Intricate Axioms as Interaction Axioms

Guillaume Aucher

Intricate Axioms as Interaction Axioms*

Guillaume Aucher†
University of Rennes 1
INRIA
263, Avenue du Général Leclerc
35042 Rennes Cedex, France

Abstract

In epistemic logic, some axioms dealing with the notion of knowledge are rather convoluted and difficult to interpret intuitively, even though some of them, such as the axioms .2 and .3, are considered to be key axioms by some epistemic logicians. We show that they can be characterized in terms of understandable interaction axioms relating knowledge and belief or knowledge and conditional belief. In order to show it, we first sketch a theory dealing with the characterization of axioms in terms of interaction axioms in modal logic. We then apply the main results and methods of this theory to obtain specific results related to epistemic and doxastic logics.

Keywords: Modal Logic, Epistemic logic, Interaction Axiom, Definability of Modalities.

1 Introduction

One of the goals of modern epistemic logic is to elucidate the nature of the interaction between knowledge and belief by means of formal and logical methods. On the basis of a semantics very close to the Kripke semantics of modal logic, Hintikka [18] and subsequent philosophers and logicians tried to formulate explicit principles governing and relating expressions of the form "a knows that φ" (subsequently formalized as $K\varphi$) and "a believes that φ" (subsequently formalized as $B\varphi$), where a is a human agent and φ is a proposition. In other words, they sought to determine 'the' logic of knowledge and belief, or at least of idealized versions of these notions. Their quest was grounded in the observation that our intuitions about these epistemic notions comply with some systematic reasoning properties, and was driven by the attempt to better *understand* and *elucidate* them [24, p. 15]. For example, the interaction axioms $K\varphi \rightarrow B\varphi$ and

*A short version of this article appears in [2]. This short version only deals with axioms .2 and .4 and does not deal with conditional beliefs.

†Email: guillaume.aucher@irisa.fr

$B\varphi \rightarrow KB\varphi$ are often considered to be intuitive principles: if agent a knows φ then (s)he also believes φ, and if agent a believes φ, then (s)he knows that (s)he believes φ. As a matter of fact, assessing whether a given principle holds true or not raises our own awareness of these epistemic notions and reveals to us some of their essential properties.

In computer science, the logic of knowledge is usually considered to be $\mathsf{S5}$, which is obtained by adding to the minimal normal modal logic K the axioms $K\varphi \rightarrow \varphi$ (T), $K\varphi \rightarrow KK\varphi$ ($\mathsf{4}$) and $\neg K\varphi \rightarrow K\neg K\varphi$ ($\mathsf{5}$). This last axiom $\mathsf{5}$ is falsified in a situation where the agent has mistaken beliefs. For this very reason, it has been attacked by various philosophers because it cannot hold in general: agents sometimes have mistaken beliefs.[1] Dropping this axiom $\mathsf{5}$ from $\mathsf{S5}$, we obtain the logic $\mathsf{S4}$. Between the logics $\mathsf{S4}$ and $\mathsf{S5}$, a rich variety of weaker logics of knowledge have been proposed and examined by epistemic logicians [25], such as $\mathsf{S4.2}$, $\mathsf{S4.3}$ and $\mathsf{S4.4}$. Even if these logics are characterized by axioms which are rather intricate, some of them have been proclaimed by some epistemic logicians as key axioms characterizing the notion of knowledge. For example, Lenzen claimed that "[t]here is strong evidence in favor of the assumption that $\mathsf{S4.2}$ is *the* logic of knowledge" [25, p. 33], where the axiom .2 is $\neg K\neg K\varphi \rightarrow K\neg K\neg\varphi$. Likewise, Kutschera argues for $\mathsf{S4.4}$ as the logic of knowledge, where the axiom .4 is $(\varphi \wedge \neg K\neg K\varphi) \rightarrow K\varphi$ [21]. As one can easily observe, it is difficult to provide these axioms with a natural and easily understandable reading. In fact, Lenzen derived his axiom .2 from a set of interaction axioms relating knowledge and belief (viewed as some sort of conviction). Similarly, the logic $\mathsf{S4.3}$ is the logic $\mathsf{S4}$ to which is added the axiom .3: $\neg K\neg\varphi \wedge \neg K\neg\psi \rightarrow \neg K\neg(\varphi \wedge \neg K\neg\psi) \vee \neg K\neg(\varphi \wedge \psi) \vee \neg K\neg(\psi \wedge \neg K\neg\varphi)$. Again, as one can easily notice, it is difficult to provide this axiom with a natural and easily understandable reading. Stalnaker argues that a certain "defeasibility theory" of knowledge gives $\mathsf{S4.3}$ [32, p.190].

To better grasp the intuitions underlying these intricate axioms .2, .3 and .4, we show that they can be characterized equivalently in terms of interaction axioms relating knowledge and belief or knowledge and conditional belief. In order to do so, we first need to explain what we mean by "interaction axiom" and what we mean by "characterizing an axiom in terms of interaction axioms". This will lead us to develop a basic theory in modal logic dealing with these notions. Then, we will apply the general results of this theory to the specific case of epistemic logic. Note that such a theory has never been developed in the modal logic literature, neither in the context of multi-modal logics nor in the context of combinations of modal logics such as products and fusion [26, 14].

This article is divided in two parts: a "theoretical" part (Section 2) which presents our basic theory dealing with the characterization of axioms in terms of

[1] For example, assume that a university lecturer 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 $\mathsf{5}$ 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 $\mathsf{5}$ is invalidated when the agent has mistaken beliefs which can be due for example to misperceptions, lies or other forms of deception.

interaction axioms, and an "applicative" part (Sections 3, 4, 5, 6) where we apply some of the results of Section 2 to epistemic and doxastic logics. These logics are recalled in Section 3. The interaction axioms introduced in the literature are discussed in Section 4 and our main results related to epistemic logic are in Sections 5 and 6.

Note. The missing proofs of propositions and theorems can be found in the appendix.

2 Towards a Theory of Interaction Axioms

In this section, we start by recalling the basics of modal logic (Section 2.1). Then, we present our basic meta-theory of modal logic dealing with interaction axioms. Firstly, we define what we mean by "interaction axiom" and what we mean by "characterizing an axiom in terms of interaction axiom" (Section 2.2). Secondly, we give general results which provide some necessary and sufficient conditions for an axiom to be characterized by a set of interaction axioms when one of the modalities is defined in terms of the other modality (Section 2.3).

2.1 Modal Logic

In this subsection, we recall the basics of modal logic. The semantics we consider for modal logic is the standard Kripke semantics.

2.1.1 Syntax

In the rest of the article, $\mathbb{P}$ is a set of propositional letters and $\mathbb{A} \subseteq \{1, 2\}$, with $\mathbb{P}$ and $\mathbb{A}$ non-empty. We define the *modal language* $\mathcal{L}_{\mathbb{A}}$ by the following BNF grammar:

$$\mathcal{L}_{\mathbb{A}} \quad : \quad \varphi \quad ::= \quad p \quad | \quad \neg\varphi \quad | \quad (\varphi \wedge \varphi) \quad | \quad [i]\,\varphi$$

where p ranges over $\mathbb{P}$ and i ranges over $\mathbb{A}$. The formula $\top$ is an abbreviation for $p \vee \neg p$ (for some fixed $p \in \mathbb{P}$), the formula $\bot$ is an abbreviation for $\neg\top$, $\varphi \vee \psi$ an abbreviation for $\neg(\neg\varphi \wedge \neg\psi)$, $\varphi \to \psi$ an abbreviation for $\neg\varphi \vee \psi$, and $\langle i\rangle\varphi$ an abbreviation for $\neg\,[i]\,\neg\varphi$. An occurrence of a proposition letter p is a *positive* occurrence if it is in the scope of an even number of negation signs $\neg$. A formula φ is *positive in* p if all occurrences of p in φ are positive. If $\Gamma := \{\varphi_1, \ldots, \varphi_n\} \subseteq \mathcal{L}_{\mathbb{A}}$, then $\bigwedge\Gamma$ is an abbreviation for $\varphi_1 \wedge \ldots \wedge \varphi_n$.

A *(modal) logic* L for the modal language $\mathcal{L}_{\mathbb{A}}$ is a set of formulas of $\mathcal{L}_{\mathbb{A}}$ called *theorems* which contains all propositional tautologies and which 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 closed under uniform substitution, that is, if φ belongs to L then so do all of its substitution instances (see [7, Def. 1.18] for the definition of a substitution instance).

A modal logic is usually defined by a set of *inference rules* and of formulas called *axioms*. A formula is a *theorem* of 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 axiom schema $([i](\varphi \to \psi) \wedge [i]\varphi) \to [i]\psi$, and the inference rule of necessitation: from $\varphi \in \mathsf{L}$, infer $[i]\varphi \in \mathsf{L}$, for all $i \in \mathbb{A}$. Let $A \subseteq \mathcal{L}_{\mathbb{A}}$. A modal logic for $\mathcal{L}_{\mathbb{A}}$ *generated* by the set A is the smallest normal modal logic for $\mathcal{L}_{\mathbb{A}}$ containing A. In that case, the formulas of A are called *axioms*. The smallest of all normal modal logics for $\mathcal{L}_{\mathbb{A}}$ is denoted K.

If L and L' are two sets of formulas of $\mathcal{L}_{\mathbb{A}}$ (possibly logics), we denote by $\mathsf{L}+\mathsf{L}'$ the modal logic for $\mathcal{L}_{\mathbb{A}}$ generated by $\mathsf{L} \cup \mathsf{L}'$ (it is very similar to the *fusion* of L and L' [26, 14]). If x is a formula of $\mathcal{L}_{\mathbb{A}}$, then $\mathsf{L}+\mathsf{x}$ abusively denotes $\mathsf{L}+\{\mathsf{x}\}$. 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.

Let $\mathsf{x} \in \mathcal{L}_{\mathbb{A}}$ and let $X, X' \subseteq \mathcal{L}_{\mathbb{A}}$. We say that x is *derivable from* X *in* L when $\mathsf{x} \in \mathsf{L}+X$ and in that case we write $X \vdash_{\mathsf{L}} \mathsf{x}$. We also write $X \vdash_{\mathsf{L}} X'$ when $X \vdash_{\mathsf{L}} \mathsf{x}'$ for all $\mathsf{x}' \in X'$, and $X >_{\mathsf{L}} X'$ when it holds that $X \vdash_{\mathsf{L}} X'$ but it does not hold that $X' \vdash_{\mathsf{L}} X$.

2.1.2 Kripke Semantics.

The Kripke semantics will be used only in the proof of Theorem 1. A *(bi-modal) Kripke model* $\mathcal{M}$ is a tuple $\mathcal{M} = (W, R_1, R_2, V)$ where W is a non-empty set of possible worlds, $R_1, R_2 \in 2^{W \times W}$ are binary relations over W called *accessibility relations*, and $V : \mathbb{P} \to 2^W$ is called a *valuation* and assigns to each propositional letter $p \in \mathbb{P}$ a subset of W. A *Kripke frame* $\mathcal{F}$ is a Kripke model without a valuation. We often denote by $R_i(w)$ the set $R_i(w) := \{v \in W \mid wR_iv\}$ and we abusively write $w \in \mathcal{M}$ when $w \in W$. In that case, $(\mathcal{M}, w)$ is called a *pointed Kripke model* and we denote by $\mathcal{K}$ the set of all pointed Kripke models.

Let $\varphi \in \mathcal{L}_{\mathbb{A}}$, let $\mathcal{M}$ be a Kripke model and let $w \in \mathcal{M}$. The *satisfaction relation* $\mathcal{M}, w \models \varphi$ is defined inductively as follows:

$$\begin{array}{lll} \mathcal{M}, w \models p & \text{iff} & w \in V(p) \\ \mathcal{M}, w \models \varphi \wedge \varphi' & \text{iff} & \mathcal{M}, w \models \varphi \text{ and } \mathcal{M}, w \models \varphi' \\ \mathcal{M}, w \models \neg\varphi & \text{iff} & \text{not } \mathcal{M}, w \models \varphi \\ \mathcal{M}, w \models [i]\varphi & \text{iff} & \text{for all } v \in R_i(w), \mathcal{M}, v \models \varphi. \end{array}$$

If $\Gamma \subseteq \mathcal{L}_{\mathbb{A}}$, then we write $\mathcal{M}, w \models \Gamma$ when $\mathcal{M}, w \models \varphi$ for all $\varphi \in \Gamma$. Likewise, if $\mathcal{F} = (W, R_1, R_2)$ is an epistemic-doxastic frame, then we abusively write $w \in \mathcal{F}$ for $w \in W$. We also write $\Gamma \models \varphi$ when for all $(\mathcal{M}, w) \in \mathcal{K}$, $\mathcal{M}, w \models \Gamma$ implies that $\mathcal{M}, w \models \varphi$. If Γ is a set of formulas of $\mathcal{L}_{1,2}$, then we write $\mathcal{F} \models \Gamma$ when for all $\varphi \in \Gamma$ and all valuation V, $(\mathcal{F}, V) \models \varphi$, and we say that Γ is *valid in* $\mathcal{F}$.

2.2 Interaction Axioms

In the sequel, L_1 and L_2 are two normal modal logics for $\mathcal{L}_1$ and $\mathcal{L}_2$ respectively, and $\mathsf{L}_{1,2}$ is a normal modal logic for $\mathcal{L}_{1,2}$ (we abusively write $\mathcal{L}_{1,2}$ for $\mathcal{L}_{\{1,2\}}$ and $\mathcal{L}_i$ for $\mathcal{L}_{\{i\}}$, where $i \in \{1, 2\}$). Intuitively, an interaction axiom is a formula

which cannot be equivalent to any formula of the modal language with a single modality.

Definition 1. A *set of interaction axioms* w.r.t. a pair of logics $(\mathsf{L}_1, \mathsf{L}_2)$ is a finite set of formulas $\Gamma \subseteq \mathcal{L}_{1,2}$ for which there is no $\chi \in \mathcal{L}_1 \cup \mathcal{L}_2$ such that

$$\chi \leftrightarrow \bigwedge \Gamma \in \mathsf{L}_1 + \mathsf{L}_2. \tag{1}$$

In the sequel, x is a formula of $\mathcal{L}_1$ and Γ is a set of interaction axioms w.r.t. $(\mathsf{L}_1, \mathsf{L}_2)$.

Definition 2. We say that x is *characterized* by the set of interaction axioms Γ w.r.t. $(\mathsf{L}_1, \mathsf{L}_2)$ when

$$\mathsf{L}_1 + \mathsf{x} = (\mathsf{L}_1 + \mathsf{L}_2 + \Gamma) \cap \mathcal{L}_1. \tag{2}$$

Moreover, x is *conservatively characterized* by Γ w.r.t. $(\mathsf{L}_1, \mathsf{L}_2)$ when the set of interaction axioms Γ satisfies the following condition as well:

$$\mathsf{L}_2 = (\mathsf{L}_1 + \mathsf{L}_2 + \Gamma) \cap \mathcal{L}_2. \tag{3}$$

Intuitively, Definition 2 tells us that an axiom x is characterized by a set of interaction axioms Γ if, when we add the interaction axioms to the base logics, we derive exactly the theorems for the language $\mathcal{L}_1$ obtained by only adding axiom x to L_1, and nothing else. For the case of conservative characterization, by adding these interaction axioms Γ, we do not even obtain new theorems for the language $\mathcal{L}_2$ as 'side effects', the theorems for this language just remain the same as before.

Finally, we define a notion of minimality among the sets of interaction axioms characterizing an axiom x.

Definition 3. The axiom x is *minimally characterized* by the set of interaction axioms Γ w.r.t. $(\mathsf{L}_1, \mathsf{L}_2)$ when x is characterized by Γ w.r.t. $(\mathsf{L}_1, \mathsf{L}_2)$ and there is no set of interaction axioms Γ' such that $\Gamma >_{\mathsf{L}_1+\mathsf{L}_2} \Gamma'$ and x is still characterized by Γ' w.r.t. $(\mathsf{L}_1, \mathsf{L}_2)$.

2.3 Definability of Modalities and Characterization of Axioms

The definability of modalities in terms of other modalities is studied from a theoretical point of view by Halpern et al. [17]. This study is subsequently applied to epistemic logic by the same authors in [16]. 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 article, we are interested only in the notion of *explicit* definability, which is also used in [25].

Definition 4. Let $\{i,j\} = \{1,2\}$. The modality $\langle i \rangle$ is *explicitly defined* in the logic $\mathsf{L}_{i,j}$ in terms of the modality $\langle j \rangle$ by a formula $\mathsf{def}_i(p) \in \mathcal{L}_j$ if, and only if,

$$\langle i \rangle p \leftrightarrow \mathsf{def}_i(p) \in \mathsf{L}_{i,j}. \qquad (\mathsf{Def}\ \langle i \rangle)$$

The following key theorem will play an important role in the last section. It provides necessary and sufficient conditions for an axiom to be (conservatively) characterized by a set of interaction axioms. It also states that in case one of the modalities is definable in terms of the other, then this characterization is actually *minimal* (in the sense of Definition 3).

Theorem 1. *Assume that* $\langle 2 \rangle$ *is explicitly defined in* $\mathsf{L}_1 + \mathsf{L}_2 + \Gamma$ *in terms of* $\langle 1 \rangle$ *by a formula* $\mathsf{def}_2(p) \in \mathcal{L}_1$ *positive in* p*. Then, the following are equivalent:*

- x *is characterized by* Γ *w.r.t.* $(\mathsf{L}_1, \mathsf{L}_2)$*;*
- $\mathsf{L}_1 + \mathsf{L}_2 + \Gamma = \mathsf{L}_1 + \mathsf{x} + \{\langle 2 \rangle p \leftrightarrow \mathsf{def}_2(p)\}$.

Moreover, assume that $\langle 1 \rangle$ *is also explicitly defined in* $\mathsf{L}_1 + \mathsf{L}_2 + \Gamma$ *in terms of* $\langle 2 \rangle$ *by a formula* $\mathsf{def}_1(p) \in \mathcal{L}_2$ *positive in* p*. Then, the following are equivalent:*

- x *is conservatively characterized by* Γ *w.r.t.* $(\mathsf{L}_1, \mathsf{L}_2)$*;*
- $\mathsf{L}_1 + \mathsf{L}_2 + \Gamma = \mathsf{L}_1 + \mathsf{x} + \{\langle 2 \rangle p \leftrightarrow \mathsf{def}_2(p)\}$ *and*
  $\mathsf{L}_1 + \mathsf{L}_2 + \Gamma = \mathsf{L}_2 + \{\langle 1 \rangle p \leftrightarrow \mathsf{def}_1(p)\}$.

Finally, in both cases, the axiom x *is (conservatively) characterized by* Γ *w.r.t.* $(\mathsf{L}_1, \mathsf{L}_2)$ *if, and only if, it is* minimally *(conservatively) characterized by* Γ *w.r.t.* $(\mathsf{L}_1, \mathsf{L}_2)$.

3 Epistemic and Doxastic Logics

In this section, we will see the standard formal semantics of knowledge, belief and conditional belief. For examples and applications of these semantics in computer science, the interested reader can consult [10] or [27]. We will also introduce the convoluted axioms .2, .3 and .4 together with the classes of frames they define.

3.1 Logics of Knowledge and Belief

3.1.1 Epistemic and doxastic languages

We define the epistemic-doxastic language $\mathcal{L}_{KB}$ by the following BNF grammar:

$$\mathcal{L}_{KB} \ : \ \varphi \ ::= \ p \ \mid \ \neg\varphi \ \mid \ (\varphi \wedge \varphi) \ \mid \ B\varphi \ \mid \ K\varphi$$

where p ranges over $\mathbb{P}$. The propositional language $\mathcal{L}_0$ is the language $\mathcal{L}_{KB}$ without the knowledge and belief operators K and B. The language $\mathcal{L}_K$ is the language $\mathcal{L}_{KB}$ without the belief operator B, and the language $\mathcal{L}_B$ is the

language $\mathcal{L}_{KB}$ 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 $\langle B\rangle\varphi$ and $\langle K\rangle\varphi$ are abbreviations of $\neg B\neg\varphi$ and $\neg K\neg\varphi$ respectively.

Remark 1. We have to be careful with the notion of belief, since the term 'belief' has 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'. Lenzen argues that there are two different kinds of belief, which he calls *weak* and *strong* belief (or *conviction*) [24]. A relatively detailed analysis distinguishing weak from strong belief is presented by Shoham and Leyton-Brown [29, p. 414-415]. Also see [24, 3]. In this paper, we only deal with the notion of strong belief (or conviction) and $B\varphi$ stands for this notion. We will sometimes denote B_w for the notion of *weak* belief.

3.1.2 Semantics

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 was proposed by Kraus and Lehmann [19].

An *epistemic-doxastic model* $\mathcal{M}$ is a (bi-modal) Kripke model as defined in Section 2.1 where R_1 is interpreted as the accessibility relation for knowledge and R_2 is interpreted as the accessibility relation for belief. The truth conditions for $\mathcal{M}, w \models \varphi$ are then defined as in Section 2.1. An *epistemic-doxastic frame* $\mathcal{F}$ is an epistemic-doxastic model without a valuation. We often denote $R_K(w) = \{v \in W \mid wR_Kv\}$ and $R_B(w) = \{v \in W \mid wR_Bv\}$.

3.1.3 Logics

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 article. We also give, below each property, the axiom which *defines* the class of epistemic-doxastic frames that fulfill this property (see [7, 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.

- *serial*: $R_K(w) \neq \emptyset$
  Axiom D: $K\varphi \rightarrow \langle K\rangle\varphi$;
- *transitive*: If $w' \in R_K(w)$ and $w'' \in R_K(w')$, then $w'' \in R_K(w)$
  Axiom $\mathsf{4}$: $K\varphi \rightarrow KK\varphi$;
- *Euclidean*: If $w' \in R_K(w)$ and $w'' \in R_K(w)$, then $w' \in R_K(w'')$
  Axiom $\mathsf{5}$: $\neg K\varphi \rightarrow K\neg K\varphi$;
- *reflexive*: $w \in R_K(w)$
  Axiom T: $K\varphi \rightarrow \varphi$;

- *symetric*: If $w' \in R_K(w)$, then $w \in R_K(w')$
  Axiom B: $\varphi \to K\neg K\neg\varphi$;
- *confluent*: If $w' \in R_K(w)$ and $w'' \in R_K(w)$, then there is v such that $v \in R_K(w')$ and $v \in R_K(w'')$
  Axiom .2: $\langle K\rangle K\varphi \to K\langle K\rangle\varphi$;
- *weakly connected*: If $w' \in R_K(w)$ and $w'' \in R_K(w)$, then $w' = w''$ or $w' \in R_K(w'')$ or $w'' \in R_K(w')$
  Axiom .3: $\langle K\rangle\varphi \wedge \langle K\rangle\psi \to \langle K\rangle(\varphi \wedge \psi) \vee \langle K\rangle(\psi \wedge \langle K\rangle\varphi) \vee \langle K\rangle(\varphi \wedge \langle K\rangle\psi)$;
- *R1*: If $w' \in R_K(w)$ and $w \neq w'$ and $w'' \in R_K(w)$, then $w' \in R_K(w'')$
  Axiom .4: $(\varphi \wedge \langle K\rangle K\varphi) \to K\varphi$.

The logic $\mathsf{KD45}_B$ is the smallest normal modal logic for $\mathcal{L}_B$ generated by the set of axioms $\{\mathsf{D}, \mathsf{4}, \mathsf{5}\}$. For any $\mathsf{x} \in \{.2, .3, .4\}$, the logic $\mathsf{S4.x}_K$ is the smallest normal modal logic for $\mathcal{L}_K$ generated by the set of axioms $\{\mathsf{T}, \mathsf{4}, \mathsf{x}\}$. We have the following relationship between these logics:

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

3.2 Logics of Knowledge and Conditional Belief

3.2.1 A language for knowledge and conditional belief

Taking up the work of Friedman and Halpern [13], we define the syntax of the language $\mathcal{L}_{KB^q}$ inductively by the following BNF grammar:

$$\mathcal{L}_{KB^q}: \quad \varphi \quad ::= \quad p \quad | \quad \neg\varphi \quad | \quad (\varphi \wedge \varphi) \quad | \quad B^{\varphi}\varphi \quad | \quad K\varphi$$

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

3.2.2 Semantics

Numerous semantics have been proposed for conditional beliefs, such as preferential structures [20], ϵ-semantics [1], possibilistic structures [9], and κ-ranking [31, 30]. They all have in common that they validate the axiomatic system P originally introduced in [20]. This remarkable fact is explained by Friedman and Halpern [13], where a general framework based on *plausibility measures* is proposed. As proved in that article, plausibility measures generalize all these semantics. So, we adopt the general framework of plausibility measures to provide a semantics for $\mathcal{L}_{KB^q}$. Plausibility spaces and epistemic-plausibility spaces were introduced by Friedman and Halpern respectively in [12] and [13]. We can nevertheless mention that other logical formalisms dealing with conditional beliefs are proposed in the economics literature [8]. These other formalisms have been taken up in the field of *dynamic* epistemic logic [4, 5, 6].

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 article, 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 $d_1 \neq d_2$.

Definition 5. A *(qualitative) plausibility space* is a tuple $S = (W, \mathcal{A}, Pl)$ where:

- W is a non-empty set of possible worlds;
- $\mathcal{A}$ is an algebra over W;
- $Pl : \mathcal{A} \to D$ is a function mapping sets of $\mathcal{A}$ 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 $\mathcal{S}$ the class of all (qualitative) plausibility spaces.

We can naturally introduce an accessibility relation R_K to (qualitative) plausibility space in order to give a semantics to the knowledge modality of the language $\mathcal{L}_{KB^q}$. This yields the notion of *epistemic plausibility space*:

Definition 6. An *epistemic-plausibility space* is a tuple $\mathcal{M} = (W, R_K, V, \mathcal{P})$ 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*;
- $V : \mathbb{P} \to 2^W$ is a function called a *valuation* mapping propositional variables to subsets of W;
- $\mathcal{P} : W \to \mathcal{S}$ is a function called a *plausibility assignment* mapping each world $w \in W$ to a (qualitative) plausibility space $(W_w, \mathcal{A}_w, Pl_w)$ such that $W_w \subseteq W$.

We denote by $\mathcal{S}_K$ the class of all (qualitative) epistemic-plausibility spaces.

Definition 7. Let $\varphi \in \mathcal{L}_{KB^q}$, let $\mathcal{M}$ be an epistemic-plausibility space and let $w \in \mathcal{M}$. The satisfaction relation $\mathcal{M}, w \models \varphi$ is defined inductively as follows:

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

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

3.2.3 Logics

The following theorem presents a slightly different version of the axiomatic system P originally introduced by Kraus, Lehmann and Magidor [20] for non-monotonic logics.

Theorem 2. *[13] The following logic K_{B^q} is a sound and complete axiomatization of $\mathcal{L}_{B^q}$ with respect to all (qualitative) plausibility spaces.*

Prop	:	*All inferences rules of propositional logic and substitution instances of propositional tautologies*
C1	:	$B^{\psi}\psi$
C2	:	$(B^{\psi}\varphi_1 \wedge B^{\psi}\varphi_2) \rightarrow B^{\psi}(\varphi_1 \wedge \varphi_2)$
C3	:	$(B^{\psi_1}\varphi \wedge B^{\psi_2}\varphi) \rightarrow B^{\psi_1 \vee \psi_2}\varphi$
C4	:	$(B^{\psi}\varphi \wedge B^{\psi}\chi) \rightarrow B^{\psi \wedge \varphi}\chi$
RC1	:	*If* $\psi \leftrightarrow \psi'$ *then* $B^{\psi}\varphi \leftrightarrow B^{\psi'}\varphi$
RC2	:	*If* $\varphi \rightarrow \varphi'$ *then* $B^{\psi}\varphi \rightarrow B^{\psi}\varphi'$

Moreover, $\mathsf{K}_{KB^q} := \mathsf{K}_K + \mathsf{K}_{B^q}$ is a sound and strongly *complete axiomatization for $\mathcal{L}_{KB^q}$ with respect to all (qualitative) epistemic-plausibility spaces.*

Note that the proof of Theorem 2 in [13] only considers accessibility relations which are equivalence relations (it corresponds to Theorem 11 of [13]). The proof can easily be adapted to our more general setting and the proof shows in fact that the logics are *strongly* complete. Compactness of K_{KB^q} follows from this strong completeness result.

Corollary 1. *K_{KB^q} is compact,* i.e.*, for all $\Gamma \subseteq \mathcal{L}_{KB^q}$ and all $\varphi \in \mathcal{L}_{KB^q}$, if $\Gamma \models \varphi$, then for some* finite *$\Gamma_0 \subseteq \Gamma$, we have that $\Gamma_0 \models \varphi$.*

The axiom $(B^q(\varphi \rightarrow \varphi') \wedge B^q\varphi) \rightarrow B^q\varphi'$ and the inference rule *from φ infer $B^q\varphi$* are both derivable in K_{B^q}. Therefore, K_{B^q} is also a *normal* modal logic. Hence, the language $\mathcal{L}_{KB^q}$ can also be given a Kripke semantics: the models would be of the form $\mathcal{M} = (W, R_K, \{R_\psi \mid \psi \in \mathcal{L}_{KB^q}\}, V)$, where R_ψ is an accessibility relation for each $\psi \in \mathcal{L}_{KB^q}$. This entails that the first part of Theorem 1 (and its proof) can easily be adapted to the setting of conditional beliefs, where the second modality [2] is replaced by a family of modalities $\{B^q \mid \psi \in \mathcal{L}_{KB^q}\}$. We obtain the following theorem:

Theorem 3. *Let $\Gamma \subseteq \mathcal{L}_{KB^q}$ and $\mathsf{x} \in \mathcal{L}_{KB^q}$. Assume that for all $\psi \in \mathcal{L}_{KB^q}$ the modality $B^{\psi}p$ is explicitly defined in $\mathsf{S4}_K + \mathsf{K}_{B^q} + \Gamma$ in terms of K by a formula $\mathsf{def}_{B^{\psi}p} \in \mathcal{L}_K$ positive in p. Then, the following are equivalent:*

- *x is characterized by Γ w.r.t. $(\mathsf{S4}_K, \mathsf{K}_{B^q})$;*
- *$\mathsf{S4}_K + \mathsf{K}_{B^q} + \Gamma = \mathsf{S4}_K + \mathsf{x} + \{B^q p \leftrightarrow \mathsf{def}_{B^q p}\}$.*

4 Interaction Axioms

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 belief, and then we will consider interaction axioms that deal with *conditional* belief. Note that a classification of certain interaction principles has been proposed by van der Hoek [33].[2]

4.1 Interaction Axioms between Knowledge and Belief

The following interaction axioms were suggested by Hintikka [18]:

$$Kp \to Bp \tag{I$_1$}$$

$$Bp \to KBp \tag{I$_2$}$$

Axiom I_1 is a cornerstone of epistemic logic. Just as axiom T, it follows from the classical analysis of knowledge of Plato presented in the Theaetetus. Axiom I_2 highlights the fact that the agent has "privileged access" to his doxastic state.

$$Bp \to BKp \tag{I$_3$}$$

Axiom I_3 above was suggested by Lenzen [24]. It characterizes a notion of belief corresponding to some sort of conviction or certainty. This kind of belief is therefore different from the notion of *weak* belief which can be represented by a probability superior to 0.5, like my belief that "it will rain tomorrow".

Another interaction axiom also introduced by [24] defines belief in terms of knowledge:

$$Bp \leftrightarrow \langle K\rangle Kp \tag{I$_4$}$$

Although this definition might seem a bit mysterious at first sight, it actually makes perfect sense, as explained in [24]. Indeed, the left to right direction $Bp \to \langle K\rangle Kp$ can be rewritten $K\neg Kp \to \neg Bp$, that is, $\neg(K\neg Kp \wedge Bp)$. 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 $\langle K\rangle Kp \to Bp$ can be rewritten $\langle B\rangle\neg p \to K\neg Kp$. This second implication states that, if the agent considers it possible that p might be false, then she knows that she does not know p.

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

$$Kp \leftrightarrow (p \wedge Bp) \tag{I$_5$}$$

This list of interaction axioms is incomplete. See [3] for more information about interaction axioms and axioms of epistemic logic.

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

Proposition 4. *The sets* $\{I_1\}, \{I_2\}, \{I_3\}, \{I_4\}, \{I_5\}$ *and* $\{I_1, I_2, I_3\}$ *are sets of interaction axioms with respect to* $(\mathsf{S4}_K, \mathsf{KD45}_B)$.[3]

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 I_1 as the only interaction axiom, then we end up with counterintuitive properties. First, as noted by Voorbraak [34], we can derive the theorem $BKp \rightarrow Kp$.[4] This theorem entails that everything one believes to know is in fact true. As it turns out, these axioms are adopted in the first logical framework combining modalities of knowledge and belief [19]. Moreover, if we add the axiom I_3, we can also prove that $Bp \rightarrow Kp$. 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 [33]. He showed, thanks to correspondence theory, that any multimodal logic with both knowledge and belief modalities that includes the set of axioms $\{\mathsf{D}, \mathsf{5}, I_1, I_3\}$ entails the theorem $Bp \rightarrow Kp$. He also showed, however, that for each proper subset of $\{\mathsf{D}, \mathsf{5}, I_1, I_3\}$, 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 $\{\mathsf{D}, \mathsf{5}, I_1, I_3\}$. Axioms D and I_3 are hardly controversial given our understanding of the notion of strong belief. In this case we have to drop either I_1 or 5. Voorbraak proposes to drop axiom I_1. His notion of knowledge, which he calls *objective knowledge*, is therefore unusual in so far as it implies that you can know something even if you don't believe it. 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 I_1 [11], since his notion of *being informed* shares similar features with Voorbraak's notion of *objective knowledge*. Halpern also proposes in [15] to drop I_1 as a general schema, keeping only those instances of I_1 where p is propositional. 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 Footnote 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

[3] In the set $\{I_1, I_2, I_3\}$, we implicitly assume that the sets of propositional letters appearing in I_1, I_2 and I_3 respectively are disjoint so that we can uniformly and independently replace each of them by arbitrary formulas to check whether Expression (1) of Definition 1 holds.

[4] Here is the proof:

1	$Kp \rightarrow Bp$	Axiom I_1
2	$K\neg Kp \rightarrow B\neg Kp$	$I_1 : \neg Kp/p$
3	$Bp \rightarrow \neg B\neg p$	Axiom D
4	$B\neg p \rightarrow \neg Bp$	3, contraposition
5	$B\neg Kp \rightarrow \neg BKp$	$4 : Kp/p$
6	$\neg Kp \rightarrow K\neg Kp$	Axiom 5
7	$\neg Kp \rightarrow B\neg Kp$	6,2, Modus Ponens
8	$\neg Kp \rightarrow \neg BKp$	7,5, Modus Ponens
9	$BKp \rightarrow Kp$	8, contraposition.

in [25].

4.2 Interaction Axioms between Knowledge and Conditional Belief

The following axioms I_1^q, I_2^q and I_3^q are natural conditional versions of the axioms I_1, I_2, I_3: if q is replaced by $\top$, then these three axioms correspond to the axioms I_1, I_2, I_3. Axioms I_1^q and I_2^q are first introduced by Moses and Shoham [28] and are also adopted by Friedman and Halpern [12]. Axiom I_3^q is actually introduced by Lamarre and Shoham [22] in the form $B^q p \to B_w^q K(q \to p)$.

$$Kp \to B^q p \qquad (I_1^q)$$
$$B^q p \to KB^q p \qquad (I_2^q)$$
$$B^q p \to B^q K(q \to p) \qquad (I_3^q)$$

Axiom I_1^q states that, if the agent knows that p, then she also believes that p, and so on under any assumption q. Note that I_1^q entails the weaker principle $Kp \to (q \to B^q p)$, which is connected to the Lehrer and Paxton's definition of knowledge as undefeated true belief [23]. Indeed, this derived principle states that if the agent knows that p (formally Kp), then her belief in p cannot be defeated by any *true* information q (formally $q \to B^q p$). Note that this very principle entails an even weaker variant of I_1^q introduced by Moses and Shoham [28], namely $Kp \to (B^q p \vee K\neg q)$, i.e. $Kp \to (\langle K\rangle q \to B^q p)$. Axiom I_2^q is a straightforward generalization of I_2. As for I_3^q, it states that, if the agent believes p under the assumption that q, then, given this very assumption q, she also believes that she knows p conditional on q.

The axioms I_4^q and I_5^q below are also introduced by Lamarre and Shoham [22]:

$$\neg B^q p \to K(\langle K\rangle q \to \neg B^q p) \qquad (I_4^q)$$
$$\langle K\rangle q \to \neg B^q \bot \qquad (I_5^q)$$

Axiom I_4^q is a conditional version of axiom $\neg Bp \to K\neg Bp$. It is introduced by Lamarre and Shoham [22] in the form $\neg B^q p \to K(K\neg q \vee \neg B^q p)$. Another possible conditional version of axiom $\neg Bp \to K\neg Bp$ could have been $\neg B^q p \to K\neg B^q p$, and this axiom is indeed adopted in [28]. However, this simpler axiom ignores the possibility of assumptions q which are known to be false ($K\neg q$): in that case, these assumptions q should not be taken into account in the reasoning.

Axiom I_5^q states that, if q is compatible with everything the agents knows, then her beliefs given this assumption cannot be inconsistent. In particular, if q holds then the agent's doxastic state given this assumption cannot be inconsistent: $q \to \neg B^q \bot$ (because $q \to \langle K\rangle q$ is valid according to axiom T). Axiom I_5^q is introduced by Lamarre and Shoham [22] in the equivalent form $\langle K\rangle q \to (B^q p \to \neg B^q \neg p)$. Together with I_1^q 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 by Baltag and Smets [6]:

$$Kp := B^{\neg p}\bot \qquad \text{(Def K')}$$

Proposition 5. *The sets* $\{I_1^q\}, \{I_2^q\}, \{I_3^q\}, \{I_4^q\}, \{I_5^q\}$ *and* $\{I_1^q, I_2^q, I_3^q, I_4^q, I_5^q\}$ *are sets of interaction axioms with respect to* $(\mathsf{S4}_K, \mathsf{K}_{B^q})$.[5]

5 Definability of Knowledge, Belief and Conditional Belief

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 concepts will not interact. We need to add interaction axioms. Halpern et al. [16] only consider the interaction axioms I_1 and I_2 suggested by Hintikka [18]. We will also add the interaction axiom I_3, suggested by Lenzen [24].

5.1 Defining Belief in Terms of Knowledge

Lenzen is the first to note that the belief modality can be defined in terms of knowledge if we adopt $\{I_1, I_2, I_3\}$ as interaction axioms:

Theorem 6. *[25] The belief modality* B *is explicitly defined in the logic* $\mathsf{L} := \mathsf{S4}_K + \mathsf{KD45}_B + \{I_1, I_2, I_3\}$ *by the formula* $\mathsf{def}_{Bp} := \langle K\rangle Kp \in \mathcal{L}_K$*:*

$$Bp \leftrightarrow \langle K\rangle Kp \in \mathsf{L} \qquad \text{(Def B)}$$

Consequently, the belief modality B *is also defined by (Def* B*) in any logic containing* L*.*

As a consequence of this theorem, the belief modality is also explicitly defined by $Bp := \langle K\rangle Kp$ in the logics $\mathsf{S4.x}_K + \mathsf{KD45}_B + \{I_1, I_2, I_3\}$, where x ranges over $\{.2, .3, .4\}$. This result is in contrast with Theorem 4.8 of [16], from which it follows that the belief modality can*not* be explicitly defined in the logic $\mathsf{S4.x}_K + \mathsf{KD45}_B + \{I_1, I_2\}$, and for any $\mathsf{x} \in \{.2, .3, .4\}$. We see here that the increase in expressivity due to the addition of the interaction axiom I_3 plays an important role in bridging the gap between belief and knowledge. Note that the definition (Def B) of belief in terms of knowledge corresponds to the interaction axiom I_4 of Section 4.1.

[5] In the set $\{I_1^q, I_2^q, I_3^q, I_4^q, I_5^q\}$, we implicitly assume that the sets of propositional letters appearing in $I_1^q, I_2^q, I_3^q, I_4^q$ and I_5^q respectively are disjoint so that we can uniformly and independently substitute each of them by arbitrary formulas to check whether Expression (1) of Definition 1 holds.

5.2 Defining Knowledge in Terms of Belief

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 S4.4, but not if the logic of knowledge is S4 and S4.x, where x ranges over $\{.2, .3\}$.

Theorem 7. *[3]*

1. *The knowledge modality K is explicitly defined in the logic $\mathsf{L.4} := \mathsf{S4.4}_K + \mathsf{KD45}_B + \{I_1, I_2, I_3\}$ by the formula $\mathsf{def}_{Kp} := p \wedge Bp \in \mathcal{L}_B$:*

$$Kp \leftrightarrow p \wedge Bp \in \mathsf{L.4} \qquad (\mathsf{Def\ K})$$

2. *The knowledge modality K cannot be explicitly defined in the logics $\mathsf{S4.x}_K + \mathsf{KD45}_B + \{I_1, I_2, I_3\}$ for any $x \in \{.2, .3\}$.*

This result can be contrasted with Theorem 4.1 of [16], from which it follows that the knowledge modality can*not* be explicitly defined in the logic $\mathsf{S4.4}_K + \mathsf{KD45}_B + \{\mathsf{KB1}, \mathsf{KB2}\}$. Again, the increase in expressivity due to the addition of the interaction axiom I_3 plays an important role in bridging the gap between belief and knowledge.

5.3 Defining Conditional Belief in Terms of Knowledge

The conditions under which conditional belief can be defined in terms of knowledge (and reciprocally) have been less explored in the epistemic logic literature.

Theorem 8. *The conditional belief modality $B^q p$ is explicitly defined in the logic $\mathsf{L}^q := \mathsf{S4}_K + \mathsf{K}_{B^q} + \{I_1^q, I_2^q, I_3^q, I_4^q, I_5^q\}$ by the formula $\mathsf{def}_{B^q p} := \langle K \rangle q \rightarrow \langle K \rangle (q \wedge K (q \rightarrow p)) \in \mathcal{L}_K$:*

$$B^q p \leftrightarrow (\langle K \rangle q \rightarrow \langle K \rangle (q \wedge K (q \rightarrow p))) \in \mathsf{L}^q \qquad (\mathsf{Def}\ B^q p)$$

Moreover, the following also holds:

$$\mathsf{S4}_K + \mathsf{K}_{B^q} + \{I_1^q, I_2^q, I_3^q, I_4^q, I_5^q\} = \mathsf{S4.3}_K + \{B^q p \leftrightarrow \mathsf{def}_{B^q p}\} \qquad (4)$$

The definition $(\mathsf{Def}\ B^q p)$ of *conditional* belief in terms of knowledge can be viewed as a generalization of the definition of belief in terms of knowledge of Section 4.1: $Bp \leftrightarrow \langle K \rangle Kp$, that is I_4. The intuitive interpretations that we gave for I_4 can also be generalized to the conditional case. The left to right direction of $(\mathsf{Def}\ B^q p)$ can be rewritten as follows: $\neg (B^q p \wedge \langle K \rangle q \wedge K (q \rightarrow \langle K \rangle (q \wedge \neg p)))$. It states that it is impossible that the agent believes p given q while at the same time knowing that if q holds then he may consider possible that p does not hold (assuming that he already considers q possible). This seems a reasonable claim. Reciprocally, the right to left direction of $(\mathsf{Def}\ B^q p)$ can be rewritten as follows: $\neg B^q \neg p \rightarrow \langle K \rangle q \wedge K (q \rightarrow \langle K \rangle (q \wedge p))$. It states that if the agent considers that p might be true given q, then the agent considers q (epistemically) possible and knows that if q holds then it is possible that p might also hold. This seems also reasonable.

5.4 Defining Knowledge in Terms of Conditional Belief

From Equation (Def $B^q p$), we easily obtain that:

Proposition 9. *The knowledge modality Kp is explicitly defined in the logic L^q by the formula $\mathsf{def}_{Kp} := B^{\neg p}\bot \in \mathcal{L}_{B^q}$:*

$$Kp \leftrightarrow B^{\neg p}\bot \in \mathsf{L}^q \qquad \text{(Def K')}$$

Proof. To prove that $Kp \rightarrow B^{\neg p}\bot \in \mathsf{L}^q$, it suffices to observe that $Kp \rightarrow B^{\neg p}p \in \mathsf{L}^q$ by the axiom I_1^q and that $B^{\neg p}\neg p$ is also an axiom of L^q. Therefore, $Kp \rightarrow B^{\neg p}(p \wedge \neg p) \in \mathsf{L}^q$ by distributivity of B^q. So, $Kp \rightarrow B^{\neg p}\bot \in \mathsf{L}^q$. The proof that $B^{\neg p}\bot \rightarrow Kp \in \mathsf{L}^q$ follows from the contraposition of axiom I_5^q. □

The definition (Def K') can be rewritten equivalently as follows: $\langle K\rangle p \leftrightarrow \neg B^p \neg\top$. So, in a sense, the intuition underlying this definition is that the epistemic possibility that p holds can be identified with the fact that p is compatible with the beliefs of the agent.

6 Axioms .2, .3 and .4 as Interaction Axioms

The results of this final section are obtained by applying Theorem 1 either to the results of Lenzen [25] or to the results obtained in the previous section. The first item of Theorem 10 below somehow makes more explicit the fact that .2 is really characterized by the interaction axioms $\{I_1, I_2, I_3\}$. Lenzen [25] showed that $\mathsf{S4}_K + \mathsf{KD45}_B + \{I_1, I_2, I_3\} = \mathsf{S4.2}_K + \{Bp \leftrightarrow \langle K\rangle Kp\}$. However, this expression alone cannot account for the fact that .2 is really characterized by the interaction axioms $\{I_1, I_2, I_3\}$. As for the second item of Theorem 10, it shows that assuming that knowledge obeys .4 has the same consequences for the logic of knowledge as assuming that knowledge is S4, belief is KD45, and knowledge is true belief.

Theorem 10. *1. The axiom .2 is characterized w.r.t. the pair $(\mathsf{S4}_K, \mathsf{KD45}_B)$ by the set of interaction axioms $\{I_1, I_2, I_3\}$.*

2. The axiom .4 is conservatively characterized w.r.t. the pair $(\mathsf{S4}_K, \mathsf{KD45}_B)$ by the interaction axiom I_5.

Proof. It follows from a direct application of Theorem 1 to the results of [25], namely the fact that $\mathsf{S4}_K + \mathsf{KD45}_B + \{I_1, I_2, I_3\} = \mathsf{S4.2}_K + \{Bp \leftrightarrow \langle K\rangle Kp\}$ (for the first result) and $\mathsf{S4}_K + \mathsf{KD45}_B + \{I_5\} = \mathsf{KD45}_B + \{I_5\} = \mathsf{S4.4}_K + \{Bp \leftrightarrow \langle K\rangle Kp\}$ (for the second result). To apply Theorem 1, we remind that p is positive in $\langle K\rangle Kp$. □

Stalnaker [32] argued that the intuition underlying the logic S4.3 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 learning any true fact. Our interaction axiom I_1^q formalizes this intuition.

Theorem 11. *The axiom .3 is characterized w.r.t. the pair* $(\mathsf{S4}_K, \mathsf{K}_{B^q})$ *by the set of interaction axioms* $\{I_1^q, I_2^q, I_3^q, I_4^q, I_5^q\}$.

Proof. Theorem 1 can easily be adapted to a setting where the second modality [2] is replaced by a family of modalities of the form B^q. In that case, the proof of the (adapted) Theorem 1 follows the same reasoning (because of the comments that follow Theorem 2). Then, the proof of our theorem follows from a direct application of (an adaptation to the case of *conditional* belief of) Theorem 1 to the results of Theorems 8, namely the fact that $\mathsf{S4}_K + \mathsf{K}_{B^q} + \{I_1^q, I_2^q, I_3^q, I_4^q, I_5^q\} = \mathsf{K}_{B^q} + \{Kp \leftrightarrow \mathsf{def}_{Kp}\} = \mathsf{S4.3}_K + \{B^q p \leftrightarrow \mathsf{def}_{B^q p}\}$. □

7 Conclusion

The second item of Theorem 10 tells us that in the epistemic-doxastic context ($\mathsf{S4}_K$, $\mathsf{KD45}_B$), the axiom .4 characterizes the fact that knowledge is actually defined as true belief (a rather strong assumption for knowledge). Although it was acknowledged by all epistemic logicians that axiom .4 characterized knowledge as true belief, this could never be justified and explained rigorously. We claim that our meta-theory of modal logic fills this conceptual gap. Likewise, still in the context of ($\mathsf{S4}_K$, $\mathsf{KD45}_B$), the first item of Theorem 10 tells us that axiom .2 characterizes the fact that the agent knows his beliefs and disbeliefs and that his beliefs are in fact certainties, convictions, and not simply 'weak' beliefs. Theorem 11 provides similar results with conditional beliefs. Stalnaker [32] claimed that $\mathsf{S4.3}$ is a logic where knowledge is defined as true belief which cannot be defeated by any true fact. A similar definition of knowledge is propounded in epistemology by Lehrer and Paxson [23]. The claim of Stalnaker was based on informal semantic arguments. This claim is supported rigorously in our approach by the fact that the axiom .3 is characterized by the set of interaction axioms $\{I_1^q, I_2^q, I_3^q, I_4^q, I_5^q\}$ in the epistemic-doxastic context $(\mathsf{S4}_K, \mathsf{K}_{B^q})$, where we recall that I_1^q entails $Kp \rightarrow (q \rightarrow B^q p)$: if the agent knows p (formally Kp) then this belief in p cannot be defeated by any true fact q (formally $q \rightarrow B^q p$). Hence, our theory of interaction axioms provides a precise and rigorous justification of Stalnaker's 'informal' claim.

Overall, our theory of characterizing axioms via interaction axioms enables us to carry out a rigorous and fine-grained analysis of the intuitive assumptions underlying the logics of knowledge between $\mathsf{S4}$ and $\mathsf{S5}$. It can give precise reasons for choosing a specific epistemic logic for representing and reasoning about a given situation. In a sense, our theory provides meaningful logical foundations for pursuing rigorous epistemological investigations.

Acknowledgements. I thank a reviewer for her/his detailed comments.

References

[1] Ernest Adams. *The Logic of Conditionals*, volume 86 of *Synthese Library*. Springer, 1975.

[2] Guillaume Aucher. Axioms .2 and .4 as interaction axioms. In Chitta Baral, Giuseppe De Giacomo, and Thomas Eiter, editors, *Principles of Knowledge Representation and Reasoning: Proceedings of the Fourteenth International Conference, KR 2014, Vienna, Austria, July 20-24, 2014*. AAAI Press, 2014.

[3] Guillaume Aucher. *Interdisciplinary Works in Logic, Epistemology, Psychology and Linguistics*, volume 3 of *Logic, Argumentation and Reasoning*, chapter Principles of Knowledge, Belief and Conditional Belief. Springer, 2014.

[4] Alexandru Baltag and Sonja Smets. Conditional doxastic models: A qualitative approach to dynamic belief revision. *Electronic Notes in Theoretical Computer Science*, 165:5–21, 2006.

[5] Alexandru Baltag and Sonja Smets. *Texts in Logic and Games*, volume 4, chapter The Logic of Conditional Doxastic Actions, pages 9–31. Amsterdam University Press, 2008.

[6] Alexandru Baltag and Sonja Smets. *Texts in Logic and Games*, volume 3, chapter A Qualitative Theory of Dynamic Interactive Belief Revision, pages 9–58. Amsterdam University Press, 2008.

[7] Patrick Blackburn, Maarten de Rijke, and Yde Venema. *Modal Logic*, volume 53 of *Cambridge Tracts in Computer Science*. Cambridge University Press, 2001.

[8] Olivier Board. Dynamic interactive epistemology. *Games and Economic Behavior*, 49:49–80, 2004.

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

[10] Ronald Fagin, Joseph Halpern, Yoram Moses, and Moshe Vardi. *Reasoning about knowledge*. MIT Press, 1995.

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

[12] Nir Friedman and Joseph Y. Halpern. Modeling belief in dynamic systems, part i: Foundations. *Artificial Intelligence*, 95(2):257 – 316, 1997.

[13] Nir Friedman and Joseph Y. Halpern. Plausibility measures and default reasoning. *Journal of the ACM*, 48(4):648–685, 2001.

[14] DM Gabbay, A Kurucz, F Wolter, and M Zakharyaschev. *Multi-dimensional modal logic: Theory and application*, volume 148 of *Studies in logic and the foundations of mathematics*. Elsevier, 1998.

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

[16] Joseph Y. Halpern, Dov Samet, and Ella Segev. Defining knowledge in terms of belief: the modal logic perspective. *The Review of Symbolic Logic*, 2:469–487, 2009.

[17] Joseph Y. Halpern, Dov Samet, and Ella Segev. On definability in multimodal logic. *The Review of Symbolic Logic*, 2:451–468, 2009.

[18] Jaakko Hintikka. *Knowledge and Belief, An Introduction to the Logic of the Two Notions*. Cornell University Press, Ithaca and London, 1962.

[19] Sarit Kraus and Daniel Lehmann. Knowledge, belief and time. In Laurent Kott, editor, *Automata, Languages and Programming*, volume 226 of *Lecture Notes in Computer Science*, pages 186–195. Springer Berlin / Heidelberg, 1986.

[20] Sarit Kraus, Daniel J. Lehmann, and Menachem Magidor. Nonmonotonic reasoning, preferential models and cumulative logics. *Artificial Intelligence*, 44(1-2):167–207, 1990.

[21] F. von Kutschera. *Einführung in die intensional Semantik*. W. de Gruyter, Berlin, 1976.

[22] Philippe Lamarre and Yoav Shoham. Knowledge, certainty, belief, and conditionalisation (abbreviated version). In *KR*, pages 415–424, 1994.

[23] Keith Lehrer and Thomas Paxson. Knowledge: Undefeated justified true belief. *The Journal of Philosophy*, 66:225–237, 1969.

[24] Wolfgang Lenzen. *Recent Work in Epistemic Logic*. Acta Philosophica Fennica 30. North Holland Publishing Company, 1978.

[25] Wolfgang Lenzen. Epistemologische betractungen zu [S4;S5]. *Erkenntnis*, 14:33–56, 1979.

[26] Maarten Marx and Yde Venema. *Multi-dimensional modal logic*. Springer, 1997.

[27] John-Jules Ch. Meyer and Wiebe van der Hoek. *Epistemic Logic for AI and Computer Science*. Cambridge University Press, Cambridge, 1995.

[28] Yoram Moses and Yoav Shoham. Belief as defeasible knowledge. *Artificial intelligence*, 64(2):299–321, 1993.

[29] Yoav Shoham and Kevin Leyton-Brown. *Multiagent Systems: Algorithmic, Game-Theoretic, and Logical Foundations*. Cambridge University Press, 2009.

[30] Wolfgang Spohn. A general non-probabilistic theory of inductive reasoning. In Ross D. Shachter, Tod S. Levitt, Laveen N. Kanal, and John F. Lemmer, editors, *UAI*, pages 149–158. North-Holland, 1988.

[31] Wolfgang Spohn. Ordinal conditional functions: A dynamic theory of epistemic states. In W. L. Harper and B. Skyrms, editors, *Causation in Decision, Belief Change, and Statistics*, volume 2, pages 105–134. reidel, Dordrecht, 1988.

[32] Robert Stalnaker. On logics of knowledge and belief. *Philosophical studies*, 128:169–199, 2006.

[33] Wiebe van der Hoek. Systems for knowledge and belief. *Journal of Logic and Computation*, 3(2):173–195, 1993.

[34] Frans Voorbraak. *As Far as I know. Epistemic Logic and Uncertainty*. PhD thesis, Utrecht University, 1993.

A Proofs of Propositions 4 and 5

Proposition 4. *The sets* $\{I_1\},\{I_2\},\{I_3\},\{I_4\},\{I_5\}$ *and* $\{I_1,I_2,I_3\}$ *are sets of interaction axioms with respect to* $(\mathsf{S4}_K,\mathsf{KD45}_B)$.[6]

Proof. First, we prove that $\{I_1\},\{I_2\},\{I_3\},\{I_4\},\{I_5\}$ are sets of interaction axioms. It suffices to prove two things, for each $i\in\{1,\ldots,5\}$ (because of the soundness and completeness of $\mathsf{S4}_K+\mathsf{KD45}_B$ for the language $\mathcal{L}_{KB}$ with respect to the class of epistemic-doxastic models whose accessibility relation R_K is transitive and reflexive and whose accessibility relation R_B is serial, transitive and Euclidean):

1. there are two pointed epistemic-doxastic models $(\mathcal{M}^1,w^1)$ and $(\mathcal{M}^2,w^2)$ such that they are bisimilar for R_K but $\mathcal{M}^1,w^1\models I_i$ and $\mathcal{M}^2,w^2\not\models I_i$,
2. there are two other pointed epistemic-doxastic models $(\mathcal{M}^1,w^1)$ and $(\mathcal{M}^2,w^2)$ such that they are bisimilar for R_B but $\mathcal{M}^1,w^1\models I_i$ and $\mathcal{M}^2,w^2\not\models I_i$.

We do so for each axiom I_1,I_2,I_3,I_4 and I_5. For each of them and for each of the two subcases, we consider two pointed epistemic-doxastic models $(\mathcal{M}^1,w^1)=(\{w^1,v^1\},R^1_K,R^1_B,V^1,w^1)$ and $(\mathcal{M}^2,w^2)=(\{w^2,v^2\},R^2_K,R^2_B,V^2,w^2)$ such that $V^1(p)=\{w^1\}$ and $V^2(p)=\{w^2\}$ for all $p\in\mathbb{P}$. For each case, only the definitions of the accessibility relations R^1_K,R^2_K,R^1_B and R^2_B change. For better readability, we omit the superscripts 1 and 2 for the worlds w and v, as they should be clear from context.

[6] In the set $\{I_1,I_2,I_3\}$, we implicitly assume that the sets of propositional letters appearing in I_1,I_2 and I_3 respectively are disjoint so that we can uniformly and independently replace each of them by arbitrary formulas to check whether Expression (1) of Definition 1 holds.

- $I_1 : Kp \to Bp$.
  1. $R^1_K = R^2_K = R^1_B = \{(w,w),(v,v)\}$ and $R^2_B = \{(w,w),(w,v),(v,w),(v,v)\}$;
  2. $R^1_K = R^1_B = R^2_B = \{(w,w),(v,v),(w,v),(v,w)\}$ and $R^2_K = \{(w,w),(v,v)\}$.
- $I_2 : Bp \to BKp$.
  1. $R^1_K = \{(w,w),(v,v)\}, R^1_B = R^2_B = \{(w,v),(v,v)\}$ and $R^2_K = \{(w,w),(v,v),(v,w)\}$;
  2. $R^1_K = R^2_K = \{(w,w),(v,v),(w,v)\}, R^1_B = \{(w,v),(v,v)\}$ and $R^2_B = \{(w,w),(v,v)\}$.
- $I_3 : Bp \to KBp$.
  1. $R^1_K = R^1_B = R^2_B = \{(w,w),(v,v)\}$ and $R^2_K = \{(w,w),(v,v),(w,v)\}$;
  2. $R^1_K = R^2_K = \{(w,w),(v,v),(w,v)\}$, $R^1_B = \{(w,v),(v,v)\}$ and $R^2_B = \{(w,w),(v,v)\}$.
- $I_4 : Bp \to \langle K\rangle Kp$.
  1. $R^1_K = \{(w,w),(v,v),(w,v)\}$, $R^1_B = R^2_B = \{(w,v),(v,v)\}$ and $R^2_K = \{(w,w),(v,v)\}$;
  2. $R^1_K = R^2_K = \{(w,w),(v,v)\}$, $R^1_B = \{(w,v),(v,v)\}$ and $R^2_B = \{(w,w),(w,v),(v,w),(v,v)\}$.
- $I_5 : Kp \leftrightarrow (p \wedge Bp)$.
  1. $R^1_K = R^1_B = R^2_B\{(w,w),(v,v)\}$ and $R^2_K = \{(w,w),(v,v),(w,v),(v,w)\}$;
  2. $R^1_K = R^1_B = R^2_K = \{(w,w),(v,v)\}$ and $R^2_B = \{(w,w),(v,v),(w,v),(v,w)\}$.

The proof for the set of interaction axioms $\{I_1, I_2, I_3\}$ follows the same method as above. For that case, we consider the same epistemic-doxastic models as in the case I_3:

1. $R^1_K = R^1_B = R^2_B = \{(w,w),(v,v)\}$ and $R^2_K = \{(w,w),(v,v),(w,v)\}$;
2. $R^1_K = R^2_K = \{(w,w),(v,v),(w,v)\}$, $R^1_B = \{(w,v),(v,v)\}$ and $R^2_B = \{(w,w),(v,v)\}$. □

Proposition 5. *The sets* $\{I^q_1\},\{I^q_2\},\{I^q_3\},\{I^q_4\},\{I^q_5\}$ *and* $\{I^q_1, I^q_2, I^q_3, I^q_4, I^q_5\}$ *are sets of interaction axioms with respect to* $(\mathsf{S4}_K, \mathsf{K}_{B^q})$.[7]

Proof. The sets of interaction axioms $\{I^q_1\},\{I^q_2\},\{I^q_3\}$ and $\{I^q_5\}$ are dealt with similarly to the way in which we dealt with the set of interaction axioms $\{I_1\},\{I_2\},\{I_3\}$ and $\{I_5\}$. We only need to replace the accessibility relations R_B by plausibility spaces $(W_w, \mathcal{A}_w, Pl_w)$ for each $w \in W$. We do so as follows. For any two worlds $w, v \in W_i$, if there is an edge from w to v but not from v to w,

[7] In the set $\{I^q_1, I^q_2, I^q_3, I^q_4, I^q_5\}$, we implicitly assume that the sets of propositional letters appearing in $I^q_1, I^q_2, I^q_3, I^q_4$ and I^q_5 respectively are disjoint so that we can uniformly and independently substitute each of them by arbitrary formulas to check whether Expression (1) of Definition 1 holds.

then we set $W_w = \{w, v\}, Pl_w(w) = \bot$ and $Pl_w(v) = \top$. If there is an edge from w to v and an edge from v to w as well, then we set $W_w = \{w, v\}, Pl_w(w) = \top$ and $Pl_w(v) = \top$. In the first three cases, the counter-example formula is given by considering $q = \top$ and it corresponds to the same formula as for the set of interaction axioms $\{I_1\}, \{I_2\}$ and $\{I_3\}$. For the fourth case corresponding to axiom I_5^q, the counter-example is obtained by considering $q = \neg p$.

The set $\{I_4^q\}$ is dealt with using the following epistemic-plausibility models:

1. $W_w^1 = W_w^2 = \{w\}$, $Pl_w^1(w) = Pl_w^2(w) = \top$; $W_v^1 = W_v^2 = \{v\}$, $Pl_v^1(v) = Pl_v^2(v) = \top$; and $R_K^1 = \{(w, w), (v, v)\}$ and $R_K^2 = \{(w, w), (v, v), (w, v), (v, w)\}$;
2. $W_w^1 = \{w, v\}$, $Pl_w(w) = \bot$, $Pl_w(v) = \top$ and $W_v^1 = \{v\}$, $Pl_v^1(v) = \top$; $W_w^2 = \{w\}$, $W_v^2 = \{v\}$ and $Pl_w^2(w) = \{w\}$ and $Pl_v^2(v) = \{v\}$; and $R_K^1 = R_K^2 = \{(w, w), (v, v), (w, v), (v, w)\}$.

The proof for the set $\{I_1^q, I_2^q, I_3^q, I_4^q, I_5^q\}$ is dealt with by considering the same epistemic-plausibility models as for the set $\{I_4^q\}$. □

B Proof of Theorem 1

Note that the proof of this theorem appears in [2] without the following Lemma.

Lemma 12. *Let L be a normal modal logic for $\mathcal{L}_{\mathbb{A}}$ and let Γ be a set of formulas of $\mathcal{L}_{\mathbb{A}}$. Then,*

$$\Gamma \subseteq \mathsf{L} \text{ iff } \{(\mathcal{M}, w) \in \mathcal{K} \mid \mathcal{M}, w \models \mathsf{L}\} \subseteq \{(\mathcal{M}, w) \in \mathcal{K} \mid \mathcal{M}, w \models \Gamma\} \qquad (5)$$

Proof. The left to right direction is straightforward, so we only prove the right to left direction. By contraposition, assume that $\Gamma \nsubseteq \mathsf{L}$. Then, there is $\varphi \in \Gamma$ such that $\varphi \notin \mathsf{L}$. We are going to show that there is $(\mathcal{M}, w) \in \mathcal{K}$ such that $\mathcal{M}, w \not\models \varphi$ and $\mathcal{M}, w \models \mathsf{L}$, so that the right-hand side of Expression (5) does not hold. Assume towards a contradiction that there is no $(\mathcal{M}, w) \in \mathcal{K}$ such that $\mathcal{M}, w \models \mathsf{L} \cup \{\neg\varphi\}$. Then, by the compactness of K, there are $\varphi_1, \ldots, \varphi_n \in \mathsf{L}$ such that for all $(\mathcal{M}, w) \in \mathcal{K}$, $\mathcal{M}, w \models \varphi_1 \wedge \ldots \wedge \varphi_n \rightarrow \varphi$. Then, by completeness of the smallest modal logic K, we have that $\varphi_1 \wedge \ldots \wedge \varphi_n \rightarrow \varphi \in \mathsf{K}$. So, because $\mathsf{K} \subseteq \mathsf{L}$ and $\varphi_1 \wedge \ldots \wedge \varphi_n \in \mathsf{L}$, we have by application of Modus Ponens that $\varphi \in \mathsf{L}$. This contradicts the fact that $\varphi \notin \mathsf{L}$. Therefore, there is $(\mathcal{M}, w) \in \mathcal{K}$ such that $\mathcal{M}, w \models \mathsf{L}$ and $\mathcal{M}, w \not\models \varphi$. This proves our result. □

Theorem 1. *Assume that $\langle 2 \rangle$ is explicitly defined in $\mathsf{L}_1 + \mathsf{L}_2 + \Gamma$ in terms of $\langle 1 \rangle$ by a formula $\mathsf{def}_2(p) \in \mathcal{L}_1$ positive in p. Then, the following are equivalent:*

- *x is characterized by Γ w.r.t. $(\mathsf{L}_1, \mathsf{L}_2)$;*
- *$\mathsf{L}_1 + \mathsf{L}_2 + \Gamma = \mathsf{L}_1 + \mathsf{x} + \{\langle 2 \rangle p \leftrightarrow \mathsf{def}_2(p)\}$.*

Moreover, assume that $\langle 1 \rangle$ is also explicitly defined in $\mathsf{L}_1 + \mathsf{L}_2 + \Gamma$ in terms of $\langle 2 \rangle$ by a formula $\mathsf{def}_1(p) \in \mathcal{L}_2$ positive in p. Then, the following are equivalent:

- *x is conservatively characterized by Γ w.r.t. $(\mathsf{L}_1, \mathsf{L}_2)$;*
- $\mathsf{L}_1 + \mathsf{L}_2 + \Gamma = \mathsf{L}_1 + \mathsf{x} + \{\langle 2\rangle p \leftrightarrow \mathsf{def}_2(p)\}$ *and* $\mathsf{L}_1 + \mathsf{L}_2 + \Gamma = \mathsf{L}_2 + \{\langle 1\rangle p \leftrightarrow \mathsf{def}_1(p)\}$.

Finally, in both cases, the axiom x is (conservatively) characterized by Γ w.r.t. $(\mathsf{L}_1, \mathsf{L}_2)$ if, and only if, it is minimally *(conservatively) characterized by Γ w.r.t. $(\mathsf{L}_1, \mathsf{L}_2)$.*

Proof. The proof of the second part of the theorem is similar to the proof of the first part. So, we only prove the first part. Assume that x is characterized by Γ w.r.t. $(\mathsf{L}_1, \mathsf{L}_2)$. Then, $\mathsf{L}_1+\mathsf{x} = (\mathsf{L}_1+\mathsf{L}_2+\Gamma)\cap\mathcal{L}_1$, and therefore $\mathsf{L}_1+\mathsf{x} \subseteq \mathsf{L}_1+\mathsf{L}_2+\Gamma$. Moreover, $\langle 2\rangle p \leftrightarrow \mathsf{def}_2(p) \in \mathsf{L}_1 + \mathsf{L}_2 + \Gamma$ by assumption. Thus, $\mathsf{L}_1 + \mathsf{x} + \{\langle 2\rangle p \leftrightarrow \mathsf{def}_2(p)\} \subseteq \mathsf{L}_1+\mathsf{L}_2+\Gamma$. Now, we prove the converse inclusion. Assume towards a contradiction that there is $\varphi \in \mathsf{L}_1+\mathsf{L}_2+\Gamma$ such that $\varphi \notin \mathsf{L}_1+\mathsf{x}+\{\langle 2\rangle p \leftrightarrow \mathsf{def}_2(p)\}$. Then, there is $\varphi' \in \mathcal{L}_1$ such that $\varphi \leftrightarrow \varphi' \in \mathsf{L}_1 + \mathsf{L}_2 + \Gamma$, because $\langle 2\rangle$ is explicitly definable in terms of $\langle 1\rangle$ in $\mathsf{L}_1 + \mathsf{L}_2 + \Gamma$. Then, $\varphi' \in (\mathsf{L}_1 + \mathsf{L}_2 + \Gamma) \cap \mathcal{L}_1$, *i.e.*, $\varphi' \in \mathsf{L}_1 + \mathsf{x}$. Then, by performing the inverse translation that we followed to obtain φ' from φ, we conclude that $\varphi \in \mathsf{L}_1 + \mathsf{x} + \{\langle 2\rangle p \leftrightarrow \mathsf{def}_2(p)\}$. This is impossible, and therefore $\mathsf{L}_1 + \mathsf{L}_2 + \Gamma = \mathsf{L}_1 + \mathsf{x} + \{\langle 2\rangle p \leftrightarrow \mathsf{def}_2(p)\}$.

Now, assume that $\mathsf{L}_1 + \mathsf{L}_2 + \Gamma = \mathsf{L}_1 + \mathsf{x} + \{\langle 2\rangle p \leftrightarrow \mathsf{def}_2(p)\}$. We are going to prove that $\mathsf{L}_1 + \mathsf{x} = (\mathsf{L}_1 + \mathsf{L}_2 + \Gamma) \cap \mathcal{L}_1$. The right to left inclusion is immediate, because $\mathsf{x} \in \mathsf{L}_1 + \mathsf{L}_2 + \Gamma$ by assumption. Now, we prove that $(\mathsf{L}_1 + \mathsf{L}_2 + \Gamma) \cap \mathcal{L}_1 \subseteq \mathsf{L}_1 + \mathsf{x}$. Because of Lemma 12, this amounts to prove that for all Kripke models $\mathcal{M}$, for all $w \in \mathcal{M}$, if $\mathcal{M}, w \models \mathsf{L}_1 + \mathsf{x}$, then $\mathcal{M}, w \models (\mathsf{L}_1 + \mathsf{L}_2 + \Gamma) \cap \mathcal{L}_1$. In order to do so, we are going to build a Kripke model $(\mathcal{M}', w')$ such that $(\mathcal{M}, w)$ and $(\mathcal{M}', w')$ are bisimilar w.r.t. the modality $\langle 1\rangle$ and such that $\mathcal{M}', w' \models \mathsf{L}_1 + \mathsf{x} + \{\langle 2\rangle p \leftrightarrow \mathsf{def}_2(p)\}$ $(*)$, that is, $\mathcal{M}', w' \models \mathsf{L}_1 + \mathsf{L}_2 + \Gamma$ (recall the assumption). This will prove the second inclusion. If $(\mathcal{M}, w) = (W, R_1, R_2, V, w)$, then we define the (pointed) Kripke model $(\mathcal{M}', w') := (W, R_1, R_2', V, w)$, where R_2' is defined as follows. First, we define the (pointed) Kripke model $(\mathcal{M}'', w) := (W, R_1, R_2, V'', w)$ by setting V'' such that for all $q \neq p$, $V''(q) = V(q)$ and such that $V''(p) = \{v\}$. Then, for all $u, v \in W$, we set $uR_2'v$ in $\mathcal{M}'$ if, and only if, $\mathcal{M}'', u \models \mathsf{def}_2(p)$. Then, using the fact that $\mathsf{def}_2(p)$ is positive in p, one can easily show that $(*)$ holds. This proves the second inclusion.

Finally, we prove the last part of the theorem. Assume towards a contradiction that x is characterized by the set of interaction axioms Γ w.r.t. $(\mathsf{L}_1, \mathsf{L}_2)$ and that there is a set of interaction axioms Γ' such that $\Gamma >_{\mathsf{L}_1+\mathsf{L}_2} \Gamma'$ and such that x is also characterized by Γ' w.r.t. $(\mathsf{L}_1, \mathsf{L}_2)$. Because $\Gamma >_{\mathsf{L}_1+\mathsf{L}_2} \Gamma'$, we should have that $\mathsf{L}_1 + \mathsf{L}_2 + \Gamma \subset \mathsf{L}_1 + \mathsf{L}_2 + \Gamma'$. However, since x is characterized by Γ and Γ', we should also have that $\mathsf{L}_1 + \mathsf{L}_2 + \Gamma = \mathsf{L}_1 + \mathsf{L}_2 + \Gamma' = \mathsf{L}_1 + \mathsf{x} + \{\langle 2\rangle p \leftrightarrow \mathsf{def}_2(p)\}$ by the result of the first part of the theorem. This is impossible. □

C Proof of Theorem 8

We first prove a series of lemmata. We also introduce a specific definition of epistemic plausibility space.

Lemma 13. *Let L be a modal logic for $\mathcal{L}_{KB^q}$ such that $\mathsf{K}_{KB^q} \subseteq \mathsf{L}$ and let $\Gamma \subseteq \mathcal{L}_{KB^q}$. Then,*

$$\Gamma \subseteq \mathsf{L} \text{ iff } \{(\mathcal{M}, w) \in \mathcal{S}_K \mid \mathcal{M}, w \models \mathsf{L}\} \subseteq \{(\mathcal{M}, w) \in \mathcal{S}_K \mid \mathcal{M}, w \models \Gamma\} \quad (6)$$

Proof. The proof is the same as the proof of Lemma 12, except that we use the completeness and the compactness of K_{KB^q} instead of K, which was proved in Theorem 2 and Corollary 1 respectively. □

Definition 8. Let $(\mathcal{M}, w) = (W, R_K, V, \mathcal{P})$ be a pointed epistemic-plausibility space. The pointed epistemic-plausibility space denoted by $(\mathcal{M}^*, w^*) = (W^*, R_K^*, V^*, \mathcal{P}^*, w^*)$ is defined as follows:

- $W^* = W$; $R_K^* = R_K$; $V^* = V$;
- for all $w \in W^*$, $\mathcal{P}^*(w) = (W_w^*, \mathcal{A}_w^*, Pl_w^*)$ where:
  - $W_w^* = R_K(w)$;
  - $\mathcal{A}_w^* = \mathcal{A}_w \cap \{R_K(w)\}$;
  - Pl_w^* is such that
    1. for all $A, B \in \mathcal{A}_w^*$, $Pl_w^*(A) \geq Pl_w^*(B)$ iff there is $v \in A$ such that for all $u \in B$, $v \in R_K(u)$;
    2. $Pl_w^*(A) = \bot$ iff $A \cap W_w^* = \emptyset$.

Lemma 14. *Let $\mathcal{M}$ be a pointed epistemic-plausibility space such that $\mathcal{M} \models \mathsf{S4.3}_K$. Then, for all $w \in \mathcal{M}$, $Pl_w^*([\![\psi \wedge \varphi]\!]_w) \geq Pl_w^*([\![\psi \wedge \neg\varphi]\!]_w)$ iff $\mathcal{M}, w \models \langle K\rangle(\psi \wedge K(\psi \to \varphi))$.*

Proof. $Pl_w^*([\![\psi \wedge \varphi]\!]_w) \geq Pl_w^*([\![\psi \wedge \neg\varphi]\!]_w)$

iff there is $v \in R_K(w)$ such that $\mathcal{M}, v \models \psi \wedge \varphi$ and for all $u \in R_K(w)$ such that $\mathcal{M}, u \models \psi \wedge \neg\varphi$, we have that $v \in R_K(u)$;

iff there is $v \in R_K(w)$ such that $\mathcal{M}, v \models \psi \wedge \varphi$ and for all $u \in R_K(w)$, if $v \notin R_K(u)$ then $\mathcal{M}, u \not\models \psi \wedge \neg\varphi$;

iff there is $v \in R_K(w)$ such that $\mathcal{M}, v \models \psi \wedge \varphi$ and for all $u \in R_K(w)$, if $u \in R_K(v)$ then $\mathcal{M}, u \models \psi \to \varphi$ by weak connectedness of R_K;

iff there is $v \in R_K(w)$ such that $\mathcal{M}, v \models \psi \wedge \varphi$ and for all $u \in R_K(v)$, $\mathcal{M}, u \models \psi \to \varphi$ by transitivity of R_K;

iff $\mathcal{M}, w \models \langle K\rangle((\psi \wedge \varphi) \wedge K(\psi \to \varphi))$;

iff $\mathcal{M}, w \models \langle K\rangle(\psi \wedge K(\psi \to \varphi))$ by reflexivity of R_K. □

Lemma 15. *Let $\mathcal{M}$ be an epistemic-plausibility space such that $\mathcal{M}^* \models \mathsf{S4.3}_K$. Then, for all $w \in \mathcal{M}^*$ such that $\mathcal{M}^*, w \models \langle K\rangle\psi$, $Pl_w^*([\![\psi \wedge \varphi]\!]_w) \geq Pl_w^*([\![\psi \wedge \neg\varphi]\!])$ iff $Pl_w^*([\![\psi \wedge \varphi]\!]_w) > Pl_w^*([\![\psi \wedge \neg\varphi]\!])$.*

Proof. Assume that $Pl_w^*([\![\psi \wedge \varphi]\!]_w) \geq Pl_w^*([\![\psi \wedge \neg\varphi]\!]_w)$ and assume towards a contradiction that $Pl_w^*([\![\psi \wedge \varphi]\!]_w) = Pl_w^*([\![\psi \wedge \neg\varphi]\!]_w)$. Then, there are $u, v \in R_K^*(w)$ such that for all $t \in R_K^*(w)$, we have that $v \in R_K^*(t)$, $u \in R_K^*(t)$, $\mathcal{M}^*, u \models \psi \wedge \varphi$ and $\mathcal{M}^*, v \models \psi \wedge \neg\varphi$ $(*)$. However, $\mathcal{M}^*, w \models \langle K\rangle(\psi \wedge K(\psi \to \varphi))$

because $\mathcal{M}^* \models B^q p \leftrightarrow (\langle K\rangle q \rightarrow \langle K\rangle (q \wedge K (q \rightarrow p)))$ and $\mathcal{M}^*, w \models \langle K\rangle \psi$ by assumption. So, there is $t \in R_K^*(w)$ such that $\mathcal{M}^*, t \models K(\psi \rightarrow \varphi)$. This contradicts $(*)$. □

Lemma 16. *Let $(\mathcal{M}, w)$ be a pointed epistemic-plausibility model such that $\mathcal{M}, w \models \mathsf{S4.3}_K + \{B^q p \leftrightarrow (\langle K\rangle q \rightarrow \langle K\rangle (q \wedge K (q \rightarrow p)))\}$. Then, for all $\varphi \in \mathcal{L}_{KB^q}$, it holds that $\mathcal{M}, w \models \varphi$ iff $\mathcal{M}^*, w^* \models \varphi$.*

Proof. The proof is by induction on φ.

- $\varphi = p$, $\varphi = \varphi \wedge \varphi'$ and $\varphi = \neg\psi$ work by definition of V^* and by Induction Hypothesis;
- $\varphi = K\psi$ works also by Induction Hypothesis because $R_K = R_K^*$;
- $\varphi = B^\psi \varphi$.
  $\mathcal{M}^*, w^* \models B^\psi \varphi$
  iff $Pl_w^*([\![\psi]\!]_w) = \bot$ or $Pl_w^*([\![\psi \wedge \varphi]\!]_w) > Pl_w^*([\![\psi \wedge \neg\varphi]\!]_w)$
  iff $[\![\psi]\!]_w \cap R_K(w) = \emptyset$ or $Pl_w^*([\![\psi \wedge \varphi]\!]_w) > Pl_w^*([\![\psi \wedge \neg\varphi]\!]_w)$ by Definition 8, item 2;
  iff $\mathcal{M}, w \models K\neg\psi$ or $Pl_w^*([\![\psi \wedge \varphi]\!]_w) > Pl_w^*([\![\psi \wedge \neg\varphi]\!])$;
  iff $\mathcal{M}, w \models K\neg\psi$ or $Pl_w^*([\![\psi \wedge \varphi]\!]_w) \geq Pl^*([\![\psi \wedge \neg\varphi]\!]_w)$ by Fact 14.
  iff $\mathcal{M}, w \models K\neg\psi$ or $\mathcal{M}, w \models \langle K\rangle(\psi \wedge K(\psi \rightarrow \varphi))$;
  iff $\mathcal{M}, w \models \langle K\rangle\psi \rightarrow \langle K\rangle(\psi \wedge K(\psi \rightarrow \varphi))$ by Fact 15;
  iff $\mathcal{M}, w \models B^\psi \varphi$ by assumption on $\mathcal{M}$. □

Theorem 8. *The conditional belief modality $B^q p$ is explicitly defined in the logic $\mathsf{L}^q := \mathsf{S4}_K + \mathsf{K}_{B^q} + \{I_1^q, I_2^q, I_3^q, I_4^q, I_5^q\}$ by the formula $\mathsf{def}_{B^q p} := \langle K\rangle q \rightarrow \langle K\rangle (q \wedge K (q \rightarrow p)) \in \mathcal{L}_K$:*

$$B^q p \leftrightarrow (\langle K\rangle q \rightarrow \langle K\rangle (q \wedge K (q \rightarrow p))) \in \mathsf{L}^q \qquad (\mathsf{Def}\ B^q p)$$

Moreover, the following also holds:

$$\mathsf{S4}_K + \mathsf{K}_{B^q} + \{I_1^q, I_2^q, I_3^q, I_4^q, I_5^q\} = \mathsf{S4.3}_K + \{B^q p \leftrightarrow \mathsf{def}_{B^q p}\} \qquad (4)$$

Proof. We split the proof of Expression $(\mathsf{Def}\ B^q p)$ into two parts. First, we prove that $\mathsf{def}_{B^q p} \rightarrow B^q p \in \mathsf{S4}_K + \mathsf{K}_{B^q} + \{I_1^q, I_2^q, I_4^q\}$. Second, we prove that $B^q p \rightarrow \mathsf{def}_{B^q} \in \mathsf{S4}_K + \mathsf{K}_{B^q} + \{I_1^q, I_3^q, I_5^q\}$.

1. Because of Lemma 13, this amounts to proving that for all pointed epistemic plausibility space $(\mathcal{M}, w)$ such that $\mathcal{M}, w \models \mathsf{S4}_K + \mathsf{K}_{B^q} + \{I_1^q, I_2^q, I_4^q\}$, we have that $\mathcal{M}, w \models \mathsf{def}_{B^q p} \rightarrow B^q p$. Let $(\mathcal{M}, w)$ be such a pointed epistemic-plausibility space and assume that $\mathcal{M}, w \models \langle K\rangle\ q \rightarrow \langle K\rangle (q \wedge K (q \rightarrow p))$. Assume towards a contradiction that $\mathcal{M}, w \not\models B^q p$. Then, by definition, $Pl_w([\![q]\!]_w) \neq \bot$ and not $Pl_w([\![q \wedge p]\!]_w) > Pl_w([\![q \wedge \neg p]\!])$. Because $Pl_w([\![q]\!]_w) \neq \bot$, it holds that $\mathcal{M}, w \models \neg B^q \bot$. Now, because $\models B^q q$ by $\mathsf{C1}$, we have that $\models B^q \neg q \rightarrow B^q \bot$ by

$\mathsf{C2}$, i.e. $\mathcal{M}, w \models \neg B^q \bot \rightarrow \neg B^q \neg q$. Therefore, $\mathcal{M}, w \models \neg B^q \neg q$. So, by axiom I_2^q, $\mathcal{M}, w \models \langle K \rangle q$. Then, by assumption, $\mathcal{M}, w \models \langle K \rangle (q \wedge K(q \rightarrow p))$. So, there is $v \in R_K(w)$ such that $\mathcal{M}, v \models q \wedge K(q \rightarrow p)$. Therefore, $\mathcal{M}, v \models K(q \rightarrow p)$, and so $\mathcal{M}, v \models B^q(q \rightarrow p)$ by application of the rule of necessitation and Axiom I_1^q. Hence, $\mathcal{M}, v \models B^q p$, because $\models B^q q$. Now, $\mathcal{M}, w \models \neg B^q p$, and therefore $\mathcal{M}, w \models K(\langle K \rangle q \rightarrow \neg B^q p)$ by I_4^q. So, $\mathcal{M}, v \models \langle K \rangle q \rightarrow \neg B^q p$. Since, $\mathcal{M}, v \models q$, we also have that $\mathcal{M}, v \models \langle K \rangle q$. Therefore, $\mathcal{M}, w \models \neg B^q p$ which contradicts our previous deduction. So, we reach a contradiction and therefore our initial assumption was wrong. Hence, we have that $\mathcal{M}, w \models \mathsf{def}_{B^q} \rightarrow B^q p$. This holds for any pointed epistemic-plausibility space $(\mathcal{M}, w)$ such that $\mathcal{M}, w \models \mathsf{S4}_K + \mathsf{K}_{B^q} + \{I_1^q, I_2^q, I_4^q\}$, so we have proved the first part.

2. Let $(\mathcal{M}, w)$ be a pointed epistemic-plausibility space. Assume that $\mathcal{M}, w \models \mathsf{S4}_K + \mathsf{K}_{B^q} + \{I_1^q, I_3^q, I_5^q\}$ and that $\mathcal{M}, w \models B^q p$. Assume towards a contradiction that $\mathcal{M}, w \not\models \mathsf{def}_{B^q p}$, that is, $\mathcal{M}, w \models \langle K \rangle q \wedge K(q \rightarrow \langle K \rangle (q \wedge \neg p))$. Then, $\mathcal{M}, w \models B^q (q \rightarrow \langle K \rangle (q \wedge \neg p))$ by I_1^q. So, $\mathcal{M}, w \models B^q \langle K \rangle (q \wedge \neg p)$ because $\mathcal{M}, w \models B^q q$ and by distributivity of B^q. Therefore, $\mathcal{M}, w \models B^q p \wedge B^q \langle K \rangle (q \wedge \neg p)$ because by assumption $\mathcal{M}, w \models B^q p$. So, $\mathcal{M}, w \models B^q K(q \rightarrow p) \wedge B^q \langle K \rangle (q \wedge \neg p)$ by I_3^q. Then, $\mathcal{M}, w \models B^q (K(q \rightarrow p) \wedge \langle K \rangle (q \wedge \neg p))$, again by distributivity of B^q. Therefore, $\mathcal{M}, w \models B^q \langle K \rangle (p \wedge \neg p)$. So, $\mathcal{M}, w \models B^q \bot$. Thus, by I_5^q, we have that $\mathcal{M}, w \models K \neg q$, that is, $\mathcal{M}, w \models \neg \langle K \rangle q$. This contradicts our assumption that $\mathcal{M}, w \models \langle K \rangle q$. So, we reach a contradiction and therefore our initial assumption was wrong. Hence, we have that $\mathcal{M}, w \models B^q p \rightarrow \mathsf{def}_{B^q}$. This holds for any pointed epistemic-plausibility space $(\mathcal{M}, w)$ such that $\mathcal{M}, w \models \mathsf{S4}_K + \mathsf{K}_{B^q} + \{I_1^q, I_3^q, I_5^q\}$, so we have proved the second part by Lemma 13. Putting these two facts altogether, we obtain that $B^q p \leftrightarrow \mathsf{def}_{B^q} \in \mathsf{S4}_K + \mathsf{K}_{B^q} + \{I_1^q, I_2^q, I_3^q, I_4^q, I_5^q\}$. Hence, we obtain the definability result.

We prove the second part of the theorem. To prove the left to right inclusion of Expression (4), we must show that $\{I_1^q, I_2^q, I_3^q, I_4^q, I_5^q\} \subseteq \mathsf{S4.3}_K + \{B^q p \leftrightarrow (\langle K \rangle q \rightarrow \langle K \rangle (q \wedge K(q \rightarrow p)))\}$. Because of Lemma 13, this amounts to proving that for all (qualitative) epistemic-plausibility spaces $(\mathcal{M}, w) \in \mathcal{S}_K$ such that $\mathcal{M}, w \models \mathsf{S4.3}_K + \{B^q p \leftrightarrow \mathsf{def}_{B^q p}\}$ $(*)$ we have that $\mathcal{M}, w \models \{I_1^q, I_2^q, I_3^q, I_4^q, I_5^q\}$. By Lemma 16, this is equivalent to showing that for all $(\mathcal{M}, w) \in \mathcal{S}_K$ such that $(*)$ holds, we have that $\mathcal{M}^*, w^* \models \{I_1^q, I_2^q, I_3^q, I_4^q, I_5^q\}$. We only deal with the cases I_4^q and I_5^q, the other cases being rather straightforward by definition of $\mathcal{M}^*$. We first prove that for all $(\mathcal{M}, w) \in \mathcal{S}_K$ such that $(*)$ holds, we have that $\mathcal{M}^*, w^* \models I_4^q$. Assume that $\mathcal{M}^*, w^* \models \neg B^q p$. Let $v^* \in R_K^*(w^*)$ and assume that $\mathcal{M}^*, v^* \models \langle K \rangle q$. Then, we must show that $\mathcal{M}^*, v^* \models \neg B^q p$, *i.e.* $\mathcal{M}^*, v^* \models \langle K \rangle q \wedge K(q \rightarrow \langle K \rangle (q \wedge \neg p))$ because of $(*)$, the rule of necessitation and Lemma 16. We already know that $\mathcal{M}^*, v^* \models \langle K \rangle q$. Now, because $\mathcal{M}^*, w^* \models \neg B^q p$, we have that $\mathcal{M}^*, w^* \models \langle K \rangle q \wedge K(q \rightarrow \langle K \rangle (q \wedge \neg p))$, and therefore $\mathcal{M}^*, w^* \models K(q \rightarrow \langle K \rangle (q \wedge \neg p))$. Hence, $\mathcal{M}^*, v^* \models K(q \rightarrow \langle K \rangle (q \wedge \neg p))$ because $v^* \in R_K^*(w^*)$. This proves that $\mathcal{M}^*, w^* \models I_4^q$. Finally, we prove the last case, namely that for all $(\mathcal{M}, w) \in \mathcal{S}_K$, $\mathcal{M}^*, w^* \models I_5^q$. Assume that $\mathcal{M}^*, w^* \models \langle K \rangle q$. We must prove that $\mathcal{M}^*, w^* \models \neg B^q \bot$. That is, by the truth conditions of B^q, we

must prove that it is not the case that $Pl^*_w([\![q \wedge \bot]\!]_w) > Pl^*_w([\![q \wedge \top]\!]_w)$, because $Pl^*_w([\![q]\!]_w) \neq \bot$ since $\mathcal{M}^*, w^* \models \langle K \rangle q$. But $Pl^*_w([\![q \wedge \bot]\!]_w) = Pl^*_w(\emptyset) = \bot$ by Condition 2 of Definition 8. Because it is impossible that for some $d \in D$ we have that $\bot > d$, we obtain our result. Second, we prove the right to left inclusion of Expression (4). For this, we must show that $.3 \in \mathsf{L}^q := \mathsf{S4}_K + \mathsf{K}_{B^q} + \{I^q_1, I^q_2, I^q_3, I^q_4, I^q_5\}$ because we already proved in the first part of the theorem that $B^q p \leftrightarrow (\langle K \rangle q \rightarrow \langle K \rangle (q \wedge K (q \rightarrow p))) \in \mathsf{L}^q := \mathsf{S4}_K + \mathsf{K}_{B^q} + \{I^q_1, I^q_2, I^q_3, I^q_4, I^q_5\}$. This result is already proved in [3]. This completes the proof of the theorem. $\square$