On Definability in Multimodal Logic*

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

November, 2007

Abstract

Three notions of definability in multimodal logic are considered. Two are analogous to the notions of explicit definability and implicit definability introduced by Beth in the context of first-order logic. However, while by Beth's theorem the two types of definability are equivalent for first-order logic, such an equivalence does not hold for multimodal logics. A third notion of definability, *reducibility*, is introduced; it is shown that in multimodal logics, explicit definability is equivalent to the combination of implicit definability and reducibility. The three notions of definability are characterized semantically using *(modal) algebras*. The use of algebras, rather than frames, is shown to be necessary for these characterizations.

*We thank Samson Abramsky, Arnon Avron, Johan van Benthem, Sergiu Hart, Martin Meier, Jouku Väänänen, and an anonymous reviewer 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

In the context of logic, the notion of the definability of an entity, described in broad strokes, refers to the expression or determination of that entity in terms of other entities of the same type in the framework of a certain logic. A simple example is the definition of conjunction in terms of negation and disjunction in propositional calculus. Closer to this paper is the case of the definition of the diamond operator in modal logics in terms of the box operator by the formula $\Diamond p \leftrightarrow \neg\Box\neg p$.

Two notions of predicate definability, explicit and implicit definability, were first formalized by Beth [1953] for first-order logic:

> An n-ary predicate R is *explicitly defined* in a first-order logic Λ if there is a formula $R(x_1, \ldots, x_n) \leftrightarrow \varphi$ in Λ such that φ does not contain the predicate R.
>
> The predicate R is *implicitly defined* in Λ if there do not exist two models of Λ that have the same domain and agree on the meaning of all predicates other than R, but disagree on the meaning of R.

Beth's theorem states that the predicate R is explicitly defined in Λ if and only if it is implicitly defined in Λ.

In this paper, we study modal definability in the context of multimodal logic, by considering when one modality is defined in terms of others. For ease of exposition, we assume that all modal operators are unary.

> The modality M is *explicitly defined* in a multimodal logic Λ if there exists a formula $Mp \leftrightarrow \delta$ in Λ such that M does not occur in δ.
>
> The modality M is *implicitly defined* in a multimodal logic Λ if there do not exist two models of Λ that coincide except in the interpretation of M.[1]

Just as in first-order logic, if a modality is explicitly defined in a modal logic Λ, then it is implicitly defined in Λ. But the converse does not hold for modal logic. An example of this is the multimodal logic of KD45-belief

[1] Implicit definability can be defined syntactically, both in first-order and multimodal logics. By the completeness theorem, the syntactic version is equivalent to that above. In Section 3, we actually define implicit definability syntactically.

and S5-knowledge. As we show in [Halpern, Samet, and Segev 2008], in this logic knowledge is implicitly defined but not explicitly defined. Henceforth, we refer to this paper as "the companion paper".[2]

We can understand the relationship between explicit and implicit definability in multimodal logic by considering a third notion of definability. Let Λ_0 be the sublogic of Λ consisting of formulas that do not mention the modality M.

> The modality M is *reducible* to the other modalities in Λ if there is a formula $Mp \leftrightarrow \delta$ such that M does not occur in δ and the logic $\Lambda_0 + (Mp \leftrightarrow \delta)$ generated by Λ_0 and this definition of M (a) includes Λ and (b) is a conservative extension of Λ_0.

As we argue in the companion paper, reducibility comes closest to capturing our intuitions when we say that knowledge is (or is not) definable in terms of belief. The question we are asking is whether, for example, by defining knowledge as true belief, that is, by adding $Kp \leftrightarrow p \wedge Bp$ to the logic of belief, we can recover all the properties of knowledge of interest. If M is explicitly defined in Λ, then Λ contains $\Lambda_0 + (Mp \leftrightarrow \delta)$. With reducibility, the containment goes in the opposite direction.

When M is explicitly defined in Λ by the formula $Mp \leftrightarrow \delta$, then M is reducible to the other modalities in M by the same formula, and M is implicitly defined in Λ. However, neither implicit definability nor reducibility implies explicit definability. Our main result states that

> The modality M is explicitly defined in Λ if and only if it is implicitly defined in Λ and is reducible in Λ to the other modalities.

Reducibility can be defined in first-order languages analogously to the definition for multimodal logics. It is easily seen to follow from explicit definability. Thus, by Beth's theorem, implicit definability implies reducibility. However, in the context of modal logic, implicit definability and reducibility are incomparable. In the companion paper, we examine the three notions of definability in the context of logics of knowledge and belief. Among other things, we show that in the logic of KD45-belief and S5-knowledge, knowledge is implicitly defined but it is not reducible to belief; in the logic of KD45-belief and S4-knowledge, knowledge is not implicitly defined but is reducible to belief by defining knowledge as true belief (i.e., using the formula $Kp \leftrightarrow p \wedge Bp$).

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

The fact that S5 knowledge is implicitly defined by KD45 belief implies that there is a unique way to extend each frame for KD45 belief to a frame for S5 knowledge. It may seem surprising that this is the case and yet S5 knowledge is not reducible to belief. We explain this apparent disconnect between syntax and semantics by going beyond frames to *(modal) algebras* [Blackburn et al. 2001; Kracht 1999]. Although each frame for KD45 belief can be extended to a frame for S5 knowledge, we show that there is an algebra for KD45 belief that cannot be extended to an algebra for S5 knowledge.

Algebras play a significant role in this paper. The three notions of definability we consider are all defined syntactically (that is, in terms of whether formulas are in certain logics). We characterize each of them semantically, using algebras. As we show, in a precise sense, the greater generality of algebras is necessary for our characterizations.

The rest of the paper is organized as follows. In Section 2, we review the relevant definitions of modal logic that we need for this paper. definability that have been considered before in the context of modal logic. In Section 4, we give semantic characterizations of our notions in terms of algebras. We discuss the extent to which we can characterize our notions using frames; this also allows us to relate definability in modal logic to definability in first-order logic. Most proofs are relegated to the appendix.

2 Modal logic review: syntax, semantics, and axioms

In this section, we review the essentials of modal logic, including syntax, semantics, and standard axiomatizations. The reader is encouraged to consult a standard reference (e.g., [Blackburn et al. 2001; Kracht 1999]) for more details.

2.1 Modal logics

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 $\neg\varphi$, $(\varphi \rightarrow \psi)$, 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 $\rightarrow$ in the usual way; we take *true* to

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

be an abbreviation of $p \vee \neg p$. The *language* $\mathcal{L}(M_1, \ldots M_n)$ 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 $\mathcal{L}(M_1, \ldots, M_n)$ that (a) contains all tautologies of propositional logic; (b) is closed under *modus ponens*, so that if $\varphi \in \Lambda$ and $\varphi \rightarrow \psi \in \Lambda$, then $\psi \in \Lambda$; and (c) is closed under substitution, so that if $\varphi \in \Lambda$, p is a primitive proposition, and $\psi \in \mathcal{L}(M_1, \ldots, M_n)$, 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}, $M_i(p \rightarrow q) \rightarrow (M_i p \rightarrow M_i q)$, 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 modus ponens. 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.

2.2 Frames and Kripke models

Perhaps the most common approach to giving semantics to modal logic makes use of frames and Kripke models. We review this approach in this section, and consider an alternative approach, using *algebras* and *algebraic models*, in the next subsection.

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 $V \colon P \rightarrow 2^W$ 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 $\mathcal{S}$ of frames, 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 $\Lambda \supseteq \mathrm{Th}(\mathcal{S})$. A frame $\mathcal{F}$ is said to be a Λ-frame if $\Lambda \subseteq \mathrm{Th}(\mathcal{F})$.

The *canonical frame* for Λ is defined on the set W that consists of all maximally consistent sets of formulas in $\mathcal{L}$. The set W is made a frame by defining, for each modality M_i, a relation R_i such that $(w, w') \in R_i$ if, for all formulas φ, if $M_i\varphi \in w$ then $\varphi \in w'$. The *canonical model* is the model based on the canonical frame with the valuation V defined by $V(p) = \{w \colon p \in w\}$. Every normal logic Λ is sound and complete with respect to its canonical model, but may not be sound with respect to its canonical frame.

In the sequel, we consider the logic $(\mathrm{KD45})_B + (\mathrm{S5})_K + \{\mathrm{L1}, \mathrm{L2}\} \subseteq \mathcal{L}(B, K)$, where the modal operator B satisfies the axioms of KD45, K satisfies the axioms of S5, and L1 and L2 are axioms that link K and B. To make this paper self-contained, we list the relevant axioms here:

$$
\begin{array}{ll}
(\mathrm{D}_B) & Bp \to \neg B \neg p \\
(4_B) & Bp \to BBp \\
(5_B) & \neg Bp \to B\neg Bp. \\
(4_K) & Kp \to KKp \\
(5_K) & \neg Kp \to K\neg Kp \\
(\mathrm{T}_K) & Kp \to p \\
(\mathrm{L1}) & Kp \to Bp \\
(\mathrm{L2}) & Bp \to KBp.
\end{array}
$$

2.3 Algebras and algebraic models

We now consider a more general approach for giving semantics to modal logics, using algebras and algebraic models, that goes back to Jónsson and Tarski [1951, 1952]. As we shall see, syntactical notions of definability have certain semantic equivalents that can be formulated in terms of algebras but not in terms of frames.

A *modal algebra* (or *algebra* for short) $\mathcal{A}$ for the language $\mathcal{L}(M_1, \ldots, M_n)$ is a tuple

$$(\mathcal{B}, \vee, \neg, 1, \mathbf{M}_1, \ldots, \mathbf{M}_n),$$

where $(\mathcal{B}, \vee, \neg, 1)$ is a Boolean algebra, and for each $i = 1, \ldots, n$, $\mathbf{M}_i$ is a unary operator on $\mathcal{B}$. An *algebraic model* $\mathcal{M}$ based on the algebra $\mathcal{A}$ is a pair $(\mathcal{A}, V)$, where $V \colon P \to \mathcal{B}$ is a *valuation* of the primitive propositions as elements of $\mathcal{B}$. The function V is extended inductively to a *meaning* function $[\![\cdot]\!]_{\mathcal{M}}$ on all formulas: $[\![\neg\varphi]\!]_{\mathcal{M}} = \neg[\![\varphi]\!]_{\mathcal{M}}$ (where the second $\neg$ is the operator in the Boolean algebra, not set complementation), $[\![\varphi \vee \psi]\!]_{\mathcal{M}} =$

$([\![\varphi]\!])_{\mathcal{M}} \vee [\![\psi]\!]_{\mathcal{M}}$ (where the second $\vee$ is the operator in the Boolean algebra), and $[\![M_i\varphi]\!]_{\mathcal{M}} = \mathbf{M}_i([\![\varphi]\!]_{\mathcal{M}})$. We again omit the subscript $\mathcal{M}$ if no confusion results. A formula φ is valid in $\mathcal{M}$ if $[\![\varphi]\!]_{\mathcal{M}} = 1$; it is valid in $\mathcal{A}$ if it is valid in all algebraic models based on $\mathcal{A}$. Soundness and completeness are defined just as for Kripke models. We define $\mathrm{Th}(\mathcal{M})$ and $\mathrm{Th}(\mathcal{A})$ in the obvious way. $\mathcal{A}$ is a Λ-*algebra* if $\Lambda \subseteq \mathrm{Th}(\mathcal{A})$; similarly, $\mathcal{M}$ is an *algebraic model for* Λ if $\Lambda \subseteq \mathrm{Th}(\mathcal{M})$.

Each frame $\mathcal{F} = (W, R_1, \ldots, R_n)$ is associated in a natural way with the algebra $\mathcal{A} = (2^W, \vee, \neg, W, \mathbf{M}_1, \ldots, \mathbf{M}_n)$, where $\vee$ is union, $\neg$ is set-theoretic complementation, and, for $i = 1, \ldots, n$, the set operator $\mathbf{M}_i$ is defined by taking

$$\mathbf{M}_i(E) = \{x \mid R_i(x) \subseteq E\}$$

for $E \subseteq W$. Similarly, we associate with the Kripke model $(\mathcal{F}, V)$ the algebraic model $(\mathcal{A}, V)$ with the same valuation function V. It is easy to see that the meaning functions in both models coincide.

It is well known that there are algebras that are not associated with frames. We demonstrate in the sequel that, as a consequence, the set of Λ-frames may have a particular definability property that does not correspond to a property of Λ. The definability properties of Λ-algebras, on the other hand, correspond exactly to those of the logic Λ.

For a logic Λ in a language $\mathcal{L}$, define an equivalence relation $\equiv_\Lambda$ on $\mathcal{L}$ by $\varphi \equiv_\Lambda \psi$ iff $\varphi \leftrightarrow \psi \in \Lambda$. Consider the partition of Λ into equivalence classes $\mathcal{L}/\equiv_\Lambda$. The equivalence class that contains the formula φ is denoted $|\varphi|_\Lambda$. The *Lindenbaum-Tarski* Λ-*algebra* is the Boolean algebra $(\mathcal{L}/\equiv_\Lambda, \vee, \neg, |true|_\Lambda)$ where $|\varphi|_\Lambda \vee |\psi|_\Lambda = |(\varphi \vee \psi)|_\Lambda$ and $\neg\, |\varphi|_\Lambda = |\neg\varphi|_\Lambda$; we leave it to the reader to check that these definitions are independent of the choice of representative of the equivalence class, and so are well defined. The *canonical* Λ-*algebra* $\mathcal{A}_\Lambda$ is the modal algebra based on the Lindenbaum-Tarski Λ-algebra where, for each i, $\mathbf{M}_i(|\varphi|_\Lambda) = |M_i(\varphi)|_\Lambda$. It is easy to see that since Λ is a normal logic, all the operators are well defined. The *canonical algebraic model for* Λ is $\mathcal{M}_\Lambda = (\mathcal{A}_\Lambda, V_\Lambda)$, where $V_\Lambda(p) = |p|_\Lambda$. It is well known that Λ is sound and complete with respect to the class of Λ-algebras, with respect to $\{\mathcal{A}_\Lambda\}$, and with respect to $\{\mathcal{M}_\Lambda\}$ [Blackburn et al. 2001; Kracht 1999].

3 Three notions of definability

In this section we examine the three different notions of defining one modality in terms of others mentioned in the introduction.

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* $M_1, \ldots, M_{n-1}$*)*. When the only primitive proposition in δ is p we say that the definition is *simple*.

The formula DM_n is the obvious analogue of the formula used in first-order logic to define one predicate in terms of others. We also have an obvious analogue of the notion of explicit definability in first-order logic. 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 $\mathrm{DM}_n \in \Lambda$.

In the context of first-order logic, an apparently weaker notion of definability called *implicit definability* has been studied. We define what seems to be the appropriate analogue for modal 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 $M_n p \leftrightarrow M_n' p \in \Lambda + \Lambda[M_n/M_n']$.

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.

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. It also enables us to consider only simple definitions of M_n, as we state next.

Proposition 3.1 *If M_n is reducible to $M_1, \ldots, M_{n-1}$ in Λ, then it is reducible by a simple definition.*

But the main reason to require that $\Lambda_0+\mathrm{DM}_n$ be a conservative extension of Λ_0 is 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 as demonstrated in the following example.

Example 3.2 Let Λ be the minimal normal logic in $\mathcal{L}$. Obviously, Λ_0 is the minimal normal logic in $\mathcal{L}_0$. Let DM_n be the formula $M_n p \leftrightarrow \neg M_1(p \wedge \neg p)$. By the minimality of Λ, $\Lambda \subseteq \Lambda_0+\mathrm{DM}_n$. By the generalization rule, $M_n true \in \Lambda_0 + \mathrm{DM}_n$, and therefore $\neg M_1(true \wedge \neg true) \in \Lambda_0 + \mathrm{DM}_n$. But this formula is not in Λ_0. Thus, the smallest normal logic containing both DM_n and Λ_0 includes formulas in $\mathcal{L}_0$ not in Λ_0.

We further discuss reducibility and some of its variants in Section 5 of the companion paper [Halpern, Samet, and Segev 2008].

In first-order logic, Beth's theorem [1953] states that implicit and explicit definability coincide. When reducibility is defined for first-order logics, analogously to the definition for multimodal logic, then it can be shown to be implied by the other two notions of definability. However, in the context of modal logic, none of the statements above holds, as demonstrated by the following proposition, which is proved in the companion paper.

Proposition 3.3

(a) Knowledge is neither explicitly nor implicitly defined in the logic $(\mathrm{KD45})_B + (\mathrm{S4})_K + \{\mathrm{L1}, \mathrm{L2}\})$*, but it is reducible to belief in this logic.*

(b) Knowledge is neither explicitly defined nor reducible to belief in the logic $(\mathrm{KD45})_B + (\mathrm{S5})_K + \{\mathrm{L1}, \mathrm{L2}\})$*, but it is implicitly defined in this logic.*

The two parts of this proposition show that neither implicit definability nor reducibility implies explicit definability, and that neither implicit definability nor reducibility implies the other.

The following theorem describes the relations between the three notions of definability.

Theorem 3.4 *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.*

We provide a direct proof of Theorem 3.4 in the appendix. We also give an alternative proof later which uses the semantic characterizations of the three notions of definability given in the next section.

Maksimova [1992a, 1992b] studies implicit and explicit definability of primitive propositions (rather than modal operators) in unimodal logics. She shows that implicit and explicit definability of primitive propositions are equivalent for large classes of modal logics (in particular, for those containing K4). Our results show that this equivalence does not hold for our notions of implicit and explicit definability. (See Kracht [1999] for a discussion of definability of primitive propositions in modal logic.)

Lenzen [1979] also studied definability of one modality in terms of other modalities. He requires that the definition DM_n be simple (which in our framework follows from reducibility in Proposition 3.1), and calls the logic $\Lambda + \mathrm{DM}_n$ (that is, the underlying logic extended by a definition DM_n) a *definitional extension* of Λ. He calls two logics $\Lambda \subseteq \mathcal{L}(M_1, \ldots, M_{n-1}, M_n)$ and $\Lambda' \subseteq \mathcal{L}(M_1, \ldots, M_{n-1}, M_n')$ *synonymous* when there is a third logic $\Lambda^* \in \mathcal{L}(M_1, \ldots, M_{n-1}, M_n, M_n')$ that is a definitional extension of both Λ and Λ'. To relate Lenzen's definitional extension to our terminology, we note that if we add the requirement that $\Lambda + \mathrm{DM}_n$ is a conservative extension of Λ, then, in our terminology, M_n is reducible to $M_1, \ldots, M_{n-1}$ in the logic $\Lambda + \mathrm{DM}_n$ by DM_n.

There has also been relevant work on translation schemes between languages that is relevant to our work (see Pelletier and Urquhart [2003] and the references therein). A definition DM_n of the modality M_n defines a natural translation $\varphi \mapsto \varphi^t$ from the language $\mathcal{L}$ to $\mathcal{L}_0$ that is described in the appendix. When M_n is explicitly defined in Λ, then for every $\varphi \in \Lambda$, the translated formula φ^t is in Λ_0 (see Lemma A.2 in the appendix). In the terminology of Pelletier and Urquhart, this means that the translation is sound (with respect to the logics Λ and Λ_0).

4 The semantics of definability

In this section, we provide semantic characterizations of the three notions of definability we have been considering. We use the following definition. An algebra $\mathcal{A}$ for the language $\mathcal{L}$ is an *extension* of an algebra $\mathcal{A}'$ for $\mathcal{L}' \subseteq \mathcal{L}$ if $\mathcal{A}$ is obtained by adding to $\mathcal{A}'$ operators that correspond to the modalities in $\mathcal{L}$ that are not in $\mathcal{L}'$. Similarly, a frame $\mathcal{F}$ for the language $\mathcal{L}$ is an *extension* of a frame $\mathcal{F}'$ for $\mathcal{L}' \subseteq \mathcal{L}$ if $\mathcal{F}$ is obtained by adding to $\mathcal{F}'$ relations that correspond to the modalities in $\mathcal{L}$ that are not in $\mathcal{L}'$. If $\mathcal{A}$ ($\mathcal{F}$) is obtained by adding operators $(\mathbf{M_i})_{i \in I}$ to $\mathcal{A}'$ (relations $(R_i)_{i \in I}$ to $\mathcal{F}'$), we sometimes abuse notation and write $\mathcal{A} = (\mathcal{A}', (\mathbf{M_i})_{i \in I})$ ($\mathcal{F} = (\mathcal{F}', (R_i)_{i \in I})$).

Note that if $\mathcal{A}$ extends $\mathcal{A}'$, then for all models $\mathcal{M} = (\mathcal{A}, V)$ and $\mathcal{M}' = (\mathcal{A}', V)$ and each formula $\varphi \in \mathcal{L}'$, we have $[\![\varphi]\!]_{\mathcal{M}} = [\![\varphi]\!]_{\mathcal{M}'}$. For an algebra $\mathcal{A} = (\mathcal{B}, \vee, \neg, 1, \mathbf{M}_1, \ldots, \mathbf{M}_n)$, let $\mathcal{A}_0$ denote the algebra $(\mathcal{B}, \vee, \neg, 1, \mathbf{M}_1, \ldots, \mathbf{M}_{n-1})$. Clearly, $\mathcal{A}$ is an extension of $\mathcal{A}_0$. Similar remarks apply to frames and Kripke models.

We start with the characterization of implicit definability.

Theorem 4.1 *The following are equivalent:*

(a) the modality M_n is implicitly defined in Λ;

(b) if $\mathcal{A} = (\mathcal{A}_0, \mathbf{M}_n)$ and $\mathcal{A}' = (\mathcal{A}_0, \mathbf{M}_n')$ are Λ-algebras, then $\mathbf{M}_n = \mathbf{M}_n'$.

We cannot expect a characterization of implicit definability in terms of frames, since a normal logic may not be complete with respect to its frames; indeed, there may be no frames for a logic at all. In the next section we formulate a characterization of implicit definability in terms of frames for a restricted class of logics, and relate modal definability to first-order definability of relations for this class of logics.

The characterization of explicit definability and reducibility is done in terms of algebras only. In the next section we will see why an analogous characterization in terms of frames or Kripke models is impossible.

For the next two characterizations we need the following definition. An *algebra of operators* $\mathcal{O}$ on a Boolean algebra $(\mathcal{B}, \vee, \neg, 1)$ is a set $\mathcal{O}$ of unary operators on $\mathcal{B}$ that is itself a Boolean algebra and is closed under composition. Thus, for every $f, g \in \mathcal{O}$, $\neg f$, $f \vee g$, and $f \circ g$ are all in $\mathcal{O}$, where $(\neg f)(x) = \neg f(x)$, $(f \vee g)(x) = f(x) \vee g(x)$, and $(f \circ g)(x) = f(g(x))$. The top element in $\mathcal{O}$ is the constant operator that always returns the value 1 in $\mathcal{B}$.

For an algebra $\mathcal{A} = (\mathcal{B}, \vee, \neg, 1, \mathbf{M}_1, \ldots, \mathbf{M}_n)$, let $\mathcal{O}^*_{\mathcal{A}}$ be the smallest algebra of operators on $\mathcal{B}$ that contains the operators $\mathbf{M}_1, \ldots, \mathbf{M}_n$, and let $\mathcal{A}_0$ be the algebra $(\mathcal{B}, \vee, \neg, 1, \mathbf{M}_1, \ldots, \mathbf{M}_{n-1})$.

Theorem 4.2 *The modality M_n is explicitly defined in Λ if and only if, for each Λ-algebra $\mathcal{A}$, $\mathcal{O}^*_{\mathcal{A}} = \mathcal{O}^*_{\mathcal{A}_0}$.*

Theorem 4.3 *The modality M_n is reducible to $M_1, \ldots, M_{n-1}$ in Λ if and only if each Λ_0-algebra $\mathcal{A}_0$ has an extension to a Λ-algebra $\mathcal{A}$ such that $\mathcal{O}^*_{\mathcal{A}} = \mathcal{O}^*_{\mathcal{A}_0}$.*

In light of Theorems 4.1, 4.2, 4.3, the following result can be viewed as a reformulation of Theorem 3.4 in semantic terms. One of the implications in this result is significantly simpler to prove than the analogous implication in Theorem 3.4; moreover, it provides an alternative proof of this result.

Theorem 4.4 *For each Λ-algebra $\mathcal{A} = (\mathcal{B}, \vee, \neg, 1, \mathbf{M}_1, \ldots, \mathbf{M}_n)$, $\mathcal{O}^*_{\mathcal{A}} = \mathcal{O}^*_{\mathcal{A}_0}$ iff (a) every Λ_0-algebra $\mathcal{A}_0$ can be extended to a Λ-algebra $\mathcal{A}$ such that $\mathcal{O}^*_{\mathcal{A}} = \mathcal{O}^*_{\mathcal{A}_0}$ and (b) if $\mathcal{A} = (\mathcal{A}_0, \mathbf{M}_n)$ and $\mathcal{A}' = (\mathcal{A}_0, \mathbf{M}'_n)$ are Λ-algebras, then $\mathbf{M}_n = \mathbf{M}'_n$.*

5 Definability and frame semantics

The semantic characterizations of definability in Section 4 mainly use algebras rather than frames. Here we explore the relationship between modal definability and frame semantics. Of course, we cannot expect a frame-semantic characterization for all normal logics, since some normal logics are poorly described by frame semantics. We therefore restrict ourselves to what we call *orthodox logics*, to be defined shortly, for which frame semantics is adequate. Using frame semantics for orthodox logics enables us to explore the relationship between the definability of modalities and the definability of predicates in first-order logic.

Given a language $\mathcal{L}$, the *first-order frame language* of $\mathcal{L}$, denoted $\mathcal{L}^{\mathrm{fo}}$, is the first-order language with equality that includes, for every modality M_i in $\mathcal{L}$, a binary predicate R_i^{fo}. The frames of many axioms of modal logic can be described in the frame language. Thus, for example, the axiom $Kp \rightarrow p$ is valid in a frame iff the relation corresponding to K is reflexive, which is expressed in the first-order frame-language by $\forall x R_K^{\mathrm{fo}}(x,x)$. We say in this case that $Kp \rightarrow p$ and $\forall x R_K^{\mathrm{fo}}(x,x)$ correspond. In general, formulas $\varphi \in \mathcal{L}$ and $\alpha \in \mathcal{L}^{\mathrm{fo}}$ *correspond* if, for all frames $\mathcal{F}$ for $\mathcal{L}$, φ is valid in $\mathcal{F}$ iff α is valid in $\mathcal{F}$.

A formula $\varphi \in \mathcal{L}$ is *canonical* if it is valid in the canonical frame of each logic Λ that contains φ [Blackburn et al. 2001]. If the logic Λ is generated by a set of canonical formulas (i.e., if there is a set C of canonical formulas such that Λ is the smallest logic containing C), then Λ is complete with respect to its canonical frame. A logic is *orthodox* if it is generated by a set A of formulas such that each formula $f \in A$ is canonical and corresponds to a first-order formula. Let A^{fo} denote the set of first-order formulas that correspond to the formulas in A. The first-order logic Λ^{fo} generated by A^{fo} is sound and complete with respect to all Λ-frames.

5.1 Implicit definability

The implicit definability of a modality can be characterized by frame semantics in a way analogous to the algebraic characterization of Theorem 4.1. This characterization is stated in the next theorem, as well as its characterization in terms of the definability properties of the predicate corresponding to the modality in the frame language.

Theorem 5.1 *If Λ is an orthodox logic in the language $\mathcal{L}(M_1, \ldots, M_n)$. then the following are equivalent:*

(a) the modality M_n is implicitly defined in Λ;

(b) for all Λ-frames $(\mathcal{F}_0, R_n)$ and $(\mathcal{F}_0, R'_n)$, we have $R_n = R'_n$;

(c) the predicate R_n^{fo} is implicitly defined in Λ^{fo};

(d) the predicate R_n^{fo} is explicitly defined in Λ^{fo}.

The equivalence of (b) and (c) follows from the fact that the set of Λ-frames is the set of Λ^{fo}-models and the definition of implicit definability for first-order logic. The equivalence of (c) and (d) is Beth's theorem. We prove in the appendix that (a) is equivalent to (b). The latter equivalence is the frame-semantics counterpart of Theorem 4.1. Parts (b) and (c) of Theorem 5.1 show that for orthodox logics, the implicit definability of a modality is equivalent to the implicit and explicit definability of its corresponding relation.

Theorem 5.1 can be used to provide a semantic proof of Proposition 3.3(b), namely, that S5 knowledge is implicitly defined in $(\mathrm{KD45})_B + (\mathrm{S5})_K + \{\mathrm{L1}, \mathrm{L2}\}$. This logic is orthodox; thus, it suffices to show that the relation R_K is explicitly defined by the predicate R_B^{fo} associated with the relation R_B. The following proposition shows that this is indeed the case.

Proposition 5.2 *The formula $R_K^{\mathrm{fo}}(x, y) \leftrightarrow \exists z(R_B^{\mathrm{fo}}(x, z) \wedge R_B^{\mathrm{fo}}(y, z))$ is valid in all $((\mathrm{KD45})_B + (\mathrm{S5})_K + \{\mathrm{L1}, \mathrm{L2}\})$-frames.*

5.2 Explicit definability

The explicit definability of M_n is characterized semantically in Theorem 4.2 by the condition $\mathcal{O}^*_{\mathcal{A}} = \mathcal{O}^*_{\mathcal{A}_0}$, or equivalently, $\mathbf{M}_n \in \mathcal{O}^*_{\mathcal{A}_0}$. This condition says that the algebraic operator $\mathbf{M}_n$ is generated using Boolean operations and composition from the algebraic operators $\mathbf{M}_1, \dots, \mathbf{M}_{n-1}$. An analogous frame-semantics condition is that R^{fo}_n can be explicitly defined in terms of $R^{\text{fo}}_1, \dots, R^{\text{fo}}_{n-1}$. But this condition does not characterize the explicit definability of M_n. The logic $(\text{KD45})_B + (\text{S5})_K + \{\text{L1}, \text{L2}\}$ illustrates this claim. As stated in Proposition 5.2, R^{fo}_K is explicitly defined by R^{fo}_B, yet, by Proposition 3.3, the modality K is not explicitly defined by B. This gap between definability in the modal logic and definability in the first-order frame language is due to the fact that in orthodox logics, the first-order frame language is more expressive than the modal language it is associated with.

5.3 Reducibility

A frame-semantics analogue of Theorem 4.3 would state that M_n is reducible to $M_1, \dots, \mathcal{M}_{n-1}$ in an orthodox logic Λ if and only if each Λ_0-frame $(W, R_1, \dots, R_{n-1})$ can extended to a Λ-frame $(W, R_1, \dots, R_{n-1}, R_n)$ in which R^{fo}_n can be explicitly defined. But this claim is false. Consider again the logic $(\text{KD45})_B + (\text{S5})_K + \{\text{L1}, \text{L2}\}$. It is easy to show that every $(\text{KD45})_B$-frame can be extended to a $((\text{KD45})_B + (\text{S5})_K + \{\text{L1}, \text{L2}\})$-frame (see Proposition A.2 in the companion paper). Combining this result with Proposition 5.2, it follows that every $(\text{KD45})_B$-frame can be extended to a $((\text{KD45})_B + (\text{S5})_K + \{\text{L1}, \text{L2}\})$-frame in which R^{fo}_K is explicitly defined. Yet, by Proposition 3.3, K is not reducible to B in this logic. By the semantic characterization of reducibility in Theorem 4.3, it follows that $(\text{KD45})_B$-algebras should not have this property of extension that $(\text{KD45})_B$-frames have. Indeed, we next construct an example of a $(\text{KD45})_B$-algebra that cannot be extended at all to a $((\text{KD45})_B + (\text{S5})_K + \{\text{L1}, \text{L2}\})$-algebra. This example also provides a direct proof of the irreducibility of S5 knowledge to belief stated in Proposition 3.3, using the semantic characterization of reducibility in Theorem 4.3.

Example 5.3 Let $(\mathcal{B}, \cup, \neg, W)$ be the Boolean algebra of the finite and cofinite subsets of the set of nonnegative integers $W = \{0, 1, 2, \dots\}$ (recall that a *cofinite* set is the complement of a finite set), where the Boolean operations are union and set-theoretic complement, and the top element is W. Let $\mathcal{U}$ be the subset of $\mathcal{B}$ which consists of all the cofinite sets. Define an operator $\mathbf{B}$ on $\mathcal{B}$ by taking

$$\mathbf{B}(E) = \begin{cases} E \cup \{0\} & \text{if } E \in \mathcal{U} \\ E \setminus \{0\} & \text{if } E \notin \mathcal{U}. \end{cases}$$

Theorem 5.4 *The algebra* $\mathcal{A} = (\mathcal{B}, \cup, \neg, W, \mathbf{B})$ *is a* $(\mathrm{KD45})_B$*-algebra that cannot be extended to a* $(\mathrm{KD45})_B + (\mathrm{S5})_K + \{\mathrm{L1}, \mathrm{L2}\}$*)-algebra.*

Note that by the Jónsson-Tarski theorem [Blackburn et al. 2001], the $(\mathrm{KD45})_B$ set-algebra in Example 5.3 can be isomorphically embedded in a $(\mathrm{KD45})_B$ set-algebra in which the operator B is derived from a relation. However, in the algebra of the example itself, B is *not* derived from a relation on W. Indeed, if it were, then, by Proposition A.2 of the companion paper, we could extend this model to one where an S5 knowledge operator is defined.

The only properties of $\mathcal{B}$ and $\mathcal{U}$ used in the proof of Theorem 5.4 are the facts that $\mathcal{B}$ is an algebra that contains all the singletons and that $\mathcal{U}$ is a nonprincipal ultrafilter in $\mathcal{B}$.[4] Thus, the theorem also holds if we take $\mathcal{B}$ to be 2^W and $\mathcal{U}$ to be a nonprincipal ultrafilter on W. These conditions also hold if $W = [0, 1]$, $\mathcal{B}$ consists of all Borel sets in $[0, 1]$ that have Lebesgue measure either 0 or 1, and $\mathcal{U}$ consists of all the sets in $\mathcal{B}$ with Lebesgue measure 1.

Theorem 5.4 has another, somewhat surprising, application. It allows us to prove general results regarding the irreducibility of knowledge to a combination of belief and justification. In the companion paper, we show that knowledge cannot be reduced to belief in the logic $(\mathrm{KD45})_B + (\mathrm{S5})_K + \{\mathrm{L1}, \mathrm{L2}\}$. However, that does not preclude knowledge from being reducible to a combination of belief and justification. Indeed, as we observe in the companion paper, without some constraints, knowledge can be reduced to a combination of belief and justification. For example, if J satisfies all the axioms of S5 and the axioms L1 and L2 with K replaced by J, then we can reduce K to J by the definition $Kp \leftrightarrow Jp$. We now provide an arguably reasonable condition on a logic Λ of belief and justification that suffices to guarantee that knowledge is not reducible to belief and justification in Λ. Roughly speaking, the condition says that the interaction between B and J is rather weak. We give two interpretations of this condition. The first is semantic, and is expressed in terms of algebras. It requires that every $(\mathrm{KD45})_B$-algebra be extendible to an algebra of $\Lambda \cap \mathcal{L}(B, J)$; intuitively, it says that the properties of J do not put any constraints on B.

[4] Recall that a *filter* $\mathcal{C}$ in $\mathcal{B}$ is a set of sets in $\mathcal{B}$ that is closed under supersets and intersection (so that if $E_1, E_2 \in \mathcal{C}$ and $E_1 \subseteq E_3$, then $E_3 \in \mathcal{C}$ and $E_1 \cap E_2 \in \mathcal{C}$); the filter $\mathcal{C}$ is *proper* if $\emptyset \notin \mathcal{C}$; it is *nonprincipal* if there is no $E \in \mathcal{B}$ such that $\mathcal{C}$ consists of all supersets of E; an *ultrafilter* is a maximal proper filter.

Theorem 5.5 *Let Λ be a logic in $\mathcal{L}(B, J, K)$ such that $(\mathrm{KD45})_B + (\mathrm{S5})_K + \{\mathrm{L1}, \mathrm{L2}\} \subseteq \Lambda$. If every $(\mathrm{KD45})_B$-algebra can be extended to an algebra of $\Lambda \cap \mathcal{L}(B, J)$, then K is not reducible to B and J in Λ.*

Obviously, our previous example that shows how S5 knowledge *can* be reduced to belief and justification must fail the stipulation of Theorem 5.5. That is, it must be the case that some $(\mathrm{KD45})_B$-algebra cannot be extended to a $\Lambda \cap \mathcal{L}(B, J)$-algebra. But the operator J in our example is just an S5 knowledge operator, and thus it must be the case that there is a $(\mathrm{KD45})_B$-algebra that cannot be extended to an algebra of belief and S5 knowledge. But this is precisely what is shown in Theorem 5.4. Theorem 5.4 not only shows that certain logics do not satisfy the antecedent of Theorem 5.5, but is actually the key to its proof.

The following corollary gives a syntactic version of the statement that the interaction between B and J be weak. It says 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 $\mathrm{S5}_J$).

Corollary 5.6 *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.*

Thus, as long as there is no axiomatic link between belief and justification, and justification does not have any properties that go beyond S5, then knowledge is not reducible to a combination of belief and justification. See the companion paper for further discussion of this issue.

5.4 Interpolation

A logic Λ has the *interpolation property* if for any formula $\varphi_1 \to \varphi_2$ in Λ, there exists a formula χ (an *interpolant*) whose non-logical constants are common to φ_1 and φ_2 such that $\varphi_1 \to \chi$ and $\chi \to \varphi_2$ are in Λ. Craig's interpolation theorem states that first-order logic has the interpolation property. The interpolation property is used in the proof of Beth's theorem to show that implicit definability implies explicit definability. The proof makes use of the Deduction Theorem, which says that for a set of sentences Γ and a sentence φ, if ψ is in the logic generated by $\Gamma \cup \{\varphi\}$, then $\varphi \to \psi$ is in the logic generated by Γ.

In the case of modal logic, Andréka, van Bentham, and Németi [1998] sketch a proof for the following interpolation theorem. Let Λ be the minimal normal modal logic in some multimodal language $\mathcal{L}$.[5] If $\varphi_1 \to \varphi_2 \in \Lambda$, then there exists a formula χ that contains only modalities that are contained in both φ_1 and φ_2 such that $\varphi_1 \to \chi$ and $\chi \to \varphi_2$ are in Λ. However, implicit definability in normal multimodal logics does not imply explicit definability. The failure of Beth's theorem for such logics is due to the fact that there is no Deduction Theorem for modal logics.

Craig's interpolation theorem for first-order logic can be generalized as follows. Let $\mathcal{L}_1$ and $\mathcal{L}_2$ be two first-order languages, and let Λ be a logic in the language $\mathcal{L} = \mathcal{L}_1 \cup \mathcal{L}_2$. If $\varphi_1 \to \varphi_2 \in \Lambda$, then there exists a formula $\chi \in \mathcal{L}_1 \cap \mathcal{L}_2$ such that $\varphi_1 \to \chi$ and $\chi \to \varphi_2$ are in Λ. This result also makes use of the Deduction Theorem.

Again, in the modal case this generalization does not hold. The logic $(\mathrm{KD45})_B + (\mathrm{S5})_K + \{\mathrm{L1}, \mathrm{L2}\}$, discussed in Subsection 5.1, demonstrates this, and highlights the difference between an orthodox modal logic and the corresponding first-order logic. For $i = 1, 2$, consider the multimodal logic $\Lambda_i = (\mathrm{KD45})_B + (\mathrm{S5})_{K_i} + \{\mathrm{L1}, \mathrm{L2}\}$ in the language $\mathcal{L}(B, K_i)$. Let $\Lambda = \Lambda_1 + \Lambda_2$. By Proposition 3.3(b), $K_1 p \leftrightarrow K_2 p \in \Lambda$. However, there is no formula χ in the language $\mathcal{L}(B, K_1) \cap \mathcal{L}(B, K_2) = \mathcal{L}(B)$ such that $K_1 p \to \chi$ and $\chi \to K_2$ are in Λ. Indeed, if such a formula χ existed, then $K_1 p \leftrightarrow \chi$ would be in Λ, because $K_2 p \to K_1 p \in \Lambda$. Let $\mathcal{F} = (W, R_{K_1}, R_B)$ be the canonical frame for Λ_1. Obviously, the frame $\hat{\mathcal{F}} = (W, R_{K_1}, R_{K_2}, R_B)$, where $R_{K_2} = R_{K_1}$, is a frame for which $\Lambda_1 + \Lambda_2$ is sound. Thus, $K_1 p \leftrightarrow \varphi$ is valid in $\hat{\mathcal{F}}$. But the interpretation of this formula depends only on R_{K_1} and R_B. Thus, it is also valid in $\mathcal{F}$. Hence, this formula is in Λ_1, which means that it is a definition of K_1 in terms of B, contrary to Proposition 3.3(b). The corresponding first-order logic Λ^{fo} includes the formula $R^{\mathrm{fo}}_{K_1}(x, y) \leftrightarrow R^{\mathrm{fo}}_{K_2}(x, y)$ and, by Proposition 5.2, the formula $\exists z (R^{\mathrm{fo}}_B(x, z) \wedge R^{\mathrm{fo}}_B(y, z))$ is an interpolant for this equivalence. However, this interpolant has no modal equivalent.

A Proofs

In this appendix, we prove all results whose proof was omitted from the main text. We repeat the statement of the results for the reader's convenience. We start with the proof of two elementary lemmas that are used in many of the proofs.

[5] "Minimal" here means that it is the least set of formulas that contains all tautologies of propositional logic, is closed under modus ponens and substitution, and is normal.

Lemma A.1 *Let φ, ψ and χ be formulas in a language $\mathcal{L}$, and let χ' be a formula obtained by replacing some occurrences of φ in χ by ψ. If Λ is a logic in $\mathcal{L}$ such that $\varphi \leftrightarrow \psi \in \Lambda$ then $\chi \leftrightarrow \chi' \in \Lambda$.*

We omit the simple proof by induction on the structure of χ.

Given a definition DM_n, we construct a map $\varphi \mapsto \varphi^t$ that translates formulas φ in $\mathcal{L}$ to formulas φ^t in $\mathcal{L}_0$. We define φ^t by induction on the structure of φ. If φ is a primitive proposition, then $\varphi^t = \varphi$. We define $(\varphi \to \psi)^t = (\varphi^t \to \psi^t)$ and $(\neg\varphi)^t = \neg\varphi^t$. For $M_i \neq M_n$, $(M_i\varphi)^t = M_i\varphi^t$, and $(M_n\varphi)^t = \delta[p/\varphi^t]$.

Lemma A.2 *For each formula $\varphi \in \mathcal{L}$, the formula $\varphi \leftrightarrow \varphi^t$ is in every logic that contains* DM_n.

Proof: Let Λ be a logic that contains DM_n. The proof that $\varphi \leftrightarrow \varphi^t \in \Lambda$ proceeds by induction on the structure of formulas, using Lemma A.1. For the case of formulas $M_n\varphi$ we use the assumption that $\mathrm{DM}_n \in \Lambda$. From this it follows by substitution that $M_n\varphi \leftrightarrow \delta[p/\varphi] \in \Lambda$ for each φ. By the induction hypothesis and Lemma A.1, $\delta[p/\varphi] \leftrightarrow \delta[p/\varphi^t] \in \Lambda$, which implies that $(M_n\varphi)^t \leftrightarrow M_n\varphi \in \Lambda$. ∎

Proposition 3.1: *If M_n is reducible to $M_1, \ldots, M_{n-1}$ in Λ, then it is reducible by a simple definition.*

Proof: Suppose that M_n is reducible to $M_1, \ldots, M_{n-1}$ in Λ by the definition $M_np \leftrightarrow \delta$. Let DM'_n be the formula $M_np \leftrightarrow \delta'$, where δ' is the formula obtained by substituting p for all primitive propositions in δ. By substitution, $\mathrm{DM}'_n \in \Lambda_0 + \mathrm{DM}_n$. Thus $\Lambda_0 + \mathrm{DM}'_n \subseteq \Lambda_0 + \mathrm{DM}_n$. It follows that $\delta \leftrightarrow \delta' \in \Lambda_0 + \mathrm{DM}_n$. But $\delta \leftrightarrow \delta' \in \mathcal{L}_0$. Hence, since $\Lambda_0 + DM$ is a conservative extension of Λ_0, $\delta \leftrightarrow \delta' \in \Lambda_0$. This implies that $\mathrm{DM}_n \in \Lambda_0 + \mathrm{DM}'_n$, and hence $\Lambda_0 + \mathrm{DM}_n \subseteq \Lambda_0 + \mathrm{DM}'_n$. Therefore, $\Lambda_0 + \mathrm{DM}_n = \Lambda_0 + \mathrm{DM}'_n$. ∎

In the proofs of the following two theorems, we write Λ' for $\Lambda[M_n/M'_n]$.

Theorem 3.4: *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.*

Proof: Let DM'_n be the formula $M'_np \leftrightarrow \delta$ that results from replacing M_n by M'_n in DM_n. Suppose that M_n is explicitly defined in Λ by DM_n. We first show that $\Lambda = \Lambda_0 + \mathrm{DM}_n$. By definition, $\Lambda_0 \subseteq \Lambda$ and, by assumption, $\mathrm{DM}_n \in \Lambda$. Thus, $\Lambda_0 + \mathrm{DM}_n \subseteq \Lambda$. For the opposite inclusion, let $\varphi \in \mathcal{L}$. By Lemma A.2 and the explicit definability of M_n, $\varphi \leftrightarrow \varphi^t \in \Lambda_0 + \mathrm{DM}_n \subseteq \Lambda$. If $\varphi \in \Lambda$, then $\varphi^t \in \Lambda$, so $\varphi^t \in \Lambda_0$. It follows that $\varphi \in \Lambda_0 + \mathrm{DM}_n$, proving that $\Lambda \subseteq \Lambda_0 + \mathrm{DM}_n$, as desired. It immediately follows that M_n is reducible to $M_1, \ldots, M_{n-1}$ in Λ.

To see that M_n is implicitly defined in Λ, note that $\mathrm{DM}_n \in \Lambda$ and similarly $\mathrm{DM}'_n \in \Lambda'$; thus, $M_n p \leftrightarrow M'_n p \in \Lambda + \Lambda'$.

Now suppose that M_n is reducible to $M_1, \ldots, M_{n-1}$ in Λ by the definition DM_n and that M_n is implicitly defined in Λ. Consider the set Λ^* of formulas in $\mathcal{L}(M_1, \ldots, M_n, M'_n)$ defined by $\Lambda^* = \{\varphi : \varphi^t \in \Lambda'\}$. Here, φ^t is the translation of φ to the language $\mathcal{L}(M_1, \ldots, M_{n-1}, M'_n)$ using DM_n. Clearly $\Lambda' \subseteq \Lambda^*$. As we now show, we also have $\Lambda \subseteq \Lambda^*$. Indeed, if $\varphi \in \Lambda$, then by reducibility, $\varphi \in \Lambda_0 + \mathrm{DM}_n$. Since, by Lemma A.2, $\varphi \leftrightarrow \varphi^t \in \Lambda_0 + \mathrm{DM}_n$, it follows that $\varphi^t \in \Lambda_0 + \mathrm{DM}_n$. Since $\Lambda_0 + \mathrm{DM}_n$ is a conservative extension of Λ_0, $\varphi^t \in \Lambda_0 \subseteq \Lambda'$. Since $\varphi^t \in \Lambda'$, it follows that $\varphi \in \Lambda^*$, as desired.

We prove below that Λ^* is a logic. Therefore $\Lambda + \Lambda' \subseteq \Lambda^*$. Since, by assumption, M_n is implicitly defined in Λ, $M_n p \leftrightarrow M'_n p \in \Lambda^*$. Clearly $\mathrm{DM}_n \in \Lambda^*$, since $\mathrm{DM}_n^t = (\delta \leftrightarrow \delta) \in \Lambda'$. Thus, by the equivalence of M_n and M'_n, we must have $\mathrm{DM}'_n \in \Lambda^*$. But $(\mathrm{DM}'_n)^t = \mathrm{DM}'_n$, and thus $\mathrm{DM}'_n \in \Lambda'$. It follows that $\mathrm{DM}_n \in \Lambda$, as desired.

It remains to show that Λ^* is a logic. Since Λ^* contains the logics Λ and $\Lambda'[M_n/M'_n]$, it contains all tautologies of propositional logic as well as the axiom K_M for each modal operator $M \in \{M_1, \ldots, M_n, M'_n\}$.

To see that Λ^* is closed under modus ponens, suppose that $\varphi, \varphi \to \psi \in \Lambda^*$. But then φ^t and $(\varphi \to \psi)^t = \varphi^t \to \psi^t$ are in Λ'. Thus, $\psi^t \in \Lambda'$, 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, $\psi^t \in \Lambda'$. Moreover, $(M\psi)^t = M(\psi^t)$, so $M\psi^t \in \Lambda'$ by the generalization rule for M in Λ'. 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 $\delta[p/\psi^t] \in \Lambda[M_n/M'_n]$. Since $\psi^t \in \Lambda[M_n/M'_n]$, it follows that $\psi^t \leftrightarrow \mathit{true} \in \Lambda[M_n/M'_n]$. By Lemma A.1, $\delta[p/\psi^t] \leftrightarrow \delta[p/\mathit{true}] \in \Lambda[M_n/M'_n]$. Thus, to complete the proof, we need to show that $\delta[p/\mathit{true}] \in \Lambda[M_n/M'_n]$. $M_n \mathit{true} \in \Lambda$ by generalization, so by reducibility, $M_n \mathit{true} \in \Lambda_0 + \mathrm{DM}_n$. Moreover, $M_n \mathit{true} \leftrightarrow \delta[p/\mathit{true}] \in \Lambda_0 + \mathrm{DM}_n$, so $\delta[p/\mathit{true}] \in \Lambda_0 + \mathrm{DM}_n$. But $\Lambda_0 + \mathrm{DM}_n$ is a conservative extension of Λ_0, so $\delta[p/\mathit{true}] \in \Lambda_0 \subseteq \Lambda[M_n/M'_n]$. ∎

Theorem 4.1: *The following are equivalent:*

(a) the modality M_n is implicitly defined in Λ;

(b) if $\mathcal{A} = (\mathcal{A}_0, \mathbf{M}_n)$ and $\mathcal{A}' = (\mathcal{A}_0, \mathbf{M}'_n)$ are Λ-algebras, then $\mathbf{M}_n = \mathbf{M}'_n$.

Proof: To prove that (a) implies (b), suppose that M_n is implicitly defined in Λ, and that both $\mathcal{A} = (\mathcal{A}_0, \mathbf{M}_n)$ and $\mathcal{A}' = (\mathcal{A}_0, \mathbf{M}'_n)$ are Λ-algebras. Let $\mathcal{A} + \mathcal{A}'$ denote the algebra $(\mathcal{A}_0, \mathbf{M}_n, \mathbf{M}'_n)$. Clearly all the formulas in $\Lambda \cup \Lambda'$ are valid in $\mathcal{A} + \mathcal{A}'$; moreover, the set of formulas valid in an algebra is easily seen to be closed under substitution and generalization, so all the formulas in $\Lambda + \Lambda'$ are also valid in $\mathcal{A} + \mathcal{A}'$. Since M_n is implicitly defined in Λ, it follows that $M_n\varphi \leftrightarrow M'_n\varphi \in \mathrm{Th}(\mathcal{A} + \mathcal{A}')$. Now suppose, by way of contradiction, that $\mathbf{M}_n \neq \mathbf{M}'_n$. Then for some x, $\mathbf{M}_n(x) \neq \mathbf{M}'_n(x)$. Consider the $\mathcal{A}+\mathcal{A}'$-model $\mathcal{M} = ((\mathcal{A}_0, \mathbf{M}_n, \mathbf{M}'), V)$ where $V(p) = x$. Clearly $[\![M_n p]\!]_{\mathcal{M}} \neq [\![M_n p]\!]_{\mathcal{M}}$, giving the desired contradiction.

To show that (b) implies (a), suppose that (b) holds. Let $\mathcal{A} = (\mathcal{A}_0, \mathbf{M}_n, \mathbf{M}'_n)$ be the canonical $(\Lambda + \Lambda')$-algebra, where now $\mathbf{M}'_n$ is taken to be the interpretation of M'_n. Note that $\varphi \in \Lambda \cup \Lambda'$ iff $\varphi \in \mathrm{Th}(\mathcal{A})$. We can view both $(\mathcal{A}_0, \mathbf{M}_n)$ and $(\mathcal{A}_0, \mathbf{M}'_\mathbf{n})$ as Λ-algebras, by taking $\mathbf{M}'_n$ to interpret M_n. By assumption, $\mathbf{M}_n = \mathbf{M}'_n$. Thus, $M_n\varphi \leftrightarrow M'_n\varphi$ must be valid in the canonical $(\Lambda + \Lambda')$-algebra for all formulas φ. Thus, $M_n\varphi \leftrightarrow M'_n\varphi \in \Lambda + \Lambda'$, so M_n is implicitly defined in Λ. ∎

To prove Theorem 4.2, we need the the following lemma, whose straightforward proof is omitted.

Lemma A.3 *Let Φ be the set of all formulas in $\mathcal{L}_0$ that contain only the primitive proposition p and let $\mathcal{A}_0$ be a Λ_0-algebra. There exists a unique function $\varphi \mapsto \varphi^{op}$ from Φ to $\mathcal{O}^*_{\mathcal{A}_0}$ that satisfies the following: p^{op} is the identity operator; for each $\varphi, \psi \in \Phi$, $(\neg\varphi)^{op} = \neg\varphi^{op}$ and $(\varphi \vee \psi)^{op} = \varphi^{op} \vee \psi^{op}$; for each $i = 1, \ldots, n-1$, $(M_i\varphi)^{op} = \mathbf{M}_i \circ \varphi^{op}$. Moreover, this function is a surjection onto $\mathcal{O}^*_{\mathcal{A}_0}$, and, for all $\varphi \in \Phi$, $\psi \in \mathcal{L}_0$, and models $\mathcal{M}_0 = (\mathcal{A}_0, V)$, $[\![\varphi[p/\psi]]\!]_{\mathcal{M}_0} = \varphi^{op}([\![\psi]\!])_{\mathcal{M}_0}$.*

Theorem 4.2: *The modality M_n is explicitly defined in Λ if and only if, for each Λ-algebra $\mathcal{A}$, $\mathcal{O}^*_{\mathcal{A}} = \mathcal{O}^*_{\mathcal{A}_0}$.*

Proof: We first note that $\mathcal{O}^*_{\mathcal{A}} = \mathcal{O}^*_{\mathcal{A}_0}$ if and only if $\mathbf{M}_n \in \mathcal{O}^*_{\mathcal{A}_0}$. Assume that the modality M_n is explicitly defined in Λ via the definition $M_n p \leftrightarrow \delta$, and let $\mathcal{A}$ be a Λ-algebra. Thus, for each model $\mathcal{M} = (\mathcal{A}, V)$, $\mathbf{M}_n([\![p]\!]_{\mathcal{M}}) = [\![\delta]\!]_{\mathcal{M}}$. By Lemma A.3, $[\![\delta]\!]_{\mathcal{M}} = \delta^{op}([\![p]\!]_{\mathcal{M}})$. For all $x \in \mathcal{A}$, let $\mathcal{M}_x = (\mathcal{A}, V_x)$ be a model such that $V(p) = x$. It is easy to check that $\mathbf{M}_n(x) = \delta^{op}(x)$. Thus, $\mathbf{M}_n = \delta^{op}$, and hence $\mathcal{O}^*_{\mathcal{A}} = \mathcal{O}^*_{\mathcal{A}_0}$.

For the converse, suppose that for each Λ-algebra $\mathcal{A}$, $\mathcal{O}^*_{\mathcal{A}} = \mathcal{O}^*_{\mathcal{A}_0}$. Let $\mathcal{A}$ be the canonical algebra of Λ. By Lemma A.3, there exists a formula $\delta \in \Phi$ such that $\mathbf{M}_n = \delta^{op}$. Moreover, for each model $\mathcal{M}$ based on $\mathcal{A}$,

$\mathbf{M}_n([\![p]\!]_{\mathcal{M}}) = \delta^{op}([\![p]\!]_{\mathcal{M}}) = [\![\delta]\!]_{\mathcal{M}}$. Thus, $M_n p \leftrightarrow \delta$ is valid in each model based on $\mathcal{A}$, and hence $M_n p \leftrightarrow \delta \in \Lambda$. ∎

To prove Theorem 4.3, we need the following lemma, which will also be useful in our later proofs.

Lemma A.4 *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 (algebra) in $\mathcal{S}$ can be extended to a Λ_2-frame (algebra), then Λ_2 is a conservative extension of Λ_1.*

Proof: Suppose that the condition in the lemma holds. Let $\mathcal{F}$ be a Λ_1-frame in $\mathcal{S}$ and $\mathcal{F}'$ an extension of $\mathcal{F}$ to a Λ_2-frame. Consider models $\mathcal{M} = (\mathcal{F}, V)$ and $\mathcal{M}' = (\mathcal{F}', V)$. Suppose that $\varphi \in \Lambda_2 \cap \mathcal{L}_1$. Then $\mathcal{M}' \models \varphi$. Since $\varphi \in \mathcal{L}_1$, it follows that $[\![\varphi]\!]_{\mathcal{M}'} = [\![\varphi]\!]_{\mathcal{M}}$. Thus, $\mathcal{M} \models \varphi$. Since this is true for any model based on a frame in $\mathcal{S}$, $\varphi \in \Lambda_1$, and hence $\Lambda_2 \cap \mathcal{L}_1 \subseteq \Lambda_1$. The converse inclusion holds since $\Lambda_1 \subseteq \Lambda_2$. The proof for algebras is similar. ∎

Theorem 4.3: *The modality M_n is reducible to $M_1, \ldots, M_{n-1}$ in Λ if and only if each Λ_0-algebra $\mathcal{A}_0$ can be extended to a Λ-algebra $\mathcal{A}$ such that $\mathcal{O}^*_{\mathcal{A}} = \mathcal{O}^*_{\mathcal{A}_0}$.*

Proof: Suppose that M_n is reducible to $M_1, \ldots, M_{n-1}$ in Λ by DM_n, which is $M_n p \leftrightarrow \delta$. Let $\mathcal{A}_0$ be a Λ_0-algebra. Extend $\mathcal{A}_0$ to $\mathcal{A}$ by defining $\mathbf{M}_n = \delta^{op}$. Thus, $\mathcal{O}^*_{\mathcal{A}} = \mathcal{O}^*_{\mathcal{A}_0}$, and we need only show that $\mathcal{A}$ is a Λ-algebra. Suppose that $\varphi \in \Lambda$. By reducibility, $\varphi \in \Lambda_0 + \mathrm{DM}_n$. By Lemma A.2, $\varphi \leftrightarrow \varphi^t \in \Lambda_0 + \mathrm{DM}_n$. Thus, $\varphi^t \in \Lambda_0 + \mathrm{DM}_n$. Since $\Lambda_0 + \mathrm{DM}_n$ is a conservative extension of Λ_0, it follows that $\varphi^t \in \Lambda_0$. Consider a model $\mathcal{M} = (\mathcal{A}, V)$ and the model $\mathcal{M}_0 = (\mathcal{A}_0, V)$. Since $\mathcal{M}$ and $\mathcal{M}_0$ agree on formulas in Λ_0, and $\mathcal{M}_0$ is a model of Λ_0, we must have $[\![\varphi^t]\!]_{\mathcal{M}} = 1$. It thus suffices to show that for every formula $\varphi \in \mathcal{L}$, $[\![\varphi]\!]_{\mathcal{M}} = [\![\varphi^t]\!]_{\mathcal{M}}$. This is proved by induction on the structure of φ. We show here only the case that $\varphi = \mathcal{M}_n\psi$. In this case, $(\mathcal{M}_n\psi)^t = \delta[p/\psi^t]$. By Lemma A.3, $[\![\delta[p/\psi^t]]\!]_{\mathcal{M}} = \delta^{op}([\![\psi^t]\!]_{\mathcal{M}})$. By the induction hypothesis, this is $\delta^{op}([\![\psi]\!]_{\mathcal{M}})$, which is $[\![M_n\psi]\!]_{\mathcal{M}}$.

For the converse, suppose that the condition in the theorem holds. Let $\mathcal{A}_0$ be the canonical algebra of Λ_0 and $\mathcal{A}$ its extension to a Λ-algebra that satisfies $\mathcal{O}^*_{\mathcal{A}} = \mathcal{O}^*_{\mathcal{A}_0}$. Then $\mathbf{M}_n \in \mathcal{O}^*_{\mathcal{A}_0}$, and hence there exists a formula $\delta \in \Phi$ such that $\mathbf{M}_n = \delta^{op}$. Let DM_n be the formula $M_n p \leftrightarrow \delta$.

We show first that $\mathcal{A}$ is a $\Lambda_0 + \mathrm{DM}_n$-algebra. Since $\mathrm{Th}(\mathcal{A})$ is a logic (the theory of any algebra is a logic), it suffices to show that $\mathrm{DM}_n \in \mathrm{Th}(\mathcal{A})$.

To see that this is the case, note that if $\mathcal{M}$ is a model based on $\mathcal{A}$, then $\mathbf{M}_n([\![p]\!]_{\mathcal{M}}) = \delta^{op}([\![p]\!]_{\mathcal{M}})$, which, by Lemma A.3, is $[\![\delta]\!]_{\mathcal{M}}$.

Since Λ_0 is complete for $\mathcal{A}_0$, it follows by Lemma A.4 that $\Lambda_0 + \mathrm{DM}_n$ is a conservative extension of Λ_0. It remains to show that $\Lambda \subseteq \Lambda_0 + \mathrm{DM}_n$. Suppose that $\varphi \in \Lambda$. Then, for any model $\mathcal{M} = (\mathcal{A}, V)$, $[\![\varphi]\!]_{\mathcal{M}} = 1$. By Lemma A.2, $\varphi \leftrightarrow \varphi^t \in \Lambda_0 + \mathrm{DM}_n$. Since $\mathcal{A}$ is a $(\Lambda_0 + \mathrm{DM}_n)$-algebra, $[\![\varphi]\!]_{\mathcal{M}} = [\![\varphi^t]\!]_{\mathcal{M}}$. But $[\![\varphi^t]\!]_{\mathcal{M}} = [\![\varphi^t]\!]_{\mathcal{M}_0}$ for the model $\mathcal{M}_0 = (\mathcal{A}_0, V)$. Thus, φ^t is valid in every model based on $\mathcal{A}_0$. Since $\mathcal{A}_0$ is canonical, it follows that $\varphi^t \in \Lambda_0$. Since $\varphi \leftrightarrow \varphi^t \in \Lambda_0 + \mathrm{DM}_n$, we have that that $\varphi \in \Lambda_0 + \mathrm{DM}_n$. ∎

Theorem 4.4: *For each Λ-algebra $\mathcal{A} = (\mathcal{B}, \vee, \neg, 1, \mathbf{M}_1, \ldots, \mathbf{M}_n)$, $\mathcal{O}^*_{\mathcal{A}} = \mathcal{O}^*_{\mathcal{A}_0}$ iff (a) every Λ_0-algebra $\mathcal{A}_0$ can be extended to a Λ-algebra $\mathcal{A}$ such that $\mathcal{O}^*_{\mathcal{A}} = \mathcal{O}^*_{\mathcal{A}_0}$ and (b) if $\mathcal{A} = (\mathcal{A}_0, \mathbf{M}_n)$ and $\mathcal{A}' = (\mathcal{A}_0, \mathbf{M}'_n)$ are Λ-algebras, then $\mathbf{M}_n = \mathbf{M}'_n$.*

Proof: The proof follows from Theorem 3.4, using Theorems 4.1, 4.2, and 4.3. Nevertheless, we prove here that (a) and (b) imply the first condition because the proof is simpler than the syntactic proof that reducibility and implicit definability imply explicit definability.

Suppose that (a) M_n is reducible to $M_1, \ldots, M_{n-1}$ in Λ by the definition DM_n, and (b) M_n is implicitly defined in Λ. By Theorem 4.2 we need to show that for each Λ-algebra $\mathcal{A}$, $\mathbf{M}_n \in \mathcal{O}^*_{\mathcal{A}_0}$. Let $\mathcal{A}$ be a Λ-algebra. By (a) and Theorem 4.3, applied to the logic $\Lambda' = \Lambda[\mathbf{M}_n/\mathbf{M}'_n]$, the algebra $\mathcal{A}_0$ can be extended to an algebra $\mathcal{A}'$ with operators $\mathbf{M}_1, \ldots, \mathbf{M}_{n-1}\mathbf{M}'_n$, such that $\mathbf{M}'_n \in \mathcal{A}^*_0$. By (b) and Theorem 4.1, $\mathbf{M}_n = \mathbf{M}'_n$, which completes the proof. ∎

Theorem 5.1: *If Λ is an orthodox logic in the language $\mathcal{L}(M_1, \ldots, M_n)$, then the following are equivalent:*

(a) the modality M_n is implicitly defined in Λ;

(b) for all Λ-frames $(\mathcal{F}_0, R_n)$ and $(\mathcal{F}_0, R'_n)$, we have $R_n = R'_n$;

(c) the predicate R^{fo}_n is implicitly defined in Λ^{fo};

(d) the predicate R^{fo}_n is explicitly defined in Λ^{fo}.

Proof: As noted in Section 5, we need to show only that (a) is equivalent to (b). Suppose that (a) holds and that $\mathcal{F} = (\mathcal{F}_0, R_n)$ and $\mathcal{F}' = (\mathcal{F}_0, R_n)$ are both Λ-frames. We can view $\mathcal{F} + \mathcal{F}' = (\mathcal{F}_0, R_n, R'_n)$ as a $\Lambda + \Lambda'$-frame by taking R'_n to be the interpretation of M'_n. (It is easy to check that $\mathrm{Th}(\mathcal{F} + \mathcal{F}')$

is normal.) Since M_n is implicitly defined in Λ, we must have $M_n\varphi \leftrightarrow M'_n\varphi \in \mathrm{Th}(\mathcal{F} + \mathcal{F}')$ for all formulas. This implies that $R_n = R'_n$, because if $R_n(w) \neq R'_n(w)$ for some w, then, without loss of generality, $R_n(w) \not\subseteq R'_n(w)$ and therefore $M_n p \leftrightarrow M'_n p$ is not valid in a model $(\mathcal{A} + \mathcal{A}', V)$ where $V(p) = R'_n(w)$. Now suppose that (b) holds. Then $M_n\varphi \leftrightarrow M'_n\varphi$ holds for all formulas φ in all $(\Lambda + \Lambda')$-frames. As Λ is orthodox, so is $(\Lambda + \Lambda')$, and therefore it is complete with respect to its canonical frame. Thus, $M_n\varphi \leftrightarrow M'_n\varphi \in \Lambda + \Lambda'$. ∎

Proposition 5.2: *The formula* $R_K^{\mathrm{fo}}(x, y) \leftrightarrow \exists z(R_B^{\mathrm{fo}}(x, z) \wedge R_B^{\mathrm{fo}}(y, z))$ *is valid in all* $((\mathrm{KD45})_B + (\mathrm{S5})_K + \{\mathrm{L1}, \mathrm{L2}\})$*-frames.*

Proof: It is well known [Hoek 1993] that $(\mathrm{KD45})_B + (\mathrm{S5})_K + \{\mathrm{L1}, \mathrm{L2}\}$ is sound and complete with respect to frames where (1) the R_B relation is serial, transitive, and Euclidean;[6] (2) the R_K is an equivalence relation; (3) $R_B \subseteq R_K$; and (4) for all x, y, and z in W, if $(x, y) \in R_K$ and $(y, z) \in R_B$, then $(x, z) \in R_B$ [Hoek 1993]. The last two conditions correspond to L1 and L2, respectively.

Let (W, R_B, R_K) be a $((\mathrm{KD45})_B + (\mathrm{S5})_K + \{\mathrm{L1}, \mathrm{L2}\})$-frame. If $(x, y) \in R_K$ then, since R_B is serial, there exists some z such that $(y, z) \in R_B$. By the semantic condition corresponding to L2, we also have $(x, z) \in R_B$. For the converse, suppose that there exists some z such that $(x, z) \in R_B$ and $(y, z) \in R_B$. Then, by the semantic condition corresponding to L1, $(x, z) \in R_K$ and $(y, z) \in R_K$. By the symmetry and transitivity of R_K, $(x, y) \in R_K$. ∎

Theorem 5.4: *The algebra* $\mathcal{A} = (W, \mathcal{B}, \cup, \neg, W, \mathbf{B})$ *is a* $\mathrm{KD45}_B$*-algebra that cannot be extended to a* $(\mathrm{KD45}_B + \mathrm{S5}_K + \{$*L1,L2*$\})$*-algebra.*

Proof: We first show that $\mathcal{A}$ is a $(\mathrm{KD45})_B$-algebra.

In order to see that axiom K_B is valid in $\mathcal{A}$, we need to show that for each E and F in $\mathcal{B}$, $\neg\mathbf{B}(\neg E \cup F) \cup (\neg\mathbf{B}(E) \cup \mathbf{B}(F)) = W$, or equivalently, $\mathbf{B}(\neg E \cup F) \subseteq \neg\mathbf{B}(E) \cup \mathbf{B}(F)$. The left and right sides of this inclusion can differ only by 0. Suppose that $0 \in \neg\mathbf{B}(E \cup \neg F)$. Then it must be the case that $\neg E \cup F \in \mathcal{U}$. Now either $E \notin \mathcal{U}$, in which case $0 \in \neg\mathbf{B}(E)$ and we are done, or $E \in \mathcal{U}$. In the latter case, since $\mathcal{U}$ is closed under intersection, it follows that $(\neg E \cup F) \cap E = F \cap E \in \mathcal{U}$, and thus F must be in $\mathcal{U}$, and $0 \in \mathbf{B}(F)$. In either case, it follows that $0 \in \neg\mathbf{B}(E) \cup \mathbf{B}(F)$, as desired.

For axiom D_B, we need to show that for each set E, $\mathbf{B}(E) \subseteq \neg\mathbf{B}(\neg E)$. Again, the two sides of the inclusion can differ only by 0. If $0 \in \mathbf{B}(E)$ then

[6] 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 $(y, z) \in R$.

$E \in \mathcal{U}$. But then $\neg E \notin \mathcal{U}$, It easily follows that $0 \in \neg \mathbf{B}(\neg E)$.

Axiom 4_B requires that $\mathbf{B}(E) \subseteq \mathbf{B}(\mathbf{B}(E))$. If $0 \in \mathbf{B}(E)$ then $E \in \mathcal{U}$, and $\mathbf{B}(E) = E \cup \{0\}$, which is also in $\mathcal{U}$. Hence, $\mathbf{B}(\mathbf{B}(E)) = \mathbf{B}(E \cup \{0\}) = E \cup \{0\}$.

For 5_B, we have to prove that $\neg \mathbf{B}(E) \subseteq \mathbf{B}(\neg \mathbf{B}(E))$. If $0 \in \neg \mathbf{B}(E)$ then $E \notin \mathcal{U}$, and $\neg \mathbf{B}(E) = \neg E \cup \{0\}$. It follows that both $\neg E$ and hence $\neg E \cup \{0\}$ are in $\mathcal{U}$, and therefore $\mathbf{B}(\neg E \cup \{0\}) = \neg E \cup \{0\}$. This complete the proof that $\mathcal{A}$ is a $(\mathrm{KD45})_B$-algebra.

Suppose, by way of contradiction, that $\mathcal{A}$ can be extended to a $((\mathrm{KD45})_B + (\mathrm{S5})_K + \{\mathrm{L1}, \mathrm{L2}\})$-algebra $(\mathcal{A}, \mathbf{K})$. Let $E = W \setminus \{0\}$. We first show that $\mathbf{K}(E) = E$. By T_K, it is enough to show that $E \subseteq \mathbf{K}(E)$. Obviously, for each $x \in E$, $\mathbf{B}(\{x\}) = \{x\}$. By L2 it follows that, for each $x \in E$, $\mathbf{B}(\{x\}) \subseteq \mathbf{K}(\mathbf{B}(\{x\}))$. Substituting $\{x\}$ for $\mathbf{B}(\{x\})$, we conclude that $\{x\} \subseteq \mathbf{K}(\{x\})$ for each $x \in E$. It is easy to see that the validity of axiom K_K implies that $\mathbf{K}$ is monotonic, and hence $\mathbf{K}(\{x\}) \subseteq \mathbf{K}(E)$, from which we conclude as required that $\{x\} \subseteq \mathbf{K}(E)$. Thus, $E \subseteq \mathbf{K}(E)$. Moreover, since $\mathbf{K}(E) = E$, $\neg \mathbf{K}(E) = \{0\}$. By L1, $\mathbf{K}(\{0\}) \subseteq \mathbf{B}(\{0\})$. By the definition of $\mathbf{B}$, $\mathbf{B}(\{0\}) = \emptyset$. Substituting $\neg \mathbf{K}(E)$ for $\{0\}$ in $\mathbf{K}(\{0\})$, we have $\mathbf{K}(\neg \mathbf{K}(\{E\})) = \emptyset \neq \neg \mathbf{K}(\{E\})$, contradicting 5_K. ∎

We remark that Theorem 5.4 shows that the converse of Lemma A.4 does not hold.

Theorem 5.5: *Let Λ be a logic in $\mathcal{L}(B, J, K)$ such that $(\mathrm{KD45})_B + (\mathrm{S5})_K + \{\mathrm{L1}, \mathrm{L2}\} \subseteq \Lambda$. If every $(\mathrm{KD45})_B$-algebra can be extended to an algebra of $\Lambda \cap \mathcal{L}(B, J)$, then K is not reducible to B and J in Λ.*

Proof: Suppose, by way of contradiction, that every $(\mathrm{KD45})_B$-algebra can be extended to an algebra of $\Lambda \cap \mathcal{L}(B, J)$ and that K is reducible to B and J in Λ. Consider the $(\mathrm{KD45})_B$-algebra $\mathcal{A}$ constructed in the proof of Theorem 5.4. By assumption, it can be extended to a $(\Lambda \cap \mathcal{L}(B, J))$-algebra $\mathcal{A}'$. Since K is reducible to B and J, by Theorem 4.3, $\mathcal{A}'$ can be extended to a Λ-algebra $\mathcal{A}''$. In particular, $\mathcal{A}$ is a $((\mathrm{KD45})_B + (\mathrm{S5})_K + \{\mathrm{L1}, \mathrm{L2}\})$-algebra. But this contradicts Theorem 5.4. ∎

Corollary 5.6: *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.*

Proof: It is easy to show that every $(\mathrm{KD45})_B$-algebra $\mathcal{A}'$ can be extended to a $((\mathrm{KD45})_B + \Lambda_J)$-algebra, simply by defining an operator $\mathbf{J}$ on $\mathcal{A}'$ by taking $\mathbf{J}(1) = 1$ and $\mathbf{J}(x) = 0$ if $x \neq 1$. This makes the resulting algebra an $(\mathrm{S5})_J$-algebra. The result now follows from Theorem 5.5. ∎

References

Andréka, H., J. van Benthem, and I. Németi (1998). Modal languages and bounded fragments of predicate logic. *Journal of Philosophical Logic 27*(3), 217–274.

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.

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

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

Jónsson, B. and A. Tarski (1951). Boolean algebras with operators, Part I. *American Journal of Mathematics 73*, 891–939.

Jónsson, B. and A. Tarski (1952). Boolean algebras with operators, Part II. *American Journal of Mathematics 74*, 127–162.

Kracht, M. (1999). *Tools and Techniques in Modal Logic*, Volume 142 of *Studies in Logic and the Foundations of Mathematics*. Amsterdam: Elsevier.

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

Maksimova, L. L. (1992a). An analogue of Beth's theorem in normal extensions of the modal logic K4. *Siberian Mathematical Journal 33*(6), 1052–1065.

Maksimova, L. L. (1992b). Modal logics and varieties of modal algebras: the Beth properties, interpolation, and amalgamation. *Algebra and Logic 31*(2), 90–105.

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