J.-J. Ch. Meyer and W. van der Hoek, **Epistemic Logic for AI and Computer Science**, Cambridge Tracts in Theoretical Computer Science, No. 41, Cambridge, Cambridge University Press, 1995, ISBN 0-52146014-X.

Epistemic logic is the logic of knowledge: how do you reason about the question whether your neighbor knows that you know that he plays his radio too loud? This area of logic began to flourish with the publication of Hintikka's **Knowledge and Belief**, Cornell University Press, 1962. In the early years, philosophers debated which axioms are suitable for modeling human knowledge. However, in Hintikka's opinion, the possible worlds semantics of epistemic logic presents more interesting problems and solutions than the axiomatic side of the subject. In the book under review, which started out as a set of course notes, semantical questions are indeed predominant. In the context of epistemic logic, one can view possible worlds that are compatible with the information of agent i at world w as *epistemic alternatives*. Such epistemic alternatives are said to be accessible from w by the relation R_i. The more an agent knows, the better it can distinguish other worlds from the real one, so the fewer worlds are epistemic alternatives for it. This corresponds to the intuitive idea that information means elimination of uncertainty. One says that an agent i *knows* φ (formally $K_i\varphi$) in world w iff φ is true in all worlds that the agent thinks possible, i.e. all worlds that are R_i-accessible from w.

Since the eighties, there has been a flurry of activity in the field of epistemic logic. Theoretical computer scientists have applied it to distributed systems and game theorists to negotiation. In 1995, two books about epistemic logic appeared: **Reasoning about Knowledge** by Fagin, Halpern, Moses and Vardi (LXII 1484) and the book by Meyer and Van der Hoek, discussed here. In the review below, inevitably some aspects of the book will be compared with its rival.

In their introductory chapter 0, Meyer and Van der Hoek state the purpose of their book. They want to investigate the logical properties of knowledge and belief and the differences between the two notions, while avoiding a "deep philosophical discussion". Moreover, they set out to show how epistemic logic may be applied to problems in computer science and AI. The reviewer regrets that the authors have shunned many intriguing philosophical aspects of their subject. However, let us see in which degree they succeed in achieving their own stated aims.

The first chapter gives the basic definitions of an epistemic language and epistemic Kripke models for m agents, where every agent may know different things and see different epistemic alternatives than its colleagues do. The authors introduce the systems $\mathbf{K}_{(m)}$ and $\mathbf{S5}_{(m)}$, the epistemic m-agent counterparts of the well-known modal logics **K** and **S5**. Unfortunately, they do not adequately defend their viewpoint that $\mathbf{S5}_{(m)}$ is an appropriate system for reasoning about distributed systems and multi-agent systems with finite storage of information. Also, it is a pity that they do not give any of the easy correspondence proofs between classes of Kripke frames and epistemic axioms. A well-known example of such a correspondence is that the positive introspection axiom $K_i\varphi \to K_iK_i\varphi$ is valid precisely in those Kripke frames in which R_i is transitive. In a separate section, some interesting properties of the single agent system $\mathbf{S5}_{(1)}$ are given. However, the authors' proof that it is sufficient to consider models in which all worlds have different truth assignments is two and a half page long and uses much too heavy artillery, including bisimulations, where a straightforward proof of only 13 lines would have sufficed (see http://tcw2.ppsw.rug.nl/r̃ineke/). Moreover, the authors' inductive proof that every epistemic formula is equivalent in $\mathbf{S5}_{(1)}$ to a formula without nestings of epistemic operators uses an equivalence that is so general that students have difficulties when putting concrete formulas into this normal form. The section about representing the set of global states of distributed systems as a Kripke structure, on the other hand, is very clear. Also, the authors give a well worked-out example of using epistemic logic to verify protocols for error-free transmission of sequences over distributed systems, based on original work by Halpern and Zuck. In fact, in this case the book under review clearly surpasses its rival by Fagin et al.

Chapter 2 treats various interesting epistemic notions like common knowledge and implicit knowledge in a group. The notion of *common knowledge* arises from David Lewis' **Convention: A Philosophical Study**, Cambridge (MA), Harvard University Press, 1969. One of the questions in his book is about the convention of driving on a certain side of the road. What kind of knowledge is needed for every driver to feel reasonably safe? The fact that all Dutch drivers drive on the right side of the road by itself is not enough to make them feel safe: they would want to *know* that all other drivers drive on the right side, as well. Now imagine that everyone drives on the right because they know that all others do, but that everyone holds the following false belief: "except for myself, everyone else drives on the right just by habit, and would continue to do so no matter what he expected others to do". Lewis argues that in this imaginary situation one cannot really say that there is a convention to drive on the right. Lewis proposes that if there is a convention among a group that φ, then everyone knows φ, everyone knows that everyone knows φ, and so on ad infinitum. In such a case we say that the group has common knowledge of φ.

About this subject, the authors first give a semantic characterization of common knowledge, and define analogs of $\mathbf{K}_{(m)}$ and $\mathbf{S5}_{(m)}$ with extra axioms and rules for the operator denoting common knowledge. Then they provide completeness proofs for these logics with respect to the appropriate Kripke semantics in sections that are marked as rather technical and to be skipped without loss of continuity. Indeed their proofs are too hard, using quasi-models and filtrations. The reviewer's preference would be a direct proof using maximally consistent sets in a finite set of adequate formulas, similar to the one given in **Reasoning**

about Knowledge for somewhat different logics with common knowledge. The authors' description of common knowledge in distributed systems, including the famous problem about the two generals who didn't manage to be sure enough of each other to stage a coordinated attack, is short but quite clear.

The second subject treated in Chapter 2 is *implicit knowledge* in a group. This is intuitively described as the knowledge that a group would have if they could pool their individual knowledge together. Van der Hoek and Meyer have contributed some original and surprising results about this topic themselves, both before and after writing the book. Thus it is not surprising that their treatment is quite well-informed; they wisely give only a proof sketch of their own completeness results, and refer the reader to their paper to fill in the details. Incidentally, one wonders whether during the few years before 1995, it was common or merely implicit knowledge among the two sets of authors of **Epistemic Logic for AI and Computer Science** and **Reasoning about knowledge** that there were competitors undertaking a similar project!

Then chapter 2 continues with an extensive treatment of approaches to *logical omniscience*, the fact that in standard epistemic logics agents know all logical consequences of their knowledge, and in particular all logical truths. Logical omniscience does not hold in actual situations where agents have bounded time available. For example in public key cryptography, the code contains the same information as the original text. However, without the key, the receiver cannot deduce this information due to the intractability of factorization. The authors are not partial and leave the choice between the many solutions to the logical omniscience problem to the reader. However, the reviewer would have liked to read more about the advantages and disadvantages of each approach.

The chapter is rounded off with good introductions to some logics that have been developed by AI researchers – the authors among them – to combine notions like knowledge, belief, time, and action. Unfortunately, the important philosophical questions about the relation between knowledge and belief are given extremely short shrift: it is not even mentioned explicitly that knowledge is more than simply true belief. Some references to the literature would have been appropriate, for example to W. Lenzen's overview *Recent work in epistemic logic*, **Acta Philosophica Fennica** vol. 30 (1978), pp. 1-219.

In chapter 3, the spotlight is on the flipside of knowledge, namely ignorance. For example, in Halpern and Moses' theory of epistemic states, one can express that an agent *only knows* the **S5**-consequences of atom p. One can then infer in their theory that atom q is not known to the agent. The authors show the interesting fact that an introspective agent cannot honestly claim that it only knows $Kp \vee Kq$ and its logical consequences, without knowing either Kp or Kq themselves. The chapter also contains clear treatments of preferential entailment and Moore's auto-epistemic logic.

The last chapter, chapter 4, is devoted to default reasoning. The authors describe how default reasoning may be based on epistemic logic, and introduce their own logic **EDL** (epistemic default logic) as the best candidate for this. The chapter is written in a lively style with a number of good examples. In particular, Meyer and Van der Hoek show that **EDL** has the nice property of cumulativity. This is the influential yardstick for good non-monotonic logics that was introduced in S. Kraus, D. Lehmann and M. Magidor, *Nonmonotonic reasoning, preferential models and cumulative logics*, **Artificial Intelligence** vol. 44, 1990, pp. 167–207. Unfortunately, the authors do not explain the results of Kraus, Lehmann and Magidor, so that novices to the area of nonmonotonic logics will not be able to assess the importance of cumulativity, and of the authors' own result.

Epistemic Logic for AI and Computer Science has obviously been proofread quite carefully, but still contains a number of minor mistakes and misprints, especially in the appendices. I will mention only two of them here; the reader may find a (partial) list of errata at http://tcw2.ppsw.rug.nl/r̃ineke/. First, in the analysis of the muddy children puzzle on page 57, it is not sufficient that the children are logically omniscient: this omniscience must be common knowledge among them. (Interestingly, the same omission occurs in **Reasoning about Knowledge**). A second erratum occurs in the definition of the accessibility relation R_i on the last line of page 57, where the first conjunct on the right-hand side should be dropped. As a last quibble, the typography is not easy on the reader's eyes, but this may easily be remedied in a future edition.

In general, the style of the book is rather dry, especially where the basic results of epistemic logic are explained by a sequence of definitions, lemmas and proofs with relatively little intuitive introduction. The enthusiasm which the authors definitely have for their subject shows mostly in the later parts of the book, where they treat more advanced topics, often ones to which they have contributed results. This is unfortunate for students who read this book as their first textbook. For them, it seems to the reviewer that the first three chapters of **Reasoning about Knowledge** provide a slower-paced, better motivated introduction, including entertaining examples and especially fine explanations of completeness and complexity results. On the other hand, a definite advantage of **Epistemic Logic for AI and Computer Science** is the large number of worked-out exercises. The reviewer has used the first two chapters of the book for an introductory course on epistemic logic for students of Philosophy and AI, but has complemented the book with a study guide, including some more motivation and intuitive explanation, historical remarks, examples, and exercises. This worked quite well. For more advanced students and researchers, the choice between the two textbooks depends on their field of interest. **Reasoning about Knowledge**, with its emphasis on interpreted systems that are geared to distributed systems, should be the book of choice for theoretical computer scientists. AI-students and researchers, on the other hand, may prefer the book under review and will appreciate the thorough treatment of logics modeling knowledge, belief, action and time, as well as the emphasis on relations with nonmonotonic logics. Philosophers will probably find both books interesting for the technical results but infuriatingly short on philosophical argument. They may like to complement their reading of either book with W. Lenzen's paper mentioned above.