiom .3.2

With a language without conditional belief operator, Lenzen provides, in (Lenzen, 1979), a derivation of .3.2 by resorting to the interaction axiom KB5 below:

$$.3.2 \in (\mathsf{S4})_K + (\mathsf{KD45})_B + \{\mathsf{KB1},\mathsf{KB2},\mathsf{KB3},\mathsf{KB5}\}$$

where

$$(K\varphi \to K\psi) \wedge B(K\varphi \to K\psi) \to K(K\varphi \to K\psi) \tag{KB5}$$

As it turns out, Lenzen proves, in (Lenzen, 1979), an even stronger result, which is the following:

$$(\mathsf{S4.3.2})_K + (\mathsf{KD45})_B + \{\mathsf{KB1},\mathsf{KB2},\mathsf{KB3}\} = (\mathsf{S4})_K + (\mathsf{KD45})_B + \{\mathsf{KB1},\mathsf{KB2},\mathsf{KB3},\mathsf{KB5}\}$$

Note that the interaction axiom KB5 is a special instance of the definition of knowledge as true belief, $p \wedge Bp \to Kp$, since p is substituted here by $K\varphi \to K\psi$. Even with this observation, it is still difficult to provide an intuitive reading of this interaction axiom. Instead, we can show that .3.2 is derivable in a logic *with* conditional belief by means of the interaction axioms KB5, which is easier to grasp.

$$.3.2 \in (\mathsf{S4})_K + (\mathsf{P})_{B^\psi} + \{\mathsf{KB1}^\psi, \mathsf{KB5}^\psi, \mathsf{KB4}^\psi, \mathsf{KB6}^\psi\} \tag{.3.2}$$

We can recall that the key interaction axiom $\mathsf{KB6}^\psi$ stands for $B\neg\psi \to (B^\psi\varphi \to K(\psi \to \varphi))$.

7.4 Derivation of axiom .4

Axiom .4 can be seen as a weakening of axiom 5 since it can be rewritten as follows: $p \rightarrow (\neg K\varphi \rightarrow K\neg K\varphi)$. The logic S4.4 is sometimes called the logic of 'true belief'. This denomination is indeed very appropriate. Lenzen proves, in (Lenzen, 1979), the following equation:

$$(\mathsf{S4.4})_K + (\mathsf{KD45})_B + \{\mathsf{KB1},\mathsf{KB2},\mathsf{KB3}\} = (\mathsf{S4})_K + (\mathsf{KD45})_B + \{\mathsf{KB4}\}$$

where we recall that the interaction axiom KB4 is $K\varphi \leftrightarrow \varphi \wedge B\varphi$. From this equation, one can easily derive the following result:

$$.4 \in (\mathsf{S4})_K + (\mathsf{KD45})_B + \{\mathsf{KB1},\mathsf{KB2},\mathsf{KB3},\mathsf{KB5}\} \qquad (.4)$$

Kutschera argues for S4.4 as the logic of knowledge (Kutschera, 1976).

8 Concluding remarks

In this paper, we have reviewed the most prominent principles of logics of knowledge and belief, and the principles relating knowledge, belief and conditional belief to one another. In doing so, we have encountered most of the problems that have beset epistemic logic during its relatively short (modern) history. We have shown that the convoluted axioms .3 and .3.2 for knowledge, which can hardly be understood in terms of interaction axioms dealing with (strong) belief only, can be expressed in terms of interaction axioms dealing with *conditional* beliefs, which are easier to grasp. We have also demonstrated that the addition of the interaction axiom $B\varphi \rightarrow BK\varphi$, which is characteristic of the notion of (strong) belief, plays an important role in bridging the gap between the notions of belief and knowledge.

As we explained in Section 3.2, the term "belief" has different meanings: my (weak) belief that it will be sunny tomorrow is different from my (strong) belief that the Fermat-Wilson theorem holds true. In this paper, we have only focused on the notion of strong belief. To deal with the notion of weak belief, we could enrich our language either with a probabilistic-doxastic operator $Prob(\varphi) \geq r$ (where r ranges over $]0.5;1[$), or with a graded belief modality $B^n\varphi$ (where n ranges over $\mathbb{N}$), or simply with a weak belief operator $B_w\varphi$. This latter language actually corresponds to a language introduced by Lenzen in (Lenzen, 2004). Its conditionalized version corresponds to the full language of (Lamarre and Shoham, 1994), which the authors of this paper have completely axiomatized.

Even if our aim was not to argue in favour of a particular logic of knowledge, it is nevertheless clear from our discussion that, on the one hand, logics like S4.2 or S4.3 are better suited to reasoning about the knowledge of agents in the most general kinds of situations; on the other hand, the simple and widely used logic S5 is more appropriate for dealing with particular situations where agents cannot have erroneous beliefs, as we have already argued at the end of Section 4.1. As a matter of fact, the logic S5 is an enrichement of these logics with extra assumptions (it is actually a superset of them). More work is needed to fully understand the logics between S4 and S5 (exclusive) and in particular to investigate and study their dynamic extensions.

Acknowledgements I thank Manuel Rebuschi and Franck Lihoreau for helpful comments on this paper. I also thank the anonymous English native speaker referee for detailed comments.

A Plausibility space and epistemic-plausibility space

A.1 Plausibility space

If W is a non-empty set of possible worlds, then an *algebra over* W is a set of subsets of W closed under union and complementation. In the rest of the paper, D is a non-empty set partially ordered by a relation $\leq$ (so that $\leq$ is reflexive, transitive and anti-symmetric). We further assume that D contains two special elements $\top$ and $\bot$ such that for all $d \in D$, $\bot \leq d \leq \top$. As usual, we define the ordering $<$ by taking $d_1 < d_2$ if and only if $d_1 \leq d_2$ and $d_1 \neq d_2$. A *(qualitative) plausibility space* is a tuple $S = (W, \mathscr{A}, Pl)$ where:

- W is a non-empty set of possible worlds;
- $\mathscr{A}$ is an algebra over W;
- $Pl : \mathscr{A} \to D$ is a function mapping sets of $\mathscr{A}$ into D and satisfying the following conditions:
  - A0 $Pl(W) = \top$ and $Pl(\emptyset) = \bot$;
  - A1 If $A \subseteq B$, then $Pl(A) \leq Pl(B)$;
  - A2 If A, B, and C are pairwise disjoint sets, $Pl(A \cup B) > Pl(C)$, and $Pl(A \cup C) > Pl(B)$, then $Pl(A) > Pl(B \cup C)$;
  - A3 If $Pl(A) = Pl(B) = \bot$, then $Pl(A \cup B) = \bot$.

We denote by $\mathscr{S}$ the class of all (qualitative) plausibility spaces.

A.2 Epistemic-plausibility space and truth conditions

An *epistemic-plausibility space* is a tuple $\mathscr{M} = (W, R, V, \mathscr{P})$ where:

- W is a non-empty set of possible worlds;
- $R_K \in 2^{W \times W}$ is a binary relation over W called an *accessibility relation*;
- $V : \Phi \to 2^W$ is a function called a *valuation* mapping propositional variables to subsets of W;
- $\mathscr{P} : W \to \mathscr{S}$ is a function called a *plausibility assignment* mapping each world $w \in W$ to a (qualitative) plausiblity space $(W_w, \mathscr{A}_w, Pl_w)$ such that $W_w \subseteq W$.

Let $\varphi \in \mathscr{L}_{KB\psi}$, let $\mathscr{M}$ be an epistemic-plausibility space and let $w \in \mathscr{M}$. The satisfaction relation $\mathscr{M}, w \models \varphi$ is defined inductively as follows:

$$
\begin{array}{lll}
\mathscr{M}, w \models p & \text{iff} & w \in V(p) \\
\mathscr{M}, w \models \varphi \wedge \varphi' & \text{iff} & \mathscr{M}, w \models \varphi \text{ and } \mathscr{M}, w \models \varphi' \\
\mathscr{M}, w \models \neg\varphi & \text{iff} & \text{not } \mathscr{M}, w \models \varphi \\
\mathscr{M}, w \models B^{\psi}\varphi & \text{iff} & \text{either } Pl_w([\![\psi]\!]_w) = \bot \text{ or } Pl_w([\![\psi \wedge \varphi]\!]_w) > Pl_w([\![\psi \wedge \neg\varphi]\!]_w) \\
\mathscr{M}, w \models K\varphi & \text{iff} & \text{for all } v \in R_K(w), \mathscr{M}, v \models \varphi
\end{array}
$$

where $[\![\varphi]\!]_w = \{v \in W_w \mid \mathscr{M}, v \models \varphi\}$. We abusively write $w \in \mathscr{M}$ for $w \in W$, and we also write $\mathscr{M} \models \varphi$ when for all $w \in \mathscr{M}$, $\mathscr{M}, w \models \varphi$. If Γ is a set of formulae (possibly infinite), we write $\mathscr{M} \models \Gamma$ when $\mathscr{M} \models \varphi$ for all $\varphi \in \Gamma$.

B Proofs of Theorems 2 and 3

B.1 Proof of Theorem 2

Theorem 4 *Let $\mathscr{F}$ be a frame such that $\mathscr{F} \models (S4)_K + (KD45)_B + \{KB1, KB2, KB3\}$. Then, for all $w, v \in \mathscr{F}$, it holds that*

$$wR_Bv \text{ iff for all } u \in \mathscr{F},\ wR_Ku \text{ implies } uR_Kv \qquad (\text{Def } \mathsf{R_B})$$

Proof Let $\mathscr{F} = (W, R_B, R_K)$ be an epistemic-doxastic frame such that $\mathscr{F} \models (\mathsf{S4})_K + (\mathsf{KD45})_B + \{\mathsf{KB1}, \mathsf{KB2}, \mathsf{KB3}\}$. Then, because the axioms T, D, $\mathsf{4}$ and $\mathsf{5}$ define, respectively, the properties of reflexivity, seriality, transitivity and Euclideanity, R_B is serial, transitive and Euclidean, and R_K is reflexive and transitive. Moreover, by the validity of $\mathsf{KB1}$, $R_B \subseteq R_K$. We can now prove that ($\mathsf{Def\ R_B}$) holds.

– From left to right: assume towards a contradiction that there are $w,v,u \in \mathscr{F}$ such that $v \in R_B(w)$ and $u \in R_K(w)$ and not $v \in R_K(u)$.
Let $p \in \Phi$. We define a valuation V over W such that $V(p) = R_B(w)$. Let $\mathscr{M}$ be the epistemic-doxastic model defined by $\mathscr{M} = (\mathscr{F},V)$. Then, $\mathscr{M},w \models Bp$. So, $\mathscr{M},w \models KBp$ by the validity of $\mathsf{KB2}$. Therefore, $\mathscr{M},u \models Bp$ because $u \in R_K(w)$. So, $\mathscr{M},u \models \hat{B}p$ because R_B is serial. Then, there is $t \in R_B(u)$ such that $\mathscr{M},t \models p$. That is, there is $t \in R_B(u)$ such that $t \in R_B(w)$, because $V(p) = R_B(w)$. However, by assumption, $v \in R_B(w)$. Therefore, because R_B is Euclidean, $v \in R_B(t)$. So, $t \in R_B(u)$ and $v \in R_B(t)$. Therefore, by the transitivity of R_B, $v \in R_B(u)$. Then, $v \in R_K(u)$, because $R_B \subseteq R_K$. This is impossible by assumption. We therefore reach a contradiction.
– From right to left: assume towards a contradiction that there are $w,v \in \mathscr{F}$ such that $v \notin R_B(w)$ and for all $u \in \mathscr{F}$, $u \in R_K(w)$ implies $v \in R_K(u)$.
Let $p \in \Phi$. We define a valuation V such that $V(p) = R_B(w)$. Let $\mathscr{M}$ be the epistemic-doxastic model defined by $\mathscr{M} = (\mathscr{F},V)$. Then, $\mathscr{M},w \models Bp$. Then, $\mathscr{M},w \models BKp$ by validity of $\mathsf{KB3}$. Because R_B is serial, there is $u \in R_B(w)$ such that $\mathscr{M},u \models Kp$. Now, because $R_B \subseteq R_K$, we also have that $u \in R_K(w)$. Then, by assumption $v \in R_K(u)$. Therefore, $\mathscr{M},v \models p$. Then, by the definition of V, we have that $v \in R_B(w)$. This is impossible by assumption. We therefore reach a contradiction.

B.2 Proof of Theorem 3

Theorem 5 – *The knowledge modality K is explicitly defined in the logic* $(\mathsf{S4.4})_K + (\mathsf{KD45})_B + \{\mathsf{KB1},\mathsf{KB2},\mathsf{KB3}\}$ *by* $K\varphi \triangleq \varphi \wedge B\varphi$.
– *The knowledge modality K cannot be explicitly defined in the logics* $(\mathsf{S4.x})_K + (\mathsf{KD45})_B + \{\mathsf{KB1},\mathsf{KB2},\mathsf{KB3}\}$ *for any* $x \in \{.2,.3,.3.2\}$.

Proof The first item in this theorem is owed to (Lenzen, 1979). We will only prove the second item. The proof method is similar to the proof method for Theorem 4.1 in (Halpern et al., 2009a). If K is explicitly defined in $\mathsf{L} = (\mathsf{S4.3.2})_K + (\mathsf{KD45})_B + \{\mathsf{KB1},\mathsf{KB2},\mathsf{KB3}\}$ by $Kp \leftrightarrow \delta$, then for every epistemic-doxastic model $\mathscr{M}$ such that $\mathscr{M} \models \mathsf{L}$, it holds that $[\![Kp]\!]_{\mathscr{M}} = [\![\delta]\!]_{\mathscr{M}}$, and therefore $[\![Kp]\!]_{\mathscr{M}} \in \{[\![\varphi]\!]_{\mathscr{M}} \mid \varphi \in \mathscr{L}_B\}$. We prove the theorem by constructing an epistemic-doxastic model $\mathscr{M}$ such that $\mathscr{M} \models \mathsf{L}$ and such that $[\![Kp]\!]_{\mathscr{M}} \notin \{[\![\varphi]\!]_{\mathscr{M}} \mid \varphi \in \mathscr{L}_B\}$.

Consider the following epistemic-doxastic frame $F = (W,R)$ where $W = \{w_1,w_2,w_3,w_4\}$, $R_B = \{(w_1,w_1),(w_2,w_2),(w_3,w_2),(w_4,w_2)\}$, and $R_K = R_B \cup \{(w_3,w_3),(w_4,w_4),(w_3,w_4),(w_4,w_3)\}$. Let $\mathscr{M} = (F,V)$ be the epistemic-doxastic model based on F such that V maps each primitive proposition to $\{w_1,w_2,w_4\}$. Clearly, $\mathscr{M} \models \mathsf{L}$. One can also show by induction on the structure of formulas in $\mathscr{L}_B$ that $\{[\![\varphi]\!]_{\mathscr{M}} \mid \varphi \in \mathscr{L}_B\} = \{\{w_1,w_2,w_4\}, \{w_3\},\emptyset,W\}$, but $[\![Kp]\!]_{\mathscr{M}} = \{w_1,w_2\}$.

C Proofs of Equations .3 and .3.2

C.1 Proof of Equation .3

$$.3 \in (\mathsf{S4})_K + (\mathsf{P})_{B^\psi} + \{\mathsf{KB1}^\psi,\mathsf{KB5}^\psi,\mathsf{KB4}^\psi\} \tag{.3}$$

Proof The proof of Equation .3 is purely syntactic. Note first that

$$B^\psi\bot \leftrightarrow B^\psi\neg\psi \in (\mathsf{P})_{B^\psi} \tag{3}$$

This fact will be used in the following proof:

1	$\hat{K}\varphi \wedge \hat{K}\psi$	Hypothesis
2	$\hat{K}(\varphi \vee \psi)$	$1,\mathsf{K}$
3	$\neg B^{\psi\vee\varphi}\bot$	$2,\mathsf{KB5}^\psi$
4	$\neg B^{\varphi\vee\psi}\neg(\varphi \vee \psi)$	3, Equation 3
5	$\neg B^{\varphi\vee\psi}(\neg\varphi \wedge \neg\psi)$	4, rewriting
6	$\neg(B^{\varphi\vee\psi}\neg\varphi \wedge B^{\varphi\vee\psi}\neg\psi)$	$5,\mathsf{C2}$
7	$\neg B^{\varphi\vee\psi}\neg\varphi \vee \neg B^{\varphi\vee\psi}\neg\psi$	rewriting
8	$K\left(\hat{K}(\varphi \vee \psi) \rightarrow \neg B^{\varphi\vee\psi}\neg\varphi\right) \vee K\left(\hat{K}(\varphi \vee \psi) \rightarrow \neg B^{\varphi\vee\psi}\neg\psi\right)$	$7,\mathsf{KB4}^\psi$
9	$K(\psi \rightarrow \neg B^{\varphi\vee\psi}\neg\varphi) \vee K(\varphi \rightarrow \neg B^{\varphi\vee\psi}\neg\psi)$	$8,\mathsf{T}$
10	$K\left(\psi \rightarrow \hat{K}\varphi\right) \vee K\left(\varphi \rightarrow \hat{K}\psi\right)$	$9,\mathsf{KB1}^\psi$
11	$\hat{K}(\psi \wedge \hat{K}\varphi) \vee \hat{K}(\varphi \wedge \hat{K}\psi)$	$1,10,\mathsf{K}$
12	$\hat{K}(\psi \wedge \hat{K}\varphi) \vee \hat{K}(\varphi \wedge \hat{K}\psi) \vee \hat{K}(\varphi \wedge \psi)$	$11,\mathsf{K}$

C.2 Proof of Equation .3.2

$$.3.2 \in (\mathsf{S4})_K + (\mathsf{P})_{B^\psi} + \{\mathsf{KB1}^\psi, \mathsf{KB5}^\psi, \mathsf{KB4}^\psi, \mathsf{KB6}^\psi\} \tag{.3.2}$$

We first prove a lemma:

Lemma 1 *Let $\mathscr{M}$ be an epistemic-plausibility space. If $\mathscr{M} \models (S4)_K + (P)_{B^\psi} + \{KB1^\psi, KB4^\psi\}$, then $\mathscr{M} \models \big(\hat{K}\psi \to \hat{K}(\psi \wedge K(\psi \to \varphi))\big) \to B^\psi\varphi$.*

Proof Let $w \in \mathscr{M}$ and assume that $\mathscr{M}, w \models \hat{K}\psi \to \hat{K}(\psi \wedge K(\psi \to \varphi))$. Assume towards a contradiction that $\mathscr{M}, w \models \neg B^\psi\varphi$. Then, by definition, $Pl_w([\![\psi]\!]_w) \neq \bot$ and $Pl_w([\![\psi \wedge \varphi]\!]_w) \not> Pl_w([\![\psi \wedge \varphi]\!]_w)$. Because $Pl_w([\![\psi]\!]_w) \neq \bot$, it holds that $\mathscr{M}, w \models \neg B^\psi \bot$. Now, because $\models B^\psi\psi$ by Ref, we have that $\models B^\psi\neg\psi \to B^\psi\bot$ by axiom $\mathsf{C2}$, i.e. $\models \neg B^\psi\bot \to \neg B^\psi\neg\psi$. Therefore, $\mathscr{M}, w \models \neg B^\psi\neg\psi$. So, by axiom $\mathsf{KB1}^\psi$, $\mathscr{M}, w \models \hat{K}\psi$. Then, by assumption, $\mathscr{M}, w \models \hat{K}(\psi \wedge K(\psi \to \varphi))$. So, there is $v \in R_K(w)$ such that $\mathscr{M}, w \models \psi \wedge K(\psi \to \varphi)$. Therefore, $\mathscr{M}, v \models K(\psi \to \varphi)$, and so $\mathscr{M}, v \models B^\psi(\psi \to \varphi)$ by application of axiom $\mathsf{KB1}^\psi$. Therefore, $\mathscr{M}, w \models B^\psi\varphi$ because $\models B^\psi\psi$. Now, $\mathscr{M}, w \models \neg B^\psi\varphi$, and so $\mathscr{M}, w \models K\big(\hat{K}\psi \to \neg B^\psi\varphi\big)$ by axiom $\mathsf{KB4}^\psi$. So, $\mathscr{M}, v \models \hat{K}\psi \to \neg B^\psi\varphi$. Since $\mathscr{M}, v \models \psi$, we also have that $\mathscr{M}, v \models \hat{K}\psi$ by axiom T. Therefore, $\mathscr{M}, v \models \neg B^\psi\varphi$, which contradicts our previous deduction. So, we reach a contradiction, and then $\mathscr{M}, w \models B^\psi\varphi$.

We can now prove Equation .3.2.

Proof Let $\mathscr{M}$ be a model and $w \in \mathscr{M}$. Assume that $\mathscr{M} \models (\mathsf{S4})_K + (\mathsf{K})_{B^\psi} + \{\mathsf{KB1}^\psi, \mathsf{KB5}^\psi, \mathsf{KB4}^\psi, \mathsf{KB6}^\psi\}$.

Then, by Lemma 1, $\mathscr{M} \models \big(\hat{K}\psi \to \hat{K}(\psi \wedge K(\psi \to \varphi))\big) \to B^\psi\varphi$, i.e. $\mathscr{M} \models \neg B^\psi\varphi \to \big(\hat{K} \wedge K\big(\psi \to \hat{K}(\psi \wedge \neg\varphi)\big)\big)$. Now, because $\mathscr{M} \models \{\mathsf{KB1}^\psi, \mathsf{KB6}^\psi\}$, it holds that $\mathscr{M} \models B\neg\psi \to (K(\psi \to \varphi) \leftrightarrow B^\psi\varphi)$. Therefore, $\mathscr{M} \models B\neg\psi \wedge \hat{K}(\psi \wedge \neg\varphi) \to \big(\hat{K}\psi \wedge K\big(\psi \to \hat{K}(\psi \wedge \neg\varphi)\big)\big)$. So,

$$\mathscr{M} \models \big(\hat{K}K\neg\psi \wedge \hat{K}(\psi \wedge \varphi)\big) \to \hat{K}\psi \wedge K\big(\psi \to \hat{K}(\psi \wedge \varphi)\big) \tag{4}$$

Now, assume that $\mathscr{M}, w \models \hat{K}\varphi \wedge \hat{K}K\neg\psi$. We will show that

$$\mathscr{M}, w \models K\big(\hat{K}\varphi \vee \neg\psi\big) \tag{5}$$

1. If $\mathscr{M}, w \models K\neg\psi$, then 5 holds.
2. If $\mathscr{M}, w \models \hat{K}\psi$, then, because $\mathscr{M} \models (\mathsf{S4})_K + (\mathsf{K})_{B^\psi} + \{\mathsf{KB1}^\psi, \mathsf{KB5}^\psi, \mathsf{KB4}^\psi\}$, it holds that $\mathscr{M} \models .3$ by Equation .3.
   Now, because $\mathscr{M}, w \models \hat{K}\varphi$ by assumption, by application of .3, it holds that either $\mathscr{M}, w \models \hat{K}(\psi \wedge \hat{K}\varphi)$ or $\mathscr{M}, w \models \hat{K}(\varphi \wedge \hat{K}\psi)$.
   (a) If $\mathscr{M}, w \models \hat{K}(\psi \wedge \hat{K}\varphi)$, then by application of 4, it holds that
   $\mathscr{M}, w \models \hat{K}\psi \wedge K\big(\psi \to \hat{K}(\psi \wedge \hat{K}\varphi)\big)$,
   then $\mathscr{M}, w \models K(\psi \to \hat{K}\hat{K}\varphi)$
   i.e. $\mathscr{M}, w \models K(\psi \to \hat{K}\varphi)$
   i.e. $\mathscr{M}, w \models K(\neg\psi \vee \hat{K}\varphi)$
   (b) If $\mathscr{M}, w \models \hat{K}(\varphi \wedge \hat{K}\psi)$, then, because $\models \hat{K}K\neg\hat{K}\psi \leftrightarrow \hat{K}K\neg\psi$, we have that $\mathscr{M}, w \models \hat{K}K\neg\hat{K}\psi$. Therefore, by application of 4, it holds that
   $\mathscr{M}, w \models \hat{K}\hat{K}\psi \wedge K\big(\hat{K}\psi \to \hat{K}\big(\hat{K}\psi \wedge \varphi\big)\big)$.
   So, $\mathscr{M}, w \models K\big(\hat{K}\psi \to \hat{K}\varphi\big)$
   i.e. $\mathscr{M}, w \models K(\psi \to \hat{K}\varphi)$ because $\models \psi \to \hat{K}\psi$,
   i.e. $\mathscr{M}, w \models K(\neg\psi \vee \hat{K}\varphi)$.

References

Adams, E. (1975). *The Logic of Conditionals*, volume 86 of *Synthese Library*. Springer.

Alchourrón, C. E., Gärdenfors, P., and Makinson, D. (1985). On the logic of theory change: Partial meet contraction and revision functions. *J. Symb. Log.*, 50(2):510–530.

Arlo-Costa, H. (1999). Qualitative and probabilistic models of full belief. In Buss, S., P.Hajek, and Pudlak, P., editors, *Logic Colloquium'98*, Lecture Notes on Logic 13.

Artemov, S. and Fitting, M. (2011). Justification logic. In Zalta, E. N., editor, *The Stanford Encyclopedia of Philosophy*. Fall 2011 edition.

Aucher, G. (2004). A combined system for update logic and belief revision. In Barley, M. and Kasabov, N. K., editors, *PRIMA*, volume 3371 of *Lecture Notes in Computer Science*, pages 1–17. Springer.

Aucher, G. (2007). Interpreting an action from what we perceive and what we expect. *Journal of Applied Non-Classical Logics*, 17(1):9–38.

Aucher, G. (2010). An internal version of epistemic logic. *Studia Logica*, 1:1–22.

Ayer, A. J. (1956). *The Problem of Knowledge*. Penguin books, London.

Baltag, A. and Smets, S. (2006). Conditional doxastic models: A qualitative approach to dynamic belief revision. *Electronic Notes in Theoretical Computer Science*, 165:5–21.

Baltag, A. and Smets, S. (2008a). *Texts in Logic and Games*, volume 4, chapter The Logic of Conditional Doxastic Actions, pages 9–31. Amsterdam University Press.

Baltag, A. and Smets, S. (2008b). *Texts in Logic and Games*, volume 3, chapter A Qualitative Theory of Dynamic Interactive Belief Revision, pages 9–58. Amsterdam University Press.

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

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

Bonnay, D. and Egré, P. (2008). Inexact knowledge with introspection. *Journal of Philosophical Logic*, 38(2):179–227.

Boutilier, C. (1994). Conditional logics of normality: A modal approach. *Artif. Intell.*, 68(1):87–154.

Broersen, J., Dastani, M., Hulstijn, J., Huang, Z., and van der Torre, L. W. N. (2001). The boid architecture: conflicts between beliefs, obligations, intentions and desires. In *Proceedings of the fifth international conference on Autonomous agents*, pages 9–16.

Castañeda, H.-N. (1964). Review of 'knowledge and belief'. *Journal of Symbolic Logic*, 29:132–134.

Cohen, P. and Levesque, H. (1990). Intention is choice with commitment. *Artificial intelligence*, 42:213–261.

de Rijke, M. (2000). A note on graded modal logic. *Studia Logica*, 64(2):271–283.

Deschene, F. and Wang, Y. (2010). To know or not to know: epistemic approach to security protocol verification. *Synthese*, 177(Supplement 1):51–76.

Dubois, D. and Prade, H. (1991). Possibilistic logic, preferential model and related issue. In *Proceedings of the 12th International Conference on Artificial Intelligence (IJCAI)*, pages 419–425. Morgan Kaufman.

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

Fagin, R. and Halpern, J. Y. (1987). Belief, awareness, and limited reasoning. *Artificial Intelligence*, 34(1):39 – 76.

Fine, K. (1972). In so many possible worlds. *Notre Dame Journal of Formal Logic*, 13(4):516–520.

Floridi, L. (2006). The logic of being informed. *Logique et Analyse*, 49(196):433–460.

Friedman, N. and Halpern, J. Y. (1997). Modeling belief in dynamic systems, part i: Foundations. *Artificial Intelligence*, 95(2):257 – 316.

Friedman, N. and Halpern, J. Y. (2001). Plausibility measures and default reasoning. *Journal of the ACM*, 48(4):648–685.

Gärdenfors, P. (1988). *Knowledge in Flux (Modeling the Dynamics of Epistemic States)*. Bradford/MIT Press, Cambridge, Massachusetts.

Georgeff, M. and Rao, A. (1991). Asymmetry thesis and side-effect problems in linear time and branching time intention logics. In *Proceedings of the Twelfth International Joint Conference on Artificial Intelligence (IJCAI-91)*, pages 498–504, (Sydney, Australia).

Gettier, E. (1963). Is justified true belief knowledge? *Analysis*, 25:121–123.

Gochet, P. (2007). *Logique épistémique et philosophie des mathématiques*, chapter Un problème ouvert en épistémologie : la formalisation du savoir-faire. Philosophie des sciences. Vuibert.

Gochet, P. and Gribomont, P. (2006). *Handbook of the History of Logic*, volume 7, Twentieth Century Modalities, chapter Epistemic Logic, pages 99–195. Elsevier, Amsterdam.

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

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

Halpern, J. and Moses, Y. (1992). A guide to completeness and complexity for modal logics of knowledge and belief. *Artificial Intelligence*, 54:311–379.

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

Halpern, J. Y. and Pucella, R. (2002). Modeling adversaries in a logic for security protocol analysis. In Abdallah, A. E., Ryan, P., and Schneider, S., editors, *Formal Aspects of Security*, volume 2629 of *Lecture Notes in Computer Science*, pages 115–132. Springer.

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

Halpern, J. Y. and Rêgo, L. C. (2009). Reasoning about knowledge of unawareness revisited. In Heifetz, A., editor, *TARK*, pages 166–173.

Halpern, J. Y., Samet, D., and Segev, E. (2009a). Defining knowledge in terms of belief: the modal logic perspective. *The Review of Symbolic Logic*, 2:469–487.

Halpern, J. Y., Samet, D., and Segev, E. (2009b). On definability in multimodal logic. *The Review of Symbolic Logic*, 2:451–468.

Heifetz, A., Meier, M., and Schipper, B. (2006). Interactive unawareness. *Journal of Economic Theory*, 130(1):78–94.

Hemp, D. (2006). The KK (knowing that one knows) principle. *The Internet Encyclopedia of Philosophy*, `http://www.iep.utm.edu/kk-princ/print/`.

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

Hendricks, V. and Symons, J., editors (2006). *8 bridges between formal and mainstream epistemology*, volume 128. Springer.

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

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

Kraus, S. and Lehmann, D. (1986). Knowledge, belief and time. In Kott, L., editor, *Automata, Languages and Programming*, volume 226 of *Lecture Notes in Computer Science*, pages 186–195. Springer Berlin / Heidelberg.

Kraus, S., Lehmann, D. J., and Magidor, M. (1990). Nonmonotonic reasoning, preferential models and cumulative logics. *Artificial Intelligence*, 44(1-2):167–207.

Kutschera, F. v. (1976). *Einführung in die intensional Semantik*. W. de Gruyter, Berlin.

Kyburg, H. (1961). *Probability and the Logic of Rational Belief*. Wesleyan University Press, Middletown, CT.

Lamarre, P. and Shoham, Y. (1994). Knowledge, certainty, belief, and conditionalisation (abbreviated version). In *KR*, pages 415–424.

Laverny, N. and Lang, J. (2005). From knowledge-based programs to graded belief-based programs, part ii: off-line reasoning. In *IJCAI*, pages 497 – 502.

Lehrer, K. and Paxson, T. (1969). Knowledge: Undefeated justified true belief. *The Journal of Philosophy*, 66:225–237.

Lenzen, W. (1978). *Recent Work in Epistemic Logic*. Acta Philosophica Fennica 30. North Holland Publishing Company.

Lenzen, W. (1979). Epistemologische betractungen zu [S4;S5]. *Erkenntnis*, 14:33–56.

Lenzen, W. (2004). Knowledge, belief, and subjective probability: Outlines of a unified system of epistemic/doxastic logic. In Hendricks, V. F., Jørgensen, K. F., Pedersen, S. A., Hendricks, V. F., Symons, J., Dalen, D., Kuipers, T. A., Seidenfeld, T., Suppes, P., and WoleŃski, J., editors, *Knowledge Contributors*, volume 322 of *Synthese Library*, pages 17–31. Springer Netherlands.

Levesque, H. (1984). A logic of implicit and explicit knowledge. In *AAAI-84*, pages 198–202, Austin Texas.

Levi, I. (1997). *The covenant of reason: rationality and the commitments of thought*, chapter The Logic of Full Belief, pages 40–69. Cambridge University Press.

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

Lorini, E. and Castelfranchi, C. (2007). The cognitive structure of surprise: looking for basic principles. *Topoi: An International Review of Philosophy*, 26(1):133–149.

Lycan, W. (2006). *Epistemology Futures*, chapter On the Gettier Problem Problem, pages 148–168. Oxford University Press.

Makinson, D. and Gärdenfors, P. (1989). Relations between the logic of theory change and nonmonotonic logic. In Fuhrmann, A. and Morreau, M., editors, *The Logic of Theory Change*, volume 465 of *Lecture Notes in Computer Science*, pages 185–205. Springer.

Meyer, J.-J. C., de Boer, F., van Eijk, R., Hindriks, K., and van der Hoek, W. (2001). On programming KARO agents. *Logic Journal of the IGPL*, 9(2).

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

Moses, Y. and Shoham, Y. (1993). Belief as defeasible knowledge. *Artificial intelligence*, 64(2):299–321.

Pacuit, E. (2012). *Dialogue, Rationality, Formalism*, chapter Procedural Information and the Dynamics of Belief. LEUS. Springer.

Prichard, D. (2004). Epistemic luck. *Journal of Philosophical Research*, 29:193–222.

Rao, A. and Georgeff, M. (1991). Modeling rational agents within a BDI-architecture. In Fikes, R. and Sandewall, E., editors, *Proceedings of Knowledge Representation and Reasoning (KR & R-91)*, pages 473–484. Morgan Kaufmann Publishers.

R.C.Moore (1984). Possible-world semantics for autoepistemic logic. In *Proceedings of the Non-Monotonic Reasoning Workshop*, pages 344–354, New Paltz NY.

R.C.Moore (1995). *Logic and Representation*. CSLI Lecture Notes.

Rott, H. (2004). Stability, strength and sensitivity: Converting belief into knowledge. *Erkenntnis*, 61:469–493. 10.1007/s10670-004-9287-1.

Sakama, C., Caminada, M., and Herzig, A. (2010). A logical account of lying. In Janhunen, T. and Niemelä, I., editors, *Logics in Artificial Intelligence*, volume 6341 of *Lecture Notes in Computer Science*, pages 286–299. Springer Berlin / Heidelberg.

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

Spohn, W. (1988a). *Causation in Decision, Belief Change, and Statistics*, volume 2, chapter Ordinal conditional functions: A dynamic theory of epistemic states, pages 105–134. reidel, Dordrecht.

Spohn, W. (1988b). A general non-probabilistic theory of inductive reasoning. In Shachter, R. D., Levitt, T. S., Kanal, L. N., and Lemmer, J. F., editors, *UAI*, pages 149–158. North-Holland.

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

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

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

van der Hoek, W. (1993). Sytems for knowledge and belief. *Journal of Logic and Computation*, 3(2):173–195.

van der Hoek, W. and Meyer, J.-J. C. (1992). Graded modalities in epistemic logic. In Nerode, A. and Taitslin, M. A., editors, *LFCS*, volume 620 of *Lecture Notes in Computer Science*, pages 503–514. Springer.

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

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

van Ditmarsch, H. P., van Eijck, J., Sietsma, F., and Wang, Y. (2011). On the logic of lying. In van Eijck, J. and Verbrugge, R., editors, *Games, Actions and Social Software*, Texts in Logic and Games, LNAI-FoLLI. Springer.

van Linder, B., van der Hoek, W., and Meyer, J.-J. C. (1998). Formalising abilities and opportunities of agents. *Fundamenta Informaticae*, 34(1-2):53–101.

Voorbraak, F. (1993). *As Far as I know. Epistemic Logic and Uncertainty*. PhD thesis, Utrecht University.

Wheeler, G. (2012). Is there a logic of information? unpublished manuscript.

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

Wooldridge, M. (2000). *Reasoning About Rational Agents*. MIT Press.