The Blackwell Guide to
Philosophical Logic

Edited by

Lou Goble

Blackwell Publishing

350 Main Street, Malden, MA 02148-5018, USA
108 Cowley Road, Oxford OX4 1JF, UK
550 Swanston Street, Carlton, Victoria 3053, Australia
Kurfürstendamm 57, 10707 Berlin, Germany

First published 2001 by Blackwell Publishers Ltd
Reprinted 2002

Library of Congress Cataloging-in-Publication Data

The Blackwell guide to philosophical logic / edited by Lou Goble.
p. cm. — (Blackwell philosophy guides)
Includes bibliographical references and index.
ISBN 0-631-20692-2 (alk. paper) — ISBN 0-631-20693-0 (pb. : alk. paper)
1. Logic. I. Title: Guide to philosophical logic. II. Goble, Lou. III. Series.

BC71 B565 2001
160—dc21 00-069788

A catalogue record for this title is available from the British Library.

Set in 10 on 12.5pt Galliard
by Graphicraft Ltd., Hong Kong

For further information on
Blackwell Publishing, visit our website:
http://www.blackwellpublishing.com

Contents

Chapter 9

Epistemic Logic

J.-J. Ch. Meyer

9.1. Introduction: A Brief History of Knowledge

Knowledge has been a subject of philosophical study since ancient times. This is not surprising since knowledge is crucial for humans to control their actions and the appetite for acquiring it seems innate to the human race. Philosophy, therefore, has always occupied itself with the question as to the nature of knowledge. This area of philosophy is generally referred to as epistemology from the Greek word for knowledge: *episteme*. Plato defined knowledge as "*justified true belief*," and this definition has influenced philosophers ever since; cf. Gettier (1963) and Pollock (1986). Although sensible, this definition does not yet explain the nature of knowledge, since all of the three notions of 'justification', 'truth', and 'belief' are not yet clear and still subject to discussion. It would go beyond the scope of our purposes here to go into this at this moment, but it is touched on later in this chapter.

Further issues concerning knowledge include the question of how it comes to us. There is the controversy between rationalists, such as Plato and Descartes, who argued that knowledge only comes via reason(ing), and empiricists, such as Locke and Hume, who maintained that knowledge derives from sense experience. Kant considered categories of *analytical* knowledge ('derivable by purely logical argument') versus *synthetic* knowledge (where this is not the case) and of *a posteriori* knowledge (based on experience) versus *a priori* knowledge (where this is not the case), which led to a big debate whether *synthetic a priori* knowledge is possible.

As is the case with so many things, in the twentieth century, the notion of knowledge became amenable to formal-logical analysis. With the development of formal mathematical logic in the second half of the nineteenth century, the formal approach also became available to the study of philosophical notions such as time, necessity, obligation, and also knowledge itself. Most of these logics are collected under the heading of modal logics, namely logics of certain modalities such as necessity and possibility. [See chapter 7.] While (formal) modal logics had been around since the publication of C. I. Lewis' (1912) paper on an axiomatic approach of strict implication, the inception of formal modal *epistemic* logic is often taken to

The Blackwell Guide to Philosophical Logic, First Edition. Edited by Lou Goble.

Published 2001 by Blackwell Publishers Ltd.

be Hintikka's (1962). The period 1912 up to the 1950s are referred to by Bull and Segerberg (1984) as the 'First Wave' of (formal) modal logic, where syntactic and algebraic approaches were prevailing, while the period of roughly 1950–80, where the focus shifted towards model-theoretic semantic approaches, is referred to as the 'Second Wave'. Hintikka's work marks the beginning of this Second Wave. In this chapter, however, epistemic logic is treated as a particular modal logic and models are considered that have become standard for modal logics in general, namely so-called Kripke models, based on the work by Kripke (1963), another leading figure in the Second Wave of formal modal logic. These models employ the notion of a *possible world* dating back to the philosopher and mathematician Leibniz. Carnap, Prior and Kanger also contributed to coining the notion of a possible world model (Bull and Segerberg, 1984).

In the 1980s, computer scientists and researchers in the area of artificial intelligence (AI) picked up the subject of epistemic logic as a means to reason about the knowledge ascribed to processors in processes of computation and that of knowledge-based systems, such as advanced databases, expert systems and so-called agent-oriented systems, respectively. (In an important sense, this work belongs to a kind of 'Third Wave' of modal logic: the use of modal logics in application areas such as computer science, linguistics and AI.) This chapter reviews briefly their contributions to epistemic logic and its application, since these concentrate on slightly different but also quite interesting aspects of knowledge, and their work also, in its turn, has influenced philosophers again. (Moreover, links were established with another interesting area of AI – nonmonotonic reasoning – which has some definite relations with philosophy as well [see chapter 15].)

9.2. The Modal Logic Approach to Knowledge

This section looks at the basic idea behind modal epistemic logic: modeling knowledge or rather ignorance (as shall be seen) by means of accessibility relations as they are present in Kripke models.

To prepare for the formal treatment, first consider the following situation. Imagine a person in Amsterdam wondering what the weather is like in New York (possibly since a friend of his is there on holiday), in particular whether it is raining in New York. Since he has no information pertaining to this (and clearly cannot obtain this information by direct observation – unless he is clairvoyant or has access to an internet site with this information, which is assumed not to be the case), this person will consider two possible situations, one in which it rains in New York, and one in which this is not the case. Note that the lack of knowledge of an agent can be represented as the agent's considering a number of situations as possible. In this example, there are only two such possible situations, resulting from being ignorant about one propositional item, but clearly, if one lacks knowledge about more items, the number of possible situations that are held possible will increase. Generally, if one has ignorance about the truth of n propositional atoms, one has to consider 2^n situations. For example, if one totally lacks knowledge about whether it rains in New

York (p) and whether it rains in Los Angeles (q), one has to reckon with four situations: one in which both p and q are true, one in which p is true and q is false, one in which p is false and q is true, and one in which both p and q are false. Since the situations to be considered stem from (lack of) knowledge, they are called *epistemically* alternative worlds or shortly *epistemic alternatives.*

The idea of considering several epistemic alternatives in case one has not complete knowledge about the situation at hand can be molded perfectly into the framework of Kripke-style possible world semantics. Assume a set $\mathcal{P}$ of propositional atoms. Use the symbols T and F for the truth values (true and false, respectively). Formally a Kripke model is a structure of the form:

Definition 9.1 A Kripke model is a structure $\mathcal{M}$ of the form $\langle S, \pi, R \rangle$, where

- S is a non-empty set (the set of possible worlds);
- $\pi : S \to \mathcal{P} \to \{\mathrm{T}, \mathrm{F}\}$ is a truth assignment function to the atoms per possible world;
- $R \subseteq S \times S$ is the knowledge accessibility relation.

By means of a Kripke model, one can represent exactly what an agent considers as the epistemic alternatives in a certain situation: given a situation (represented again by a possible world $s \in S$), the epistemic alternatives for the agent are given by the set $\{t \in S \mid R(s, t)\}$, i.e. all possible worlds t that are accessible from s by means of the relation R.

Thus the example above can be represented in a Kripke model as follows; see figure 9.1. Suppose that the actual situation at hand (which the agent does not have complete knowledge about) is that it rains in New York but not in LA, represented by a state $s_1 \in S$ for which it holds that $\pi(s_1)(p) = \mathrm{T}$ and $\pi(s_1)(q) = \mathrm{F}$. Now the model can be represented by taking $S = \{s_0, s_1, s_2, s_3\}$, where s_0 is such that $\pi(s_0)(p) = \pi(s_0)(q) = \mathrm{T}$, s_1 is as above, s_2 is such that $\pi(s_2)(p) = \mathrm{F}$ and $\pi(s_2)(q) = \mathrm{T}$, and s_3 is such that $\pi(s_3)(p) = \pi(s_3)(q) = \mathrm{F}$. The relation R of the model is given by $R(s_1, t)$ for every $t \in S$. To represent that the agent has more information in situation s_2, e.g., that the agent knows that it is raining in LA (perhaps because the situation is so unusual that it has been on the news), the relation R in the model can be extended by stipulating $R(s_2, t)$ for $t = s_0, s_2$. Now the agent has no doubt anymore about the truth of proposition q, but is still ignorant about the truth value of proposition p.

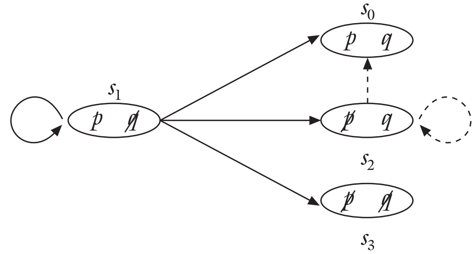

Figure 9.1

On the basis of the Kripke models a modal logic of knowledge can be devised. To this end, introduce a modal operator K, to be interpreted as 'it is known that,' and give it a formal semantic by a clause: for Kripke model $\mathcal{M} = \langle S, \pi, R \rangle$ and $s \in S$,

$$\mathcal{M}, s \models K\varphi \text{ iff (if and only if) for all } t \text{ with } R(s, t) \text{ it holds that } \mathcal{M}, s \models \varphi$$

This clause states that, in a possible world s, it is known (by the agent) that the formula φ is true iff φ is true in all the worlds t that the agent deems epistemic alternatives. In other words, although one may have doubts about the true nature of the world (if one considers more than one epistemic alternative as possible), one has no doubts about the truth of φ: this formula holds in all epistemic alternatives. Thus, it can really be said that, in this case, one *knows* the formula φ.

To complete the logic, assume that, besides propositional atoms from $\mathcal{P}$, formulas can also be composed by means of the usual propositional connectives $\neg$ (not), $\wedge$ (and), $\vee$ (or), $\rightarrow$ (implication) and $\leftrightarrow$ (bi-implication), with their usual semantics, such as, e.g.,

$$\mathcal{M}, s \models \neg\varphi \text{ iff not } \mathcal{M}, s \models \varphi$$

$$\mathcal{M}, s \models \varphi \wedge \psi \text{ iff } \mathcal{M}, s \models \varphi \text{ and } \mathcal{M}, s \models \psi$$

Propositional atoms $p \in \mathcal{P}$ are, of course, interpreted by using the truth assignment function π:

$$\mathcal{M}, s \models p \text{ iff } \pi(s)(p) = \mathrm{T}$$

Finally, a formula φ in this logic is said to be valid, notation $\models \varphi$, if $\mathcal{M}, s \models \varphi$ for all Kripke models $\mathcal{M} = \langle S, \pi, R \rangle$ and all $s \in S$.

By interpreting the operator K in the above way, one directly obtains a number of validities:

Proposition 9.1

1. $\models K(\varphi \rightarrow \psi) \rightarrow (K\varphi \rightarrow K\psi)$
2. If $\models \varphi$ then $\models K\varphi$

This proposition says that by modeling knowledge in this way, it is closed under logical consequence. Furthermore, validities are always known. With respect to an idealized notion of knowledge, these properties are certainly defensible. For more practical purposes (when using the notion of knowledge in certain applications, e.g., describing the knowledge of human or artificial beings such as robots) they may be undesirable. In this case, one may speak of the so-called problem of *logical omniscience*, discussed in section 9.5 below. For the time being, one accepts these properties of knowledge, and wonders what other properties knowledge should satisfy.

Finally in this section, note that the valid formulas with respect to the class of Kripke models that have been introduced here can be axiomatized by the following

Hilbert-style system (called system **K** in the literature of modal logic), consisting of the axioms:

(P) any axiomatization of propositional logic

(K) $K(\varphi \to \psi) \to (K\varphi \to K\psi)$

and rules *modus ponens* (MP) and

(N_K) $\varphi / K\varphi$

The validity (K) is generally referred to as the *K-axiom*, while rule N_K is called the *necessitation rule*.

Technically, one can show that this system **K** is *sound* and *complete* with respect to the class of all Kripke models, which states that the set of theorems in this system is exactly the set of validities (with respect to the class of all Kripke models). Since the proof of this is rather technical, it is omitted here, but it can be found in many textbooks on modal logic; see, for example, Chellas (1980), Hughes and Cresswell (1984) and Meyer and van der Hoek (1995), [and chapter 7].

9.3. The Systems T, S4, and S5

As seen in section 9.2, the notion of knowledge, as captured by a modal logic based on Kripke models of the form introduced there, satisfies certain properties. However, some properties that intuitively hold of knowledge are not validities in this setting. For instance, one of the defining properties of knowledge is that it is *true*! That is, in a formula: $K\varphi \to \varphi$, if it is known that φ then φ must be true. This formula, however, is not a validity in the framework given thus far [see chapter 7].

This can be remedied, however, by putting constraints on the class of Kripke models being considered. By stipulating that the accessibility relation R is *reflexive*, i.e., satisfies the constraint $R(s, s)$ for all $s \in S$, then the formula $K\varphi \to \varphi$ becomes valid with respect to this new class of models.

Proposition 9.2 Any Kripke model $\mathcal{M} = \langle S, \pi, R \rangle$ where R is reflexive, satisfies $\mathcal{M} \models K\varphi \to \varphi$.

Extending system **K** of the previous section with axiom $K\varphi \to \varphi$ gives the system referred to as system **T**. This system can be shown to be *sound* and *complete* with respect to the new class of models, i.e. the class of all Kripke models in which the accessibility relation is reflexive ([see chapter 7]; see also Chellas (1980), Hughes and Cresswell (1984), and Meyer and van der Hoek (1995).)

Furthermore, it would also be reasonable to have a property stating that knowledge is known itself, expressed by the formula $K\varphi \to KK\varphi$: if one knows φ, then one also knows that one knows φ. This formula is not a validity in the setting presented thus far either. But, again, constraints on the class of Kripke models can be introduced to overcome this difficulty. If the accessibility relation R in a model $\mathcal{M} = \langle S, \pi, R \rangle$ is required to be transitive, i.e., satisfies the constraint that

$$R(s, t) \wedge R(t, u) \Rightarrow R(s, u)$$

for all $s, t, u \in S$, then the formula $K\varphi \rightarrow KK\varphi$ becomes a validity with respect to this class of models.

Proposition 9.3 Any Kripke model $\mathcal{M} = \langle S, \pi, R \rangle$ where R is transitive, satisfies $\mathcal{M} \vDash K\varphi \rightarrow KK\varphi$.

Extending the system **T** with axiom $K\varphi \rightarrow KK\varphi$ gives a system called **S4**, which is a well-known axiom system for knowledge (at least in philosophy). The axiom is called the *positive introspection* axiom, since it states something about the agent's own knowledge about knowledge. Again, it can be shown that **S4** is *sound* and *complete* with respect to models with accessibility relations that are reflexive and transitive ([see chapter 7]; see also Chellas (1980), Hughes and Cresswell (1984) and Meyer and van der Hoek (1995).)

Now one can ask the question whether there is more to knowledge? Can further properties of knowledge be identified? This issue will be discussed later, but first it is relevant to mention here that in computer science and AI, where epistemic logic is employed to describe the 'knowledge' of artificial systems like (distributed) computer systems, information systems and 'intelligent' systems such as 'agent systems' and robots (Meyer and van der Hoek, 1995), it is customary to also add another axiom, which says something about knowledge of ignorance.

This axiom is called the *negative introspection* axiom:

$$\neg K\varphi \rightarrow K\neg K\varphi$$

It states that if the agent does *not* know formula φ, then it *knows* that it does not know φ. Of course, for human agents this axiom is highly unlikely to hold in general, since one may not even be aware of one's not knowing φ. However, for some artficial agents, dealing with *finite* information, like only a finite set $\mathcal{P}$ of propositional atoms and a finite set of formulas that it knows, the truth of this axiom may be argued (informally) like this: if the artificial agent does not know a formula, then this formula does not follow from the agent's finite information, and the agent is able to detect this, so that it knows that it does not know the formula. Also, in some cases, the validity of the axiom follows directly from the special kind of models that is used in applications – as in the case of using epistemic logic in distributed systems, cf. Halpern and Moses (1990) and Meyer and van der Hoek (1995).

To cater to the validity of the negative introspection axiom, one has to constrain the (accessibility relations of the) Kripke models even further. One can show that by requiring the relation R to be an *equivalence* relation, namely a relation that satisfies the three properties:

- *Reflexivity*: $R(s, s)$ for all $s \in S$
- *Transitivity*: $R(s, t) \wedge R(t, u) \Rightarrow R(s, u)$ for all $s, t, u \in S$
- *Symmetry*: $R(s, t) \Rightarrow R(t, s)$ for all $s, t \in S$

one obtains that the negative introspection axiom as well as all axioms of system **S4** are valid with respect to this new class of Kripke models. (And, of course, also the rules of *modus ponens* and necessitation remain sound.)

The new system is known as **S5**, and, as noted above, is very popular among computer scientists who use epistemic logic.[1] One of the reasons is the very intuitive interpretation of the models with equivalence relations as accessibility relations that are briefly discussed below. For the reason given before, philosophers do not regard **S5** as a correct logic for knowledge. They usually stick to **S4**, and possibly some logics in between **S4** and **S5**. The system **S5** can be shown *sound* and *complete* with respect to the class of Kripke models in which the accessibility relations are equivalence relations ([see chapter 7]; see also Chellas (1980), Hughes and Cresswell (1984), and Meyer and van der Hoek (1995).)

Equivalence relations divide the set of possible worlds into equivalence classes, the members of which are all mutually accessible. An equivalence class is, so to speak, a bunch of worlds that are epistemic alternatives of each other. One can show that in the case that one has only one knowledge operator as here, one can restrict oneself to equivalence relations with only *one* equivalence class without losing soundness and completeness of the logic. Such a model is particularly simple: it just consists of a set of states which are *all* mutually accessible, or speaking in epistemic terms: are all each other's epistemic alternative. So, in such a case, it does not matter what is the actual world where one is considering alternatives: for each world there is exactly the same set of alternatives: the *whole* set S of possible worlds (Meyer and van der Hoek, 1995).

A final note: as can be verified – see, for example Meyer and van der Hoek (1995) – the system **S5** contains a redundancy: the positive introspection axiom can be deleted since it can be derived from the other axioms together with the rules. Nevertheless, in the sequel when speaking about the system **S5**, it is convenient to include the positive introspection axiom as well.

9.4. Belief: The Systems K45 and KD45

Belief is mostly regarded as a weaker form of knowledge (but see later in section 9.6.2). The crucial difference between knowledge and belief is that the former must be true whereas the latter need not. When considering properties (axioms) of belief rather than knowledge, those of knowledge can be copied except for the one stating that knowledge is true. Mostly, the modal operator for belief is denoted B: $B\varphi$ is read as 'it is believed that φ' or 'the agent believes that φ.' Copying the system **S5** without the 'truth axiom' for belief gives the system known as **K45**:

- any axiomatization of propositional logic
- $B(\varphi \rightarrow \psi) \rightarrow (B\varphi \rightarrow B\psi)$

- $B\varphi \to BB\varphi$
- $\neg B\varphi \to B\neg B\varphi$

and rules *modus ponens* (MP) and

$$(\mathrm{N}_B) \qquad \varphi / B\varphi$$

The class of Kripke models with respect to which system **K45** is *sound* and *complete* consists of those models in which the accessibility relation R (now used to interpret the operator B, of course) is *transitive* and *Euclidean*, the latter meaning that R satisfies:

$$R(s, t) \wedge R(s, u) \Rightarrow R(t, u)$$

for all s, t, u, $\in S$; ([see chapter 7], also Meyer and van der Hoek (1995).)

Mostly, also it is stipulated that beliefs should be consistent, in a formula: $\neg B(p \wedge \neg p)$, for some $p \in \mathcal{P}$. Adding this formula to the system **K45** as an axiom (often called the D-axiom, since it was held a typical axiom of *d*eontic logic [see chapter 8]) yields the system **KD45**, or **weak S5**. This system can be proven *sound* and *complete* with respect to Kripke models in which the accessibility relation is *transitive*, *Euclidean*, and *serial*, where seriality of a relation R means that for all $s \in S$ there exists $t \in S$ such that $R(s, t)$. This property expresses that, in any possible world, the agent considers at least one epistemic alternative.

As with **S5**, it can be shown (Meyer and van der Hoek, 1995) that **K(D)45** is (sound and) complete with respect to a class of simpler models, in this case models consisting of an 'actual' world s_0, and a set S of worlds not including s_0 such that the accessibility relation R satisfies $R(s_0, s)$ for each $s \in S$ and $R(s, t)$ for any $s, t \in S$. In case one considers **KD45**, the set S is non-empty, whereas, in the case of **K45**, it may be empty. This provides a neat picture which can be interpreted philosophically in a very intuitive way: these simple models for **K(D)45**-belief consist of an actual world (representing the current state of the external world) together with a set of epistemic alternatives, or put differently, an actual world and an (**S5**-)epistemic model, which in the case of **K45** may be empty (representing inconsistent belief). In general, the actual world may have nothing to do with the epistemic model, reflecting the fact that beliefs may be 'counterfactual' in the sense that they may be false in reality.

Note: contrary to the case of **S5** the positive introspection axiom for belief is *not* redundant in the systems **K45** and **KD45**!

9.5. Logical Omniscience: The Problem and Some Solutions

As seen in a previous section, a modal approach to knowledge (and belief) based on Kripke models of the kind as defined thus far yields that knowledge (belief) is closed under logical consequence and that validities are known (believed) (proposition 9.1).

In fact, there are a number of further properties, collectively called properties of *logical omniscience*, since they have to do with some idealizations on the part of the knowing (believing) agent (here $\Box$ stands for either the knowledge operator K or the belief operator B):

Proposition 9.4

LO1: $\models \Box\varphi \wedge \Box(\varphi \to \psi) \to \Box\psi$

LO2: $\models \varphi \Rightarrow \models \Box\varphi$

LO3: $\models \varphi \to \psi \Rightarrow \models \Box\varphi \to \Box\psi$

LO4: $\models \varphi \leftrightarrow \psi \Rightarrow \models \Box\varphi \leftrightarrow \Box\psi$

LO5: $\models (\Box\varphi \vee \Box\psi) \to \Box(\varphi \vee \psi)$

LO6: $\models \Box\ \varphi \to \Box(\varphi \vee \psi)$

LO7: $\models \neg\ (\Box\varphi \wedge \Box\ \neg\varphi)$

Properties LO1 and LO2 were already mentioned in proposition 9.1 for K. More precisely, LO1 says that if both φ and the implication $\varphi \to \psi$ is known (believed) then also ψ is known (believed). LO3 is a similar property but slightly different: if some formula ϕ is known (believed) then also everything (ψ) that is a logical consequence is known (believed). LO4 says that logically equivalent formulas are either both known (believed) or both not known (believed). LO5 says that if both φ is known (believed) and ψ is known (believed) then also the conjunction of φ and ψ is known. LO6 says that if ϕ is known (believed) then it is also known (believed) that φ or ψ. (In fact, this is a direct consequence of property LO3.) LO7 says that it cannot be the case that both a formula and its negation is known (believed).

Sometimes, it is very convenient to consider these properties as true, but in some more practical situations these formulas might be deemed unrealistic. For instance, it is very unlikely that human agents will know (believe) all logical consequences of their knowledge (beliefs) including all validities. Although at first sight a reasonable property, even LO4 is unlikely to hold for human agents: imagine two logically equivalent formulas both of length greater than 10 million characters. These formulas are not even parsable for the unfortunate agent, let alone known to be equivalent! So, sometimes it is argued that on the grounds of the *resource boundedness* of an agent one has to deny or at least weaken the properties LO1–LO7. However, this is not as simple as it sounds. Recall that these validities are the very properties of Kripke-style modal logic as expounded thus far. The formulas LO1–LO6 are true of all Kripke models. (LO7 can be denied by taking accessibility relations that are not serial, as seen before. However, in the case of knowledge one is still stuck with this property: the models that are associated with the systems **S4** and **S5** have reflexive, and thus serial, accessibility relations.)

Therefore, to deny the above properties something 'non-standard' is needed. In the literature (Thijsse, 1992; Meyer and van der Hoek, 1995) there appear quite a number of approaches varying in 'drasticality.' Three such approaches are presented here, starting with a rather radical method considering 'non-standard' Kripke models in which nonstandard ('impossible') worlds are present. Here, the focus will be on solving the logical omniscience problem for *belief* rather than knowledge, since, in the context of belief, the problem seems to be more pregnant.

9.5.1. Rantala models

Rantala models are a non-standard type of Kripke models in which, besides the possible worlds, also so-called '*impossible worlds*' are incorporated (Rantala, 1982). The idea behind these 'impossible' worlds is that, as the name suggests, strange things may hold there: in these impossible worlds, anything may be the case, even contradictions may be true there! Thus these worlds are impossible in the true sense of the word. However, they can nevertheless be regarded as epistemic alternatives by agents which are not ideal reasoners (are less rational). And this is exactly what is needed to avoid the agent's logical omniscience.

Formally, (epistemic) Rantala models are structures of the following kind (here $\mathcal{L}$ stands for the whole logical language):

Definition 9.2 A(n epistemic) Rantala model is a structure $\mathcal{M}$ of the form $\langle S, \sigma, T, S^* \rangle$, where

- S is a non-empty set, the set of (possible *and impossible*) worlds;
- $S^* \subseteq S$ is the set of *impossible* worlds;
- $\sigma : (S \backslash S^* \rightarrow (\mathcal{P} \rightarrow \{\mathrm{T}, \mathrm{F}\})) \cup (S^* \rightarrow (\mathcal{L} \rightarrow \{\mathrm{T}, \mathrm{F}\}))$, a function assigning truth to atoms on possible worlds, and truth to *arbitrary formulas* on impossible worlds, is a truth assignment function to the atoms per state;
- $T \subseteq S \times S$ is the belief accessibility relation, for which seriality and *Rantala-model versions of transitivity and Euclidicity* are required:[2]

  1 for all $s, t, u \in S \backslash S^* : R(s, t)$ & $R(t, u)$ implies $R(s, u)$, and for all $s \in S \backslash S^*$, $t^* \in S^* : R(s, t^*)$ & $\sigma(t^*)(\neg B \neg \varphi) = \mathrm{T}$ implies $\mathcal{M}, t' \models \varphi$ for some $t' \in S$ with $R(s, t')$.

  2 for all $s, t, u \in S \backslash S^* : R(s, t)$ & $R(s, u)$ implies $R(t, u)$, and for all $s \in S \backslash S^*$, $t^* \in S^* : R(s, t^*)$ & $\sigma(t^*)(B\varphi) = \mathrm{T}$ implies $\mathcal{M}, t' \models \varphi$ for all $t' \in S$ with $R(s, t')$.

Formulas in *possible* worlds $s \in S \backslash S^*$ are interpreted in exactly the same way as in Kripke models including the clause for the modal operator $\Box$:

$$\mathcal{M}, s \models \Box\varphi \text{ iff } \mathcal{M}, t \models \varphi \text{ for every } t \in S \text{ such that } T(s, t)$$

However, in *impossible* worlds $s^* \in S^*$ every formula is regarded as atomic, and given its truth value by means of the truth assignment function σ:

$$\mathcal{M}, s^* \models \varphi \text{ iff } \sigma(s^*)(\varphi) = \mathrm{T}$$

Thus, it may happen that for example, the formula $p \wedge \neg p$ is assigned the value T by the function σ in an impossible world $s^* \in S^*$. Formulas are valid if they are true in every *possible world* $s \in S \backslash S^*$ in any Rantala model $\mathcal{M} = \langle S, \sigma, T, S^* \rangle$. This is very understandable: the worlds in which one evaluates are the worlds from which one takes up a stance and considers epistemic alternatives. Although these alternatives may be 'impossible,' the worlds of evaluation represent the actual world and thus must be 'possible'!

The feature of allowing for these impossible worlds gives one the possibility to deny all of the formulas LO1–LO7, so that none of them are validities with respect to Rantala models. For instance, consider the possibility of denying LO7 in Rantala models. This is very easy: just by taking a model $\mathcal{M} = \langle S, \sigma, T, S^* \rangle$ with $S = \{s, s^*\}$, $S^* = \{s^*\}$, $T = \{(s, s^*), (s^*, s^*)\}$, and $\sigma(s^*)(\varphi) = \sigma(s^*)(\neg\varphi) = \mathrm{T}$. Then $\mathcal{M}, s \models \Box\varphi \wedge \Box\neg\varphi$, i.e., $\mathcal{M}, s \models \neg$LO7. Moreover by stipulating that $\sigma(s^*)(\varphi \vee \neg\varphi) = \mathrm{F}$ one can deny the validity of LO2, since now $\mathcal{M}, s \not\models \Box(\varphi \vee \neg\varphi)$. In the same way, the other logical omniscience properties can be denied. Finally, note too that, due to the condition of 'Rantala-transitivity' on the model, the positive introspection axiom is a validity again, as can be easily verified.

9.5.2. Sieve models

The second approach to avoiding logical omniscience of the agent is quite different. Again a variation of a standard Kripke model is employed, but now instead of introducing nonstandard worlds, the model is endowed with a function $\mathcal{A}$ that acts as a kind of sieve: it determines whether some formula is allowed to be known (believed) (Fagin and Halpern, 1988). Intuitively, the function $\mathcal{A}$ expresses some kind of awareness on the agent's part: it indicates whether the agent is *aware* of the formula at hand in a particular situation (world), and thus is amenable to be known (believed) by the agent in that world.

Formally these models have the following form (again, use $\mathcal{L}$ for the whole logical language):

Definition 9.3 A(n epistemic) sieve model is a structure $\mathcal{M}$ of the form $\langle S, \pi, T, \mathcal{A} \rangle$, where

- S is a non-empty set (the set of *states* or *possible worlds*);
- $\pi : S \rightarrow (\mathcal{P} \rightarrow \{\mathrm{T}, \mathrm{F}\})$ is a truth assignment function to the atoms per state;
- $T \subseteq S \times S$ is the belief accessibility relation, which is assumed to be serial, transitive and Euclidean again;
- $\mathcal{A} : S \rightarrow \wp(\mathcal{L})$ is the awareness function, assigning per state the set of formulas that the agent is aware of; for any $s \in S$, $\mathcal{A}(s)$ is assumed to contain all instances of the D-axiom and the introspection axioms.

Let the language contain an 'awareness' operator A as well as the epistemic operator $\Box$. These are interpreted on a sieve model $\mathcal{M} = \langle S, \pi, T, \mathcal{A} \rangle$ and a state $s \in S$ as follows:

$\mathcal{M}, s \models A\varphi$ iff $\varphi \in \mathcal{A}(s)$

and

$\mathcal{M}, s \models \Box\varphi$ iff $\varphi \in \mathcal{A}(s)$ & $\mathcal{M}, t \models \varphi$ for all t such that $T(s, t)$

So from the definition, one can see how indeed the function $\mathcal{A}$ acts as a sieve: only those formulas are considered as knowledge (beliefs) that are indicated as being aware of by it. By the condition put on this function (which states something like that the agent is aware of the D-axiom and both introspection axioms), and the fact that the rest of the model is a standard **KD45** model, it is easy to see that these axioms are validities again.

Since the sieve model approach can only filter out formulas to be known (believed), by this approach only the validities LO1–LO6 can be avoided. This is obvious by taking a model that contains a possible world s where the formula to be denied, say ψ, is not being aware of, namely take $\mathcal{A}$ such that $\psi \notin \mathcal{A}(s)$. Then immediately $\mathcal{M}, s \not\models A\psi$, and hence $\mathcal{M}, s \not\models \Box\psi$. This can be used to show that LO1–LO6 are not valid.

9.5.3. *Cluster models*

Finally, here is a method with which one can avoid LO7 while still keeping the axiom D (or, in semantical terms, keeping serial accessibility relations). This method is strongly related to the use of what Chellas (1980) calls *minimal models* for so-called non-normal modal logic, and goes back to so-called neighborhood semantics by Scott (1970) and Montague (1974 [1968]). Chellas was mainly interested in applying it to deontic logic, but something very similar was re-invented by Fagin and Halpern (1988) in the context of epistemic logic and dubbed *local reasoning* by means of *cluster models*.

Cluster models are variants of standard Kripke models in the sense that instead of a set of epistemic alternatives a set of *sets of* epistemic alternatives is incorporated in the models. The intuition behind this is that what is normally the set of epistemic alternatives (as viewed from an actual world) is partitioned in subsets ('clusters'), where these clusters correspond to coherent bodies of knowledge while two clusters can be mutually incoherent. The typical example of such a partition of knowledge (represented by a set of epistemic alternatives) is the theory of mechanics in physics which can be partitioned into classical mechanics and quantum mechanics, where these two subtheories of mechanics are mutually inconsistent. Nevertheless, and this is very important, it is perfectly rational for a physicist to consider both theories and apply them when appropriate. [See also chapter 13, section 13.4.]

Formally, cluster models are defined as follows:

Definition 9.4 A cluster model is a structure $\mathcal{M}$ of the form $\langle S, \pi, \mathcal{C}\rangle$, where

- S is a non-empty set (the set of *possible worlds*);
- $\pi : S \to (\mathcal{P} \to \{T, F\})$ is a truth assignment function to the atoms per state;
- $\mathcal{C} : S \to \wp(\wp(S))$, such that, for every s, $\mathcal{C}(s)$ is a non-empty collection of non-empty subsets (clusters) of S.

The belief operator may now be interpreted as follows:

$\mathcal{M}, s \models B\varphi$ iff $\exists T \in \mathcal{C}(s) \forall t \in T\ \mathcal{M}, t \models \varphi$

Validity is defined as usual again.

With this interpretation of belief, one may now indeed deny LO7: take a model $\mathcal{M} = \langle S, \pi, \mathcal{C} \rangle$, with $S = \{s, t\}$, $\mathcal{C}(s) = \{\{s\}, \{t\}\}$, and $\pi(s)(p) = \mathrm{T}$, $\pi(t)(p) = \mathrm{F}$. Then $\mathcal{M}, s \models Bp \wedge B\neg p$, thus falsifying LO7. Note that, on the other hand, it is still the case that $\mathcal{M}, s \not\models B(p \wedge \neg p)$!

Cluster models as defined above are not yet models of epistemic logic: they do not yet satisfy the D-axiom and the two introspection axioms. Thijsse (1992) gives necessary and sufficient conditions for turning cluster models into 'epistemic cluster models.' Since these are rather technical (they follow from a correspondence to neighborhood semantics in the style of Scott–Montague, and have a topological meaning), they are mentioned here without further comments.

Let $\mathcal{C}^{\uparrow}(s)$ be defined as the set $\{X \mid T \subseteq X$ for some $T \in \mathcal{C}(s)\}$. Then to cater for the positive introspection axiom, impose the condition that

$$X \in \mathcal{C}^{\uparrow}(s) \Rightarrow \{v \mid X \in \mathcal{C}^{\uparrow}(v)\} \in \mathcal{C}^{\uparrow}(s)$$

and for the negative introspection axiom impose the condition

$$X \notin \mathcal{C}^{\uparrow}(s) \Rightarrow \{v \mid X \notin \mathcal{C}^{\uparrow}(v)\} \in \mathcal{C}^{\uparrow}(s)$$

9.6. Further Refinements and Extensions

Having looked at the 'standard' treatment of knowledge by means of the systems **S4** and **S5**, this section discusses some of the more advanced systems that have been proposed to deal with knowledge more adequately.

9.6.1. Other systems for knowledge

Philosophers, who do not judge the system **S5** as an adequate formalization of knowledge, have asked whether it is possible to find a suitable system for capturing the properties of knowledge that goes beyond **S4**, but stays 'below' **S5**, so to speak. Indeed, such systems between **S4** and **S5** have been proposed (Lenzen, 1980; Voorbraak, 1991, 1993). For instance, Lenzen (1980) observed that if one takes knowledge to be *true belief*, i.e., defining

$$K'\varphi = B\varphi \wedge \varphi$$

where B satisfies the logic **KD45**, one obtains a logic for K' that is known under the name **S4.4**, which is the logic **S4** together with the axiom

$$\varphi \rightarrow (\neg K'\neg K'\varphi \rightarrow K'\varphi)$$

It is not directly obvious whether this property is intuitively a suitable one for knowledge. Other candidate logics of knowledge include that of *rationally believed objective knowledge* K'', defined as

$$K''\varphi = K\varphi \wedge B\varphi$$

where K is the usual **S5**-style type of knowledge and B is a **KD45**-type of belief. This operator K'' appears to be axiomatized by the logic **S4F**, which consists of the system **S4** extended with the axiom (using M'' as the dual of K'')

$$(M''\varphi \wedge M''K''\psi) \rightarrow K''(M''\varphi \vee \psi)$$

See Voorbraak (1993). Finally; the concept of *justified knowledge* (K^j) that is considered by Voorbraak (1993). By giving a careful and rather ingenious semantic analysis of this notion by means of a generalized form of Kripke models, he argues that the logic for this type of knowledge should be the system **S4.2**, which consists of **S4** together with the axiom

$$M^jK^j\varphi \rightarrow K^jM^j\varphi$$

(using M^j as the dual of K^j).

9.6.2. *Systems for combining knowledge and belief*

After having looked at the notions of knowledge and belief separately, it is natural to question what the relations between these two notions are, and whether these relations may be formalized in a logical system. Such a system might then be used in cases where it is important to distinguish between an agent's knowledge and beliefs, and reason about both these notions. A starting point of such a combined logical system would be to take the logic **S5** for knowledge (K) and add to it the logic **KD45** for belief (B). Of course, to make this a little more exciting, one should also add some connecting axioms. Kraus and Lehmann (1986) have done so by adding the axioms[3]

$$K\varphi \rightarrow B\varphi$$

$$B\varphi \rightarrow KB\varphi$$

The former expresses that knowledge is stronger than belief, whereas the latter expresses that if one believes something then one knows that one believes this (a kind of generalized form of introspection). In itself, these two axioms are rather intuitive and seem innocuous. However, as Kraus and Lehmann (1986) themselves observe, they would have also liked to include another intuitive axiom – $B\varphi \rightarrow BK\varphi$, stating that one believes to know what one believes – but this would cause the notions of knowledge and belief to collapse, since then $K\varphi \leftrightarrow B\varphi$ would become derivable! This indicates that something is wrong with the intuitions. Voorbraak (1993) blames it on having the axiom $K\varphi \rightarrow B\varphi$, in line with his views on **S5** being a weak form of objective knowledge that cannot be stronger than rational belief (as represented by the logic **KD45**). Van der Hoek (1993) offers a different solution to the problem; he sacrifices the negative introspection axiom for knowledge – thus essentially adopting an **S4**-type of knowledge – which allows him to add the above mentioned formula $B\varphi \rightarrow BK\varphi$ as an axiom as well as the two *KB*-connecting axioms of Kraus and Lehmann above. (In fact, Van der Hoek shows that by thus dropping the negative introspection axiom for knowledge, some room is created for an unproblematic (simultaneous) addition of some more axioms like $\neg B\varphi \rightarrow K\neg B\varphi$ and $\neg K\varphi \rightarrow B\neg K\varphi$, expressing a kind of cross-over negative introspection.)

9.6.3. Knowledge in a group of agents

So far, only the notion of knowledge (and belief) of a single agent has been discussed. When considering a group of agents, one can, of course, consider the knowledge (K_i) of every individual agent i, so that a Kripke model of the form to describe the knowledge of the various agents may be used:

Definition 9.5 An (n agents) Kripke model is a structure $\mathcal{M}$ of the form $\langle S, \pi, R_1, \ldots, R_n \rangle$, where

- S is a non-empty set (the set of *possible worlds*);
- $\pi : S \rightarrow (\mathcal{P} \rightarrow \{T, F\})$ is a truth assignment function to the atoms per possible world;
- for $1 \leq i \leq n$, $R_i \subseteq S \times S$ is the knowledge accessibility relation for agent i, assumed to be an equivalence relation.

The validities with respect to these (multi-modal) models are simply axiomatized by a multi-modal version of **S5**: for each K_i one takes an **S5**-axiomatization.

However, it is also worthwhile examining notions of knowledge that have to do with the group as a whole. This has been done by Halpern and Moses (1985). At least two such notions come to mind immediately. The first is knowledge that is shared by everyone: the facts that *every* agent in the group knows, denoted by E ('everybody knows'). The axiomatization of E-knowledge is trivial: just take as an axiom (assuming there are n agents in the group):

$$E\varphi \leftrightarrow K_1\varphi \wedge \cdots \wedge K_n\varphi$$

Semantically E can be associated with an accessibility relation $R_E = \cup_{i=1}^{n} R_i$ (intuitively this means that all agents put their sets of epistemic alternatives together in one big set), and define

$$\mathcal{M}, s \vDash E\varphi \text{ iff } \mathcal{M}, t \vDash \varphi \text{ for every } t \text{ with } R_E(s, t)$$

Now the above axiom becomes a validity. Intuitively, this is because since the agents have collected their epistemic alternatives, the only things they can be sure of as a group are those formulas that are true in *all* of these alternatives.

Moreover, E satisfies the basic properties of a modal (necessity-type) operator, namely the K-axiom and necessitation rule:

$$E(\varphi \to \psi) \to (E\varphi \to E\psi)$$

$$\varphi / E\varphi$$

It also satisfies the T-axiom

$$E\varphi \to \varphi$$

E does *not*, however, satisfy the introspection axioms. The technical reason is that the union of reflexive, transitive, Euclidean relations is again reflexive, but not in general transitive or Euclidean. That E fails to satisfy introspection should not come as a surprise: It might very well be that every agent of a group (consisting, say, of agents 1 and 2) knows some fact p, while, for example, agent 2 does not know that 1 knows p. This situation is described by the formula

$$K_1 p \wedge K_2 p \wedge \neg K_2 K_1 p$$

As one might verify easily, this formula implies $Ep \wedge \neg EEp$.

The second kind of group knowledge that comes to mind is perhaps the knowledge of *some* agent in the group, write F for this notion, axiomatized by the axiom

$$F\varphi \leftrightarrow K_1\varphi \vee \cdots \vee K_n\varphi$$

However, this is not such an interesting notion. It does not even satisfy the K-axiom. A better idea is to look at knowledge that is implicit in the group in the sense that if everyone shares his knowledge with everyone, it becomes knowledge in the group. Semantically, this can be obtained as follows. Consider the sets of the epistemic alternatives regarded by the agents separately. If there is communication between the agents, they can help each other to rule out epistemic alternatives. In fact, what remains after such a group communication, is the intersection of the sets of epistemic alternatives. Thus, one can directly define an accessibility relation $R_G = \cap_{i=1}^{n} R_i$, and associate a modal operator G with it by means of

$$\mathcal{M}, s \models G\varphi \text{ iff } \mathcal{M}, t \models \varphi \text{ for every } t \text{ with } R_G(s, t)$$

Of course, one may wonder as to the properties/axiomatization of such an operator. It is directly clear that apart from K-axiom and necessitation it also satisfies

$$K_i\varphi \to G\varphi$$

However, somewhat surprisingly, since this axiom appears to express only that $R_G \subseteq \cap_{i=1}^{n} R_i$, this is already sound and complete, as shown by, for example, van der

Hoek and Meyer (1992). The rather technical details are omitted here, but the secret is that this type of modal logic is too coarse to distinguish between models where $R_G = \bigcap_{i=1}^{n} R_i$ and those where $R_G \subseteq \bigcap_{i=1}^{n} R_i$, so that due to this 'deficiency' one can still obtain completeness!

Although the above notion of group knowledge seems intuitively clear at first sight, it is not completely evident what it amounts to exactly. This is also reflected somewhat in the history of the naming of the operator: the operator was first called 'implicit' knowledge by, for example, Halpern and Moses (1985), then renamed 'distributed' knowledge by Halpern and Moses (1992). However, as it is shown in van der Hoek et al. (1999 [1995]), both the properties of implicitness and of 'distributedness' are debatable for the notion of group knowledge as defined above. In particular, it is shown that, without further restrictions on the models, it may happen that group knowledge is really stronger than what can be derived from the agents' individual knowledge, when pooled together by means of communication, which is rather counter-intuitive!

Another very interesting notion that has been introduced and studied in the literature is that of *common knowledge*. Something is common knowledge within a group of agents if not only everybody in the group knows it but also the fact that it is known by everyone is known by everyone, and the same for this fact, *ad infinitum*. Thus, intuitively one would define common knowledge of φ, denoted $C\varphi$, as $E\varphi \wedge EE\varphi \wedge \cdots$. However, infinite formulas are not part of our logical language.

Formally, given an n agent Kripke model $\mathcal{M} = \langle S, \pi, R_1, \ldots, R_n \rangle$, the accessibility relation R_C associated with the modal operator C is given as the (reflexive) transitive closure of the relation R_E : $R_C = R_E^*$. This means that $R_C(s, t)$ iff there is a sequence $s = s_0, s_1, \ldots, s_m = t$ such that $R_E(s_i, s_{i+1})$ for all $0 \leq i \leq m - 1$. This means that the relation R_C connects all those possible worlds that are in 0 or more steps accessible via the relation R_E, or in other words, via some relation R_i, where at each step a different R_i may be chosen.

If the relations R_i are assumed to be equivalence relations (**S5**), this definition amounts to the following validities, which are taken as axioms for the modality C:

(K_C)	$C(\varphi \rightarrow \psi) \rightarrow (C\varphi \rightarrow C\psi)$
(T_C)	$C\varphi \rightarrow \varphi$
(K_G)	$G(\varphi \rightarrow \psi) \rightarrow (G\varphi \rightarrow G\psi)$
(T_G)	$G\varphi \rightarrow \varphi$
(4_G)	$G\varphi \rightarrow GG\varphi$
(5_G)	$\neg G\varphi \rightarrow G\neg G\varphi$
(KE)	$E\varphi \leftrightarrow (K_i\varphi \wedge \cdots \wedge K_m\varphi)$
(EC)	$C\varphi \rightarrow EC\varphi$
(C-ind)	$C(\varphi \rightarrow E\varphi) \rightarrow (\varphi \rightarrow C\varphi)$

and to complete the system one takes rules *modus ponens* (MP) and

(N_i) $\varphi / K_i\varphi$

(N_C) $\varphi / C\varphi$

In addition to a multi-agent version of the system **S5** (with the axioms and rules for each operator K_i) the resulting system can again be proven to be sound and complete, which due to the rather complex notion of C is not exactly obtained *sine cura* (Meyer and van der Hoek, 1995). Note that the modality C satisfies the same basic **S5**-like axioms and rules. Furthermore, note the axiom C-ind, which, as its name suggests, is a kind of induction axiom to capture the infinite behaviour of the C-modality in a finite axiom! In semantical terms, it really is about induction along the R_C relation. It states that if anywhere along a chain of R_E-related worlds it holds that if φ holds somewhere, it also holds one R_E-related world further, then if φ holds at the beginning of such a chain then it holds also at every world along the chain (which is exactly the same as saying that in the intial world it is common knowledge that φ).

9.7. Conclusion

This chapter has taken a peek into epistemic (and doxastic) logic, the logic of knowledge and belief. More accurately, it has looked at epistemic logic as a special branch of *modal* logic. This has led to consideration of possible world models as a suitable semantics for epistemic logic, and the modal systems **S4** and **S5** for knowledge, and **K**(**D**)**45** for belief. As seen, this sometimes gave too idealized properties of knowledge and belief, giving rise to the problem(s) of logical omniscience. This, in turn, gave rise to approaches in the literature where the possible world semantics was modified (or 'polluted' if one prefers this term) to cope with that problem. The more properties one wants to avoid, the more one has to deviate (pollute) from Kripke-style possible world semantics with non-standard elements. Finally some more sophisticated notions and issues have been discussed, such as other systems for knowledge that have been proposed in the literature, systems in which knowledge and belief can be reasoned with at the same time, and epistemic notions that are related to a group of agents.

Suggested further reading

First of all, Lenzen (1980) is a 'classic' comprehensive textbook on epistemic logic in German; it is written from a philosophical perspective and also covers the notion of probability. A number of issues touched on in this chapter are elaborated much more extensively by Meyer and van der Hoek (1995). For example, much more attention is paid to the formal aspects of the logics of knowledge and belief, such as the issue of completeness, while the logical omniscience problem and various ways of dealing with it are also treated in more depth. Here too one may find material on the relation of knowledge (and epistemic logic more in particular) with defeasible (or 'nonmonotonic') reasoning in AI. Some of the more technical material on that will also appear in a compact form in the author's forthcoming chapter (Meyer, 2001), of the new edition of the *Handbook of Philosophical Logic*. Fagin et al. (1995) is an influential book, which emphasizes employing epistemic logic for reasoning about dynamic (computer-based) systems. Many fundamental results are presented on how knowledge may evolve (e.g., be obtained) within computer networks where the communication links are not completely secure, in the sense that information may be lost or mutilated in the communication process. The successful series of proceedings of the TARK (Theoretical Aspects of Reasoning about Knowledge, and later Theoretical Aspects of Rationality and Knowledge) and LOFT (LOgic and the Foundations of game and decision Theory) conferences on the multi-disciplinary use of epistemic logic (especially in computer science and economic theory) are also worth mentioning; see, for example, Bacharach et al. (1997), Gilboa (1998), and Halpern (1986). Finally, Laux and Wansing (1995) offers a recent collection of papers on modern topics in epistemic logic.

Acknowledgments

The author wishes to thank Rogier van Eijk and Wiebe van der Hoek for discussions on the topics treated in this chapter. Rogier is also greatly thanked for his help with the figure that appears in it.

Notes

1 Voorbraak (1991, 1993) generalizes the argument to defend **S5** as the logic of distributed systems to refer to **S5** as the logic of *objective knowledge*, a weak kind of knowledge that may be ascribed to artificial systems, like computer-based systems or even a thermometer. In this case, the so-called introspective axioms have little to do with true introspection by an agent, but rather are a way of expressing that nested forms of knowledge (like KK, or $K\neg K$) can always be eliminated by reducing it to non-nested forms of knowledge (like K and $\neg K$, respectively).

2 Admittedly, these conditions lack the elegance and beauty of those for standard Kripke models in order to deal with impossible worlds where truth is defined rather syntactically by means of the function σ. However, it is rather natural to still demand the validity of the introspection axioms, and therefore these conditions are added for the sake of completeness.

3 Actually, Kraus and Lehmann (1986) propose a much richer system involving notions like common knowledge, which we will encounter later on. Here we consider the part of the system involving only the modal operators K and B.

References

Bacharach, M. O. L., Gérard-Varet, L.-A., Mongin, P. and Shin, H. S. (eds.) 1997: *Epistemic Logic and the Theory of Games and Decisions*, (Kluwer, Dordrecht).

Bull, R. A. and Segerberg, K. 1984: "Basic Modal Logic," in *Handbook of Philosophical Logic*, Vol. II, D. Gabbay and F. Guenthner, eds., (D. Reidel, Dordrecht), 1–88.

Chellas, B. F. 1980: *Modal Logic: An Introduction*, (Cambridge University Press, Cambridge).

Fagin, R. and Halpern, J. Y. 1988: "Belief, Awareness and Limited Reasoning," *Artificial Intelligence*, 34, 39–76.

Fagin, R., Halpern, J. Y., Moses, Y. and Vardi, M. Y. 1995: *Reasoning about Knowledge*, (The MIT Press, Cambridge, Massachusetts).

Gettier, E. 1963: "Is Justified True Belief Knowledge?," *Analysis*, 23, 121–3.
Gilboa, I. (ed.) 1998: *Theoretical Aspects of Rationality and Knowledge, Proceedings of the 7th Conference*, (Morgan Kaufmann, San Francisco).
Halpern, J. Y. (ed.) 1986: *Theoretical Aspects of Reasoning about Knowledge, Proceedings of the 1986 Conference*, (Morgan Kaufmann, Los Altos, CA).
Halpern, J. Y. and Moses, Y. O. 1985: "A Guide to the Modal Logics of Knowledge and Belief," in *Proceedings of the 9th International Joint Conference on Artificial Intelligence*, A. Joshi, (ed.), (Morgan Kaufmann, Los Altos, CA), 480–90.
Halpern, J. Y. and Moses, Y. O. 1990: "Knowledge and Common Knowledge in a Distributed Environment," *J. ACM*, 37, 549–87.
Halpern, J. Y. and Moses, Y. O. 1992: "A Guide to Completeness and Complexity for Modal Logics of Knowledge and Belief," *Artificial Intelligence*, 54, 319–79.
Hintikka, J. 1962: *Knowledge and Belief*, (Cornell University Press, Ithaca, NY).
van der Hoek, W. 1993: "Systems for Knowledge and Belief" *J. of Logic and Comp.*, 3, 173–95.
van der Hoek, W. and Meyer, J.-J. Ch. 1992: "Making Some Issues of Implicit Knowledge Explicit," *Int. J. of Foundations of Computer Science*, 3, 193–223.
van der Hoek, W., van Linder, B. and Meyer, J.-J. Ch. 1999: "Group Knowledge Isn't Always Distributed (Neither Is It Always Implicit)," *Mathematical Social Sciences*, 38, 215–40.
Hughes, G. E. and Cresswell, M. J. 1984: *A Companion to Modal Logic*, (Methuen, London).
Kraus, S. and Lehmann, D. 1986: "Knowledge, Belief and Time," in *Proceedings of the 13th Int. Colloquium on Automata, Languages and Programming*, Rennes, LNCS 226, L. Kott, ed., (Springer-Verlag, Berlin), 186–95.
Kripke, S. 1963: "Semantic Analysis of Modal Logic," *Zeitschrift für Mathematische Logik und Grundlagen der Mathematik*, 9, 67–96.
Laux, A. and Wansing, H. (eds.) 1995: *Knowledge and Belief in Philosophy and Artificial Intelligence*, (Akademie Verlag, Berlin).
Lenzen, W. 1980: *Glauben, Wissen und Wahrscheinlichkeit, Systeme der epistemische Logik*, (Springer-Verlag, Vienna).
Lewis, C. I. 1912: "Implication and the Algebra of Logic," *Mind* (N.S.), 21, 522–31.
Meyer, J.-J. Ch. 2001: "Modal Epistemic and Doxastic Logic," in *Handbook of Philosophical Logic*, 2nd edn., Vol. 10, D. Gabbay, ed. (Kluwer Academic Publishers, Dordrecht).
Meyer, J.-J. Ch. and van der Hoek, W. 1995: *Epistemic Logic for AI and Computer Science*, (Cambridge University Press, Cambridge).
Montague, R. M. 1974: "Pragmatics," in *Contemporary Philosophy: A Survey*, R. Klibansky, ed., (La Nuovo Italia Editrice, Florence), 1968, 102–122; reprinted in R. Montague, *Formal Philosophy*, R. Thomason, ed., (Yale University Press, New Haven), 95–118.
Pollock, J. L. 1986: *Contemporary Theories of Knowledge*, (Hutchinson, London).
Rantala, V. 1982: "Impossible World Semantics and Logical Omniscience," *Acta Philosophica Fennica*, 35, 106–15.
Scott, S. D. 1970: "Advice on Modal Logic," in *Philosophical Problems in Logic*, K. Lambert, ed., (D. Reidel, Dordrecht), 143–73.
Thijsse, E. 1992: "Partial Logic and Knowledge Representation," PhD Thesis, Tilburg University, 1992.
Voorbraak, F. 1991: "The Logic of Objective Knowledge and Rational Belief," in *Logics in AI (Proceedings of JELIA '90)*, LNCS 478, J. van Eijck, ed., (Springer-Verlag, Berlin), 499–516.
Voorbraak, F. 1993: "As Far as I Know: Epistemic Logic and Uncertainty," PhD Thesis, Utrecht University, Utrecht.