Cognitieve Kunstmatige Intelligentie

Cognitieve Kunstmatige Intelligentie

Cognitieve Kunstmatige Intelligentie

Cognitieve Kunstmatige Intelligentie

Cognitieve Kunstmatige Intelligentie

Cognitieve Kunstmatige Intelligentie

# Epistemic Logic

*John-Jules Meyer*

Preprint nr. 010 October 1999

*Artificial Intelligence Preprint Series*

# Epistemic Logic

J.-J.Ch. Meyer
Utrecht University
Department of Computing Science
Intelligent Systems Group
P.O. Box 80.089
3508 TB Utrecht, The Netherlands

September 1, 1999

## 1 Introduction: A Brief History of Knowledge

Knowledge has been subject of philosophical study since ancient times. This is not surprising since knowledge is crucial to humans in order 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, viz. episteme. Plato defined knowledge as *"justified true belief"*, and this definition has influenced philosophers ever since (cf. [6, 25]); for instance, it is still held by the contemporary philosopher Alfred Ayer. 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 we will touch upon this later on again.

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

As is the case with so many things, in the 20th century the notion of knowledge became amenable to formal-logical analysis. With the development of formal mathematical logic in the second half of the 19th 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, viz. logics of certain modalities such as necessity and possibility. While (formal) modal logics had been around since the publication of C.I. Lewis' paper in 1912 on an axiomatic approach of strict implication [21], the inception of formal modal *epistemic* logic is often taken to be Jaakko Hintikka's publication [12]. The period 1912 upto the 1950's are referred to by Bull & Segerberg [2] as the 'First Wave' of (formal) modal logic, where syntactic and algebraic approaches were prevailing, while the period of roughly 1950 - 1980, where the focus shifted towards model-theoretic semantic approaches, is referred to as the 'Second Wave' in [2]. Hintikka's work marks the beginning of this Second Wave. In this chapter, however, we will treat epistemic logic as a particular modal logic and consider models that have become standard for modal logics in general, viz. so-called Kripke models, based on the work by Kripke [18], 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. (In fact, also Carnap, Prior and Kanger contributed to coining the notion of a possible world model, cf. [2].)

In the 1980's 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 some sense this work is part of a kind of 'Third Wave' of modal logic in the terminology of Bull & Segerberg: the use of modal logics in application areas such as computer science, linguistics and AI.) We will review 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, viz. nonmonotonic reasoning, which has some definite relations with philosophy as well. However, this will be beyond the scope of this paper.)

# 2 The Modal Logic Approach to Knowledge

## 2.1 The Basic Idea: Using Accessibility Relations to Model (Incomplete) Knowledge

In this section we will look at the basic idea behind modal epistemic logic, viz. modelling knowledge or rather ignorance (as we shall see) by means of accessibility relations as they are present in Kripke models.

In order to get prepared 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 holidays), in particular whether it is raining in New York. Since he has no information pertaining to this (and clearly cannot get this information by direct observation unless he is clairvoyant or has access to an internet site with this information, and for the time being we assume both are not 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. We see here 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 4 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 moulded perfectly into the framework of Kripke-style possible world semantics. Assume a set $\mathcal{P}$ of propositional atoms. Below we will use the symbols $tt$ and $ff$ for the truth values (true and false, respectively). Formally a Kripke model is a structure of the following form:

**Definition 2.1** *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*);*
- *$\pi : S \to (\mathcal{P} \to \{tt, ff\})$ 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 his 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 $\{t \in S \mid R(s,t)\}$, that is all possible worlds $t$ that are accessible from $s$ by means of the relation $R$.

For example, the example above can be represented in a Kripke model as follows. 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 $\pi(s_1)(p) = tt$ and $\pi(s_1)(q) = ff$. Now the model can be represented by taking $S = \{s_0, s_1, s_2, s_3\}$, where $s_0$ is such that $\pi(s_0)(p) = \pi(s_0)(q) = tt$, $s_1$ is as above, $s_2$ is such that $\pi(s_2)(p) = ff$

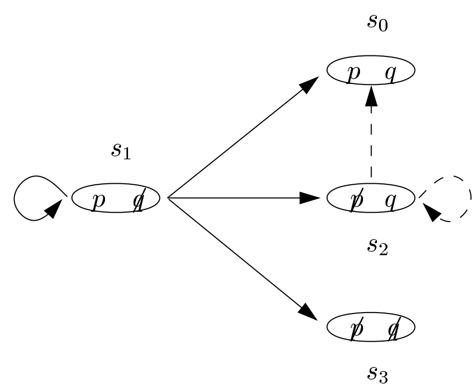


Figure 1:

and $\pi(s_2)(q) = tt$, and $s_3$ is such that $\pi(s_3)(p) = \pi(s_3)(q) = ff$. The relation $R$ of the model is given by $R(s_1, t)$ for every $t \in S$. When we want to represent that the agent, when in situation $s_2$, has more information, e.g. that it is known to him that it is raining in LA (perhaps because the situation is so unusual that it has been on the news), we can extend the relation $R$ in the model 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 he is still ignorant about the truth value of proposition $p$.

## 2.2 A Modal Logic of Knowledge

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

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

This clause states that in a possible world $s$ it is known (by the agent) that the formula $\varphi$ is true if and only if in all the worlds $t$ that are deemed epistemic alternatives by the agent it holds that $\varphi$ is true. In other words, although the agent may have doubts about the true nature of the world (if it consideres more than one epistemic alternative as possible), it has no doubts about the truth of $\varphi$: this formula holds in all epistemic alternatives. Thus, we can really say that in this case the agent *knows* the formula $\varphi$.

To complement the logic we assume that besides propositional atoms from $\mathcal{P}$ we can also compose formulas 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.

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

and

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

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

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

Finally, we say that a formula $\varphi$ in this logic is 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 2.2** *1.* $\models K(\varphi \to \psi) \to (K\varphi \to K\psi)$*;*

*2. If* $\models \varphi$ *then* $\models K\varphi$*.*

This proposition says that by modelling knowledge in this way, it is closed under logical consequence. Furthermore, validities are always known. With respect to an idealised 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*, to which we will return in a later section. For the time being we accept these properties of knowledge, and wonder what other properties knowledge should satisfy.

Finally in this section we note that the valid formulas with respect to the class of Kripke models that we have introduced can be axiomatized by the following Hilbert-style system (called system **K** in the literature of modal logic), consisting of the axioms:

(P) any axiomatisation of propositional logic;

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

and rules *modus ponens* (MP) and

($N_K$) $\frac{\varphi}{K\varphi}$

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). The proof of this can be found in many textbooks on modal logic (e.g. [3, 16, 23]). Since it is rather technical it is omitted here.

# 3 The Systems T, S4, and S5

As we have seen in the previous section 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*! In a formula: $K\varphi \to \varphi$, if it is known that $\varphi$ then $\varphi$ must be true. This formula, however, is not a validity in the framework given thusfar.

**Example 3.1** *Take a Kripke model* $\mathcal{M} = \langle S, \pi, R \rangle$ *over* $\mathcal{P} = \{p\}$ *with* $S = \{s_0, s_1\}$, $R = \{(s_0, s_1)\}$, *and* $\pi$ *such that* $\pi(s_1)(p) = tt$ *and* $\pi(s_0)(p) = ff$. *Now it is easy to verify that* $\mathcal{M}, s_0 \models Kp$, *but* $\mathcal{M}, s_0 \not\models p$. *So,* $\mathcal{M}, s_0 \not\models Kp \to p$, *and thus* $\not\models Kp \to p$.

However, we can remedy this by putting constraints on the class of Kripke models that we considering. By stipulating that the accessibility relation $R$ is *reflexive*, that is, 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 3.2** *Any Kripke model* $\mathcal{M} = \langle S, \pi, R \rangle$ *where* $R$ *is reflexive, satisfies* $\mathcal{M} \models K\varphi \to \varphi$.

**Proof.** Let $\mathcal{M} = \langle S, \pi, R \rangle$ with $R$ a reflexive relation. Suppose, for some $s \in S$, $\mathcal{M}, s \models K\varphi$. Then $\mathcal{M}, t \models \varphi$ for all $t$ with $R(s, t)$. Since we know that $R(s, s)$, we immediately find that $\mathcal{M}, s \models \varphi$. Consequently, for any $s \in S$, $\mathcal{M}, s \models K\varphi \to \varphi$. □

If we extend system **K** of the previous section with axiom

$$K\varphi \to \varphi$$

we get 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, that is the class of all Kripke models in which the accessibility relation is reflexive (cf. [3, 16, 23]).

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 the agent knows $\varphi$, then the agent also knows that it knows $\varphi$. As can be easily seen by considering a counter-model this formula is not a validity in the setting presented thusfar either.

**Example 3.3** *Take a Kripke model* $\mathcal{M} = \langle S, \pi, R \rangle$ *over* $\mathcal{P} = \{p\}$ *with* $S = \{s_0, s_1, s_2\}$, $R = \{(s_0, s_1), (s_1, s_2)\}$, *and* $\pi$ *such that* $\pi(s_1)(p) = tt$ *and* $\pi(s_2)(p) =$

*ff. (We do not care about the value of $\pi(s_0)(p)$; it does not matter in what follows.) Now it is easy to verify that $\mathcal{M}, s_0 \models Kp$, but $\mathcal{M}, s_0 \not\models KKp$. So, $\mathcal{M}, s_0 \not\models Kp \to KKp$, and thus $\not\models Kp \to KKp$.*

Again, we can introduce constraints on the class of Kripke models in order to overcome this difficulty. If we require that the accessibility relation $R$ in a model $\mathcal{M} = \langle S, \pi, R\rangle$ is *transitive*, that is, 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 $K\varphi \to KK\varphi$ becomes a validity with respect to this class of models.

**Proposition 3.4** *Any Kripke model $\mathcal{M} = \langle S, \pi, R\rangle$ where $R$ is transitive, satisfies $\mathcal{M} \models K\varphi \to KK\varphi$.*

**Proof.** Let $\mathcal{M} = \langle S, \pi, R\rangle$ with $R$ a transitive relation. Suppose, for some $s \in S$, $\mathcal{M}, s \models K\varphi$. Then $\mathcal{M}, t \models \varphi$ for all $t$ with $R(s,t)$. Take some $t$ with $R(s,t)$. Now consider an arbitrary $u$ with $R(t,u)$. By transitivity, also $R(s,u)$, and so $\mathcal{M}, u \models \varphi$. Since this holds for every $u$ with $R(t,u)$, we have that $\mathcal{M}, t \models K\varphi$, and since this in its turn holds true for every $t$ with $R(s,t)$, we obtain that $\mathcal{M}, s \models KK\varphi$. Thus it holds that, for any $s \in S$, $\mathcal{M}, s \models K\varphi \to KK\varphi$. □

Extending the system **T** with axiom

$$K\varphi \to KK\varphi$$

we obtain a system called **S4**, which is a well-known axioma system for knowledge (at least in philosophy). The axiom is called the *positive introspection* axiom, since it states something about the agent's knowledge about (its own) knowledge. Again, it can be shown that **S4** is *sound* and *complete* with respect to models with accessibility relations that are reflexive and transitive (cf. [3, 16, 23]).

Now we can ask ourselves the question whether there is more to knowledge? Can we identify further properties of knowledge? We shall come back to this later, but let us mention here that in computer science and artificial intelligence, 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 ([23]), it is customary to also add another axiom, which says something about knowledge about ignorance.

This axiom, called the *negative introspection* axiom, is the following:

$$\neg K\varphi \to K\neg K\varphi$$

It states that if the agent does *not* know formula $\varphi$, then it *knows* that it does not know $\varphi$. Of course, for human agents this axiom is highly unlikely to hold in general, since the agent may not even be aware of its not knowing $\varphi$.

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 (like in the case of using epistemic logic in distributed systems, cf. [10, 23]).

In order to cater for 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, viz. a relation that satisfies the following 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 we said 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 we will briefly discuss 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**. We return to this in a later section. 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 (cf. [3, 16, 23]).

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 words that are epistemic alternatives of each other. One can show that in the case that one has only one knowledge operator as we do 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

[1] Voorbraak [29, 30] generalises 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).

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, viz. the *whole* set $S$ of possible worlds (cf. [23]).

A final note: as can be verified (cf. e.g. [23]), in the system **S5** actually 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 we speak about the system **S5** we include the positive introspection axiom as well.

# 4 Belief: the Systems K45 and KD45

Belief is mostly regarded as a weaker form of knowledge (but see later in section 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 we can copy those of knowledge 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 $\varphi$' or 'the agent believes that $\varphi$'. Copying the system **S5** without the 'truth axiom' for belief gives us the system known as **K45**:

- any axiomatisation of propositional logic;
- $B(\varphi \to \psi) \to (B\varphi \to B\psi)$;
- $B\varphi \to BB\varphi$;
- $\neg B\varphi \to B\neg B\varphi$

and rules *modus ponens* (MP) and

$(\mathrm{N}_B)$ $\frac{\varphi}{B\varphi}$

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 $s, t, u, \in S$ (cf. [23]).

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) 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 ([23]) 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 we consider **KD45** the set $S$ is non-empty, whereas in the case of **K45** it may be empty. This provides us with 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**!

# 5 Logical Omniscience: The Problem and Some Solutions

As we have seen in a previous section, a modal approach to knowledge (and belief) based on Kripke models of the kind that we have defined thusfar yields that knowledge (belief) is closed under logical consequence and that validities are known (believed) (Proposition 2.2):

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

**Proposition 5.1**

- $\models \Box\varphi \land \Box(\varphi \to \psi) \to \Box\psi$ *LO1*
- $\models \varphi \Rightarrow \models \Box\varphi$ *LO2*
- $\models \varphi \to \psi \Rightarrow \models \Box\varphi \to \Box\psi$ *LO3*
- $\models \varphi \leftrightarrow \psi \Rightarrow \models \Box\varphi \leftrightarrow \Box\psi$ *LO4*
- $\models (\Box\varphi \land \Box\psi) \to \Box(\varphi \land \psi)$ *LO5*
- $\models \Box\varphi \to \Box(\varphi \lor \psi)$ *LO6*
- $\models \neg(\Box\varphi \land \Box\neg\varphi)$ *LO7*

Properties LO1 and LO2 are the already mentioned ones of Proposition 2.2. More precisely, LO1 says that if both $\varphi$ and the implication $\varphi \rightarrow \psi$ is known (believed) then also $\psi$ is known (believed). LO3 is a similar property but slightly different: if some formula $\phi$ is known (believed) then also everything ($\psi$) 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 is both $\varphi$ is known (believed) and $\psi$ is known (believed) then also the conjunction of $\varphi$ and $\psi$ is known. LO6 says that if $\phi$ is known (believed) then it is also known (believed) that $\varphi$ or $\psi$. (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, for a human agent it is very unlikely that he will know (believe) all logical consequences of his 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 is 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 we have expounded thusfar. The formulas LO1 - LO6 are true of all Kripke models (LO7 can be denied by taking accessibility relations that are not serial, as we have seen before. However, in the case of knowledge we are still stuck with this property: the models that we associated with the systems **S4** and **S5** have reflexive, and thus serial(!), accessibility relations.)

Therefore if we want to deny the above properties we should do something 'non-standard'. In the literature [28, 23] there appear quite a number of approaches varying in 'drasticality'. We present three such approaches here, starting off with a rather radical method considering 'non-standard' Kripke models in which nonstandard ('impossible') worlds are present. We will restrict attention to solving the logical omniscience problem for *belief* rather than knowledge, since in the context of belief the problems seems to be a more pregnant.

## 5.1 Rantala Models

Rantala models [26] are a non-standard type of Kripke models in which, besides the possible worlds, also so-called *'impossible worlds'* are incorporated. 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 we like to achieve in avoiding 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 5.2** *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;*
- *$\sigma : (S \backslash S^* \to (\mathcal{P} \to \{tt, ff\})) \cup (S^* \to (\mathcal{L} \to \{tt, ff\}))$, 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 we require seriality and Rantala-model versions of transitivity and euclidicity:*[2]
    1. *for all $s, t, u \in S \backslash S^*$: $R(s,t)$ & $R(t,u)$ implies $R(s,u)$, and for all $s \in S \backslash S^*, t^* \in S^*$: $R(s,t^*)$ & $\sigma(t^*)(\neg B \neg \varphi) = tt$ 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 $s \in S \backslash S^*, t^* \in S^*$: $R(s,t^*)$ & $\sigma(t^*)(B\varphi) = tt$ 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 $\Box$:

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

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 $\sigma$:

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

Thus, it may happen that e.g. the formula $p \wedge \neg p$ is assigned the value $tt$ by the function $\sigma$ 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 we evaluate are the worlds from which we take up a stance and consider epistemic alternatives. Although these alternatives may be 'impossible', the worlds of evaluation represent the actual world and thus must be 'possible'!

[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 $\sigma$. 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.

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, let us discuss 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 $\sigma(s^*)(\varphi) = \sigma(s^*)(\neg\varphi) = tt$. Then we have that $\mathcal{M}, s \models \Box\varphi \wedge \Box\neg\varphi$, i.e. $\mathcal{M}, s \models \neg$LO7. Moreover by stipulating that $\sigma(s^*)(\varphi \vee \neg\varphi) = ff$ we 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 we note that, due to the condition on the model of 'Rantala-transitivity', the positive introspection axiom is an validity again as it can be verified easily.

## 5.2 Sieve Models

The second approach to avoiding logical omniscience of the agent is quite different. Again we employ a variation of a standard Kripke model, but now instead of introducing non-standard worlds, we endow the model with a function $\mathcal{A}$ that acts as a kind of sieve: it determines whether some formula is allowed to be known (believed) [4]. Intuitively the function $\mathcal{A}$ expresses some kind of awarenes 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 (we use $\mathcal{L}$ for the whole logical language again):

**Definition 5.3** *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*);*
- *$\pi : S \to (\mathcal{P} \to \{tt, ff\})$ 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;*
- *$\mathcal{A} : S \to \wp(\mathcal{L})$ 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.*

The awareness operator $A$ and the epistemic operator $\Box$ are now interpreted on a sieve model $\mathcal{M} = \langle S, \pi, T, \mathcal{A} \rangle$ and a state $s \in S$ as follows:

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

and

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

So from the definition we 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 we have 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 $\psi$, is not being aware of, viz. take $\mathcal{A}$ such that $\psi \notin \mathcal{A}(s)$. Then immediately we have that $\mathcal{M}, s \not\models A\psi$, and hence $\mathcal{M}, s \not\models \Box\psi$. This can be used to show that LO1 - LO6 are not valid.

## 5.3 Cluster Models

Finally we show 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 [3] calls *minimal models* for so-called non-normal modal logic, and goes back to so-called neighbourhood semantics by Scott [27] and Montague [24]. Chellas was mainly interested in applying it to deontic logic, but something very similar was re-invented by Fagin and Halpern [4] 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.

Formally, cluster models are defined as follows:

**Definition 5.4** *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*);*

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

We may now interpret the belief operator as follows:

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

Validity is defined as usual again.

With this interpretation of belief we 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 $\pi(s)(p) = tt, \pi(t)(p) = ff$. Then $\mathcal{M}, s \models Bp \wedge B\neg p$, thus falsifying LO7. Note that on the other hand we still have here 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 [28] gives necessary and sufficient conditions for turning cluster models into 'epistemic cluster models'. Since these are rather technical (they follow from a correspondence to neighbourhood semantics in the style of Scott-Montague, and have a topological meaning), we mention them without further comments:

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

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

and for the negative introspection axiom we need to impose the condition

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

# 6 Further Refinements and Extensions

## 6.1 Other Systems for Knowledge

After having looked at the 'standard' treatment of knowledge by means of the systems **S4** and **S5**, in this subsection we like to mention some of the more advanced systems that have been proposed in the literature to deal with knowledge properly. For instance, a question that was raised by philosophers (who, as we know, do not judge the system **S5** as an adequate formalisation of knowledge) is that 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 in between **S4** and **S5** have been proposed ([20, 29, 30]). For instance, it was already observed by Lenzen [20] 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

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

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 axiomatised by the logic **S4F**, which consists of the system **S4** extended with the axiom (using $M''$ as the dual of $K''$)

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

(cf. [30]). Finally we mention the concept of *justified knowledge* ($K^j$) that is considered by Voorbraak [30]. By giving a careful and rather ingenious semantic analysis of this notion by means of a generalised 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

$$M^j K^j \varphi \to K^j M^j \varphi$$

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

## 6.2 Systems for Combining Knowledge and Belief

After having looked at the notions of knowledge and belief separately, it is a natural 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 the knowledge of an agent and his 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 we should also add some connecting axioms. Kraus & Lehmann [17] have done so by adding the axioms [3]

$$K\varphi \to B\varphi$$

and

$$B\varphi \to KB\varphi.$$

[3] Actually, Kraus & Lehmann 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 $B$.

The former expresses that knowledge is stronger than belief, whereas the latter expresses that if an agent believes something then it knows that it believes this (a kind of generalised form of introspection). In itself these two axioms are rather intuitive and seemingly innocuous. However, as Kraus & Lehmann already observe themselves in [17], they would have also liked to include another intuitive axiom, viz. $B\varphi \to BK\varphi$, stating that an agent believes to know what he believes, but this would cause the notions of knowledge and belief to collapse, since then $K\varphi \leftrightarrow B\varphi$ would become derivable! This indicated that something is wrong with the intuitions. Voorbraak [30] blames it on having the axiom $K\varphi \to B\varphi$, 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 [13] offers a different solution to the problem: he is willing to sacrifice the negative introspection axiom for knowledge (thus essentially adopting an **S4**-type of knowledge) and then one can safely add the above mentioned formula $B\varphi \to BK\varphi$ as an axiom as well as the two KB-connecting axioms of Kraus & Lehmann above. (In fact, Van der Hoek shows that in this case (so by dropping the negative introspection axiom for knowledge) some room is created for an unproblematic (simultaneous) addition of some more axioms like $\neg B\varphi \to K\neg B\varphi$ and $\neg K\varphi \to B\neg K\varphi$, expressing a kind of cross-over negative introspection.)

## 6.3 Knowledge in a Group of Agents

Until now we have discussed the notion of knowledge (and belief) of a single agent. When considering a group of agents, we can, of course, consider the knowledge ($K_i$) of every individual agent $i$, so that we may use a Kripke model of the form to describe the knowledge of the various agents:

**Definition 6.1** *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*);*
- *$\pi : S \to (\mathcal{P} \to \{tt, ff\})$ 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 axiomatised by a multi-modal version of **S5**: for each $K_i$ we take an **S5**-axiomatisation.

However, it is also worthwhile to examine notions of knowledge that have to do with the group as a whole. This has been done by Halpern & Moses [9]. 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. This type of knowledge we will denote by $E$ ('everybody knows'). The axiomatisation of E-knowledge is trivial: just take as an axiom (assuming there are $n$ agents in the group):

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

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

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

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

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

1. $\models E(\varphi \to \psi) \to (E\varphi \to E\psi)$;
2. If $\models \varphi$ then $\models E\varphi$.

It also satisfies

$$E\varphi \to \varphi.$$

However $E$ does *not* satisfy the introspection axioms. The technical reason for this is that the union of reflexive, transitive, euclidean relations is again reflexive, but not again transitive and euclidean in general. The failure of the operator $E$ of satisfying introspection should not come as a surprise to us: it might very well be that every agent of a group (say consisting of the 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 $K_1p \wedge K_2p \wedge \neg K_2K_1p$. 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, let us write $F$ for this notion, axiomatized by the axiom $F\varphi \leftrightarrow K_1\varphi \vee \ldots \vee K_n\varphi$. 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, we can directly define a accessibility relation $R_G = \bigcap_{i=1}^n R_i$, 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 / axiomatisation 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 only express the property that $R_G \subseteq \bigcap_{i=1}^n R_i$, this is already sound and complete, as shown in e.g. [15]. We omit the rather technical details 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 in e.g. [9], then renamed 'distributed' knowledge in [11]. However, as it is shown in [14], 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 $\varphi$, denoted $C\varphi$, as $E\varphi \wedge EE\varphi \wedge \ldots$. 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 $R_E$: $R_C = R_E^\star$. 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 $C$:

($K_C$) $C(\varphi \to \psi) \to (C\varphi \to C\psi)$;

($T_C$) $C\varphi \to \varphi$;

($K_G$) $G(\varphi \to \psi) \to (G\varphi \to G\psi)$;

($T_G$) $G\varphi \to \varphi$;

($4_G$) $G\varphi \to GG\varphi$;

($5_G$) $\neg G\varphi \to G\neg G\varphi$;

(KE) $E\varphi \leftrightarrow (K_i\varphi \wedge \ldots \wedge K_m\varphi)$;

(EC) $C\varphi \to EC\varphi$

(C-ind) $C(\varphi \to E\varphi) \to (\varphi \to C\varphi)$;

and to complement the system we take rules modus ponens (MP) and

($N_i$) $\frac{\varphi}{K_i\varphi}$

($N_C$) $\frac{\varphi}{C\varphi}$

In addition to a multi-agent version of the system **S5** (where we have the axioms and rules for each modal 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* (cf. e.g. [23]). 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 axioms 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 $\varphi$ holds somewhere, it also holds one $R_E$-related world further, then if $\varphi$ holds at the beginning of such a chain then it holds also at every world along the chain (and this is exactly the same as saying that in the intial world it is common knowledge that $\varphi$).

# 7 Conclusion

In this chapter we have taken a peek into epistemic (and doxastic) logic, the logic of knowledge and belief. More accurately, we have looked at epistemic logic as a special branch of *modal* logic. This has led us to consider 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 we have seen this gave us sometimes too idealized properties of knowledge and belief, giving rise to the problem(s) of logical omniscience. This has given rise to approaches in the literature where the possible world semantics was modified (or 'polluted' if one prefers this term) to cope with the logical omniscience problem. The more properties one wants to avoid, the more one has to deviate (pollute) Kripke-style possible world semantics with non-standard elements. Finally we have discussed some more sophisticated notions and issues, 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.

## 8 Further Reading

First of all, [20] is a 'classic' comprehensive textbook on epistemic logic in the German language which is written from a philosophical perspective and which also covers the notion of probability.

A number of issues touched upon in this chapter is elaborated much more extensively in our book [23]. For example, much more attention is paid to the formal aspects of the logics of knowledge and belief such as the issue of completeness. Also the logical omniscience problem and various ways of dealing with it is treated in more depth. Furthermore, in [23] one may find material on the relation of knowledge (and epistemic logic more in particular) with defeasible (or 'nonmonotonic') reasoning in Artificial Intelligence (AI), in which in the absence of certain knowledge one may 'jump to' plausible conclusions that may have to be retracted when additional information becomes available. Some of the more technical material will also appear in a compact form in a forthcoming chapter of the new edition of the Handbook of Philosophical Logic ([22]).

Next we like to mention the influential book [5], where the emphasis is on employing epistemic logic for reasoning about dynamic (computer-based) systems. Many fundamental results are presented on the way knowledge may evolve (e.g. be obtained) within computer networks where the communication links are not completely secure in the sense that information may get lost or mutilated in the communication process. Also 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 computer science and economic theory) are worth mentioning here (e.g. [8, 1, 7]).

Finally, [19] is a recent collection of papers on modern topics in epistemic logic.

## 9 Acknowledgements

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.

## References


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

[2] R. Bull & K. Segerberg, Basic Modal Logic, in: D. Gabbay & F. Guenthner (eds.), *Handbook of Philosophical Logic*, Vol. II, 1984, pp. 1–88.

[3] B.F. Chellas, *Modal Logic: An Introduction*, Cambridge University Press, Cambridge / London, 1980.

[4] R. Fagin & J.Y. Halpern, Belief, Awareness and Limited Reasoning, *Artificial Intelligence* 34, 1988, pp. 39–76.

[5] R. Fagin, J.Y. Halpern, Y. Moses & M.Y. Vardi, *Reasoning about Knowledge*, The MIT Press, Cambridge, Massachusetts, 1995.

[6] E. Gettier, Is Justified True Belief Knowledge? *Analysis* 23, 1963, pp. 121–123.

[7] I. Gilboa (ed.), *Theoretical Aspects of Rationality and Knowledge*, Proceedings of the 7th Conference (TARK 1998), Morgan Kaufmann, San Francisco, 1998.

[8] J.Y. Halpern (ed.), *Theoretical Aspects of Reasoning about Knowledge, Proceedings of the 1986 Conference*, Monterey, Morgan Kaufmann, Los Altos, 1986.

[9] J.Y. Halpern & Y.O. Moses, A Guide to the Modal Logics of Knowledge and Belief, in: *Proceedings of the 9th International Joint Conference on Artificial Intelligence*, 1985, pp. 480–490.

[10] J.Y. Halpern & Y.O. Moses, Knowledge and Common Knowledge in a Distributed Environment, *J. ACM* 37(3), 1990, pp. 549–587.

[11] J.Y. Halpern & Y.O. Moses, A Guide to Completeness and Complexity for Modal Logics of Knowledge and Belief, *Artificial Intelligence* 54, 1992, pp. 319–379.

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

[13] W. van de Hoek, Systems for Knowledge and Belief, *J. of Logic and Comp.* 3(2), 1993, pp. 173–195.

[14] W. van der Hoek, B. van Linder & J.-J. Ch. Meyer, Group Knowledge Isn't Always Distributed (Neither Is It Always Implicit), in *Proc. BISFAI'95* (M. Koppel & E. Shamir, eds.), Ramat Gan/Jerusalem, 1995, pp. 191-200; revised and expanded version as Techn. Report UU-CS-1997-14, to appear in *Mathematical Social Sciences*.

[15] W. van der Hoek & J.-J. Ch. Meyer, Making Some Issues of Implicit Knowledge Explicit, *Int. J. of Foundations of Computer Science* 3(2), 1992, pp. 193–223.

[16] G.E. Hughes & M.J. Cresswell, *A Companion to Modal Logic*, Methuen, London, 1984.

[17] S. Kraus & D. Lehmann, Knowledge, Belief and Time, in: L. Kott (ed.), *Proceedings of the 13th Int. Colloquium on Automata, Languages and Programming*, Rennes, LNCS 226, Springer, Berlin, 1986.

[18] S. Kripke, Semantic Analysis of Modal Logic, *Zeitschrift für Mathematische Logik und Grundlagen der Mathematik* 9, 1963, pp. 67–96.

[19] A. Laux & H. Wansing (eds.), *Knowledge and Belief in Philosophy and Artificial Intelligence*, Akademie Verlag, Berlin, 1995.

[20] W. Lenzen, *Glauben, Wissen und Wahrscheinlichkeit, Systeme der epistemische Logik*, Springer, Wien, 1980.

[21] C.I. Lewis, Implication and the Algebra of Logic, *Mind* (N.S.) 21, 1912, pp. 522–531.

[22] J.-J. Ch. Meyer, Modal Epistemic and Doxastic Logic, in: *Handbook of Philosophical Logic (2nd ed.), Vol. D7, Semi Classical Logics 7* (D.M. Gabbay, ed.), to appear.

[23] J.-J. Ch. Meyer & W. van der Hoek, *Epistemic Logic for AI and Computer Science*, Cambridge University Press, Cambridge, 1995.

[24] R. Montague, Pragmatics, in Contemporary Philosophy: a Survey (R. Klibansky, ed.), La Nuovo Italia Editrice, Florence, 1968.

[25] J.L. Pollock, *Contemporary Theories of Knowledge*, Hutchinson, London, 1986.

[26] V. Rantala, Impossible World Semantics and Logical Omniscience, *Acta Philosophica Fennica* 35, 1982, pp. 106–115.

[27] D. Scott, Advice on Modal Logic, in *Philosophical Problems in Logic* (K. Lambert, ed.), Reidel, Dordrecht, 1970, pp. 143–173.

[28] E. Thijsse, *Partial Logic and Knowledge Representation*, PhD Thesis, Tilburg University, 1992.

[29] F. Voorbraak, The Logic of Objective Knowledge and Rational Belief, in: J. van Eijck (ed.), *Logics in AI (Proceedings of JELIA '90)*, LNCS 478, Springer, 1991, pp. 499–516.

[30] F. Voorbraak, *As Far as I Know: Epistemic Logic and Uncertainty*, PhD Thesis, Utrecht University, Utrecht, 1993.