Defining Knowledge in Terms of Belief: The Modal Logic Perspective*

Joseph Y. Halpern†
Computer Science Department
Cornell University, U.S.A.
halpern@cs.cornell.edu

Dov Samet‡
The Faculty of Management
Tel Aviv University, Israel
samet@post.tau.ac.il

Ella Segev
Faculty of Industrial Engineering and Management
Technion - Israel Institute of Technology, Israel
esegev@ie.technion.ac.il

Abstract

The question of whether knowledge is definable in terms of belief, which has played an important role in epistemology for the last fifty years, is studied here in the framework of epistemic and doxastic *logics*. Three notions of definability are considered: explicit definability, implicit definability, and reducibility, where explicit definability is equivalent to the combination of implicit definability and reducibility. It is shown that if knowledge satisfies any set of axioms contained in S5, then it cannot be explicitly defined in terms of belief. S5 knowledge can be implicitly defined by belief, but not reduced to it. On the other hand, S4.4 knowledge and weaker notions of knowledge cannot be implicitly defined by belief, but can be reduced to it by defining knowledge as *true belief*. It is also shown that S5 knowledge cannot be reduced to belief *and* justification, provided that there are no axioms that involve both belief and justification.

*We thank Samson Abramsky, Arnon Avron, Johan van Benthem, Sergiu Hart, Martin Meier, and Jouku Väänänen for helpful discussions.

†Supported in part by NSF under grants TR-0325453 and IIS-0534064, and by AFOSR under grant FA9550-05-1-0055.

‡Support by the Israeli Science Foundation under grant 891/04 is acknowledged.

1 Introduction

The observations that knowledge and belief are related goes back to Plato's dialogue *Theaetetus*, whose protagonist suggests that knowledge is justified true belief (JTB). Two millennia later, analytic philosophers such as Ayer [1956] and Chisholm [1957] adopted Plato's slogan. But then a three-page paper by Gettier [1963] refuted the proposed definition by means of a few counterexamples; this started a new area of epistemological study of knowledge that tried to justify and clarify the notion of knowledge as JTB.

Another flourishing new area of epistemology was started at the same time by Hintikka's book *Knowledge and Belief: An Introduction to the Logic of the Two Notions* [1962]. Hintikka studied knowledge and belief as modalities, employing and developing for this purpose the syntax and semantics of modal logic, thus laying the foundations of modern epistemic logic. As indicated by the title of Hintikka's book, knowledge and belief were intertwined in epistemic logic since its inception.

As noted recently by Stalnaker [2006] and van Benthem [2006], the formal study of epistemic logic has contributed very little to the study of knowledge as JTB. Indeed, there seems to have been very little communication between epistemic logic and epistemology; several recent monographs and articles devoted to JTB have altogether ignored the developments in epistemic logic in the past forty years (see, for example, [Alston 1989; BonJour and Sosa 2003; Swinburne 2001]) While there is certainly awareness of the theory of JTB among researchers in epistemic logic, there seems to have been very little work on it. For example, a survey of epistemic logic of Lenzen [1978] formalized the notions of JTB in modal logic, but did not apply the tools of this logic to analyze the theory. There has been a little more activity recently. For example, Stalnaker [2006] proposed and studied a new logic of knowledge and belief, addressing some of the issues that were raised in the literature on JTB. while Artemov and Nogina [2005] introduced a logic of justified knowledge. (Van Benthem [2006] provides an overview of some of the recent lines of research.)

There is a central aspect of the theory of JTB that seems particularly suited to the tools of epistemic logic: the *definability* of knowledge. Most variants of the theory of JTB share in common a quest for a definition or expression of knowledge in terms of belief and possibly other things. But what does it mean to *define* knowledge in terms of belief, or, more generally, to define one modality in terms of others? We develop a general theory of modal definability in a companion paper [Halpern, Samet, and Segev 2008]. Here we present the main results and apply them to specific logics of knowledge and belief.[1]

We consider three different notions of definability, and show that there is a surprisingly rich interplay between these notions of definability and the assumptions we make about the underlying notion of knowledge. Our results highlight the special role of the much-debated negative introspection axiom for knowledge in the context of definability. They also give us the tools to reexamine the adequacy of weak logics of knowledge and explicate the notion of epistemic luck.

We briefly describe the three notions of definability here, so as to be able to outline our results in a little more detail. In first-order logic, the notions of *implicit definability* and *explicit definability* of predicates are standard, and are known to be equivalent by Beth's theorem [1953]. These notions can be lifted to the definability of modalities in modal logics in a straightforward way. We explain the definitions in the context of knowledge and belief.

Consider a logic Λ for knowledge and belief. Knowledge is *explicitly defined* in Λ if there is a formula DK (for "definition of knowledge") in Λ of the form $Kp \leftrightarrow \delta$, where δ is a formula that does not mention the knowledge operator. Knowledge is *implicitly defined in* Λ if, roughly speaking, Λ "determines" knowledge uniquely. Syntactically, this determination means that any two modal operators for knowledge that satisfy Λ must be equivalent. Semantically, this means that two Kripke models of Λ with the same set of worlds that agree on the interpretation of belief (and on the interpretations of all primitive propositions) must agree also on the interpretation of knowledge.

Unlike the case of first-order definability, these two notions of definability do not coincide for modal logics; implicit definability is strictly weaker than explicit definability. In Part I, we define another notion of definability, *reducibility*, and prove that implicit definability combined with reducibility is equivalent to explicit definability. Knowledge is *reducible to belief* in Λ if all the theorems in Λ follow from the logic of belief included in Λ together with the definition DK.[2]

In this paper, we consider logics Λ of knowledge and belief where the logic of belief in Λ is KD45. The logic Λ is assumed to include the two axioms, suggested by Hintikka [1962], that provide the link between belief and knowledge: the axiom that whatever is known is also believed ($Kp \rightarrow Bp$), and the axiom that all beliefs are known ($Bp \rightarrow KBp$). Our first result shows

[1] We have included enough review in each paper to make them both self-contained.

[2] We also require that the logic of belief and DK be a conservative extension of the logic of belief. We discuss this issue in more detail in Section 5.

that as long as the logic of knowledge in Λ is contained in S5, knowledge is not explicitly defined in Λ. By the equivalence theorem of the companion paper, in any such logic it is impossible that knowledge is both implicitly defined in Λ and reducible to belief. Which of these properties holds for a logic of knowledge and belief depends on the properties of knowledge. When knowledge satisfies S5, it is not reducible to belief, but, somewhat surprisingly, it is implicitly defined by belief. On the other hand, when the logic of knowledge is contained in S4.4, then knowledge is not implicitly defined but is reducible to belief.

The reducibility of S4.4 knowledge to belief requires a formula that defines knowledge in terms of belief. The defining formula is perhaps the most natural candidate: the formula $Kp \leftrightarrow p \wedge Bp$, denoted TB, which defines knowledge as true belief. The connection between the formula TB and the logic of knowledge S4.4 was already pointed out by Lenzen [1979], who showed that the logic generated by adding TB to KD45 contains S4.4. The reducibility of S4.4 knowledge and all weaker logics of knowledge to belief implies that these logics do not capture the distinction between knowledge and mere (unjustified) true belief.

The irreducibility of S5 knowledge to belief raises the question of whether S5 knowledge is reducible to belief and an additional modality: justification. We addressed this question in the companion paper [Halpern, Samet, and Segev 2008], using algebraic semantics. Using the result obtained in the companion paper, we conclude here that if there are no axioms that involve both justification and belief, then S5 knowledge is not reducible to belief and justification.

Finally, we address the question of defining KD45 belief in terms of knowledge. It is shown that it can be neither explicitly not implicitly defined for any logic of knowledge contained in S5, and hence for any weaker logic. Yet KD45 belief can be reduced to S4.4 knowledge by the axiom $Bp \leftrightarrow \neg K \neg Kp$. The fact that adding this formula to S4.4 generates the logic of KD45 was already observed by Lenzen [1979].

The rest of this paper is organized as follows. In Section 2, we review the standard logics of knowledge and belief and their semantics in terms of Kripke models and frames. In Section 3, we review the definitions of the three notions of definability that we consider from the companion paper, and state the main result of the companion paper, that explicit definability is equivalent to implicit definability and reducibility. Using these definitions, we consider the extent to which knowledge can be defined in terms of belief (and vice versa) in Section 4. In Section 5 we take a closer look at the notion of reducibility, consider some variants of the definition similar in spirit to other notions considered in the literature, and examine the extent to which our results hold with these variants. In Section 6, we consider what our results have to say about *epistemic luck* [Pritchard 2005] and negative introspection. Proofs are relegated to the appendix.

2 Logics of knowledge and belief

We assume that the reader is familiar with the standard logics of knowledge and belief, and their semantics. We briefly review the relevant material here. (Some of this material is taken verbatim from the companion paper; we include it here to make this paper self-contained.)

Let P be a nonempty set of *primitive propositions*. Let $M_1, \dots M_n$ be *modal operators* or *modalities*. *Formulas* are defined by induction. Each primitive proposition is a formula. If φ and ψ are formulas then $\neg\varphi$, $(\varphi \to \psi)$, and $M_i\varphi$ for $i = 1, \dots n$, are also formulas.[3] The propositional connectives $\vee$, $\wedge$, $\leftrightarrow$ are defined in terms of $\neg$ and $\to$ in the usual way; we take *true* to be an abbreviation of $p \vee \neg p$. The *language* $\mathcal{L}(M_1, \dots M_n)$ is the set of all formulas defined in this way.

For the purposes of this paper, we take a (*modal*) *logic* Λ to be any collection of formulas in a language $\mathcal{L}(M_1, \dots, M_n)$ that (a) contains all tautologies of propositional logic; (b) is closed under *modus ponens*, so that if $\varphi \in \Lambda$ and $\varphi \to \psi \in \Lambda$, then $\psi \in \Lambda$; and (c) is closed under substitution, so that if $\varphi \in \Lambda$, p is a primitive proposition, and $\psi \in \mathcal{L}(M_1, \dots, M_n)$, then $\varphi[p/\psi] \in \Lambda$, where $\varphi[p/\psi]$ is the formula that results by replacing all occurrences of p in φ by ψ. A logic Λ is *normal* if, in addition, for each modal operator M_i, Λ contains the axiom K_{M_i}, $M_i(p \to q) \to (M_ip \to M_iq)$, and is closed under generalization, so that if $\varphi \in \Lambda$, then so is $M_i\varphi$. In this paper, we consider only normal modal logics. If Λ_1 and Λ_2 are two sets of formulas, we denote by $\Lambda_1 + \Lambda_2$ the smallest normal modal logic containing Λ_1 and Λ_2. Even if Λ_1 and Λ_2 are themselves normal modal logics, $\Lambda_1 \cup \Lambda_2$ may not be; for example, it may not be closed under the substitution rule. Thus, $\Lambda_1 + \Lambda_2$ will in general be a superset of $\Lambda_1 \cup \Lambda_2$. Note that if Λ is a normal logic and $\mathcal{L}$ is a language (which might not contain Λ), then $\Lambda \cap \mathcal{L}$ is a normal logic.

We are often interested in logics generated by axioms. The *logic generated by a set A of formulas*, typically called *axioms*, is the smallest normal logic that contains A.

[3] The modalities in this paper are unary. It is straightforward to extend our results to modal operators of higher arity.

The logic of belief we adopt here is the normal logic in the language $\mathcal{L}(B)$ generated by the following three axioms:

(D$_B$) $\quad Bp \to \neg B \neg p$
(4$_B$) $\quad Bp \to BBp$
(5$_B$) $\quad \neg Bp \to B \neg Bp.$

The first axiom states that one cannot believe a contradiction. The other two axioms require that belief is positively and negatively introspective. The logic generated by these axioms is conventionally called KD45, To emphasize that the logic is in $\mathcal{L}(B)$ we denote it by $(\mathrm{KD45})_B$.

The logics of knowledge we consider are subsets of the logic $(\mathrm{S5})_K$ in the language $\mathcal{L}(K)$, where S5 is the normal logic generated by the following three axioms:

(T$_K$) $\quad Kp \to p$
(4$_K$) $\quad Kp \to KKp$
(5$_K$) $\quad \neg Kp \to K \neg Kp.$

The logic $(\mathrm{S4})_K$ is the normal logic generated by T_K, 4_K. There are several logics of interest between $(\mathrm{S4})_K$ and $(\mathrm{S5})_K$. A key role is played here by the logic $(\mathrm{S4.4})_K$, which is generated by T_K and 4_K, and the following weakened version of 5_K:

(4.4$_K$) $\quad p \to (\neg Kp \to K \neg Kp).$

The following two axioms link knowledge and belief in logics that contain both modalities:

(L1) $\quad Kp \to Bp$
(L2) $\quad Bp \to KBp.$

We briefly review the semantics of frames and Kripke models. A *frame* $\mathcal{F}$ for the language $\mathcal{L}(M_1, \ldots, M_n)$ is a tuple $(W, R_1, \ldots, R_n)$, where W is a nonempty set of *possible worlds* (*worlds*, for short), and for each $i = 1, \ldots, n$, $R_i \subseteq W \times W$ is a binary relation on W, called the *accessibility relation* for the modality M_i. A *Kripke model* $\mathcal{M}$ based on the frame $\mathcal{F}$ is a pair $(\mathcal{F}, V)$, where $V \colon P \to 2^W$ is a *valuation* of the primitive propositions as subsets of W.

The function V is extended inductively to a *meaning* function $[\![\cdot]\!]_{\mathcal{M}}$ on all formulas. We omit the subscript $\mathcal{M}$ when it is clear from context. For each primitive formula p, $[\![p]\!] = V(p)$. For all formulas φ and ψ, $[\![\neg\varphi]\!] = \neg[\![\varphi]\!]$, where we abuse notation and use $\neg$ to denote set-theoretic complementation, $[\![\varphi \vee \psi]\!] = [\![\varphi]\!] \cup [\![\psi]\!]$, and $[\![M_i\varphi]\!] = \{x \mid R_i(x) \subseteq [\![\varphi]\!]\}$, where $R_i(x) = \{y \mid (x, y) \in R_i\}$.

We write $(\mathcal{M}, w) \models \varphi$ if $w \in [\![\varphi]\!]$. When $[\![\varphi]\!] = W$ we write $\mathcal{M} \models \varphi$ and say that φ is *valid in* $\mathcal{M}$. The formula φ is *valid in a frame* $\mathcal{F}$ if it is valid in each of the models based on $\mathcal{F}$. The set of formulas that are valid in a frame $\mathcal{F}$ is called the *theory* of $\mathcal{F}$, denoted $\mathrm{Th}(\mathcal{F})$.

For a class of frames $\mathcal{S}$, $\mathrm{Th}(\mathcal{S})$ is the set of formulas that are valid in each frame in $\mathcal{S}$. A logic Λ is *sound* for $\mathcal{S}$ if $\Lambda \subseteq \mathrm{Th}(\mathcal{S})$, and is *complete* for $\mathcal{S}$ if $\Lambda \supseteq \mathrm{Th}(\mathcal{S})$.

Let Λ be a logic. A frame $\mathcal{F}$ is said to be a Λ-frame if $\Lambda \subseteq \mathrm{Th}(\mathcal{F})$. The logics that we focus on in this paper are the logic of belief $(\mathrm{KD45})_B$, logics of knowledge that are subsets of $(\mathrm{S5})_K$, and logics of belief and knowledge that are subsets of $(\mathrm{KD45})_B + (\mathrm{S5})_K + \{\mathrm{L1}, \mathrm{L2}\}$.

In the sequel we use the following characterizations of Λ-frames for some logics Λ:

- A frame $\mathcal{F} = (W, R_B)$ is a $(\mathrm{KD45})_B$-frame iff R_B is serial, transitive, and Euclidean.[4]
- A frame $\mathcal{F} = (W, R_K)$ is an $(\mathrm{S4.4})_K$-frame iff R_K is reflexive, transitive, convergent, and remotely symmetric, where a relation R is *convergent* if $(w, w) \in R$ and $(w, y) \in R$ implies that there exists a world z such that $(x, z) \in R$ and $(y, z) \in R$, and R is *remotely symmetric* if $(x, y) \in R$ and $(y, z) \in R$ implies that $(z, y) \in R$ or $x = y$.
- A frame $\mathcal{F} = (W, R_K)$ is an $(\mathrm{S5})_K$-frame iff R_K is an equivalence relation.
- Let Λ_1 be the logic generated by L1. Then a frame (W, R_B, R_K) is a Λ_1-frame iff $R_B \subseteq R_K$.
- Let Λ_2 be the logic generated by L2. Then a frame (W, R_B, R_K) is a Λ_2-frame iff for all x, y, and z in W, if $(x, y) \in R_K$ and $(y, z) \in R_B$, then $(x, z) \in R_B$.

It is well known that for each of the logics Λ mentioned above, Λ is sound and complete with respect to Λ-frames [Georgacarakos 1976; Hoek 1993]. Moreover, if the logic Λ is generated by the union of several of the logics above, then a frame is a Λ-frame if and only if it satisfies the conditions above for each of the generating logics. In this case too, the logic Λ is sound and complete with respect to the family of Λ-frames.

[4] R is serial if for each x there exists a y such that $(x, y) \in R$; R is Euclidean if, for all x, y, and z, if $(x, y) \in R$ and $(x, z) \in R$ then $(y, z) \in R$.

3 Three notions of definability

Since we are interested in the extent to which knowledge can be defined in terms of belief, we need to clarify what we mean by "define". In this section, we review the three different notions of defining one modality in terms of others examined carefully in the companion paper, and restate the main results from that paper. Again, much of the text is taken verbatim from the companion paper.

Let δ be a formula in $\mathcal{L}(M_1, \ldots, M_{n-1})$. The formula

$$(\mathrm{DM}_n) \qquad M_n p \leftrightarrow \delta$$

is called a *definition of* M_n *(in terms of* $M_1, \ldots, M_{n-1}$*)*. Consider a logic Λ in the language $\mathcal{L}(M_1, \ldots, M_n)$.

> **Explicit definability:** M_n is *explicitly defined in* Λ if there is a definition DM_n of M_n such that $\mathrm{DM}_n \in \Lambda$.

To capture implicit definability in multimodal logic, let M'_n be a modal operator distinct from $M_1, \ldots, M_n$, and consider the language $\mathcal{L}(M_1, \ldots, M_n, M'_n)$. The logic $\Lambda[M_n/M'_n]$ is obtained by replacing all occurrences of M_n in formulas in Λ by M'_n.

> **Implicit definability:** M_n is *implicitly defined* in Λ if $M_n p \leftrightarrow M'_n p \in \Lambda + \Lambda[M_n/M'_n]$.

These definitions of explicit and implicit definability are obvious analogues of explicit and implicit definability of predicates in first-order logic, However, in first-order logic, Beth's celebrated theorem [Beth 1953] states that implicit and explicit definability coincide. In the context of modal logic, as we show by example (see Section 4), they do not.

To simplify notation, we henceforth take $\mathcal{L} = \mathcal{L}(M_1, \ldots, M_n)$, $\mathcal{L}_0 = \mathcal{L}(M_1, \ldots, M_{n-1})$, and $\Lambda_0 = \Lambda \cap \Lambda_0$. With this notation, explicit definability can be described by the inclusion $\Lambda_0 + \mathrm{DM}_n \subseteq \Lambda$.

The notion of reducibility, which we introduce next, seems to capture our intuition of defining knowledge in terms of belief better than the notion of explicit definability. When we define knowledge as true, justified belief, we do not expect this definition to follow from the logic that characterizes knowledge. We expect just the opposite: that the desired properties of knowledge follow from this definition when it is added to the logic of belief and justification. We get this effect by reversing the inclusion in the above description of explicit definability. Recall that a logic Λ in a language $\mathcal{L}$ is a *conservative extension* of a logic Λ' in a language $\mathcal{L}' \subseteq \mathcal{L}$ if $\Lambda' = \Lambda \cap \mathcal{L}'$.

Reducibility: M_n is *reducible* to $M_1, \ldots, M_{n-1}$ in Λ if there is a definition DM_n of M_n, such that $\Lambda \subseteq \Lambda_0 + \mathrm{DM}_n$, and $\Lambda_0 + \mathrm{DM}_n$ is a conservative extension of Λ_0. In this case, we say that M_n *is reducible to* $M_1, \ldots, M_{n-1}$ *by* DM_n.

The requirement that $\Lambda_0 + \mathrm{DM}_n$ be a conservative extension of Λ_0 guarantees that when Λ is consistent, then $\Lambda_0 + \mathrm{DM}_n$ is also consistent. But this requirement is needed also to ensure that the definition DM_n does not affect the operators $M_1, \ldots, M_{n-1}$. Without this requirement it is possible that the definition "sneaks in" extra properties of the defining modalities. We demonstrate this and further discuss reducibility in Section 5.

The main result of the companion paper shows that the three notions that we have considered are intimately connected: explicit definability is equivalent to the combination of implicit definability and reducibility; this is formalized in Theorem 3.1 below. Examples in Section 4 show that explicit definability is not equivalent to implicit definability or reducibility separately. Thus, the situation in modal logic is quite different from that in first-order logic, where implicit and explicit definability coincide.

Theorem 3.1 *[Halpern, Samet, and Segev 2008] The modal operator M_n is explicitly defined in Λ if and only if M_n is implicitly defined and reducible to $M_1, \ldots, M_{n-1}$ in Λ.*

4 Definability in epistemic-doxastic logics

In this section we study logics in the language $\mathcal{L}(B, K)$ that are subsets of $(\mathrm{KD45})_B + (\mathrm{S5})_K + \{\mathrm{L1}, \mathrm{L2}\}$, and apply the three notions of definability from Section 3 to the modalities B and K.

4.1 Defining knowledge in terms of belief

We show that knowledge cannot be defined explicitly in terms of belief in any of the logics in which we are interested.

Theorem 4.1 *The modality K cannot be defined explicitly in the logic* $(\mathrm{KD45})_B + (\mathrm{S5})_K + \{\mathrm{L1}, \mathrm{L2}\}$*, and hence not in any logic contained in it.*

In contrast to the universality of this theorem, the other two notions of definability depend on whether we consider S5-knowledge or S4.4. We first consider epistemic logics contained in S4.4.

4.1.1 Definability of K for S4.4 and weaker logics

True belief is the most natural candidate for a definition of knowledge in terms of belief. It is universally accepted among epistemologists that knowledge cannot be thus defined. However, as Theorem 4.2 below shows, provided we accept the linking axioms (which seem fairly noncontroversial), for all logics contained in S4.4, knowledge can be reduced to belief by defining it as true belief. This would suggest that none of these logics is a good candidate for the "true" logic of knowledge.

Consider the definition of knowledge as true belief:

(TB) $\quad Kp \leftrightarrow (p \wedge Bp).$

Theorem 4.2 *For all logics Λ such that $(\mathrm{KD45})_B \subseteq \Lambda \subseteq (\mathrm{KD45})_B + (\mathrm{S4.4})_K + \{\mathrm{L1}, \mathrm{L2}\}$, the knowledge modality K is reducible to B in Λ by* TB. *Moreover, if $\Lambda = (\mathrm{KD45})_B + \Lambda' + \{\mathrm{L1}, \mathrm{L2}\}$ for $\Lambda' \subseteq (\mathrm{S5})_K$, then K is reducible to B in Λ by* TB *if and only if $\Lambda' \subseteq (\mathrm{S4.4})_K$.*

It follows from Theorem 4.2 that $(\mathrm{S4.4})_K \subseteq (\mathrm{KD45})_B + \mathrm{TB}$. Corollary A.8 (proved in the appendix) shows that $(\mathrm{KD45})_B + \mathrm{TB}$ is a conservative extension of $(\mathrm{S4.4})_K$. Thus, $((\mathrm{KD45})_B + \mathrm{TB}) \cap \mathcal{L}(K) \subseteq (\mathrm{S4.4})_K$. Therefore, $((\mathrm{KD45})_B + \mathrm{TB}) \cap \mathcal{L}(K) = (\mathrm{S4.4})_K$, which means that S4.4 is the logic of knowledge defined as true belief.

Applying the equivalence in Theorem 3.1 to Theorems 4.1 and 4.2, we immediately obtain the following conclusion:

Theorem 4.3 *The modality K is not implicitly defined in $(\mathrm{KD45})_B + (\mathrm{S4.4})_K + \{\mathrm{L1}, \mathrm{L2}\}$, and hence also not in any logic contained in this logic.*

4.1.2 Definability of K for S5

The definability properties of K in the logic of belief and S5-knowledge are the opposite to those described in Theorems 4.2 and 4.3.

Theorem 4.4 *The modality K is implicitly defined in the logic $(\mathrm{KD45})_B + (\mathrm{S5})_K + \{\mathrm{L1}, \mathrm{L2}\}$.*

By Theorem 3.1, explicit definability is equivalent to implicit definability and reducibility. Thus, the following result, which stands in contrast to Theorem 4.2, follows immediately from Theorems 4.1 and 4.4.

Theorem 4.5 *The modality K is not reducible to B in the logic* $(\mathrm{KD45})_B + (\mathrm{S5})_K + \{L1, L2\}$.

In the appendix (see Proposition A.2) we show that there is a unique way to extend a Kripke model for KD45 belief to a Kripke model for S5 knowledge. It may thus seem somewhat surprising that S5 knowledge is not reducible to belief. As we show in the companion paper, this apparent disconnect can be explained by considering *(modal) algebras*, which provide a more general approach to giving semantics to modal logic than Kripke models [Blackburn, Rijke, and Venema 2001]. In particular, we show (see Theorem 5.4 in [Halpern, Samet, and Segev 2008]) that there is an algebraic model for $(\mathrm{KD45})_B$ that cannot be extended to an algebraic model for S5 knowledge.

Table 1 summarizes our main results about defining knowledge in terms of belief.

	reduction of knowledge to belief	implicit definition of knowledge	explicit definition of knowledge
$(\mathrm{KD45})_B + (\mathrm{S5})_K + \{\mathrm{L1}, \mathrm{L2}\}$	$-$	$+$	$-$
$(\mathrm{KD45})_B + (\mathrm{S4.4})_K + \{\mathrm{L1}, \mathrm{L2}\}$	$+$ (by TB)	$-$	$-$

Table 1: Definability of knowledge by belief.

4.2 Justification

While we have shown that S5 knowledge is not reducible to belief, our results do not preclude the possibility that it is reducible to belief and justification in some logic of these modalities. Note that, if it is, then it follows from Theorems 3.1 and 4.4 that S5 knowledge would then be explicitly defined in terms of belief and justification.

With no constraints, knowledge is trivially reducible to belief and justification. We simply assume that the modal operator J (for justification) satisfies all the S5 axioms, and that the axioms L1 and L2 hold with K replaced by J. If Λ is the resulting logic, then K is reducible to J in Λ via the definition $Kp \leftrightarrow Jp$.

In the companion paper we show that, roughly speaking, if the interaction between B and J is rather weak, then knowledge cannot be reduced to a combination of belief and justification. We provide both a semantic and more syntactic characterization of "weak interaction". The semantic characterization involves algebras. The syntactic characterization is relevant for the results of this paper, so we restate it.

Theorem 4.6 *Let Λ be a logic in $\mathcal{L}(B,J,K)$ such that $\Lambda \cap \mathcal{L}(B,J) = (\mathrm{KD45})_B + \Lambda_J$, where $\Lambda_J \subseteq \mathrm{S5}_J$. Then K is not reducible to B and J in Λ.*

The "weak interaction" between B and J is captured by saying that the axioms for B and J can be "decomposed" into axioms for B ($(\mathrm{KD45})_B$) and axioms for J (which are contained in $\mathrm{S5}_J$). Using Theorem 4.6, we can prove the following result.

Corollary 4.7 *Let $\Lambda = (\mathrm{KD45})_B + \Lambda_J + (\mathrm{S5})_K + \{\mathrm{L1}, \mathrm{L2}\}$ be a logic in $\mathcal{L}(B,J,K)$, where $\Lambda_J \subseteq (\mathrm{S5})_J$ is a logic in $\mathcal{L}(J)$. Then K is not reducible to B and J in Λ.*

These results suggest that if S5 knowledge is defined in terms of true justified belief, then there must be some interaction between justification and belief.

The philosophical interest in Corollary 4.7 depends in large part on what are viewed as reasonable characteristics of justification. We are not aware of work on logics of justification in the context of epistemic logic beyond that of Artemov and Nogina [2005]. Artemov and Nogina do not consider a logic with a J operator, but instead have formulas of the form $t : \varphi$, which can be read "φ is known for reason t". The term t can be thought of as a justification for φ. If we define $J\varphi$ to hold if there exists a t such that $t : \varphi$ holds, then Artemov and Nogina's axioms imply that J satisfies S4. However, Artemov and Nogina require that $J\varphi \rightarrow K\varphi$, so their notion of justification is stronger than the one we have implicitly been considering. It remains to be seen whether there is a logic of justification that captures the more standard philosophical intuitions and allows knowledge to be defined in terms of belief and justification; of course, this means that J cannot satisfy the hypotheses of Corollary 4.7.

4.3 Defining belief in terms of knowledge

Although the focus in the literature has been on defining knowledge in terms of belief, we can ask the same questions about defining belief in terms of knowledge. Somewhat surprisingly, we get results that parallel those for defining knowledge in terms of belief, in particular the following analogues of Theorems 4.1 and 4.2.

Theorem 4.8 *The modality B cannot be defined explicitly in the logic* $(\mathrm{KD45})_B + (\mathrm{S5})_K + \{\mathrm{L1}, \mathrm{L2}\}$*, and hence not in any logic contained in it.*

To study reducibility of belief to knowledge, we consider the definition of belief as possible knowledge given by the formula

(PK) $\quad Bp \leftrightarrow \neg K \neg Kp.$

The following proposition, due to Lenzen [1979], can be easily verified by showing that TB and all the axioms of KD45 are provable in $(\mathrm{S4.4})_K$ + PK and that PK and all the axioms of S4.4 are provable in $(\mathrm{KD45})_B$+TB.

Proposition 4.9 $(\mathrm{KD45})_B$ +TB = $(\mathrm{S4.4})_K$ + PK.

By Lenzen's definition, this equality establishes that the logics KD45 and $(\mathrm{S4.4})_K$ are *synonymous.* As suggested above, Proposition 4.9 can be viewed as saying four things: (a) PK $\in (\mathrm{KD45})_B$ +TB, (b) $(\mathrm{S4.4})_K \subseteq (\mathrm{KD45})_B$+TB, (c) TB $\in (\mathrm{S4.4})_K$ + PK, and (d) $(\mathrm{KD45})_B \subseteq (\mathrm{S4.4})_K$ + PK. The first part of Theorem 4.2 can be viewed as a strengthening of (b), since it shows that $(\mathrm{S4.4})_K$ + {L1, L2} $\subseteq (\mathrm{KD45})_B$ +TB, and that $(\mathrm{KD45})_B$ +TB is a conservative extension of $(\mathrm{KD45})_B$. The following result strengthens (d) in the corresponding way.

Theorem 4.10 *For all logics Λ such that* $(\mathrm{S4.4})_K \subseteq \Lambda \subseteq (\mathrm{KD45})_B + (\mathrm{S4.4})_K + \{\mathrm{L1}, \mathrm{L2}\}$*, the belief modality B is reducible to K in Λ by* PK.

The equivalence in Theorem 3.1 combined with the previous two theorems immediately implies the following:

Theorem 4.11 *The modality B is not defined implicitly in the logic* $(\mathrm{KD45})_B + (\mathrm{S5})_K + \{\mathrm{L1}, \mathrm{L2}\}$*, and hence not in any logic contained in* $(\mathrm{KD45})_B + (\mathrm{S5})_K + \{\mathrm{L1}, \mathrm{L2}\}$.

5 A closer look at reducibility

Our requirement in the definition of reducibility that $\Lambda_0 + \mathrm{DM}_n$ be a conservative extension of Λ_0 has no analogue in the work of Lenzen [1979] or Pelletier and Urquhart [2003]. One consequence of requiring that $\Lambda_0 + \mathrm{DM}_n$

be a conservative extension of Λ_0 is proved in the companion paper (see Proposition 3.1); we restate the result here. A definition of a modality M_n by the formula $M_n p \leftrightarrow \delta$ is *simple* if δ contains no primitive propositions other than p.

Proposition 5.1 *[Halpern, Samet, and Segev 2008] If M_n is reducible to $M_1, \ldots, M_{n-1}$ in Λ, then it is reducible by a simple definition.*

In the next subsection we consider another consequence of requiring conservative extensions; in the second subsection we explain why it is an arguably necessary requirement; finally, in the last subsection, we consider reducibility when logics are not required to be normal.

5.1 Consistency

It is easy to see that if Λ is inconsistent, then $\mathcal{M}_n$ is reducible to $\mathcal{M}_n, \ldots, M_{n-1}$ in Λ by any definition. If Λ is consistent, which is the case of interest, then Λ_0 is also consistent, but for some definitions DM_n, $\Lambda_0 + \mathrm{DM}_n$ may be inconsistent, in which case the inclusion $\Lambda \subseteq \Lambda_0 + \mathrm{DM}_n$ trivially holds. As the following example shows, without requiring that $\Lambda_0 + \mathrm{DM}_n$ be a conservative extension of Λ_0, for *every* logic Λ we can find such undesired definitions.

Example 5.2 Consider the definition DM_n given by the formula $M_n p \leftrightarrow \neg true$. Substituting *true* for p, we conclude that $M_n true \leftrightarrow \neg true \in \Lambda_0 + \mathrm{DM}_n$. By normality, $M_n true \leftrightarrow true \in \Lambda_0 + \mathrm{DM}_n$. Thus, $\Lambda_0 + \mathrm{DM}_n$ is inconsistent, and hence $\Lambda \subseteq \Lambda_0 + \mathrm{DM}_n$. ∎

Obviously, if Λ_0 is consistent, then $\Lambda_0 + \mathrm{DM}_n$ is guaranteed to be consistent if it is a conservative extension of Λ_0. But we can prevent the undesirable situation described in Example 5.2 just by requiring that $\Lambda_0 + \mathrm{DM}_n$ be consistent, rather than requiring that it be a conservative extension of Λ_0. This leads to the following definition.

> **Weak reducibility:** M_n is *weakly reducible* to $M_1, \ldots, M_{n-1}$ in Λ if either Λ is inconsistent or if there is a definition DM_n of M_n such that $\Lambda \subseteq \Lambda_0 + \mathrm{DM}_n$, and $\Lambda_0 + \mathrm{DM}_n$ is consistent.

5.2 Why require a conservative extension?

Weak reducibility does not suffice to prevent the defining formula from adding extra properties to the defining modalities, even if we use a simple definition. The key point is that the requirement that $\Lambda_0+\mathrm{DM}_n$ be a *normal* logic means that $\Lambda_0 + \mathrm{DM}_n$ must contain all substitution instances of the axiom K_{M_n}, $M_n(p \rightarrow q) \rightarrow (M_n p \rightarrow M_n q)$, and the formula $M_n \mathit{true}$. Let Φ_δ consist of all substitution instances of $\delta[p/(p \rightarrow q)] \rightarrow (\delta \rightarrow \delta[p/q])$, and the formula $\delta[p/\mathit{true}]$. Since $\Lambda_0+\mathrm{DM}_n$ contains $M_n p \leftrightarrow \delta$, it clearly must contain Φ_δ. (For if ψ is a substitution instance of $\delta[p/(p \rightarrow q)] \rightarrow (\delta \rightarrow \delta[p/q])$, let ψ' be the corresponding substitution instance of $M_n(p \rightarrow q) \rightarrow (M_n p \rightarrow M_n q)$. Since Λ_0 is normal, $\psi' \in \Lambda_0$. By applying DM_n and propositional reasoning repeatedly, it easily follows that $\psi \in \Lambda_0 + \mathrm{DM}_n$.) Thus, $\Lambda_0 + \mathrm{DM}_n$ cannot be a conservative extension of Λ_0 unless $\Phi_\delta \subseteq \Lambda_0$. As the following result shows, this is also a sufficient condition for $\Lambda_0 + \mathrm{DM}_n$ to be a conservative extension of Λ_0. Therefore, if $\Lambda_0 + \mathrm{DM}_n$ is not a conservative extension of Λ_0, then Φ_δ is not a subset of Λ_0, but is a subset of $\Lambda_0 + \mathrm{DM}_n$. Hence, adding DM_n to Λ_0 results in new formulas, those in Φ_δ which are not in Λ_0, that are satisfied by the modalities $M_1, \ldots, M_{n-1}$.

Proposition 5.3 *If* DM_n *is a simple definition, then* $\Lambda_0 + \mathrm{DM}_n$ *is a conservative extension of* Λ_0 *iff* $\Phi_\delta \subseteq \Lambda_0$.

Proposition 5.3 follows from a characterization of $\Lambda_0 + \mathrm{DM}_n$, proved as Proposition A.10 in the appendix.

Using these insights, we can show that knowledge is weakly reducible to belief by a simple definition. This is done by a definition of K that "redefines" B to be an S5 knowledge operator in $(\mathrm{KD45})_B$ + DK.

Theorem 5.4 *The modality* K *is weakly reducible to* B *in the logic* $(\mathrm{KD45})_B+ (\mathrm{S5})_K + \{\mathrm{L1}, \mathrm{L2}\}$ *by the simple definition*

$$\mathrm{DK} = Kp \leftrightarrow ((p \wedge Bp) \vee (\neg p \wedge Bp) \vee (p \wedge B\neg p)).$$

To us, Theorem 5.4 suggests that weak reducibility is an inappropriate notion, and is the major reason that we added the requirement that $\Lambda_0+\mathrm{DM}_n$ be a conservative extension of Λ_0 to the definition of reducibility.

5.3 Normality

Given the power of the assumption of normality, we conclude by considering what happens if we drop it. Let $\Lambda \oplus \mathrm{DM}$ denote the smallest logic (not necessarily normal) that contains Λ and DM; We can define the notions of (weak) reducibility$'$ just by replacing $+$ by $\oplus$ in the definition of (weak) reducibility.

The analogue of Proposition 5.1 holds with no change in proof for reducibility$'$; thus, for reducibility$'$ we can consider simple definitions without loss of generality. This is not the case for weak reducibility$'$. The next result shows that K is not weakly reducible$'$ to B by a simple definition, and hence not reducible$'$ to belief, but K is weakly reducible$'$ to B by a non-simple definition.

Theorem 5.5 *The modality K is not weakly reducible$'$ to B by a simple definition in the logic* $(\mathrm{KD45})_B + (\mathrm{S5})_K + \{\mathrm{L1}, \mathrm{L2}\}$*, and hence not reducible$'$ to B. However, K is weakly reducible$'$ to B in* $(\mathrm{KD45})_B + (\mathrm{S5})_K + \{\mathrm{L1}, \mathrm{L2}\}$ *by the definition $Kp \leftrightarrow (Bp \wedge (Bq \rightarrow q))$.*

6 Epistemic luck, negative introspection, and knowledge as true belief

When knowledge is defined as true belief, the difference between knowledge and belief is completely external; even externalists would reject such a definition of knowledge. According to Theorem 4.2, all the logics of belief and knowledge where the axioms of knowledge are contained in $(\mathrm{S4.4})_K$ can be derived from the definition of knowledge as true belief. Thus, such logics fail to express an internal aspect of knowledge that distinguishes it from true belief.

A different objection to defining knowledge as true belief was raised by Foley [1984]. He used the term "epistemic luck" to describe the case that a believed sentence φ turns out to be true. Since knowledge should be due to more than just luck, chance, or serendipity, true belief, that is, belief that turns out to be true by mere luck, should not count as knowledge. Several authors tried to explicate and analyze epistemic luck (see [Pritchard 2005] and the references therein). For example, Pritchard [2004] suggests an explanation of epistemic luck in the framework of the semantics of possible worlds. His analysis makes use of the distance between possible worlds. We now suggest a simpler explanation of epistemic luck in terms of the logic of belief adopted.

We start by defining the opposite of epistemic luck. We say that the formula $\neg\varphi \wedge B\varphi$ represents *epistemic misfortune* with respect to φ. If $\neg\varphi \wedge B\varphi$ happen to be true, the believing agent is epistemically unfortunate.

It might seem reasonable to say that $\varphi \wedge B\varphi$ represents epistemic luck with respect to φ, but whether this is indeed reasonable depends on φ and on the logic of belief we adopt, as the following examples show.

If we adopt a logic of belief in which tautologies are always believed, then believing a tautology φ should not count as epistemic luck. No luck is involved in believing something one must believe. Similarly, if the logic of belief we adopt is KD45, then believing $B\psi$ should not count as epistemic luck. This belief is the result of self-introspection, which is assumed in KD45 to always be correct. But believing $K\psi$ *does* involve epistemic luck; this belief can be incorrect. Roughly speaking, *there is luck only when there is room for misfortune*, that is, if it is consistent to believe φ even if φ is false. This suggests the following definition of epistemic luck:

> The formula $\varphi \wedge B\varphi$ represents *epistemic luck* with respect to φ in logic Λ if $\neg\varphi \wedge B\varphi$ is consistent in Λ.

The difference between the first two examples and the third should now be clear. If φ is a tautology or of the form $B\psi$, then $\neg\varphi \wedge B\varphi$ is inconsistent (in the case of $B\psi$, this follows from 5_B and D_B). On the other hand, $\neg K\varphi \wedge BK\varphi$ is consistent.

In order to properly define knowledge in terms of belief, we need to get rid of epistemic luck. Defining knowledge as true justified belief can be viewed as eliminating, or at least weakening, the luck element by requiring justification. But we can achieve the same objective directly without adding justification. We start with a KD45 belief modality, B. In order to make this belief modality a knowledge modality we add an axiom that ensures that epistemic misfortune is logically impossible. But requiring that epistemic misfortune is logically impossible amounts to saying that $\neg(\neg\varphi \wedge B\varphi)$, or equivalently $B\varphi \rightarrow \varphi$, is valid. Adding this axiom turns the belief modality B into an S5 knowledge modality. We can summarize this as follows:

> An S5-knowledge modality is a KD45-belief modality for which no formula represents epistemic luck.

These observations also shed some light on the issue of negative introspection. A standard intuitive argument against the negative introspection axiom 5_K is the following. An agent may not know φ, but still believe that he knows φ, and therefore is unable to know that he does not know φ. We can require this to be the only case that 5_K fails to hold by including the following axiom in the logic: $\neg(\neg Kp \rightarrow K\neg Kp) \leftrightarrow (\neg KP \wedge BKp)$.

A straightforward argument, whose proof we omit, shows that this axiom already follows once we define knowledge as true belief:

Proposition 6.1 $\neg(\neg Kp \to K\neg Kp) \leftrightarrow (\neg Kp \wedge BKp) \in (\mathrm{KD45})_B$+TB.

But the hollow meaning of knowledge in $(\mathrm{KD45})_B$ + TB also renders this explanation of the failure of 5_K hollow. Indeed, it is easy to see that $(\neg Kp \wedge BKp) \leftrightarrow (\neg p \wedge Bp)$ is also a theorem of $(\mathrm{KD45})_B$ + TB. Thus, in $(\mathrm{KD45})_B$ + TB, this argument for the failure of 5_K is tantamount to epistemic failure. In summary, the intuitive explanation of why 5_K should not hold does not capture the internal nature of knowledge.

Although knowledge as true belief does not imply negative introspection, it is consistent with it; that is, $(\mathrm{KD45})_B$ + TB + 5_K is consistent. However, it follows easily from Proposition 6.1 and the discussion above that $(\mathrm{KD45})_B$ + TB + 5_K implies $Bp \to p$. That is, this set of axioms precludes false beliefs, and makes B a knowledge operator. Moreover, it is immediate that both K and B satisfy L1 and L2 with respect to B, that is, L1 and L2 hold both as stated and when we replace K by B. Thus, the following result follows easily from Theorem 4.4.

Proposition 6.2 $Bp \leftrightarrow Kp \in (\mathrm{KD45})_B + \mathrm{TB} + 5_K$.

By assuming both that knowledge is true belief and that knowledge satisfies 5_K we give up the distinction between knowledge and belief, and hence also the content of the definition of knowledge as true belief.

A Proofs

For the proof of Theorem 4.1, we need the following definition. Given an algebraic model $\mathcal{M}$ for a language $\mathcal{L}$, let $[\![\mathcal{L}]\!]_{\mathcal{M}} = \{[\![\varphi]\!]_{\mathcal{M}} \mid \varphi \in \mathcal{L}\}$.

Theorem 4.1: *The modality K cannot be defined explicitly in the logic* $(\mathrm{KD45})_B + (\mathrm{S5})_K + \{\mathrm{L1}, \mathrm{L2}\}$*, and hence not in any logic contained in it.*

Proof: If K is defined explicitly in $(\mathrm{KD45})_B + (\mathrm{S5})_K + \{\mathrm{L1}, \mathrm{L2}\}$ via the definition $Kp \leftrightarrow \delta$, then for every model $\mathcal{M}$ of this logic, $[\![Kp]\!]_{\mathcal{M}} = [\![\delta]\!]_{\mathcal{M}}$, and hence $[\![Kp]\!]_{\mathcal{M}} \in [\![\mathcal{L}(B)]\!]_{\mathcal{M}}$. We prove the theorem by constructing a model $\mathcal{M}$ such that $[\![Kp]\!]_{\mathcal{M}} \notin [\![\mathcal{L}(B)]\!]_{\mathcal{M}}$.

Consider the $((\mathrm{KD45})_B + (\mathrm{S5})_K + \{\mathrm{L1}, \mathrm{L2}\})$-frame $\mathcal{F} = (W, R_B, R_K)$, where $W = \{w_1, w_2, w_3\}$, $R_B = \{(w_1, w_1), (w_2, w_2), (w_3, w_2)\}$, and $R_K = R_B \cup \{(w_2, w_3), (w_3, w_3)\}$. Let $\mathcal{M} = (\mathcal{F}, V)$ be the model based on $\mathcal{F}$

such that V maps each primitive proposition to $\{w_1, w_2\}$. It is easy to show by induction on the structure of formulas in $\mathcal{L}(B)$ that $[\![\mathcal{L}(B)]\!]_{\mathcal{M}} = \{\emptyset, W, \{w_1, w_2\}, \{w_3\}\}$, but $[\![Kp]\!]_{\mathcal{M}} = \{w_1\}$. ∎

To prove Theorems 4.2, 4.4, and 4.8, we need a number of results concerning algebraic extensions and conservative extensions.

Lemma A.1 *If $\mathcal{L}_1 \subseteq \mathcal{L}_2$, $\Lambda_1 \subseteq \Lambda_2$ are two logics in the corresponding languages such that Λ_2 is a conservative extension of Λ_1, and Λ' is a logic such that $\Lambda_1 \subseteq \Lambda' \subseteq \Lambda_2$, then Λ' is a conservative extension of Λ_1.*

Proof: If Λ_2 is a conservative extension of Λ_1. then $\Lambda_1 = \Lambda_2 \cap \mathcal{L}_1$. Since $\Lambda_1 = \Lambda_1 \cap \mathcal{L}_1 \subseteq \Lambda' \cap \mathcal{L}_1 \subseteq \Lambda_2 \cap \mathcal{L}_1 = \Lambda_1$, it follows that $\Lambda_1 = \Lambda' \cap \mathcal{L}_1$, Thus, Λ' is a conservative extension of Λ_1. ∎

Proposition A.2 *Every* $(\mathrm{KD45})_B$*-frame can be uniquely extended to a (*$(\mathrm{KD45})_B+$ $(\mathrm{S5})_K+$ $\{\mathrm{L1}, \mathrm{L2}\}$*-frame.*

Proof: Let (W, R_B) be a $(\mathrm{KD45})_B$-frame. Define R_K so that $(x, y) \in R_K$ iff there exists a world z such that $(x, z) \in R_B$ and $(y, z) \in R_B$. Obviously, R_K is reflexive and symmetric. To show transitivity, suppose that $(x, y) \in R_K$ and $(y, z) \in R_K$. Then there exist w and w' such that $(x, w) \in R_B$, $(y, w) \in R_B$, $(y, w') \in R_B$, and $(z, w') \in R_B$. Since $(y, w) \in R_B$ and $(y, w') \in R_B$, it follows from the Euclidean property that $(w', w) \in R_B$. By the transitivity of R_B, we have $(z, w) \in R_B$. Since $(x, w) \in R_B$ by assumption, it follows from the definition of R_K that $(x, z) \in R_K$, as desired.

To prove that $R_B \subseteq R_K$, note that if $(x, y) \in R_B$ then, by the Euclidean property, $(y, y) \in R_B$, and hence $(x, y) \in R_K$. If $(x, y) \in R_K$ and $(y, z) \in R_B$, then there exists a world w such that $(x, w) \in R_B$ and $(y, w) \in R_B$. By the Euclidean property, $(w, z) \in R_B$, and by transitivity, $(x, z) \in R_B$. To show that the property corresponding to L2 holds, suppose that $(x, y) \in R_K$ and $(y, z) \in R_B$. Then, by definition, there exists some w such that $(x, w) \in R_B$ and $(w, y) \in R_B$. By the transitivity of R_B, it follows that $(x, z) \in R_B$, as desired. This completes the proof that (W, R_B, R_K) is a ($(\mathrm{KD45})_B$ + $(\mathrm{S5})_K$ + {L1,L2})-frame.

Suppose that R'_K is another relation such that (W, R_B, R'_K) is a (KD45 + S5 + {L1,L2})-frame. Then, $R_K \subseteq R'_K$. Indeed, if $(x, z) \in R_B$ and $(y, z) \in R_B$, then $(x, z) \in R'_K$ and $(y, z) \in R'_K$, since $R_B \subseteq R'_K$. By the symmetry and transitivity of R'_K, $(x, y) \in R'_K$. To show that $R'_K \subseteq R_K$, suppose that $(x, y) \in R'_K$. By seriality, for some world z, $(y, z) \in R_B$, and hence, by the property corresponding to L2, we must have $(x, z) \in R_B$.

Thus, by the definition of R_K, $(x, y) \in R_K$. This completes the uniqueness part of the theorem. ∎

The following lemma, proved in the companion paper, will be useful in our later results:

Lemma A.3 *[Halpern, Samet, and Segev 2008] If $\mathcal{L}_1 \subseteq \mathcal{L}_2$, $\Lambda_1 \subseteq \Lambda_2$ are two logics in the corresponding languages such that Λ_1 is sound and complete for a family $\mathcal{S}$ of frames, and each frame in $\mathcal{S}$ can be extended to a Λ_2-frame, then Λ_2 is a conservative extension of Λ_1.*

The following corollary is immediate from Proposition A.2 and Lemmas A.3 and A.1.

Corollary A.4 *If $(\mathrm{KD45})_B \subseteq \Lambda \subseteq (\mathrm{KD45})_B + (\mathrm{S5})_K + \{\mathrm{L1}, \mathrm{L2}\}$, then Λ is a conservative extension of $(\mathrm{KD45})_B$.*

Thus, the logics of most interest to us here are conservative extensions of KD45. In particular, for each such logic Λ, the logic Λ_0 in the definition of reducibility is KD45.

Proposition A.5 *Every $(\mathrm{KD45})_B$-frame can be extended to a $((\mathrm{KD45})_B + \mathrm{TB})$-frame.*

Proof: Let (W, R_B) be a $(\mathrm{KD45})_B$-frame. Define R_K to be the reflexive closure of R_B, i.e., $R_K = R_B \cup \{(x,x) \colon x \in W\}$. If $\mathcal{M}$ is a model based on (W, R_B, R_K), then $x \in [\![K\varphi]\!]_{\mathcal{M}}$ iff $R_K(x) \subseteq [\![\varphi]\!]_{\mathcal{M}}$, which is equivalent to $R_B(x) \cup \{x\} \subseteq [\![\varphi]\!]_{\mathcal{M}}$, which holds iff $x \in [\![B\varphi]\!]_{\mathcal{M}} \cap [\![\varphi]\!]_{\mathcal{M}} = [\![\varphi \wedge B\varphi]\!]_{\mathcal{M}}$. ∎

Applying Lemma A.3, we get the following corollary.

Corollary A.6 *The logic $(\mathrm{KD45})_B + \mathrm{TB}$ is a conservative extension of $(\mathrm{KD45})_B$.*

Proposition A.7 *Every $(\mathrm{S4.4})_K$-frame can be extended to an $((\mathrm{S4.4})_K + (\mathrm{KD45})_B + \mathrm{TB} + \{\mathrm{L1,L2}\})$-frame.*

Proof: Let (W, R_K) be an $(\mathrm{S4.4})_K$-frame. Define

$$R_B = R_K \setminus \{(x,x) \colon \exists y \text{ such that } (x,y) \in R_K \text{ but } (y,x) \notin R_K\}.$$

We now show that R_B is Euclidean, transitive, and serial.

We claim that

(1) if there exists some $z \neq x$ such that $(z, x) \in R_K$, then $(x, x) \in R_B$.

To prove (1), suppose that the antecedent holds. Since R_K is reflexive, we have $(x, x) \in R_K$. Suppose that $(x, y) \in R_K$. If $y = x$, then clearly $(y, x) \in R_K$. If $y \neq x$, then by remote symmetry and the fact that $(z, x) \in R_K$, we must have $(y, x) \in R_K$. Thus, by the definition of R_B, $(x, x) \in R_B$, as desired.

To see that R_B is transitive, suppose that $(x, y) \in R_B$ and $(y, z) \in R_B$. If $y = x$, then $(x, z) \in R_B$ is immediate. So suppose that $x \neq y$. Since $(x, y) \in R_B$ and $(y, z) \in R_B$, we must have $(x, y) \in R_K$ and $(y, z) \in R_K$. Since R_K is transitive, $(x, z) \in R_K$. By definition, $(x, z) \in R_B$ if $z \neq x$. But if $z = x$, then we have $(y, z) \in R_K$, so it is immediate from (1) that $(x, x) \in R_B$. So in either case, $(x, z) \in R_B$, as desired. R_B is serial because, for each x, either $(x, x) \in R_B$ or, for some $y \neq x$, $(x, y) \in R_K$, and thus $(x, y) \in R_B$. Finally, to show the R_B is Euclidean, suppose that $(x, y) \in R_B$ and $(x, z) \in R_B$. We show that $(y, z) \in R_B$ by considering the following exhaustive list of cases.

- $x = y$: Then it is immediate that $(y, z) \in R_B$.
- $x \neq y$, $x = z$: Thus, $(x, x) \in R_B$. Since $(x, y) \in R_B$, we must have $(x, y) \in R_K$. We must also have $(y, x) \in R_K$, otherwise, by the definition of R_B, we would not have $(x, x) \in R_B$. Since $y \neq x$, we have $(y, x) \in R_B$. Since $x = z$, we have $(y, z) \in R_B$, as desired.
- $x \neq y$, $x \neq z$, $y = z$: Since $(x, y) \in R_K$ and $x \neq y$, by (1), we must have $(y, y) \in R_B$, as desired.
- $x \neq y$, $x \neq z$, $y \neq z$: By convergence, there is some w such that $(y, w) \in R_K$ and $(z, w) \in R_K$. By remote symmetry, $(w, z) \in R_K$; by transitivity, $(y, z) \in R_K$. Since $y \neq z$, it follows that $(y, z) \in R_B$.

Thus, we have shown that (W, R_B, R_K) is a $(\mathrm{KD45})_B$-frame. Since R_K is reflexive, it follows that $R_K = R_B \cup \{(x, x) : x \in W\}$. As shown in the proof of Proposition A.5, (W, R_B, R_K) is a frame for the logic generated by TB. As $R_B \subseteq R_K$, it is also a frame for the logic generated by L1. Finally, we need to show that it is a frame for the logic generated by L2. Suppose that $(x, y) \in R_K$ and $(y, z) \in R_B$. We must show that $(x, z) \in R_B$. Since $R_B \subseteq R_K$, we must have $(y, z) \in R_K$. Thus, by the transitivity of R_K, $(x, z) \in R_K$. If $x \neq z$, then it is immediate that $(x, z) \in R_B$. But if $x = z$, then the result is immediate from (1). ∎

The following two corollaries are immediate from Proposition A.7 and Lemma A.3.

Corollary A.8 *The logic* $(\mathrm{KD45})_B$ + TB *is a conservative extension of* $(\mathrm{S4.4})_K$*.*

Corollary A.9 *The logic* $(\mathrm{KD45})_B + (\mathrm{S4.4})_K + \{\mathrm{L1}, \mathrm{L2}\}$ *is a conservative extension of* $(\mathrm{S4.4})_K$*.*

We are finally ready to prove Theorem 4.2.

Theorem 4.2: *For all logics* Λ *such that* $(\mathrm{KD45})_B \subseteq \Lambda \subseteq (\mathrm{KD45})_B + (\mathrm{S4.4})_K + \{\mathrm{L1}, \mathrm{L2}\}$*, the knowledge modality* K *is reducible to* B *in* Λ *by* TB*. Moreover, if* $\Lambda = (\mathrm{KD45})_B + \Lambda' + \{\mathrm{L1}, \mathrm{L2}\}$ *for* $\Lambda' \subseteq (\mathrm{S5})_K$*, then* K *is reducible to* B *in* Λ *by* TB *if and only if* $\Lambda' \subseteq (\mathrm{S4.4})_K$*.*

Proof: Let Λ satisfy the inclusions in the theorem. By Corollary A.4, $\Lambda \cap \mathcal{L}(B) = (\mathrm{KD45})_B$. By Corollary A.6, $(\mathrm{KD45})_B$ + TB is a conservative extension of $(\mathrm{KD45})_B$. Thus, it remains to show that $\Lambda \subseteq (\mathrm{KD45})_B + \mathrm{TB}$. Since $\Lambda \subseteq (\mathrm{KD45})_B + (\mathrm{S4.4})_K + \{\mathrm{L1}, \mathrm{L2}\}$, by Lemma A.1, it suffices to show that $(\mathrm{KD45})_B + (\mathrm{S4.4})_K + \{\mathrm{L1}, \mathrm{L2}\} \subseteq (\mathrm{KD45})_B + \mathrm{TB}$. By Proposition 4.9, it suffices to show that $\{\mathrm{L1}, \mathrm{L2}\} \subseteq (\mathrm{KD45})_B + \mathrm{TB}$.

It is immediate that $Kp \rightarrow Bp \in (\mathrm{KD45})_B + \mathrm{TB}$, so $\mathrm{L1} \in (\mathrm{KD45})_B + \mathrm{TB}$. For L2, note that $Bp \rightarrow BBp \wedge Bp \in (\mathrm{KD45})_B + \mathrm{TB}$. Substituting Bp for p in TB, it follows that $BBp \wedge Bp \leftrightarrow KBp$, so $Bp \rightarrow KBp \in (\mathrm{KD45})_B + \mathrm{TB}$. This completes the first half of the theorem.

For the second half, it suffices to show that if $\Lambda' \subseteq (\mathrm{S5})_K$ and K is reducible to B in $\Lambda = (\mathrm{KD45})_B + \Lambda' + \{\mathrm{L1}, \mathrm{L2}\}$ by TB, then $\Lambda' \subseteq (\mathrm{S4.4})_K$. Indeed, $\Lambda' \subseteq \Lambda \cap \mathcal{L}(K) \subseteq ((\mathrm{KD45})_B + \mathrm{TB}) \cap \mathcal{L}(K) \subseteq (\mathrm{S4.4})_K$. The first inclusion is obvious. The second follows from the reducibility of K to B in Λ, and the third by Corollary A.8. ∎

Theorem 4.4: *The modality* K *is defined implicitly in the logic* $(\mathrm{KD45})_B + (\mathrm{S5})_K + \{\mathrm{L1}, \mathrm{L2}\}$*.*

Proof: There is a semantic proof of this result in the companion paper (see Theorem 5.1 there and the remark that follows). It is also straightforward to provide a proof based on the fact, shown in Proposition A.2, that every frame $(\mathrm{KD45})_B$ can be *uniquely* extended to a $(\mathrm{KD45}_B + (\mathrm{S5})_K + \{\mathrm{L1}, \mathrm{L2}\})$-frame. We provide a direct syntactic proof here.

We first show that for all formulas $\varphi \in \mathcal{L}(B, K)$, $\varphi \leftrightarrow B\varphi \in \Lambda$ iff $\varphi \leftrightarrow K\varphi \in \Lambda$. Suppose that $\varphi \leftrightarrow B\varphi \in \Lambda$. By generalization and K_K, we must have $K\varphi \leftrightarrow KB\varphi \in \Lambda$. By L2 and T_K, $B\varphi \leftrightarrow KB\varphi \in \Lambda$. Putting together the pieces, it easily follows that $\varphi \leftrightarrow B\varphi \in \Lambda$.

For the converse, suppose that $\varphi \leftrightarrow K\varphi \in \Lambda$. By L1, it follows that $\varphi \rightarrow B\varphi \in \Lambda$. It remains to show that $B\varphi \rightarrow \varphi \in \Lambda$. By 5_K and T_K, we have that $\neg K\varphi \leftrightarrow K\neg K\varphi \in \Lambda$. Since, by assumption, $\varphi \leftrightarrow K\varphi \in \Lambda$, it follows that $\neg\varphi \leftrightarrow \neg K\varphi \in \Lambda$, and hence that $\neg\varphi \leftrightarrow K\neg K\varphi \in \Lambda$. Applying L1 again gives us that $\neg\varphi \rightarrow B\neg\varphi \in \Lambda$. Using D_B, we get that $B\neg\varphi \rightarrow \neg B\varphi \in \Lambda$. Thus, $B\varphi \rightarrow \varphi \in \Lambda$, as desired.

It follows from what we have shown that $\psi \leftrightarrow K_1\psi \in \Lambda_1 + \Lambda_2$ iff $\psi \leftrightarrow K_2\psi \in \Lambda_1 + \Lambda_2$. By 4_K and T_K, $K_1\varphi \leftrightarrow K_1K_1\varphi \in \Lambda_1 + \Lambda_2$, and thus

$$K_1\varphi \leftrightarrow K_2K_1\varphi \in \Lambda_1 + \Lambda_2. \tag{2}$$

By the generalization rule and T_K, $K_2(K_1\varphi \rightarrow \varphi) \in \Lambda_1 + \Lambda_2$. By axiom K_{K_2},

$$K_2K_1\varphi \rightarrow K_2\varphi \in \Lambda_1 + \Lambda_2. \tag{3}$$

By (2) and (3), $K_1\varphi \rightarrow K_2\varphi \in \Lambda_1 + \Lambda_2$. The converse implication follows in the same way. ∎

Corollary 4.7: *Let* $\Lambda = (\mathrm{KD45})_B + \Lambda_J + (\mathrm{S5})_K + \{\mathrm{L1}, \mathrm{L2}\}$ *be a logic in* $\mathcal{L}(B, J, K)$*, where* $\Lambda_J \subseteq (\mathrm{S5})_J$ *is a logic in* $\mathcal{L}(J)$*. Then* K *is not reducible* B *and* J *in* Λ*.*

Proof: By Proposition A.2, every $((\mathrm{KD45})_B + \Lambda_J)$-frame can be extended to a Λ-frame, and thus by Lemma A.3, $\Lambda \cap \mathcal{L}(B, J) = (\mathrm{KD45})_B + \Lambda_J$. The result now follows from Theorem 4.6. ∎

Theorem 4.8: *The modality* B *cannot be defined explicitly in the logic* $(\mathrm{KD45})_B + (\mathrm{S5})_K + \{\mathrm{L1}, \mathrm{L2}\}$*, and hence not in any logic contained in* $(\mathrm{KD45})_B + (\mathrm{S5})_K + \{\mathrm{L1}, \mathrm{L2}\}$*.*

Proof: As in the proof of Theorem 4.1, it suffices to construct a model of $((\mathrm{KD45})_B + (\mathrm{S5})_K + \{\mathrm{L1}, \mathrm{L2}\})$ in which $[\![Bp]\!] \notin [\![\mathcal{L}(K)]\!]$. Let $\mathcal{F} = (W, R_B, R_K)$ be a frame for this logic, where $W = \{w_1, w_1', w_2, w_2'\}$, $R_B = \{(w_i, w_i), (w_i', w_i) \colon i = 1, 2\}$, and $R_K = R_B \cup \{(w_i', w_i') \colon i = 1, 2\}$. Let $\mathcal{M} = (\mathcal{F}, V)$ be the model based on $\mathcal{F}$ such that $V(p) = \{w_1, w_2'\}$, and for each other primitive formula q, $V(q) = W$. It is easy to show by induction on the structure of formulas in $\mathcal{L}(K)$ that $[\![\mathcal{L}(K)]\!] = \{\emptyset, W, \{w_1, w_2'\}, \{w_1', w_2\}\}$. But $[\![Bp]\!] = \{w_1, w_1'\}$. ∎

Theorem 4.10: *For all logics* Λ *such that* $(\mathrm{S4.4})_K \subseteq \Lambda \subseteq (\mathrm{KD45})_B + (\mathrm{S4.4})_K + \{\mathrm{L1}, \mathrm{L2}\}$*, the belief modality* B *is reducible to* K *in* Λ *by* PK*.*

Proof: Let Λ satisfy the inclusions in the theorem. By Corollary A.9, $\Lambda \cap \mathcal{L}(K) = (\mathrm{S4.4})_K$. By Corollary A.8, $(\mathrm{KD45})_B + \mathrm{TB}$ is a conservative extension of S4.4, and therefore, by Proposition 4.9, $(\mathrm{S4.4})_K + \mathrm{PK}$ is a conservative extension of S4.4.

Thus, it remains to show that $\Lambda \subseteq (\mathrm{S4.4})_K + \mathrm{PK}$. Since $\Lambda \subseteq (\mathrm{KD45})_B + (\mathrm{S4.4})_K + \{\mathrm{L1}, \mathrm{L2}\}$, by Lemma A.1, it suffices to show that $(\mathrm{KD45})_B + (\mathrm{S4.4})_K + \{\mathrm{L1}, \mathrm{L2}\} \subseteq (\mathrm{S4.4})_K + \mathrm{PK}$. By Proposition 4.9 it suffices to show that $\{\mathrm{L1}, \mathrm{L2}\} \subseteq (\mathrm{S4.4})_K + \mathrm{PK}$.

Let $\Lambda' = (\mathrm{S4.4})_K + \mathrm{PK}$. For L1, note that $Kp \rightarrow \neg K \neg Kp$ is in Λ', since it is the contrapositive of $K\neg Kp \rightarrow \neg Kp$, which follows from T_K. The desired result now follows by applying PK. For L2, it suffices to show that both $Kp \rightarrow (Bp \rightarrow KBp)$ and $\neg Kp \rightarrow (Bp \rightarrow KBp)$ are in Λ'. For the first formula, note that since $Kp \rightarrow Bp \in \Lambda'$ (by L1), by generalization, so is $KKp \rightarrow KBp$. Now using the axiom 4_K, it follows that $Kp \rightarrow KBp \in \Lambda'$. Finally, propositional reasoning shows that $Kp \rightarrow (Bp \rightarrow KBp) \in \Lambda'$. For the second formula, by axioms K_K and 4_K we conclude that $Kp \rightarrow KBp$ is in Λ'. The second formula is obtained by substituting Kp for p in 4.4_K, and then substituting Bp for $\neg K \neg Kp$, by PK. ∎

To prove Proposition 5.3, we first provide a characterization of $\Lambda_0 + \mathrm{DM}_n$.

Proposition A.10 *Suppose that* $\mathrm{DM}_n = M_n p \leftrightarrow \delta$ *is a simple definition. Then* $\Lambda_0 + \mathrm{DM}_n = \{\varphi : \varphi^t \in \Lambda_0 + \Phi_\delta\}$.

Proof: Let $\Lambda^* = \{\varphi : \varphi^t \in \Lambda_0 + \Phi_\delta\}$. As we observed earlier, $\Phi_\delta \subseteq \Lambda_0 + \mathrm{DM}_n$, so $\Lambda_0 + \Phi_\delta \subseteq \Lambda_0 + \mathrm{DM}_n$. Moreover, since $\varphi \leftrightarrow \varphi^t \in \Lambda_0 + \mathrm{DM}_n$, we must have $\Lambda^* \subseteq \Lambda_0 + \mathrm{DM}_n$. On the other hand, it is easy to see that $\Lambda_0 \subseteq \Lambda^*$ and $\mathrm{DM}_n \in \Lambda^*$ (since $(M_n p \leftrightarrow \delta)^t = (\delta \leftrightarrow \delta)$). Thus, to show that $\Lambda_0 + \mathrm{DM}_n \subseteq \Lambda^*$, it suffices to show that Λ^* is a logic. Since $((M_n p \wedge M_n(p \rightarrow q)) \rightarrow M_n q)^t \in \Phi_\delta$. it follows that $((M_n p \wedge M_n(p \rightarrow q)) \rightarrow M_n q) \in \Lambda^*$. Similarly, $M_n \mathit{true} \in \Lambda^*$. Clearly Λ^* contains all instances of propositional tautologies. The argument that Λ^* is closed under modus ponens and generalization is almost identical to an analogous argument in the proof of Theorem 3.1 in the companion paper. We sketch the details here.

To see that Λ^* is closed under modus ponens, suppose that $\varphi, \varphi \rightarrow \psi \in \Lambda^*$. But then φ^t and $(\varphi \rightarrow \psi)^t = \varphi^t \rightarrow \psi^t$ are in $\Lambda_0 + \mathrm{DM}_n$. Thus, $\psi^t \in \Lambda_0 + \mathrm{DM}_n$, so $\psi \in \Lambda^*$, as desired. Another argument in this spirit shows that Λ^* is closed under substitution. Finally, we must show that Λ^* satisfies the generalization rules. If $M \neq M_n$ and $\psi \in \Lambda^*$ then, by definition, $\psi^t \in \Lambda_0 + \mathrm{DM}_n$. Moreover, $(M\psi)^t = M(\psi^t) \in \Lambda_0 + \mathrm{DM}_n$ by the generalization rule for M in $\Lambda_0 + \mathrm{DM}_n$. Hence, $M\psi \in \Lambda^*$. If $M = M_n$, we proceed as follows. Since $(M_n\psi)^t = \delta[p/\psi^t]$, we need to show that $\delta[p/\psi^t] \in \Lambda_0 + \mathrm{DM}_n$. Since $\psi^t \in \Lambda_0 + \mathrm{DM}_n$, it follows that $\psi^t \leftrightarrow \mathit{true} \in \Lambda_0 + \mathrm{DM}_n$. It easily follows that $\delta[p/\psi^t] \leftrightarrow \delta[p/\mathit{true}] \in \Lambda_0 + \mathrm{DM}_n$ (cf. By Lemma A.1 in the companion paper). Since $M_n \mathit{true} \in \Lambda^*$, it follows that $\delta[p/\mathit{true}] \in \Lambda_0 + \mathrm{DM}_n$. Thus, $\delta[p/\psi^t] \in \Lambda_0 + \mathrm{DM}_n$, as desired. ∎

Proposition 5.3 is now almost immediate.

Proposition 5.3: *If* DM_n *is a simple definition, then* $\Lambda_0 + \mathrm{DM}_n$ *is a conservative extension of* Λ_0 *iff* $\Phi_\delta \subseteq \Lambda_0$.

Proof: We have already observed that $\Phi_\delta \subseteq \Lambda_0 + \mathrm{DM}_n$. Thus, if $\Lambda_0 + \mathrm{DM}_n$ is a conservative extension of Λ_0, we must have $\Phi_\delta \subseteq \Lambda_0$. For the converse, if $\Phi_\delta \subseteq \Lambda_0$, then it follows from Proposition A.10 that $\Lambda_0 + \mathrm{DM}_n = \{\varphi : \varphi^t \in \Lambda_0\}$. It is immediate that $(\Lambda_0 + \mathrm{DM}_n) \cap \mathcal{L}(M_1, \ldots, M_{n-1} = \Lambda_0$. ∎

Theorem 5.4: *The modality* K *is weakly reducible to* B *in the logic* $(\mathrm{KD45})_B + (\mathrm{S5})_K + \{\mathrm{L1}, \mathrm{L2}\}$ *by the simple definition*

$$\mathrm{DK} = Kp \leftrightarrow ((p \wedge Bp) \vee (\neg p \wedge Bp) \vee (p \wedge B\neg p)).$$

Proof: Denote by δ the formula $((p \wedge Bp) \vee (\neg p \wedge Bp) \vee (p \wedge B\neg p))$. As $\Phi_\delta \subseteq \Lambda_0 + \mathrm{DK}$, the formula $\delta[p/(p \rightarrow \mathit{false})] \rightarrow (\delta \rightarrow \delta[p/\mathit{false}])$ is also in this logic. The axioms of KD45 ensure that $\delta[p/\mathit{false}] \leftrightarrow \mathit{false}$. Since $p \rightarrow \mathit{false}$ is propositionally equivalent to $\neg p$, the formula $B(p \rightarrow \mathit{false}) \leftrightarrow B\neg p \in \mathrm{KD45}$. This observation together with straightforward propositional reasoning shows that $\delta \wedge \delta[p/(p \rightarrow \mathit{false})] \leftrightarrow ((p \wedge B\neg p) \vee (\neg p \wedge Bp))$. The upshot of this is that $((p \wedge B\neg p) \vee (\neg p \wedge Bp)) \rightarrow \mathit{false} \in \mathrm{KD45} + \mathrm{DK}$. Propositional reasoning now shows that $(Bp \rightarrow p) \wedge (B\neg p \rightarrow \neg p) \in \mathrm{KD45} + \mathrm{DK}$. Thus, B satisfies all the axioms of S5. Moreover, it follows that $\delta \leftrightarrow Bp \in \mathrm{KD45} + \mathrm{DK}$. Thus, $Kp \leftrightarrow Bp \in \mathrm{KD45} + \mathrm{DK}$. Hence K is equivalent to B in $\mathrm{KD45} + \mathrm{DK}$ and also satisfies the S5 axioms. The equivalence of K and B also trivially implies that $\{\mathrm{L1}, \mathrm{L2}\} \subseteq \mathrm{KD45} + \mathrm{DK}$.

It remains to show that $\mathrm{KD45} + \mathrm{DK}$ is consistent. To do this, it suffices to construct a Kripke model that satisfies $\mathrm{KD45} + \mathrm{DK}$. It is easy to see that any Kripke model where R_B is an equivalence relation and $R_K = R_B$ does so. ∎

To prove Theorem 5.5, we first provide a characterization of $\Lambda_0 \oplus \mathrm{DM}_n$, similar in spirit to that of $\Lambda_0 + \mathrm{DM}_n$ given in Proposition A.10.

Proposition A.11 *Suppose that* $\mathrm{DM}_n = M_n p \leftrightarrow \delta$ *is a simple definition. Then* $\Lambda_0 \oplus \mathrm{DM}_n = \{\varphi : \varphi^t \in \Lambda_0\}$.

Proof: Let $\Lambda^* = \{\varphi : \varphi^t \in \Lambda_0\}$. Clearly, $\Lambda_0 \subseteq \Lambda^*$ and $\mathrm{DM}_n \in \Lambda^*$ (since, as we observed earlier, $(M_n p \leftrightarrow \delta)^t = (\delta \leftrightarrow \delta)$). It is straightforward to show from the definition that Λ^* contains all instances of propositional tautologies and is closed under substitution. ∎

Theorem 5.5: *The modality K is not weakly reducible$'$ to B by a simple definition in the logic* $(\mathrm{KD45})_B + (\mathrm{S5})_K + \{\mathrm{L1}, \mathrm{L2}\}$*, and hence not reducible$'$ to B. However, K is weakly reducible$'$ to B in* $(\mathrm{KD45})_B + (\mathrm{S5})_K + \{\mathrm{L1}, \mathrm{L2}\}$ *by the definition $Kp \leftrightarrow (Bp \wedge (Bq \rightarrow q))$.*

Proof: We first show that K is not weakly reducible$'$ to B by a simple definition in the logic $\Lambda = (\mathrm{KD45})_B + (\mathrm{S5})_K + \{\mathrm{L1}, \mathrm{L2}\}$. Suppose, by way of contradiction, that K is weakly reducible$'$ to B by the simple definition DK. Then, $\Lambda \subseteq \Lambda_0 \oplus$ DK. As Λ is a normal logic, it follows that $\Lambda_0 \oplus$ DK is also normal, and thus $\Lambda_0 \oplus \mathrm{DK} = \Lambda_0 + \mathrm{DK}$. By Proposition A.11, $\Lambda_0 \oplus$ DK is a conservative extension of Λ_0. This implies that K is reducible to B in Λ, which contradicts Theorem 4.5.

It remains to show that K is weakly reducible$'$ to B in $(\mathrm{KD45})_B + (\mathrm{S5})_K + \{\mathrm{L1}, \mathrm{L2}\}$ by the definition $\mathrm{DK} = Kp \leftrightarrow (Bp \wedge (Bq \rightarrow q))$. Since $\{B\mathit{true}, (B\mathit{true} \rightarrow \mathit{true})\} \subseteq \mathrm{KD45}$, it is easy to see by substituting *true* for both p and q in DK that $K\mathit{true} \in \mathrm{KD45{+}DK}$. Substituting *true* for p in DK, we also have that $Bq \rightarrow q \in \mathrm{KD45} + \mathrm{DK}$. Thus, B satisfies all the axioms of S5. Substituting *true* for q in DK, we have that $Kp \leftrightarrow Bp \in \mathrm{KD45} + \mathrm{DK}$, which shows that $\mathrm{S5} \subseteq \mathrm{KD45} + \mathrm{DK}$. The equivalence of K and B also trivially implies that $\{\mathrm{L1}, \mathrm{L2}\} \subseteq \mathrm{KD45} + \mathrm{DK}$. The argument that KD45 + DK is consistent is identical to that used in Theorem 5.4. ∎

References

Alston, W. P. (1989). *Epistemic Justification: Essays in the Theory of Knowledge*. Ithaca, NY: Cornell University Press.

Artemov, S. and E. Nogina (2005). Introducing justification to epistemic logic. *Journal of Logic and Computation 15*(6), 1059–1073.

Ayer, A. J. (1956). *The Problem of Knowledge*. London: Macmillan.

Benthem, J. F. A. K. v. (2006). Epistemic logic and epistemology: the state of their affairs. *Philosophical Studies 128*(1), 49–76.

Beth, E. W. (1953). On Padoa's method in the theory of definition. *Indagationes Mathematicae 15*, 330–339.

Blackburn, P., M. de Rijke, and Y. Venema (2001). *Modal Logic.* Cambridge Tracts in Theoretical Computer Science, No. 53. Cambridge, U.K.: Cambridge University Press.

BonJour, L. and E. Sosa (2003). *Epistemic Justification: Internalism Vs. Externalism, Foundations Vs. Virtues.* Malden, MA: Blackwell Publishing.

Chisholm, R. M. (1957). *Perceiving: A Philosophical Study.* Ithaca, N.Y.: Cornell University Press.

Foley, R. (1984). Epistemic luck and the purely epistemic. *American Philosophical Quarterly 21*, 113–124.

Georgacarakos, G. N. (1976). Semantics for S4.04, S4.4, and S4.3.2. *Notre Dame Journal of Formal Logic 17*(2), 297–302.

Gettier, E. (1963). Is justified true belief knowledge? *Analysis 23*, 121–123.

Halpern, J. Y., D. Samet, and E. Segev (2008). On definability in multimodal logic: Basic notions. Unpublished manuscript, available at www.cs.cornell.edu/home/halpern/papers.

Hintikka, J. (1962). *Knowledge and Belief.* Ithaca, N.Y.: Cornell University Press.

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

Lenzen, W. (1978). Recent work in epistemic logic. *Acta Philosophica Fennica 30*, 1–219.

Lenzen, W. (1979). Epistemoligische betrachtungen zu [S4,S5]. *Erkenntnis 14*, 33–56.

Pelletier, F. J. and A. Urquhart (2003). Synonymous logics. *Journal of Philosophical Logic 32*(3), 259–285.

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

Pritchard, D. (2005). *Epistemic Luck.* Oxford, UK: Oxford University Press.

Stalnaker, R. (2006). On logics of knowledge and belief. *Philosophical Studies 128*(1), 169–199.

Swinburne, R. (2001). *Epistemic Justification.* Oxford, U.K.: Oxford University Press.