# Axioms .2 and .4 as Interaction Axioms

**Guillaume Aucher**
University of Rennes 1
INRIA
Rennes, France
*guillaume.aucher@irisa.fr*

**Abstract**

In epistemic logic, some axioms dealing with the notion of knowledge are rather convoluted and it is difficult to give them an intuitive interpretation, even if some of them, like .2 and .4, are considered by some epistemic logicians to be key axioms. We show that they can be characterized in terms of understandable interaction axioms relating knowledge and belief. In order to show it, we first present 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 our results related to epistemic logic.

## 1 Introduction

The development of modern epistemic logic can be viewed as an attempt 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 and subsequent philosophers and logicians tried to formulate explicit principles governing and relating expressions of the form "$a$ knows that $\varphi$" (subsequently formalized as $K\varphi$) and "$a$ believes that $\varphi$" (subsequently formalized as $B\varphi$), where $a$ is a human agent and $\varphi$ is a proposition. In other words, they 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 1978, p. 15). For example, the interaction axioms $K\varphi \to B\varphi$ and $B\varphi \to KB\varphi$ are often considered to be intuitive principles: if agent $a$ knows $\varphi$ then (s)he also believes $\varphi$, or if agent $a$ believes $\varphi$, then (s)he knows that (s)he believes $\varphi$.

In computer science, the logic of knowledge is usually considered to be S5, which is the logic S4 composed of the axioms $K\varphi \to \varphi$ and $K\varphi \to KK\varphi$ to which is added the axiom 5: $\neg K\varphi \to K\neg K\varphi$. This last axiom is validated in situations where the agent cannot have mistaken beliefs. For this very reason, it has been attacked by various philosophers because it cannot hold in general. Dropping this axiom 5, we obtain a rich variety of weaker logics of knowledge which have been proposed and examined by logicians (Lenzen 1979), such as S4.2, S4.3, S4.3.2 and 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 S4.2 is the logic of knowledge" (Lenzen 1979, p. 33), where the axiom .2 is $\neg K\neg K\varphi \to K\neg K\neg\varphi$. Likewise, Kutschera argues for S4.4 as the logic of knowledge, where the axiom .4 is $(\varphi \wedge \langle K\rangle K\varphi) \to K\varphi$ (Kutschera 1976). 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.

To better grasp the intuitions underlying these intricate axioms, we show that axioms .2 and .4 can be characterized equivalently in terms of interaction axioms relating knowledge and 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 meta-theory of 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 the problems that we address in this article have never been addressed in the logical literature, neither for multi-modal logics nor for combinations of modal logics (Marx and Venema 1997; Gabbay et al. 1998).

## 2 Towards a Theory of Interaction Axioms

In this section, we start by recalling the basics of modal logic. Then, we present our primitive meta-theory of modal logic dealing with interaction axioms.

### 2.1 Modal Logic

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

$$\mathcal{L}_{\mathbb{A}} \ : \ \varphi \ ::= \ p \ \mid \ \neg\varphi \ \mid \ \varphi \wedge \varphi \ \mid \ [i]\varphi$$

where $p$ ranges over $\Phi$ and $i$ ranges over $\mathbb{A}$. The formula $\top$ is an abbreviation for $p \vee \neg p$ (for some fixed $p \in \Phi$), the

Copyright © 2014, Association for the Advancement of Artificial Intelligence (www.aaai.org). All rights reserved.

formula $\bot$ is an abbreviation for $\neg\top$ 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 $\neg$. It is *positive in* $p$ if all occurrences of $p$ in $\varphi$ 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* $\mathsf{L}$ 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 $\varphi$ belongs to $\mathsf{L}$ then so do all of its substitution instances (see (Blackburn, de Rijke, and Venema 2001, 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*.

If $\mathsf{L}$ and $\mathsf{L}'$ 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 also called the *fusion* of $\mathsf{L}$ and $\mathsf{L}'$ (Marx and Venema 1997; Gabbay et al. 1998)). If $\mathsf{x}$ 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 $\mathsf{x}$ is *derivable from* $X$ *in* $\mathsf{L}$ 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 $X' \vdash_{\mathsf{L}} X$.

**Kripke Semantics.** The Kripke semantics will be used only in the proof of Theorem 2.2. 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 $V : \Phi \to 2^W$ is called a *valuation* and assigns to each propositional letter $p \in \Phi$ a subset of $W$. 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 $w \in W$.

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 $\varphi \in \Gamma$.

## 2.2 Interaction Axioms and Characterization of Axioms

In the sequel, $\mathsf{L}_1$ and $\mathsf{L}_2$ are two modal logics for $\mathcal{L}_1$ and $\mathcal{L}_2$ respectively, and $\mathsf{L}_{1,2}$ is a modal logic for $\mathcal{L}_{1,2}$.

**Definition 2.1 (Interaction Axioms).** 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

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

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

**Definition 2.2 (Characterization of an Axiom).** We say that $\mathsf{x}$ is *characterized* by the set of interaction axioms $\Gamma$ w.r.t. $(\mathsf{L}_1, \mathsf{L}_2)$ when

$$\mathsf{L}_1 + \mathsf{x} = (\mathsf{L}_1 + \mathsf{L}_2 + \Gamma) \cap \mathcal{L}_1. \tag{1}$$

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

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

Definition 2.2 tells us that an axiom $\mathsf{x}$ is characterized by a set of interaction axioms $\Gamma$ 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 $\mathsf{x}$ to $\mathsf{L}_1$, and nothing else. For the case of strong characterization, by adding these interaction axioms $\Gamma$, 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 initially.

**Proposition 2.1.**
- *If $\mathsf{x}$ is conservatively characterized by a set of interaction axioms w.r.t. $(\mathsf{L}_1, \mathsf{L}_2)$, then $\mathsf{x} \notin \mathsf{L}_1$.*
- *Conversely, if $\mathsf{x} \notin \mathsf{L}_1$ and $\Gamma$ is a set of formulas of $\mathcal{L}_{1,2}$ such that Equation* (1) *holds, then $\Gamma$ is a set of interaction axioms w.r.t. $(\mathsf{L}_1, \mathsf{L}_2)$.*

*Proof.* It follows easily from Definitions 2.1 and 2.2. $\square$

Finally, we define a notion of minimality among the sets of interaction axioms characterizing an axiom $\mathsf{x}$.

**Definition 2.3 (Minimal Characterization).** The axiom $\mathsf{x}$ is *minimally characterized* by the set of interaction axioms $\Gamma$ w.r.t. $(\mathsf{L}_1, \mathsf{L}_2)$ when $\mathsf{x}$ is characterized by $\Gamma$ w.r.t. $(\mathsf{L}_1, \mathsf{L}_2)$ and there is no set of interaction axioms $\Gamma'$ such that $\Gamma >_{\mathsf{L}_1+\mathsf{L}_2} \Gamma'$ and $\mathsf{x}$ is still characterized by $\Gamma'$ 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 in (Halpern, Samet, and Segev 2009b). This study is subsequently applied to epistemic logic in (Halpern, Samet, and Segev 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 article, we are interested only in the notion of *explicit* definability, which is also used in (Lenzen 1979).

**Definition 2.4 (Explicit Definability of a Modality).** 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.

**Theorem 2.2.** *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:*

- *$\mathsf{x}$ is characterized by $\Gamma$ 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:*

- *$\mathsf{x}$ is conservatively characterized by $\Gamma$ 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 $\mathsf{x}$ is (conservatively) characterized by $\Gamma$ w.r.t. $(\mathsf{L}_1, \mathsf{L}_2)$ if, and only if, it is* minimally *(conservatively) characterized by $\Gamma$ 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 $\mathsf{x}$ is characterized by $\Gamma$ 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 $\varphi'$ from $\varphi$, 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 $(\mathsf{L}_1 + \mathsf{L}_2 + \Gamma) \cap \mathcal{L}_1 \subseteq \mathsf{L}_1 + \mathsf{x}$, *i.e.*, 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 $u R_2' v$ 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 $\mathsf{x}$ is characterized by the set of interaction axioms $\Gamma$ w.r.t. $(\mathsf{L}_1, \mathsf{L}_2)$ and that there is a set of interaction axioms $\Gamma'$ such that $\Gamma >_{\mathsf{L}_1 + \mathsf{L}_2} \Gamma'$ and such that $\mathsf{x}$ is also characterized by $\Gamma'$ 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 $\mathsf{x}$ is characterized by $\Gamma$ and $\Gamma'$, 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. $\square$

## 3 Epistemic Logic and Interaction Axioms

We introduce the basics of epistemic logic (see (Fagin et al. 1995; Meyer and van der Hoek 1995) for more details).

### 3.1 Logics of Knowledge and Belief

We define the epistemic-doxastic language $\mathcal{L}_{KB}$ inductively as follows:

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

where $p$ ranges over $\Phi$. 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 $B$elieves $\varphi$' and $K\varphi$ reads as 'the agent $K$nows $\varphi$'. 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. Below, we give a list of axioms that will be used in the rest of the article.

| | | | |
|---|---|---|---|
| D | $: K\varphi \to \langle K \rangle\varphi$ | 4 | $: K\varphi \to KK\varphi$ |
| 5 | $: \neg K\varphi \to K\neg K\varphi$ | T | $: K\varphi \to \varphi$ |
| .2 | $: \langle K \rangle K\varphi \to K\langle K \rangle\varphi$ | .4 | $: (\varphi \wedge \langle K \rangle K\varphi) \to K\varphi.$ |

The logic $\mathsf{KD45}_B$ is the modal logic for $\mathcal{L}_B$ generated by the set of axioms $\{\mathsf{D}, \mathsf{4}, \mathsf{5}\}$. For any $\mathsf{x} \in \{.2, .4\}$, the logic $\mathsf{S4.x}_K$ is the 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:

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

### 3.2 Interaction Axioms for Epistemic Logic

The following interaction axioms were suggested in (Hintikka 1962):

$$K\varphi \to B\varphi \qquad (\mathrm{I}_1)$$
$$B\varphi \to KB\varphi \qquad (\mathrm{I}_2)$$

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

$$B\varphi \to BK\varphi \qquad (\mathrm{I}_3)$$

Axiom $I_3$ above was suggested in (Lenzen 1978). 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".

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

$$K\varphi \leftrightarrow (\varphi \wedge B\varphi) \qquad (\mathrm{I}_4)$$

This list of interaction axioms is incomplete, see (Aucher 2014) for more information about interaction axioms and axioms of epistemic logic.

## 4 Applying our Theory to Epistemic Logic

In (Halpern, Samet, and Segev 2009a), only the interaction axioms $\mathrm{I}_1$ and $\mathrm{I}_2$ suggested in (Hintikka 1962) are considered. In this section, we also add the interaction axiom $I_3$. 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 4.1.** *(Lenzen 1979)*

*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 following definition:*

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

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

This result can be contrasted with Theorem 4.8 of (Halpern, Samet, and Segev 2009a), from which it follows that the belief modality can*not* be explicitly defined in the logic $(\mathsf{S4.x})_K + (\mathsf{KD45})_B + \{I_1, I_2\}$, for any $\mathsf{x} \in \{.2, .4\}$. On the other hand, as Theorem 4.2 below shows, knowledge can be defined in terms of belief only if the logic of knowledge is S4.4, but not if the logic of knowledge is S4 or S4.2.

**Theorem 4.2.** *(Aucher 2014)*

- *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 following definition:*

$$K\varphi \leftrightarrow \varphi \wedge B\varphi \in \mathsf{L.4} \qquad (\mathsf{Def\ K})$$

- *The knowledge modality $K$ cannot be explicitly defined in the logic $\mathsf{S4.2}_K + \mathsf{KD45}_B + \{I_1, I_2, I_3\}$.*

Again, this result can be contrasted with Theorem 4.1 of (Halpern, Samet, and Segev 2009a), from which it follows that the knowledge modality can*not* be explicitly defined in the logic $(\mathsf{S4.4})_K + (\mathsf{KD45})_B + \{I_1, I_2\}$. We see that in both cases the increase in expressivity due to the addition of the interaction axiom $\mathrm{I}_3$ plays an important role in bridging the gap between belief and knowledge.

**Theorem 4.3.** • *The axiom .2 is characterized w.r.t. the pair $(\mathsf{S4}_K, \mathsf{KD45}_B)$ by the set of interaction axioms $\{I_1, I_2, I_3\}$.*

- *The axiom .4 is conservatively characterized w.r.t. the pair $(\mathsf{S4}_K, \mathsf{KD45}_B)$ by the interaction axiom $I_4$.*

*Proof.* It follows from a direct application of Theorem 2.2 to the results of (Lenzen 1979), namely the facts that $\mathsf{S4}_K + \mathsf{KD45}_B + \{I_1, I_2, I_3\} = \mathsf{S4.2}_K + \{B\varphi \leftrightarrow \langle K\rangle K\varphi\}$ and $\mathsf{S4}_K + \mathsf{KD45}_B + \{I_4\} = \mathsf{KD45}_B + \{I_4\} = \mathsf{S4.4}_K + \{B\varphi \leftrightarrow \langle K\rangle K\varphi\}$. □

## 5 Conclusion

Theorem 4.3 tells us that if we use S4.4 as the logic of knowledge (and KD45 as the logic of belief), then we implicitly assume that knowledge is in fact 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, if we use S4.2 as the logic of knowledge (and KD45 as the logic for belief), then Theorem 4.3 tells us that, by doing so, we are only assuming that the agent knows his beliefs and disbeliefs and that his beliefs are in fact certainties, convictions, and not simply weak beliefs.

Overall, our meta-theory of modal logic enables to carry out a rigorous and fine-grained analysis of the intuitive assumptions underlying the logics of knowledge between S4 and S5. This theory provides a meaningful logical foundation for these analysis and can serve as a means to justify and explain the intuitive arguments employed.

## References


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

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

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

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

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.

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

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

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.

Marx, M., and Venema, Y. 1997. *Multi-dimensional modal logic*. Springer.

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