Chapter 9

Logical Omniscience

The animal knows, of course. But it certainly does not know that it knows.

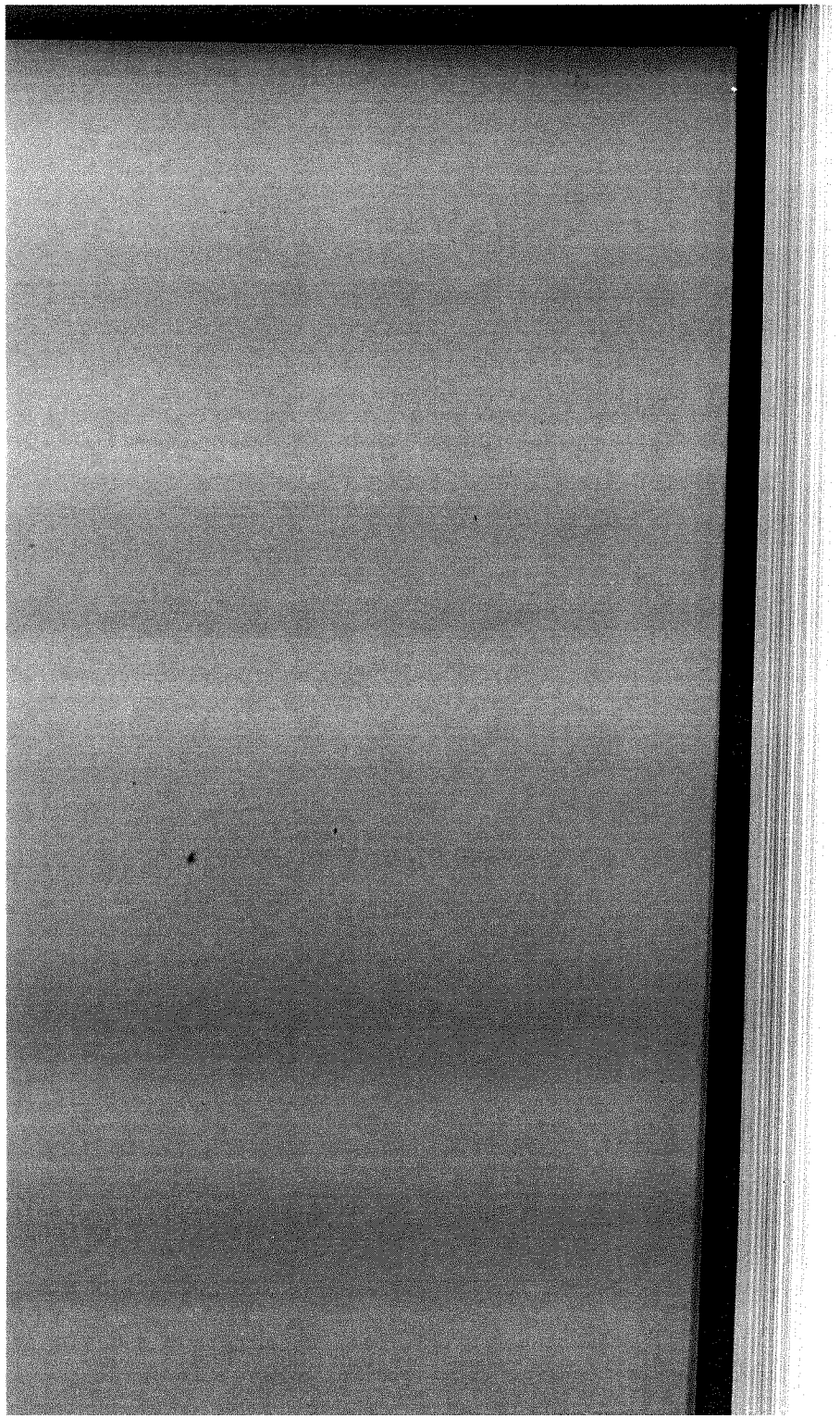
Teilhard de Chardin

A person who knows anything, by that very fact knows that he knows and knows that he knows that he knows, and so on ad infinitum.

Baruch Spinoza, *Ethics*, II, Prop. 21, 1677

Throughout this book we have found the possible-worlds model to be a very useful tool. As we already saw in Chapter 2, however, the possible-worlds model gives rise to a notion of knowledge that seems to require that agents be very powerful reasoners, since they know all consequences of their knowledge and, in particular, they know all tautologies. Thus, the agents could be described as *logically omniscient*. This does not especially trouble us in the context of multi-agent systems, since in that context we view our notion of knowledge as *external*. That is, knowledge is *ascribed* by the system designer to the agents. There is no notion of the agents computing their knowledge, and no requirement that the agents be able to answer questions based on their knowledge.

Nevertheless, there are situations where the assumption of logical omniscience seems completely inappropriate. Perhaps the most obvious example occurs when we consider human reasoning. People are simply not logically omniscient; a person can know a set of facts without knowing all of the logical consequences of this set of facts. For example, a person can know the rules of chess without knowing whether or not White has a winning strategy.



Lack of logical omniscience may stem from many sources. One obvious source is lack of computational power; for example, an agent simply may not have the computational resources to compute whether White has a winning strategy in chess. But there are other causes of lack of logical omniscience that are quite independent of computational power. For example, people may do faulty reasoning or refuse to acknowledge some of the logical consequences of what they know, even in cases where they do not lack the computational resources to compute those logical consequences. Our goal in this chapter is to develop formal models of knowledge that do not suffer from the logical-omniscience problem to the same extent that the standard possible-worlds approach does. We consider a number of approaches that are appropriate to capture different sources of the lack of logical omniscience. In Chapter 10, we describe a computational model of knowledge, which addresses issues such as capturing what a resource-bounded agent knows. As we shall see, that model also addresses the logical-omniscience problem in a way that is closely related to some of the approaches described in this chapter.

9.1 Logical Omniscience

A careful examination of the logical-omniscience problem must begin with the notion of logical omniscience itself. Exactly what do we mean by logical omniscience?

Underlying the notion of logical omniscience is the notion of *logical implication* (or *logical consequence*). Roughly speaking, a formula ψ logically implies the formula φ if φ holds whenever ψ holds; a set Ψ of formulas logically implies the formula φ if φ holds whenever all members of Ψ hold. Clearly, a formula is valid precisely if it is a logical consequence of the empty set of formulas. Like validity, logical implication is not an absolute notion, but is relative to a class of structures (and to a notion of truth, or satisfaction). Thus, more formally, we say that Ψ *logically implies* φ *with respect to a class* $\mathcal{M}$ *of structures* if, for all $M \in \mathcal{M}$ and all states s in M, whenever $(M, s) \models \psi$ for every $\psi \in \Psi$, then $(M, s) \models \varphi$. If Ψ is finite, it follows from Theorem 3.1.3 of Chapter 3 that Ψ logically implies φ with respect to $\mathcal{M}_n$ if and only if the formula $(\bigwedge \Psi) \Rightarrow \varphi$ is provable in K_n, where $\bigwedge \Psi$ is the conjunction of all formulas in Ψ. Thus, in the standard modal logic studied so far, logical and material implication coincide (where φ *materially implies* ψ if the formula $\varphi \Rightarrow \psi$ is valid). As we shall see, logical and material implication do not coincide in some of the logics considered in this chapter.

We can define *logical equivalence* in terms of logical implication. Two formulas are logically equivalent if each logically implies the other. Thus, if two formulas are

Lack of logical omniscience may stem from many sources. One obvious source ick of computational power; for example, an agent simply may not have the iputational resources to compute whether White has a winning strategy in chess. there are other causes of lack of logical omniscience that are quite independent omputational power. For example, people may do faulty reasoning or refuse to nowledge some of the logical consequences of what they know, even in cases re they do not lack the computational resources to compute those logical consences. Our goal in this chapter is to develop formal models of knowledge that do suffer from the logical-omniscience problem to the same extent that the standard sible-worlds approach does. We consider a number of approaches that are appro-ite to capture different sources of the lack of logical omniscience. In Chapter 10, describe a computational model of knowledge, which addresses issues such as turing what a resource-bounded agent knows. As we shall see, that model also resses the logical-omniscience problem in a way that is closely related to some he approaches described in this chapter.

Logical Omniscience

areful examination of the logical-omniscience problem must begin with the notion ogical omniscience itself. Exactly what do we mean by logical omniscience? Underlying the notion of logical omniscience is the notion of *logical implication logical consequence*). Roughly speaking, a formula ψ logically implies the mula φ if φ holds whenever ψ holds; a set Ψ of formulas logically implies the mula φ if φ holds whenever all members of Ψ hold. Clearly, a formula is valid cisely if it is a logical consequence of the empty set of formulas. Like validity, ical implication is not an absolute notion, but is relative to a class of structures d to a notion of truth, or satisfaction). Thus, more formally, we say that Ψ *ically implies* φ *with respect to a class* $\mathcal{M}$ *of structures* if, for all $M \in \mathcal{M}$ and states s in M, whenever $(M, s) \models \psi$ for every $\psi \in \Psi$, then $(M, s) \models \varphi$. If Ψ inite, it follows from Theorem 3.1.3 of Chapter 3 that Ψ logically implies φ with pect to $\mathcal{M}_n$ if and only if the formula $(\bigwedge \Psi) \Rightarrow \varphi$ is provable in K_n, where $\bigwedge \Psi$ he conjunction of all formulas in Ψ. Thus, in the standard modal logic studied so , logical and material implication coincide (where φ *materially implies* ψ if the mula $\varphi \Rightarrow \psi$ is valid). As we shall see, logical and material implication do not incide in some of the logics considered in this chapter.

We can define *logical equivalence* in terms of logical implication. Two formulas logically equivalent if each logically implies the other. Thus, if two formulas are

logically equivalent, then one is true precisely when the other is. Note that logical equivalence is also a relative notion, just as logical implication is.

Logical omniscience can be viewed as a certain closure property of an agent's knowledge; it says that if an agent knows certain facts and if certain conditions hold, then the agent must also know some other facts. The term logical omniscience actually refers to a family of related closure conditions. As we shall see, some of our models will eliminate certain strong forms of logical omniscience, but not necessarily all of its forms. The strongest form is what we call full logical omniscience.

- An agent is *fully logically omniscient* with respect to a class $\mathcal{M}$ of structures if, whenever he knows all of the formulas in a set Ψ, and Ψ logically implies the formula φ with respect to $\mathcal{M}$, then the agent also knows φ.

It is easy to see that we have full logical omniscience with respect to $\mathcal{M}_n$ (see Exercise 9.1). Full logical omniscience encompasses several weaker forms of omniscience. All of these notions also depend on the class of structures under consideration; we do not state that dependence here, to avoid clutter.

- *Knowledge of valid formulas*: if φ is valid, then agent i knows φ.
- *Closure under logical implication*: if agent i knows φ and if φ logically implies ψ, then agent i knows ψ.
- *Closure under logical equivalence*: if agent i knows φ and if φ and ψ are logically equivalent, then agent i knows ψ.

Because all of the above are special cases of full logical omniscience, they automatically hold for $\mathcal{M}_n$. As we shall see, however, it is possible to define classes of structures (and notions of truth) for which we do not have full logical omniscience, but do have some of these weaker notions.

There are also forms of omniscience that do not necessarily follow from full logical omniscience:

- *Closure under material implication*: if agent i knows φ and if agent i knows $\varphi \Rightarrow \psi$, then agent i also knows ψ.
- *Closure under valid implication*: if agent i knows φ and if $\varphi \Rightarrow \psi$ is valid, then agent i knows ψ.
- *Closure under conjunction*: if agent i knows both φ and ψ, then agent i knows $\varphi \wedge \psi$.

All these forms of logical omniscience in fact do hold for $\mathcal{M}_n$. Indeed, if $\{\varphi, \varphi \Rightarrow \psi\}$ logically implies ψ, as is it does with respect to $\mathcal{M}_n$, and we give $\Rightarrow$ its standard interpretation, then closure under material implication is a special case of full logical omniscience. One of the approaches described in this chapter is based on a *nonstandard* propositional semantics. In this approach, $\Rightarrow$ does not get its standard interpretation. So logical and material implication do not coincide; we get full logical omniscience but not closure under material implication. Similarly, closure under valid implication is a special case of full logical omniscience (and in fact is equivalent to closure under logical implication) if $\varphi \Rightarrow \psi$ is valid precisely when φ logically implies ψ. Again, while this is true under standard propositional semantics, it is not true in our nonstandard approach. Finally, closure under conjunction is a special case of full logical omniscience if the set $\{\varphi, \psi\}$ logically implies $\varphi \wedge \psi$. This indeed will be the case in all the logics we consider, including the nonstandard one. It is clear that there are other relationships among the types of omniscience previously discussed. For example, closure under logical equivalence follows from closure under logical implication. See Exercise 9.2 for further examples.

Full logical omniscience seems to be unavoidable, given that knowledge is defined as truth in all possible worlds, as it is in Kripke structures. If an agent knows all of the formulas in Ψ, then every formula in Ψ is true in every world he considers possible. Therefore, if Ψ logically implies φ, then φ is true in every world he considers possible. Hence, the agent must also know φ. Thus, any line of attack on the logical-omniscience problem has to start by addressing this basic definition of knowledge as truth in all possible worlds. The most radical approach is to abandon the definition altogether. We describe two approaches that do this, one syntactic and one semantic, in Section 9.2.

It is possible, however, to attack the logical-omniscience problem without completely abandoning the idea that knowledge means truth in all possible worlds. One approach to dealing with logical omniscience is to change the notion of truth. This is the direction pursued in Section 9.3, where we consider a nonstandard notion of truth. Another approach is to change the notion of possible world. This is what we do in Section 9.4, where we consider "impossible" worlds. Yet another approach is to have truth in all possible worlds be a necessary but not sufficient condition for knowledge. For example, in Section 9.5, we consider a notion of knowledge in which an agent is said to know a formula φ if φ is true in all worlds he considers possible and if, in addition, he is "aware" of the formula φ. Finally, we consider an approach in which knowledge is defined as truth in a *subset* of the possible worlds. In Section 9.6, we describe a notion of knowledge in which an agent is said to know

a formula φ if φ is true in all worlds he considers possible in some particular "frame of mind."

As we saw earlier, the possible-worlds approach gives rise to many different notions of knowledge whose appropriateness depends on the situation under consideration. For example, at the beginning of Chapter 3 we described a situation where the $\mathcal{K}_i$ relation need not be reflexive. Similarly, our goal in this chapter is to demonstrate that the logical-omniscience problem can be attacked in a variety of ways. Rather than prescribe the "correct" way to deal with logical omniscience, we describe several ways in which the problem can be dealt with. Because the issue of dealing with logical omniscience is orthogonal to the issues of dealing with distributed knowledge and common knowledge, we do not deal with distributed and common knowledge in this chapter.

In the chapters dealing with knowledge in multi-agent systems we used the term *knowledge* with a specific sense in mind: knowledge was defined by the possible-worlds semantics. The properties of this notion of knowledge were described in Chapters 2 and 3. But it is precisely this notion of knowledge that creates the logical-omniscience problem. Thus, in this chapter we start with a notion of knowledge that involves no prior assumptions on its properties. What exactly we mean by knowledge in this chapter will depend on the approach discussed and will differ from approach to approach.

9.2 Explicit Representation of Knowledge

As we already said, the simplest and most radical approach to the logical-omniscience problem is to abandon the definition of knowledge as truth in all possible worlds. This does not mean that we also have to abandon the notion of possible worlds. After all, the notion that the world can be in any one of a number of different states is independent of the concept of knowledge, and it arises naturally in a number of contexts, in particular, as we have seen, in our model of multi-agent systems. In this section, we intend to keep the possible-worlds framework, but change the way we define knowledge. Instead of defining knowledge *in terms* of possible worlds, we let knowledge be defined directly. Intuitively, we think of each agent's knowledge as being explicitly stored in a database of formulas. We now describe two ways to capture this intuition: a syntactic approach and its semantic analogue.

9.2.1 The Syntactic Approach

As we saw in Chapter 3, a Kripke structure $M = (S, \pi, \mathcal{K}_1, \ldots, \mathcal{K}_n)$ consists of frame $F = (S, \mathcal{K}_1, \ldots, \mathcal{K}_n)$ and an assignment π of truth values to the primitive propositions in each state. Our definition of the satisfaction relation $\models$ then gives truth values to all formulas in all states. Here we replace the truth assignment π by a *syntactic assignment*. A syntactic assignment simply assigns truth values to all formulas in all states. For example, a syntactic assignment σ can assign both p and $\neg p$ to be true in a state s.

Since in this section we are interested in changing the definition of knowledge but not the underlying propositional semantics, we restrict our syntactic assignments here to be standard. A *standard* syntactic assignment σ is a syntactic assignment that obeys the following constraints for all formulas φ and ψ:

$\sigma(s)(\varphi) =$ **true** if and only if $\sigma(s)(\neg\varphi) =$ **false**, and

$\sigma(s)(\varphi \wedge \psi) =$ **true** if and only if $\sigma(s)(\varphi) =$ **true** and $\sigma(s)(\psi) =$ **true**.

Thus, in syntactic structures we replace truth assignments by standard syntactic assignments. In addition, we discard the possibility relations, because these relations were needed just to define satisfaction for formulas of the form $K_i\varphi$. Formally, a *syntactic structure* M is a pair (S, σ), consisting of a set S of states and a standard syntactic assignment σ. We can now define the truth of a formula φ in a syntactic structure in a straightforward way: $(M, s) \models \varphi$ precisely when $\sigma(s)(\varphi) =$ **true**. Notice that we can identify every Kripke structure $M = (S, \pi, \mathcal{K}_1, \ldots, \mathcal{K}_n)$ with the syntactic structure (S, σ), where $\sigma(s)(\varphi) =$ **true** if $(M, s) \models \varphi$. Thus, syntactic structures can be viewed as a generalization of Kripke structures. In fact, syntactic structures provide the most general model of knowledge. (More precisely, they provide a most general model for knowledge among models that are based on standard propositional semantics.)

It is easy to see that no form of logical omniscience holds for syntactic structures. For example, knowledge of valid formulas fails, because there is no requirement that a standard syntactic assignment assign the truth value **true** to formulas of the form $K_i\varphi$ where φ is a valid formula. Similarly, closure under logical equivalence fails, since φ and ψ could be logically equivalent, but a standard syntactic assignment may assign the truth value **true** to $K_i\varphi$ and the truth value **false** to $K_i\psi$. In fact, knowledge in syntactic structures does not have any interesting properties; the only formulas that are valid in all syntactic structures are substitution instances of propositional tautologies. If we want to use syntactic structures to model a notion of knowledge that does

obey certain properties, then we have to impose some constraints on the allowable standard syntactic assignments. For example, if we want to capture the fact that we are modeling knowledge rather than belief, then we can enforce the Knowledge Axiom ($K_i\varphi \Rightarrow \varphi$) by requiring that $\sigma(s)(\varphi) =$ **true** whenever $\sigma(s)(K_i\varphi) =$ **true**.

One particular property of knowledge that syntactic structures fail to capture is a property that played a central role in our study of multi-agent systems in Chapter 4. In that framework we assumed that every agent in the system is in some local state at any point in time. An important feature of knowledge in the framework of Chapter 4 is its *locality*. That is, if s and s' are states of the system such that agent i has the same local state in both of them, i.e., $s \sim_i s'$, and agent i knows φ in state s, then i also knows φ in state s'. In Chapter 4, this property was a consequence of the definition of knowledge as truth in all possible worlds, but it is a property that we may want to keep even if we abandon that definition. For example, a natural interpretation of $\sigma(s)(K_i\varphi) =$ **true**, which we pursue in Chapter 10, is that agent i can decide, say by using some algorithm A, whether φ follows from the information in i's local state. Unfortunately, syntactic structures have no notion of local state; we cannot get locality of knowledge just by imposing further restrictions on the syntactic assignments. If we want to capture locality of knowledge, then we need to reintroduce possibility relations, because we can use them to express locality. If the possibility relation $\mathcal{K}_i$ is an equivalence relation $\sim_i$, for $i = 1, \ldots, n$, then we can say that an agent i is in the same local state in states s and t precisely when $s \sim_i t$. For knowledge to depend only on the agent's local state, we should require that if $s \sim_i t$, then $\sigma(s)(K_i\varphi) = \sigma(t)(K_i\varphi)$.

Syntactic structures can be used to model fairly realistic situations. Consider the situation where each agent has a base set of formulas from which the agent's knowledge is derived using a sound but possibly incomplete set of inference rules. Formally, we could interpret this to mean that for each agent i, there is a formal system R_i of inference rules, and for each state $s \in S$, there is a set $B_i(s)$ (the base set of formulas) such that $\sigma(s)(K_i\varphi) =$ **true** iff φ is derivable from $B_i(s)$ using R_i. Intuitively, agent i knows φ if she can deduce φ from her base formulas using her inference rules. For example, a student might know that $x + a = b$ but not conclude that $x = b - a$, because he might not know the rule allowing subtraction of equal quantities from both sides. In this case, we would have $\sigma(s)(K_i(x+a=b)) =$ **true**, but $\sigma(s)(K_i(x = b-a)) =$ **false**. As another example, a deduction system might be capable of certain limited reasoning about equality. For example, from $A = B$ and $B = C$, it might be able to deduce that $A = C$; however, given the information that $f(1) = 1$ and that $f(x) = x \cdot f(x-1)$, it might not be able to deduce that $f(4) = 24$. In both of these cases, agents have a base set of formulas and an incomplete set of

inference rules. Notice that, in these examples, if we view the base set $B_i(s)$ of formulas as agent i's local state, then agent i's knowledge indeed depends only on his local state, so knowledge has the locality property.

9.2.2 The Semantic Approach

In a syntactic structure $M = (S, \sigma)$, for each state s the function $\sigma(s)$ tells which formulas are true at state s. In particular, agent i knows φ at state s precisely if $\sigma(s)(K_i\varphi) = \mathbf{true}$. Essentially, an agent's knowledge is *explicitly* described at a state by giving a list of the formulas that he knows. While this approach indeed avoids the logical-omniscience problem, its main weakness is its syntactic flavor. After all, the separation between syntax and semantics is one of the major strengths of modern logic. Is it possible to "semanticize" this approach? That is, is it possible to model knowledge explicitly on a semantic level?

To do that, we first need to find the semantic counterpart of formulas. We can identify the semantic "content" of a formula φ with its *intension* (see Section 2.5), i.e., the set of states in which φ holds. The motivation for this identification is as follows. Let φ and ψ be two formulas with the same intension in a structure M. Then for all $s \in S$ we have that $(M, s) \models \varphi$ if and only if $(M, s) \models \psi$. That is, if φ and ψ have the same intension in M, then they are semantically indistinguishable in M. Consequently, we can take sets of states as the semantic counterpart of formulas. Put another way, we can think of a set W of states as a "proposition" p_W that is true precisely at the states of W. Thus, we can represent an agent's semantic knowledge by simply listing the propositions that he knows, instead of representing his knowledge syntactically by listing the formulas that he knows. Since a proposition is a set of states, we can describe agent i's semantic knowledge explicitly by a set of sets of states.

The previous discussion motivates the following definition. A *Montague-Scott structure* M is a tuple $(S, \pi, \mathcal{C}_1, \ldots, \mathcal{C}_n)$ where S is a set of states, $\pi(s)$ is a truth assignment to the primitive propositions for each state $s \in S$, and $\mathcal{C}_i(s)$ is a set of subsets of S, for $i = 1, \ldots, n$. For the sake of brevity, we refer to Montague-Scott structures as *MS structures*. In an MS structure, we describe agent i's knowledge (in state s) by a set of sets of states; this is given to us by $\mathcal{C}_i(s)$. The members of $\mathcal{C}_i(s)$ are the propositions that agent i knows.

We can now define $\models$ for all formulas. The clauses for primitive propositions, conjunctions, and negations are identical to the corresponding clauses for Kripke structures. The clause for formulas $K_i\varphi$ is different:

$(M, s) \models K_i\varphi$ iff $\{t \mid (M, t) \models \varphi\} \in \mathcal{C}_i(s)$.

As in Section 2.5, we denote the intension of a formula φ in the structure M by φ^M. That is, $\varphi^M = \{s \mid (M, s) \models \varphi\}$ is the set of states in M where φ is true. The clause above says that agent i knows φ at state s if the intension of φ is one of the propositions that he knows, that is, if $\varphi^M \in \mathcal{C}_i(s)$.

Example 9.2.1 These definitions are perhaps best illustrated by a simple example. Suppose $\Phi = \{p\}$ and $n = 2$, so that our language has one primitive proposition p and there are two agents. Further suppose that $M' = (S, \pi, \mathcal{C}_1, \mathcal{C}_2)$, where $S = \{s, t, u\}$, and that the primitive proposition p is true at states s and u, but false at t (so that $\pi(s)(p) = \pi(u)(p) = \mathbf{true}$ and $\pi(t)(p) = \mathbf{false}$). Suppose $\mathcal{C}_1(s) = \mathcal{C}_1(t) = \{\{s, t\}, \{s, t, u\}\}$ and $\mathcal{C}_1(u) = \{\{u\}, \{s, u\}, \{t, u\}, \{s, t, u\}\}$. Suppose also that $\mathcal{C}_2(s) = \mathcal{C}_2(u) = \{\{s, u\}, \{s, t, u\}\}$ and $\mathcal{C}_2(t) = \{\{t\}, \{s, t\}, \{t, u\}, \{s, t, u\}\}$. Consider agent 1. In states s and t agent 1 knows the propositions $\{s, t\}$ and $\{s, t, u\}$, and in state u he knows the propositions $\{u\}$, $\{s, u\}$, $\{t, u\}$, and $\{s, t, u\}$. In some sense, one could say that agent 1 cannot distinguish between the states s and t, since $\mathcal{C}_1(s) = \mathcal{C}_1(t)$, and s and t play symmetric roles in $\mathcal{C}_1(s)$, $\mathcal{C}_1(t)$, and $\mathcal{C}_1(u)$. On the other hand, agent 1 can distinguish between the state u and each of s and t, since $\mathcal{C}_1(s) = \mathcal{C}_1(t) \neq \mathcal{C}_1(u)$. The situation for agent 2 is analogous.

Consider the formulas K_1p and $K_1\neg p$. The intension of p is $\{s, u\}$ and the intension of $\neg p$ is $\{t\}$. Since $\{s, u\} \notin \mathcal{C}_1(s)$ and $\{t\} \notin \mathcal{C}_1(s)$, in state s agent 1 does not know whether or not p holds. That is,

$$(M', s) \not\models K_1p \vee K_1\neg p.$$

Consider the formulas K_2p and $K_2\neg p$. Because the intension of p is $\{s, u\}$, and since $\{s, u\} \in \mathcal{C}_2(s)$, $\{s, u\} \in \mathcal{C}_2(u)$, and $\{s, u\} \notin \mathcal{C}_2(t)$, the intension of K_2p is $\{s, u\}$. Similarly, the intension of $K_2\neg p$ is $\{t\}$. Thus, the intension of $K_2p \vee K_2\neg p$ is $\{s, t, u\}$. Because $\{s, t, u\} \in \mathcal{C}_1(s)$, it follows that in state s agent 1 knows that agent 2 knows whether or not p is true. That is,

$$(M', s) \models K_1(K_2p \vee K_2\neg p).$$

It is instructive to compare this example with the example described by Figure 2.1. In that example, we considered a Kripke structure $M = (S, \pi, \mathcal{K}_1, \mathcal{K}_2)$ with the same state space as M'. There are a number of other more significant similarities between M and M'. Just as with M', agent 1 cannot distinguish s and t in M, although he can distinguish u from both of them. Similarly, agent 2 cannot distinguish s and u in M, although she can distinguish s and t. Notice, however, that the way we captured indistinguishability in M (using the relations $\mathcal{K}_1$ and $\mathcal{K}_2$) is very different

from the way we capture it in M' (in terms of the sets of propositions the agents know). Nevertheless, it can be shown that precisely the same formulas are true at corresponding states in the two structures. As we shall see, this similarity between the MS structure M' and the Kripke structure M is not a coincidence. ∎

We observed in Section 9.2.1 that syntactic structures generalize Kripke structures. Similarly, MS structures can also be viewed as a generalization of Kripke structures. Thus, let $M = (S, \pi, \mathcal{K}_1, \ldots, \mathcal{K}_n)$ be a Kripke structure. Let $\mathcal{K}_i(s)$ be the set of all "i-neighbors" of a state s, i.e.,

$$\mathcal{K}_i(s) = \{t \mid (s, t) \in \mathcal{K}_i\}.$$

Let M' be the MS structure $(S, \pi, \mathcal{C}_1, \ldots, \mathcal{C}_n)$, where $\mathcal{C}_i(s)$ is the set of all supersets of $\mathcal{K}_i(s)$. Intuitively, in a state s of M, an agent i knows that the actual state is one of the states in $\mathcal{K}_i(s)$. Thus, i knows all the propositions that contain $\mathcal{K}_i(s)$. It can now be shown that for each formula φ we have $(M, s) \models \varphi$ iff $(M', s) \models \varphi$ (Exercise 9.4). This explains the tight correspondence between the MS structure M' and the Kripke structure M observed in Example 9.2.1. The reader can verify that in that example we indeed had that $\mathcal{C}_i(s)$ is the set of all supersets of $\mathcal{K}_i(s)$.

Earlier we observed that syntactic structures strip away all properties of knowledge. This is not quite the case for MS structures. Suppose that φ and ψ are equivalent formulas. Then, by definition, they must have the same intension in every MS structure. It follows that $K_i\varphi$ is true in a state precisely when $K_i\psi$ is true in that state. Thus, knowledge in MS structures is closed under logical equivalence. It is easy to verify, however, that all other forms of logical omniscience fail here (see Exercise 9.5). In fact, closure under logical equivalence is in some formal sense the only necessary property of knowledge in MS structures.

Theorem 9.2.2 *The following is a sound and complete axiomatization for validity with respect to MS structures:*

A1. All instances of tautologies of propositional logic

R1. From φ and $\varphi \Rightarrow \psi$ infer ψ (modus ponens)

LE. From $\varphi \Leftrightarrow \psi$ infer $K_i\varphi \Leftrightarrow K_i\psi$

Proof See Exercise 9.6. ∎

Thus, propositional reasoning and closure under logical equivalence completely characterize knowledge in MS structures. This suggests that reasoning about knowledge in Montague-Scott semantics may not be any harder than propositional reasoning. This indeed is the case, as the following result shows.

Theorem 9.2.3 *The satisfiability problem with respect to MS structures is NP-complete.*

Proof The lower bound is immediate, since the satisfiability problem for propositional logic is already *NP*-hard. For the upper bound, we proceed along similar lines to the proof of Proposition 3.6.2: we show that if a formula φ is satisfiable, then it is satisfiable in an MS structure with at most $|\varphi|^2$ states. We leave details to the reader (Exercise 9.7); however, the following example might provide some intuition. Consider the formula $\varphi = K_i\varphi_1 \wedge \ldots \wedge K_i\varphi_r \wedge \neg K_i\psi$. Clearly, for φ to be satisfiable, ψ cannot be logically equivalent to any of the φ_j, for $j = 1, \ldots, r$. In other words, not knowing ψ has to be consistent with knowing φ_j, for $j = 1, \ldots, r$. It turns out that the consistency of not knowing ψ with knowing φ_j, for $j = 1, \ldots, r$, is also a sufficient condition for the satisfiability of φ. This means that we can test if φ is satisfiable by testing if $K_i\varphi_j \wedge \neg K_i\psi$ is satisfiable, for $j = 1, \ldots, r$. Moreover, it is easy to see that $K_i\varphi_j \wedge \neg K_i\psi$ is satisfiable exactly if at least one of $\neg\varphi_j \wedge \psi$ or $\varphi_j \wedge \neg\psi$ is satisfiable. Thus, we can decompose the problem of testing if φ is satisfiable into testing a number of smaller satisfiability problems. In fact, it can be shown that there are only quadratically many such problems to test (at most two for every pair of subformulas of φ), so the problem is in *NP*. ∎

Do MS structures avoid logical omniscience? The answer is "almost." As we observed, all forms of logical omniscience fail except for closure under logical equivalence. In other words, while agents need not know all logical consequences of their knowledge, they are unable to distinguish between logically equivalent formulas. This is as much as we can expect to accomplish in a purely semantic model, since logically equivalent formulas are by definition semantically indistinguishable. Thus, just as syntactic structures provide the most general model of knowledge, MS structures provide the most general *semantic* model of knowledge. (Of course, just as in the case of syntactic structures, MS structures provide the most general semantic model of knowledge only among models that are based on standard propositional semantics, since Montague-Scott semantics is based on standard propositional semantics.)

We saw in Chapter 3 how certain properties of knowledge in Kripke structures correspond to certain conditions on the possibility relations $\mathcal{K}_i$. Similarly, certain properties of knowledge in MS structures correspond to certain conditions on the $\mathcal{C}_i$'s. Especially interesting are properties of knowledge that correspond to various forms of logical omniscience. For example, knowledge of valid formulas corresponds to the condition that $S \in \mathcal{C}_i(s)$, where S is the set of all states, since the intension of *true* is the set S. For another example, consider closure under conjunction. This

roperty corresponds to $\mathcal{C}_i(s)$ being closed under intersection; that is, if U and V re in $\mathcal{C}_i(s)$, then $U \cap V$ is also in $\mathcal{C}_i(s)$. The reason for this is that the intension f $\varphi \wedge \psi$ is the intersection of the intensions of φ and ψ. Exercise 9.8 provides a recise statement of the equivalence between various properties of knowledge and orresponding restrictions on the $\mathcal{C}_i$'s.

We already saw in Chapter 3 that imposing certain restrictions on Kripke structures, or, equivalently, assuming that knowledge satisfies some additional properties ıay sometimes (but not always) have an effect on the computational complexity of easoning about knowledge. A similar phenomenon occurs in the context of MS tructures. It turns out that we can capture several properties of knowledge without ıcreasing the complexity of reasoning beyond that of propositional reasoning (i.e., NP-complete). Once, however, we capture closure under conjunction by insisting ıat each $\mathcal{C}_i(s)$ be closed under intersection, then the complexity of the satisfiability roblem rises to *PSPACE*-complete. To understand the intuitive reason for this difference, consider again the formula $\varphi = K_i\varphi_1 \wedge \ldots \wedge K_i\varphi_r \wedge \neg K_i\psi$. As we observed ı the proof of Theorem 9.2.3, if we do not assume closure under conjunction, then a ecessary and sufficient condition for φ to be satisfiable is that ψ cannot be logically quivalent to any of the φ_j, for $j = 1, \ldots, r$. The situation is quite different when we o assume closure under conjunction. Now it is not sufficient that ψ not be logically quivalent to any of the φ_j's; it also cannot be equivalent to any conjunction of φ_j's. n other words, in the presence of closure under conjunction we have to show that ot knowing ψ is simultaneously consistent with knowing any conjunction of φ_j's. hus, to test whether φ is satisfiable, we have to consider sets of subformulas of φ ather than only pairs of subformulas of φ. Since there are exponentially many such ets, the problem is *PSPACE*-hard.

.2.3 Discussion

he two approaches described in this section overcome the logical omniscience prob- em by explicitly modeling an agent's knowledge, either as a set of formulas (the ormulas the agent knows) or as a set of sets of possible worlds (the intensions of he formulas the agent knows). These approaches are very powerful. They solve the ogical-omniscience problem by giving us direct fine-grained control over an agent's nowledge. This power, however, comes at a price. One gains very little intuition bout knowledge from studying syntactic structures or MS structures; in these ap- roaches knowledge is a primitive construct (much like the primitive propositions n a Kripke structure). Arguably, these approaches give us ways of *representing*

knowledge in Kripke structures *explains* knowledge as truth in all possible worlds. Unfortunately, this "explanation" does not fit certain applications, because it forces logical omniscience.

In the following sections, we try to steer a middle course, by keeping the flavor of the possible-worlds approach, while trying to mitigate its side effects.

9.3 Nonstandard Logic

If knowledge is truth in all possible worlds, then one way to deal with logical omniscience is to change the notion of truth. The underlying idea is to weaken the "logical" aspect of the logical-omniscience problem, thus reducing the acuteness of the problem. Indeed, as we saw in Section 9.1, certain forms of logical omniscience follow from full logical omniscience only under standard propositional semantics. The nonstandard semantics for knowledge we are about to describe is based on a nonstandard propositional semantics. Knowledge is still defined to be truth in all possible worlds, so we still have logical omniscience, but this time with respect to the nonstandard logic. The hope is that the logical-omniscience problem can be alleviated somewhat by appropriately choosing the nonstandard logic.

There are many ways in which one can define a nonstandard propositional semantics. We describe here one approach that changes the treatment of negation. We do not mean to argue that this is the "right" propositional semantics to deal with knowledge, but rather we mean to demonstrate how knowledge can be modeled on the basis of a nonstandard propositional semantics.

9.3.1 Nonstandard Structures

Standard propositional logic has several undesirable and counterintuitive properties. Often people first introduced to propositional logic are somewhat uncomfortable when they learn that "$\varphi \Rightarrow \psi$" is taken to be simply an abbreviation for $\neg\varphi \vee \psi$. Why should the fact that either $\neg\varphi$ is true or ψ is true correspond to "if φ is true, then ψ is true"?

Another problem with standard propositional logic is that it is fragile: a false statement implies everything. In particular, the formula $(p \wedge \neg p) \Rightarrow q$ is valid, even when p and q are unrelated primitive propositions; for example, p could say that Alice graduated from college in 1987 and q could say that Bob's salary is $500,000. This could be a serious problem if we have a large database of formulas obtained

xample, someone may have input the datum that Alice graduated in 1987, and omeone else may have input the datum that Alice graduated in 1986. If the database ontains a constraint that each person's graduation year is unique, then, using standard propositional reasoning, any arbitrary fact about Bob's salary can be derived from he database. Many alternatives to the standard semantics have been proposed over he years, designed to deal with various aspects of these problems. We focus on one particular alternative here, and consider its consequences.

The idea is to allow formulas φ and $\neg\varphi$ to have "independent" truth values. Thus, rather than requiring that $\neg\varphi$ be true if and only if φ is false, we wish instead to allow the possibility that $\neg\varphi$ can be either true or false, regardless of whether φ is true or false. Intuitively, the truth of formulas can be thought of as being determined by some database of formulas. We can think of φ being true as meaning that the fact φ is in a database of true formulas, and we can think of $\neg\varphi$ being true as meaning that the fact φ is in a database of false formulas. Since it is possible for φ to be in both databases, it is possible for both φ and $\neg\varphi$ to be true. Similarly, if φ is in neither database, then neither φ nor $\neg\varphi$ would be true.

There are several ways to capture this intuition formally. We now discuss one approach; some closely related approaches are discussed in Exercises 9.9 and 9.10. For each state s, there is an *adjunct* state s^*, which is used for giving semantics to negated formulas. Rather than defining $\neg\varphi$ to hold at s iff φ does not hold at s, we instead define $\neg\varphi$ to hold at s iff φ does not hold at s^*. Note that if $s = s^*$, then this gives our usual notion of negation. Very roughly, we can think of a state s as consisting of a pair $\langle B_T, B_F\rangle$ of databases; B_T is the database of true facts, while B_F is the database of false facts. The state s^* should be thought of as the adjunct pair $\langle \overline{B_F}, \overline{B_T}\rangle$ (where, if X is a set of formulas, then $\overline{X}$ is the set consisting of all formulas not in X). Continuing this intuition, to see if φ holds at s, we check if $\varphi \in B_T$; to see if $\neg\varphi$ holds at s, we check if $\varphi \in B_F$. Notice that $\varphi \in B_F$ iff $\varphi \notin \overline{B_F}$. Since $\overline{B_F}$ is the database of true facts at s^*, we have an alternative way of checking if $\neg\varphi$ holds at s: we can check if φ does not hold at s^*. Note that if $B_T = \overline{B_F}$, then $s = s^*$ and we get the standard semantics of negation.

Under this interpretation, not only is s^* the adjunct state of s, but s is the adjunct state of s^*; i.e., $s^{**} = s$ (where $s^{**} = (s^*)^*$). To support this intuitive view of s as a pair of databases and s^* as its adjunct, we make $s^{**} = s$ a general requirement in our framework.

A *nonstandard (Kripke) structure* M is a tuple $(S, \pi, \mathcal{K}_1, \ldots, \mathcal{K}_n, {}^*)$, where the tuple $(S, \pi, \mathcal{K}_1, \ldots, \mathcal{K}_n)$ is a Kripke structure, and * is a unary function from the set S of worlds to itself (where we write s^* for the result of applying the function * to the state s) such that $s^{**} = s$ for each $s \in S$. We call these structures nonstandard,

since we think of a world where φ and $\neg\varphi$ are both true or both false as nonstandard. We denote the class of nonstandard structures for n agents over Φ by $\mathcal{NM}_n(\Phi)$ (or by $\mathcal{NM}_n$ when Φ is clear from the context).

The definition of $\models$ is the same as for standard Kripke structures, except for the clause for negation. In this case, we have

$(M, s) \models \neg\varphi$ iff $(M, s^*) \not\models \varphi$.

Note that it is possible for neither φ nor $\neg\varphi$ to be true at state s (if $(M, s) \not\models \varphi$ and $(M, s^*) \models \varphi$) and for both φ and $\neg\varphi$ to be true at state s (if $(M, s) \models \varphi$ and $(M, s^*) \not\models \varphi$). We call a state s where neither φ nor $\neg\varphi$ is true *incomplete* (with respect to φ); otherwise, we call s *complete* (with respect to φ). The intuition behind an incomplete state is that there is not enough information to determine whether φ is true or whether $\neg\varphi$ is true. We call a state s where both φ and $\neg\varphi$ are true *incoherent* (with respect to φ); otherwise, s is *coherent* (with respect to φ). The intuition behind an incoherent state is that it is overdetermined; it might correspond to a situation where several people have provided mutually inconsistent information.

A state s is *standard* if $s = s^*$. Note that for a standard state, the semantics of negation is equivalent to the standard semantics. In particular, a standard state s is both complete and coherent with respect to all formulas: for each formula φ exactly one of φ or $\neg\varphi$ is true at s. (See also Exercise 9.11.)

In standard propositional logic, disjunction ($\vee$) and material implication ($\Rightarrow$) can be defined in terms of conjunction and negation, that is, $\varphi_1 \vee \varphi_2$ can be defined as $\neg(\neg\varphi_1 \wedge \neg\varphi_2)$, and $\varphi_1 \Rightarrow \varphi_2$ can be defined as $\neg\varphi_1 \vee \varphi_2$. We retain these definitions in the nonstandard framework. Since, however, the semantics of negation is now nonstandard, it is not *a priori* clear how the propositional connectives behave in our nonstandard semantics. For example, while $p \wedge q$ holds by definition precisely when p and q both hold, it is not clear that $p \vee q$ holds precisely when at least one of p or q holds. It is even less clear how negation interacts with conjunction and disjunction in our nonstandard semantics. The next proposition shows that even though we have decoupled the semantics for φ and $\neg\varphi$, the propositional connectives $\neg$, $\wedge$, and $\vee$ still behave in a fairly standard way.

Proposition 9.3.1 *Let M be a nonstandard structure. Then*

(a) $(M, s) \models \neg\neg\varphi$ *iff* $(M, s) \models \varphi$.

(b) $(M, s) \models \varphi \vee \psi$ *iff* $(M, s) \models \varphi$ *or* $(M, s) \models \psi$.

(c) $(M, s) \models \neg(\varphi \wedge \psi)$ *iff* $(M, s) \models \neg\varphi \vee \neg\psi$.

(d) $(M, s) \models \neg(\varphi \vee \psi)$ *iff* $(M, s) \models \neg\varphi \wedge \neg\psi$.

(e) $(M, s) \models \varphi \wedge (\psi_1 \vee \psi_2)$ *iff* $(M, s) \models (\varphi \wedge \psi_1) \vee (\varphi \wedge \psi_2)$.

(f) $(M, s) \models \varphi \vee (\psi_1 \wedge \psi_2)$ *iff* $(M, s) \models (\varphi \vee \psi_1) \wedge (\varphi \vee \psi_2)$.

Proof See Exercise 9.12. ∎

In contrast to $\wedge$ and $\vee$, the connective $\Rightarrow$ behaves in a nonstandard fashion. In particular, both p and $p \Rightarrow q$ can be true at a state without q being true, so $\Rightarrow$ does not capture our usual notion of logical implication (see Exercise 9.14).

What are the properties of knowledge in nonstandard structures? So far, our approach to understanding the properties of knowledge in some semantic model has been to consider all the valid formulas under that semantics. What are the valid formulas with respect to $\mathcal{NM}_n$? It is easy to verify that certain tautologies of standard propositional logic are not valid. For example, the formula $(p \wedge \neg p) \Rightarrow q$, which wreaked havoc in deriving consequences from a database, is not valid. How about even simpler tautologies of standard propositional logic, such as $p \Rightarrow p$? This formula, too, is not valid. One might think that these formulas are not valid because of the nonstandard behavior of $\Rightarrow$, but observe that $p \Rightarrow p$ is just an abbreviation for $\neg p \vee p$ (which is not valid). In fact, no formula is valid with respect to $\mathcal{NM}_n$! Furthermore, there is a single structure that simultaneously shows that no formula is valid!

Theorem 9.3.2 *No formula of $\mathcal{L}_n$ is valid with respect to $\mathcal{NM}_n$. In fact, there is a nonstandard structure M and a state s of M such that every formula of $\mathcal{L}_n$ is false at s, and a state t of M such that every formula of $\mathcal{L}_n$ is true at t.*

Proof Let $M = (S, \pi, \mathcal{K}_1, \ldots, \mathcal{K}_n, {}^*)$ be a special nonstandard structure, defined as follows. Let S contain only two states s and t, where $t = s^*$ (and so $s = t^*$). Define π by taking $\pi(s)$ be the truth assignment where $\pi(s)(p) =$ **false** for every primitive proposition p, and taking $\pi(t)$ be the truth assignment where $\pi(t)(p) =$ **true** for every primitive proposition p. Define $\mathcal{K}_i$ to be $\{(s, s), (t, t)\}$, for $i = 1, \ldots, n$. By a straightforward induction on the structure of formulas (Exercise 9.13), it follows that for every formula φ of $\mathcal{L}_n$, we have $(M, s) \not\models \varphi$ and $(M, t) \models \varphi$. In particular, every formula of $\mathcal{L}_n$ is false at s and every formula of $\mathcal{L}_n$ is true at t. Since every formula of $\mathcal{L}_n$ is false at s, no formula of $\mathcal{L}_n$ is valid with respect to $\mathcal{NM}_n$. ∎

It follows from Theorem 9.3.2 that the validity problem with respect to $\mathcal{NM}_n$ is very easy: the answer is always, "No, the formula is not valid!" Thus, the notion

of validity is trivially uninteresting in our logic. In particular, we cannot use valid formulas to characterize the properties of knowledge in nonstandard structures, since there are no valid formulas.

In contrast to validity, there are many nontrivial logical implications with respect to $\mathcal{NM}_n$. For example, as we see from Proposition 9.3.1, $\neg\neg\varphi$ logically implies φ and $\neg(\varphi_1 \wedge \varphi_2)$ logically implies $\neg\varphi_1 \vee \neg\varphi_2$. The reader may be puzzled why Proposition 9.3.1 does not provide us with some tautologies. For example, Proposition 9.3.1 tells us that $\neg\neg\varphi$ logically implies φ. Doesn't this mean that $\neg\neg\varphi \Rightarrow \varphi$ is a tautology? This does not follow. With *standard* Kripke structures, φ logically implies ψ iff the formula $\varphi \Rightarrow \psi$ is valid. This is not the case for *nonstandard* structures; here, logical and material implication do not coincide. For example, φ logically implies φ, yet we have already observed that $\varphi \Rightarrow \varphi$ (i.e., $\neg\varphi \vee \varphi$) is not valid with respect to $\mathcal{NM}_n$. In Section 9.3.2, we define a new connective that allows us to express logical implication *in the language*, just as $\Rightarrow$ does for standard Kripke structures.

What about logical omniscience? Full logical omniscience holds, just as with ordinary Kripke structures. For example, it follows from Proposition 9.3.1(b) that φ logically implies $\varphi \vee \psi$; hence, by full logical omniscience, $K_i\varphi$ logically implies $K_i(\varphi \vee \psi)$. Moreover, closure under conjunction holds, since $\{\varphi, \psi\}$ logically implies $\varphi \wedge \psi$. Nevertheless, since it is not the case here that $\{\varphi, \varphi \Rightarrow \psi\}$ logically implies ψ (Exercise 9.14), we might expect that closure under material implication would fail. This is indeed the case: it is possible for $K_i\varphi$ and $K_i(\varphi \Rightarrow \psi)$ to hold, without $K_i\psi$ holding (Exercise 9.16). Finally, note that although knowledge of valid formulas holds, it is completely innocuous here; there are no valid formulas!

9.3.2 Strong Implication

In the previous subsection we introduced nonstandard semantics, motivated by our discomfort with the tautology $(p \wedge \neg p) \Rightarrow q$, and, indeed, under this semantics $(p \wedge \neg p) \Rightarrow q$ is no longer valid. Unfortunately, the bad news is that other formulas, such as $\varphi \Rightarrow \varphi$, that seem as if they should be valid, are not valid either. In fact, as we saw, no formula is valid in the nonstandard approach. It seems that we have thrown out the baby with the bath water.

To get better insight into this problem, let us look more closely at why the formula $\varphi \Rightarrow \varphi$ is not valid. Our intuition about implication tells us that $\varphi_1 \Rightarrow \varphi_2$ should say "if φ_1 is true, then φ_2 is true," but $\varphi_1 \Rightarrow \varphi_2$ is defined to be $\neg\varphi_1 \vee \varphi_2$, which says "either $\neg\varphi_1$ is true or φ_2 is true." In standard propositional logic, this is the same as "if φ_1 is true, then φ_2 is true," since $\neg\varphi_1$ is false in standard logic iff φ_1 is true. In

nonstandard structures, however, these are not equivalent. In particular, $\varphi \Rightarrow \varphi$ is simply an abbreviation for $\neg\varphi \vee \varphi$. Since our semantics decouples the meaning of φ and $\neg\varphi$, the formula $\neg\varphi \vee \varphi$ *should not* be valid.

While the above explains why $\neg\varphi \vee \varphi$ is not valid, it still seems that the statement "if φ is true then φ is true" ought to be valid. Unfortunately, our definition of $\Rightarrow$ does not capture the intuition of "if ... then ...." This motivates the definition of a new propositional connective $\hookrightarrow$, which we call *strong implication*, where $\varphi_1 \hookrightarrow \varphi_2$ is defined to be true if whenever φ_1 is true, then φ_2 is true. Formally,

$(M, s) \models \varphi_1 \hookrightarrow \varphi_2$ iff $(M, s) \models \varphi_2$ holds whenever $(M, s) \models \varphi_1$ does.

That is, $(M, s) \models \varphi_1 \hookrightarrow \varphi_2$ iff either $(M, s) \not\models \varphi_1$ or $(M, s) \models \varphi_2$.

We denote by $\mathcal{L}_n^{\hookrightarrow}(\Phi)$, or $\mathcal{L}_n^{\hookrightarrow}$ when Φ is obvious from the context, the set of formulas obtained by closing off $\mathcal{L}_n(\Phi)$ under the connective $\hookrightarrow$ (i.e., if φ and ψ are formulas of $\mathcal{L}_n^{\hookrightarrow}(\Phi)$, then so is $\varphi \hookrightarrow \psi$). We call the formulas of $\mathcal{L}_n^{\hookrightarrow}$ *nonstandard formulas*, to distinguish them from the *standard formulas* of $\mathcal{L}_n$. We call the propositional fragment of $\mathcal{L}_n^{\hookrightarrow}$ and its interpretation by nonstandard structures *nonstandard propositional logic*, to distinguish it from *standard propositional logic*. We redefine *true* to be an abbreviation for some fixed nonstandard tautology such as $p \hookrightarrow p$; again, we abbreviate $\neg$*true* by *false*.

Strong implication is indeed a new connective, that is, it cannot be defined using (nonstandard) $\neg$ and $\wedge$. For example, there are no valid formulas using only $\neg$ and $\wedge$, whereas by using $\hookrightarrow$, there are valid formulas: $\varphi \hookrightarrow \varphi$ is valid, as is $\varphi_1 \hookrightarrow (\varphi_1 \vee \varphi_2)$. Strong implication is indeed stronger than implication, in the sense that if φ_1 and φ_2 are standard formulas (formulas of $\mathcal{L}_n$), and if $\varphi_1 \hookrightarrow \varphi_2$ is valid with respect to nonstandard Kripke structures, then $\varphi_1 \Rightarrow \varphi_2$ is valid with respect to standard Kripke structures (Exercise 9.17). The converse, however, is false. For example, the formula $(p \wedge \neg p) \Rightarrow q$ is valid with respect to standard propositional logic, whereas the formula $(p \wedge \neg p) \hookrightarrow q$ is not valid with respect to nonstandard propositional logic (Exercise 9.17).

As we promised in the previous section, we can now express nonstandard logical implication using $\hookrightarrow$, just as we can express standard logical implication using $\Rightarrow$.

Proposition 9.3.3 *Let φ_1 and φ_2 be formulas in $\mathcal{L}_n^{\hookrightarrow}$. Then φ_1 logically implies φ_2 with respect to $\mathcal{NM}_n$ iff $\varphi_1 \hookrightarrow \varphi_2$ is valid with respect to $\mathcal{NM}_n$.*

Proof See Exercise 9.18. ∎

It turns out that it was the lack of ability to express logical implication within the language that prevented us from getting interesting formulas valid in $\mathcal{NM}_n$. Now

that we have introduced $\hookrightarrow$, there are many valid formulas. Thus, we can try again to characterize the properties of knowledge by getting a sound and complete axiomatization for $\mathcal{L}_n^{\hookrightarrow}$. Such an axiomatization can be obtained by modifying the axiom system K_n by (a) replacing propositional reasoning by nonstandard propositional reasoning, and (b) replacing standard implication ($\Rightarrow$) in the other axioms and rules by strong implication ($\hookrightarrow$). Thus, we obtain the axiom system $\mathrm{K}_n^{\hookrightarrow}$, which consists of all instances (for the language $\mathcal{L}_n^{\hookrightarrow}$) of the following axiom scheme and inference rules:

A2$^{\hookrightarrow}$. $(K_i\varphi \wedge K_i(\varphi \hookrightarrow \psi)) \hookrightarrow K_i\psi$ (Distribution Axiom)

NPR. All sound inference rules of nonstandard propositional logic

R2. From φ infer $K_i\varphi$ (Knowledge Generalization)

Note that the way we include nonstandard propositional reasoning in our axiomatization is quite different from the way we include standard propositional reasoning in the axiomatizations of Chapter 3. In the axiomatizations of Chapter 3, we took from the underlying propositional logic only the tautologies and modus ponens, while here we include all sound inference rules of nonstandard propositional logic. An example of a sound inference rule of nonstandard propositional logic is (nonstandard) modus ponens: From φ and $\varphi \hookrightarrow \psi$ infer ψ (this is a "nonstandard" rule, since it uses $\hookrightarrow$ instead of $\Rightarrow$). Other examples of sound inference rules of nonstandard propositional logic are given in Exercise 9.20. It can be shown that the axiomatization would not be complete had we included only the valid formulas of our nonstandard propositional logic as axioms, along with nonstandard modus ponens as the sole nonstandard propositional inference rule, rather than including all the sound inference rules of nonstandard propositional logic.

Theorem 9.3.4 *$\mathrm{K}_n^{\hookrightarrow}$ is a sound and complete axiomatization with respect to $\mathcal{NM}_n$ for formulas in the language $\mathcal{L}_n^{\hookrightarrow}$.*

Proof See Exercise 9.19. ∎

Theorem 9.3.4 shows that we can in some sense separate the properties of knowledge from the properties of the underlying propositional semantics. Even though logical omniscience is an essential feature of the possible-worlds approach (it is easy to see that in our enlarged language $\mathcal{L}_n^{\hookrightarrow}$, all of our types of logical omniscience hold, where we use $\hookrightarrow$ instead of $\Rightarrow$; see Exercise 9.21), it can be controlled to a certain degree by varying the propositional component of the semantics. Thus, one can say that in the nonstandard approach agents are "nonstandardly" logically omniscient.

In Chapter 3, we saw that under the standard semantics we can capture additional properties of knowledge by imposing suitable restrictions on the possibility relations $\mathcal{K}_i$. For example, restricting $\mathcal{K}_i$ to be reflexive captures the property that agents know only true facts ($K_i\varphi \Rightarrow \varphi$), and restricting $\mathcal{K}_i$ to be transitive captures the property of positive introspection ($K_i\varphi \Rightarrow K_iK_i\varphi$). The same phenomenon occurs under the nonstandard semantics. For example, restricting $\mathcal{K}_i$ to be reflexive captures the property that agents know only true facts, and analogously for positive introspection. Note, however, that to express these properties by valid formulas in the nonstandard approach, we need to use strong implication instead of standard implication. That is, the property that agents know only true facts is expressed by the axiom $K_i\varphi \hookrightarrow \varphi$, and the property of positive introspection is expressed by the axiom $K_i\varphi \hookrightarrow K_iK_i\varphi$. In the standard approach we captured negative introspection ($\neg K_i\varphi \Rightarrow K_i\neg K_i\varphi$) by requiring $\mathcal{K}_i$ to be Euclidean. It turns out that this is not sufficient to capture negative introspection in the nonstandard approach (as expressed by the axiom $\neg K_i\varphi \hookrightarrow K_i\neg K_i\varphi$), because of the nonstandard behavior of negation. To capture negative introspection we have to impose further restrictions on $\mathcal{K}_i$. For example, a sufficient additional restriction is that $(s,t) \in \mathcal{K}_i$ implies that $(s^*, t^*) \in \mathcal{K}_i$ (see Exercise 9.22).

Now that we have characterized the properties of knowledge in the nonstandard approach (in Theorem 9.3.4), we can consider the computational complexity of reasoning about knowledge, i.e., the complexity of determining validity in the nonstandard approach. Clearly, without $\hookrightarrow$, determining validity is very easy, since no formula is valid. As we saw in Proposition 9.3.3, however, $\hookrightarrow$ enables us to express logical implication. It turns out that once we introduce $\hookrightarrow$, reasoning about knowledge in the nonstandard approach is just as hard as reasoning about knowledge in the standard approach. The reason for this is the ability to "emulate" the standard semantics within the nonstandard semantics. Let φ be a formula of $\mathcal{L}_n^{\hookrightarrow}$. Then $(M,s) \models \varphi \hookrightarrow \mathit{false}$ iff $(M,s) \not\models \varphi$ for all nonstandard structures M and states s (see Exercise 9.23). Thus, standard negation can be expressed using strong implication.

Theorem 9.3.5 *The validity problem for $\mathcal{L}_n^{\hookrightarrow}$-formulas with respect to $\mathcal{NM}_n$ is PSPACE-complete.*

Proof The polynomial-space upper bound holds for much the same reasons that it does in the case of the logic K_n (see Section 3.5); a proof is beyond the scope of this book. For the lower bound, see Exercise 9.24. ∎

As we observed, logical omniscience still holds in the nonstandard approach. We also observed that the computational complexity of reasoning about knowledge does not improve. Nevertheless, our goal, which was to weaken the "logical" aspect in the logical-omniscience problem, is accomplished. For example, under the nonstandard semantics, agents do not know all *standard* tautologies. In the next section we provide an additional payoff of this approach: we show that in a certain important application we can obtain polynomial-time algorithms for reasoning about knowledge.

9.3.3 A Payoff: Querying Knowledge Bases

In Section 4.4.1, we introduced and discussed knowledge bases. An interesting application of our approach here concerns query evaluation in knowledge bases. Recall that the knowledge base (KB for short) is told facts about an external world, and is asked queries about the world. As we saw in Section 4.4.1, after the KB is told a sequence of standard propositional facts whose conjunction is κ, it then answers the propositional query ψ positively precisely when $\kappa \Rightarrow \psi$ is valid with respect to standard propositional logic, which is precisely when $K_{KB}\kappa \Rightarrow K_{KB}\psi$ is valid with respect to $\mathcal{M}_n^{rst}$. Thus, the formula κ completely characterizes the KB's knowledge in this case.

Our focus in this section is on the computational complexity of query evaluation in knowledge bases. We know that in the standard propositional approach, determining whether κ logically implies ψ or, equivalently, whether $K_i\kappa$ logically implies $K_i\psi$, is co-*NP*-complete. We show now that the nonstandard approach can yield a more efficient algorithm.

Consider the query-evaluation problem from the nonstandard perspective. Is the problem of determining the consequences of a knowledge base in the nonstandard approach any easier than in the standard approach? It turns out that, just as in the standard case, in the nonstandard approach, determining whether κ logically implies ψ is equivalent to determining whether $K_i\kappa$ logically implies $K_i\psi$ and both problems are still co-*NP*-complete (see Exercises 9.25 and 9.26). There is, however, an important special case where using the nonstandard semantics does make the problem easier.

Define a *literal* to be a primitive proposition p or its negation $\neg p$, and a *clause* to be a disjunction of literals. For example, a typical clause is $p \vee \neg q \vee r$. A formula that is a conjunction of clauses is said to be in *conjunctive normal form* (*CNF*). We assume here that κ (the formula that characterizes the KB's knowledge) is in CNF. This is not so unreasonable in practice, since once we have a knowledge base in CNF, it is easy to maintain it in CNF; before telling a new fact to the KB, we simply convert it to CNF. If φ^{CNF} is the result of converting φ to CNF, then the result of adding φ to κ is $\kappa \wedge \varphi^{CNF}$. Note that $\kappa \wedge \varphi^{CNF}$ is in CNF if κ is. Every

tandard propositional formula is equivalent to a formula in CNF (this is true even n our nonstandard semantics; see Exercise 9.27), so the transformation from φ to φ^{CNF} can always be carried out. Now, in general, this transformation can result in n exponential blowup; i.e., the length of φ^{CNF} can be exponential in the length of φ. We typically expect each fact φ that the KB is told to be small relative to the size of the KB, so even this exponential blowup is not unreasonable in practice. (On the other hand, it would not be reasonable to convert to CNF a whole knowledge base hat had not been maintained in CNF.) For similar reasons, we can safely assume that he query ψ has been transformed to CNF.

Let us now reconsider the query evaluation problem, where both the KB and he query are in CNF. Under the standard semantics, the problem is no easier than he general problem of logical implication in propositional logic, that is, co-*NP*-complete (Exercise 9.28). By contrast, the problem is feasible under the nonstandard semantics.

Theorem 9.3.6 *There is a polynomial-time algorithm for deciding whether κ logically implies ψ with respect to $\mathcal{NM}_n$ for CNF formulas κ and ψ.*

Proof We say that clause α_1 *includes* clause α_2 if every literal that is a disjunct of α_2 is also a disjunct of α_1. For example, the clause $p \vee \neg q \vee \neg r$ includes the clause $p \vee \neg q$. We can now characterize when κ logically implies ψ with respect to nonstandard propositional logic, for CNF formulas κ and ψ.

Let κ and ψ be propositional formulas in CNF. We claim that κ logically implies ψ with respect to nonstandard propositional logic iff every clause of ψ includes a clause of κ. (This claim is false in standard propositional logic. For example, let κ be $q \vee \neg q$, and let ψ be $p \vee \neg p$. Then $\kappa \Rightarrow \psi$ is valid, but the single clause $p \vee \neg p$ of ψ does not include the single clause of κ.)

The "if" direction, which is fairly straightforward, is left to the reader (Exercise 9.29). We now prove the other direction. Assume that some clause α of ψ includes no clause of κ. We need only show that there is a nonstandard structure $M = (S, \pi, \lambda_1, \ldots, \lambda_n, ^*)$ and state $s \in S$ such that $(M, s) \models \kappa$ but $(M, s) \not\models \psi$. Let S contain precisely two states s and t, and let $s^* = t$. Define $\pi(s)(p) =$ **false** iff p is a disjunct of α, and $\pi(t)(p) =$ **true** iff $\neg p$ is a disjunct of α, for each primitive proposition p. The λ_i's are arbitrary. We now show that $(M, s) \not\models \alpha'$, for each disjunct α' of α. Notice that α' is a literal. If α' is a primitive proposition p, then $\pi(s)(p) =$ **false**, so $(M, s) \not\models \alpha'$; if α' is $\neg p$, where p is a primitive proposition, then $\pi(t)(p) =$ **true**, so $(M, t) \models p$, so again $(M, s) \not\models \alpha'$. Hence, $(M, s) \not\models \alpha$. Since α is one of the conjuncts of ψ, it follows that $(M, s) \not\models \psi$. We next show that

$(M, s) \models \beta$ if α does not include β. For if α does not include β, then there is some literal β' that is a disjunct of β but not of α. It is easy to see that $(M, s) \models \beta'$, and hence that $(M, s) \models \beta$. It follows that $(M, s) \models \kappa$, since by assumption, α does not include any of the conjuncts of κ.

It is clear that this characterization of nonstandard implication gives us a polynomial-time decision procedure for deciding whether one CNF formula implies another in the nonstandard approach. ∎

Theorem 9.3.6 gives us a real payoff of the nonstandard approach. It shows that even though the nonstandard approach does not improve the complexity of reasoning about knowledge in general, there are practical applications for which reasoning about knowledge can be feasible. As the following proposition shows, this also has implications for reasoning in standard logic.

Proposition 9.3.7 *Let κ and ψ be propositional formulas in CNF. If κ logically implies ψ with respect to $\mathcal{NM}_n$, then κ logically implies ψ with respect to standard propositional logic.*

Proof See Exercise 9.30. ∎

Theorem 9.3.6 and Proposition 9.3.7 yield an efficient algorithm for the evaluation of a CNF query ψ with respect to a CNF knowledge base κ: answer "Yes" if κ logically implies ψ with respect to $\mathcal{NM}_n$. By Theorem 9.3.6, logical implication of CNF formulas with respect to $\mathcal{NM}_n$ can be checked in polynomial time. Proposition 9.3.7 implies that any positive answer we obtain from testing logical implication between CNF formulas in nonstandard semantics will provide us with a correct positive answer for standard semantics as well. This means that even if we are ultimately interested only in conclusions that are derivable from standard reasoning, we can safely use the positive conclusions we obtain using nonstandard reasoning. Thus, the nonstandard approach yields a feasible query-answering algorithm for knowledge bases. Notice that the algorithm need not be correct with respect to negative answers. It is possible that κ does not logically imply ψ with respect to $\mathcal{NM}_n$, even though κ logically implies ψ with respect to standard propositional logic (see Exercise 9.30).

9.3.4 Discussion

The goal of our approach in this section was to gain some control over logical omniscience rather than to eliminate it. To this end, we tried to decouple the knowledge part of the semantics from its propositional part by keeping the definition of knowledge as truth in all possible worlds but changing the underlying notion of truth (the

structure $M = (S, \pi, \mathcal{K}_1, \ldots, \mathcal{K}_n, ^*)$, we can identify it with the impossible-worlds structure $M' = (S, W, \sigma, \mathcal{K}_1, \ldots, \mathcal{K}_n)$, where W is the set of standard states, i.e., the states s such that $s^* = s$, and, for all states $s \in S$, we have that $\sigma(s)(\varphi) = \mathbf{true}$ iff $(M, s) \models \varphi$. We can therefore view a nonstandard structure M as implicitly defining the impossible-worlds structure M' obtained by this translation. We shall abuse language slightly and say that we view M as an impossible-worlds structure. When M is viewed as a nonstandard structure, the distinction between standard and nonstandard states does not play any role. In contrast, when M is viewed as an impossible-worlds structure, the standard states have a special status. Intuitively, although an agent (who is not a perfect reasoner) might consider nonstandard states possible (where, for example, $p \wedge \neg p$ or $K_i p \wedge \neg K_i p$ holds), we do not consider such states possible; surely in the real world a formula is either true or false, but not both.

Nonstandard structures can be viewed both from the perspective of the nonstandard-logic approach and from the perspective of the impossible-worlds approach. When we view a nonstandard structure as an impossible-worlds structure, we consider nonstandard states to be impossible states, and thus consider a formula φ to be valid if it is true in all of the possible states, that is, in all of the standard states. Formally, define a formula of $\mathcal{L}_n$ to be *standard-state valid* if it is true at every standard state of every nonstandard structure. The definition for *standard-state logical implication* is analogous.

We demonstrate the difference between logical implication and standard-state logical implication by reconsidering the knowledge base example discussed in Section 9.3.3, where the knowledge base is characterized by the formula κ and the query is the formula φ. We saw in Section 9.3.3 that in the nonstandard approach, φ is a consequence of κ precisely when knowledge of φ is a consequence of knowledge of κ. This is not the case in the impossible-worlds approach; it is possible to find φ_1 and φ_2 in $\mathcal{L}_n$ such that φ_1 standard-state logically implies φ_2, but $K_i\varphi_1$ does not standard-state logically imply $K_i\varphi_2$ (Exercise 9.32). The reason for this difference is that φ_1's standard-state logically implying φ_2 deals with logical implication in standard states, whereas $K_i\varphi_1$'s standard-state logically implying $K_i\varphi_2$ deals with logical implication in states agents consider possible, which can include nonstandard states. Interestingly, logical implication of knowledge formulas coincides in the nonstandard approach and the impossible-worlds approach; that is, $K_i\varphi_1$ standard-state logically implies $K_i\varphi_2$ iff $K_i\varphi_1$ logically implies $K_i\varphi_2$ with respect to $\mathcal{NM}_n$ (Exercise 9.33).

The reader may recall that under the nonstandard semantics, $\Rightarrow$ behaves in a nonstandard way. In particular, $\Rightarrow$ does not capture the notion of logical implication. In fact, that was part of the motivation for the introduction of strong implication. In standard states, however, $\Rightarrow$ and $\hookrightarrow$ coincide; that is, $\varphi_1 \Rightarrow \varphi_2$ holds at a standard state precisely if $\varphi_1 \hookrightarrow \varphi_2$ holds. It follows that even though $\Rightarrow$ does not capture logical implication, it does capture standard-state logical implication. The following analogue to Proposition 9.3.3 is immediate.

Proposition 9.4.1 *Let φ_1 and φ_2 be formulas in $\mathcal{L}_n$. Then φ_1 standard-state logically implies φ_2 iff $\varphi_1 \Rightarrow \varphi_2$ is standard-state valid.*

The main feature of the impossible-worlds approach is the fact that knowledge is evaluated with respect to all states, while logical implication is evaluated only with respect to standard states. As a result, we avoid logical omniscience. For example, an agent does not necessarily know valid formulas of standard propositional logic. Although the classical tautology $\varphi \vee \neg\varphi$ is standard-state valid, $K_i(\varphi \vee \neg\varphi)$ may not hold at a standard state s, since agent i might consider an incomplete state possible. (Recall that in an incomplete state of a nonstandard structure both φ and $\neg\varphi$ may fail to hold.) On the other hand, as we now show, incompleteness is all that prevents an agent from knowing valid formulas. In particular, we show that if an agent knows that the state is complete, then he does know all tautologies.

What does it mean for an agent to know that a state is complete? Let φ be a propositional formula that contains precisely the primitive propositions $p_1, \ldots, p_k$. Define *complete*(φ) to be the formula

$$(p_1 \vee \neg p_1) \wedge \ldots \wedge (p_k \vee \neg p_k).$$

Thus, *complete*(φ) is true at a state s precisely if s is complete as far as all the primitive propositions in φ are concerned. In particular, if *complete*(φ) is true at s, then s is complete with respect to φ (see Exercise 9.34). Thus, if an agent knows *complete*(φ), then he knows that he is in a state that is complete with respect to φ.

The following result makes precise our earlier claim that incompleteness is all that prevents an agent from knowing tautologies.

Theorem 9.4.2 *Let φ be a tautology of standard propositional logic. Then K_i(complete(φ)) $\Rightarrow K_i\varphi$ is standard-state valid.*

Proof By Exercise 9.35 (see also Exercise 9.36), *complete*(φ) logically implies φ with respect to $\mathcal{NM}_n$. From Exercise 9.25 it follows that K_i(*complete*(φ)) logically implies $K_i\varphi$ with respect to $\mathcal{NM}_n$. In particular, K_i(*complete*(φ)) standard-state logically implies $K_i\varphi$. It follows by Proposition 9.4.1 that K_i(*complete*(φ)) $\Rightarrow K_i\varphi$ is standard-state valid. ∎

:opositional semantics). With this approach, we still have closure under logical nplication. Since knowledge is still defined as truth in all possible worlds, it is still ıe case that if φ logically implies ψ, then an agent that knows φ will also know ψ. ievertheless, as a result of the change in the definition of truth, the notion of logical nplication has changed. It may not be so unreasonable for an agent's knowledge) be closed under logical implication if we have a weaker notion of logical impli- ation. As a particular example of this approach, we considered a nonstandard logic ı which the truth values of φ and $\neg\varphi$ are independent, and logical implication is aptured using $\hookrightarrow$ rather than $\Rightarrow$. While this particular nonstandard approach does ot improve the complexity of reasoning about knowledge in general, we gave one pplication where it does yield a significant improvement.

We should stress that we considered only one particular nonstandard logic in this ection. Many nonstandard propositional logics have been studied. (See the notes at he end of the chapter for references.) It would be interesting to explore how these ther nonstandard propositional logics could be combined with epistemic operators, nd what the consequences of doing so would be.

9.4 Impossible Worlds

ogical omniscience arises from considering knowledge as truth in all possible vorlds. In the previous section, we modified logical omniscience by changing the ıotion of truth. In this section, we modify logical omniscience by changing the ıotion of possible world. The idea is to augment the possible worlds by *impossible worlds*, where the customary rules of logic do not hold. For example, we may have ɔoth φ and ψ holding in an impossible world without having $\varphi \wedge \psi$ hold in that world. Even though these worlds are logically impossible, the agents nevertheless may con- sider them possible. Unlike our approach in the previous section, where nonstandard worlds had the same status as standard worlds, under the current approach the im- possible worlds are only a figment of the agents' imagination; they serve only as epistemic alternatives. Thus, logical implication and validity are determined solely with respect to the standard worlds.

Formally, an *impossible-worlds structure* M is a tuple $(S, W, \sigma, \mathcal{K}_1, \ldots, \mathcal{K}_n)$, where $(S, \mathcal{K}_1, \ldots, \mathcal{K}_n)$ is a Kripke frame, $W \subseteq S$ is the set of *possible* states or worlds, and σ is a syntactic assignment (recall that syntactic assignments assign truth values to all formulas in all states). We require that σ behaves standardly on possible states, that is, if $s \in W$, then

$\sigma(s)(\varphi \wedge \psi) = \textbf{true}$ iff $\sigma(s)(\varphi) = \textbf{true}$ and $\sigma(s)(\psi) = \textbf{true}$,

$\sigma(s)(\neg\varphi) = \textbf{true}$ iff $\sigma(s)(\varphi) = \textbf{false}$, and

$\sigma(s)(K_i\varphi) = \textbf{true}$ iff $\sigma(t)(\varphi) = \textbf{true}$ for all t such that $(s, t) \in \mathcal{K}_i$.

Note that σ can behave in an arbitrary way on the impossible states, i.e, the states in $S - W$. We use σ to define satisfaction in the obvious way: $(M, s) \models \varphi$ precisely when $\sigma(s)(\varphi) = \textbf{true}$.

As mentioned earlier, logical implication and validity are determined only with respect to possible states, that is, the states in W. Formally, a set Ψ of formulas *logically implies* the formula φ *with respect to impossible-worlds structures* if for each impossible-worlds structure $M = (S, W, \sigma, \mathcal{K}_1, \ldots, \mathcal{K}_n)$ and possible state $s \in W$ we have that whenever $(M, s) \models \psi$ for all $\psi \in \Psi$, then $(M, s) \models \varphi$. Similarly, φ is *valid with respect to impossible-worlds structures* if for each impossible-worlds structure $M = (S, W, \sigma, \mathcal{K}_1, \ldots, \mathcal{K}_n)$ and possible state $s \in W$ we have that $(M, s) \models \varphi$.

Since agents consider the impossible states when determining their knowledge, but impossible states are not considered when determining logical implication, logical omniscience need not hold. Consider, for example, full logical omniscience. Suppose that an agent knows all formulas in Ψ, and Ψ logically implies φ. Since the agent knows all formulas in Ψ, all formulas in Ψ must hold in all the states that the agent considers epistemically possible. But in an impossible state, φ may fail even though Ψ holds. The reason for this is that logical implication is determined by us, rational logicians, for whom impossible states are simply impossible and are therefore not taken into account. Thus, the agent need not know φ, since φ may fail to hold in some impossible state that the agent considers possible.

The impossible-worlds approach is very general; it can capture different properties of knowledge by imposing certain conditions on the behavior of syntactic assignment σ in the impossible states. For example, to capture closure under conjunction we have to demand that in an impossible state if both φ and ψ are true, then $\varphi \wedge \psi$ is also true. (See also Exercise 9.31.)

We now consider one instance of the impossible-worlds approach, which will enable us to contrast the impossible-worlds approach with the nonstandard-logic approach of Section 9.3. Essentially, the idea is to view nonstandard structures as impossible-worlds structures, where the nonstandard states are the impossible worlds. Recall that nonstandard structures are Kripke structures with a * function. This function associates with a state s an adjunct state s^*. If $s = s^*$, then s is a standard state and therefore a possible world. If $s \neq s^*$, then s and s^* are nonstandard states and therefore considered to be impossible worlds. More formally, given a nonstandard

In addition to the failure of knowledge of valid formulas, another form of logical omniscience that fails under the impossible-worlds approach is closure under logical implication: the formula $K_i\varphi \wedge K_i(\varphi \Rightarrow \psi) \Rightarrow K_i\psi$ is not standard-state valid (Exercise 9.37). This lack of closure results from considering incoherent states impossible: indeed, $K_i\varphi \wedge K_i(\varphi \Rightarrow \psi) \Rightarrow K_i(\psi \vee (\varphi \wedge \neg\varphi))$ is standard-state valid (Exercise 9.37). That is, if an agent knows that φ holds and also knows that $\varphi \Rightarrow \psi$ holds, then she knows that either ψ holds or the state is incoherent. This observation generalizes. As we now show, as long as the agent knows that the state is coherent, then her knowledge is closed under logical implication.

Recall that *true* is an abbreviation for some fixed nonstandard tautology such as $p \hookrightarrow p$, and *false* is an abbreviation for $\neg$*true*. We use the fact that the formula $\varphi \hookrightarrow$ *false* asserts the falsehood of φ (see Exercise 9.23). Let φ be a formula that contains precisely the primitive propositions $p_1, \ldots, p_k$. Define *coherent*(φ) to be the formula

$$\big((p_1 \wedge \neg p_1) \hookrightarrow \mathit{false}\big) \wedge \ldots \wedge \big((p_k \wedge \neg p_k) \hookrightarrow \mathit{false}\big).$$

Thus, *coherent*(φ) is true at a state s precisely if s is coherent as far as the primitive propositions in φ are concerned. In particular, if *coherent*(φ) holds at s, then s is coherent with respect to φ. (Note that *coherent*(φ) is not definable in $\mathcal{L}_n$ but only in $\mathcal{L}_n^{\hookrightarrow}$; see Exercise 9.38.) Knowledge of coherence implies that knowledge is closed under material implication.

Theorem 9.4.3 *Let φ and ψ be standard propositional formulas. Then $(K_i(\mathit{coherent}(\varphi)) \wedge K_i\varphi \wedge K_i(\varphi \Rightarrow \psi)) \Rightarrow K_i\psi$ is standard-state valid.*

Proof Denote $K_i(\mathit{coherent}(\varphi)) \wedge K_i\varphi \wedge K_i(\varphi \Rightarrow \psi)$ by τ. By Proposition 9.4.1, it is sufficient to show that τ standard-state logically implies $K_i\psi$. We shall show the stronger fact that τ logically implies $K_i\psi$. Let $M = (S, \pi, \mathcal{K}_1, \ldots, \mathcal{K}_n, {}^*)$ be a nonstandard structure, and s a state of M. Assume that τ is true at s, and that $(s, t) \in \mathcal{K}_i$. So *coherent*(φ) is true at t. By a straightforward induction on the structure of formulas, we can show that for every propositional formula γ all of whose primitive propositions are contained in φ, it is not the case that both γ and $\neg\gamma$ are true at t. Now φ and $\varphi \Rightarrow \psi$ are both true at t, since $K_i\varphi$ and $K_i(\varphi \Rightarrow \psi)$ are true at s. Since φ is true at t, it follows from what we just showed that $\neg\varphi$ is not true at t. Since $\varphi \Rightarrow \psi$ is an abbreviation for $\neg\varphi \vee \psi$, it follows that ψ is true at t. Hence, $K_i\psi$ is true at s. ∎

Theorems 9.4.2 and 9.4.3 explain why agents are not logically omniscient: when we view nonstandard structures as impossible-worlds structures, "logically" is defined with respect to standard states, but the agents may consider nonstandard states possible. If an agent considers only standard states possible, so that both $K_i(\mathit{complete}(\varphi))$ and $K_i(\mathit{coherent}(\varphi))$ hold, then by Theorems 9.4.2 and 9.4.3, this agent is logically omniscient (more accurately, he knows every tautology of standard propositional logic and his knowledge is closed under material implication).

9.5 Awareness

In Section 9.2, we described syntactic and semantic approaches to dealing with omniscience. In Section 9.4, we described what can be viewed as a mixed approach, i.e., an approach that has both semantic and syntactic components: in impossible-worlds structures, truth is defined semantically in the possible states and syntactically in the impossible states. We now describe another approach that has both semantic and syntactic components.

The underlying idea is that it is necessary to be *aware* of a concept before one can have beliefs about it. One cannot know something of which one is unaware. Indeed, how can someone say that he knows or doesn't know about p if p is a concept of which he is completely unaware? One can imagine the puzzled response of someone not up on the latest computer jargon when asked if he knows that the price of SIMMs is going down! (For the benefit of the reader who is not fluent in the computer-speak of the early 1990's, a SIMM is a Single In-line Memory Module, a basic component in current-day computer memories.) In fact, even a sentence such as "He doesn't even know that he doesn't know p!" is often best understood as saying "He's not even *aware* that he doesn't know p."

In this section we augment the possible-worlds approach with a syntactic notion of *awareness*. This will be reflected in the language by a new modal operator A_i for each agent i. The intended interpretation of $A_i\varphi$ is "i is aware of φ." We do not wish to attach any fixed cognitive meaning to the notion of awareness; $A_i\varphi$ may mean "i is familiar with all the propositions mentioned in φ," "i is able to figure out the truth of φ," or perhaps "i is able to compute the truth of φ within time T." (We return to a computational notion of knowledge later in this chapter and also in Chapter 10.) The power of the approach comes from the flexibility of the notion of awareness.

To represent the knowledge of agent i, we allow two modal operators K_i and X_i, standing for *implicit knowledge* and *explicit knowledge* of agent i, respectively. Implicit knowledge is the notion we have been considering up to now: truth in all worlds that the agent considers possible. On the other hand, an agent explicitly knows a formula φ if he is aware of φ and implicitly knows φ. Intuitively, an agent's implicit

knowledge includes all the logical consequences of his explicit knowledge. We denote by $\mathcal{L}_n^A(\Phi)$, or $\mathcal{L}_n^A$ for short, the set of formulas obtained by enlarging $\mathcal{L}_n(\Phi)$ to include the new modal operators A_i and X_i.

An *awareness structure* is a tuple $M = (S, \pi, \mathcal{K}_1, \ldots, \mathcal{K}_n, \mathcal{A}_1, \ldots, \mathcal{A}_n)$, where the tuple $(S, \pi, \mathcal{K}_1, \ldots, \mathcal{K}_n)$ is a Kripke structure and $\mathcal{A}_i$ is a function associating a set of formulas with each state, for $i = 1, \ldots, n$. Intuitively, $\mathcal{A}_i(s)$ is the set of formulas that agent i is aware of at state s. The awareness functions $\mathcal{A}_i$ form the syntactic component of the semantics. The formulas in $\mathcal{A}_i(s)$ are those that the agent is "aware of," not necessarily those he knows. The set of formulas that the agent is aware of can be arbitrary. It is possible for both φ and $\neg\varphi$ to be in $\mathcal{A}_i(s)$, for only one of φ and $\neg\varphi$ to be in $\mathcal{A}_i(s)$, or for neither φ nor $\neg\varphi$ to be in $\mathcal{A}_i(s)$. It is also possible, for example, that $\varphi \vee \psi$ is in $\mathcal{A}_i(s)$ but $\psi \vee \varphi$ is not in $\mathcal{A}_i(s)$.

The semantics for primitive propositions, conjunctions, negations, and for formulas $K_i\varphi$ is just as for standard Kripke structures. We only need to add new clauses for formulas of the form $A_i\varphi$ and $X_i\varphi$:

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

$(M, s) \models X_i\varphi$ iff $(M, s) \models A_i\varphi$ and $(M, t) \models K_i\varphi$

The first clause states that agent i is aware of φ at state s exactly if φ is in $\mathcal{A}_i(s)$. The second clause states that agent i explicitly knows φ iff (1) agent i is aware of φ, and (2) agent i implicitly knows φ (i.e., φ is true in all the worlds he considers possible). We see immediately that $X_i\varphi \Leftrightarrow A_i\varphi \wedge K_i\varphi$ is valid. You cannot have explicit knowledge about formulas of which you are not aware! If we assume that agents are aware of all formulas, then explicit knowledge reduces to implicit knowledge.

By definition, the implicit-knowledge operator K_i behaves just as it does in a Kripke structure. Thus, as in Chapter 3, implicit knowledge is closed under material implication (that is, $(K_i\varphi \wedge K_i(\varphi \Rightarrow \psi)) \Rightarrow K_i\psi$ is valid) and $K_i\varphi$ is valid for every valid formula φ. The explicit-knowledge operator X_i, however, may behave differently. Agents do not explicitly know all valid formulas; for example, $\neg X_i(p \vee \neg p)$ is satisfiable, because the agent might not be aware of the formula $p \vee \neg p$. Also, an agent's explicit knowledge is not necessarily closed under material implication; $X_i p \wedge X_i(p \Rightarrow q) \wedge \neg X_i q$ is satisfiable, because i might not be aware of q. Since awareness is essentially a syntactic operator, this approach shares some of the features of the syntactic approach. For example, order of presentation matters; there is no reason to suppose that the formula $X_i(\varphi \vee \psi)$ is equivalent to $X_i(\psi \vee \varphi)$, since $A_i(\varphi \vee \psi)$ might hold without $A_i(\psi \vee \varphi)$ holding. A computer program that can determine in time T whether $\varphi \vee \psi$ follows from some initial premises might not

be able to determine in time T whether $\psi \vee \varphi$ follows from those premises. (The program might work on, say, the left disjunct first, and be able to determine quickly that φ is true, but get stuck working on ψ.) And people do *not* necessarily identify formulas such as $\varphi \vee \psi$ and $\psi \vee \varphi$. The reader can validate the idea that the order matters by computing the product $1 \times 2 \times 3 \times 4 \times 5 \times 6 \times 7 \times 8 \times 9 \times 0$.

Up to now we have placed no restrictions on the set of formulas that an agent may be aware of. Once we have a concrete interpretation in mind, we may well want to add some restrictions to the awareness function to capture certain types of "awareness." The clean separation in our framework between knowledge (captured by the binary relations $\mathcal{K}_i$) and awareness (captured by the syntactic functions $\mathcal{A}_i$) makes this easy to do. Some typical restrictions we may want to add to $\mathcal{A}_i$ include the following:

- Awareness could be closed under subformulas; i.e., if $\varphi \in \mathcal{A}_i(s)$ and ψ is a subformula of φ, then $\psi \in \mathcal{A}_i(s)$. Note that this makes sense if we are reasoning about a computer program that will never compute the truth of a formula unless it has computed the truth of all its subformulas. But it is also easy to imagine a program that knows that $\varphi \vee \neg\varphi$ is true without needing to compute the truth of φ. Perhaps a more reasonable restriction is simply to require that if $\varphi \wedge \psi \in \mathcal{A}_i(s)$ then both $\varphi, \psi \in \mathcal{A}_i(s)$ (see Exercise 9.39).

- Agent i might be aware of only a certain subset of the primitive propositions, say Ψ. In this case we could take $\mathcal{A}_i(s)$ to consist of exactly those formulas that mention only primitive propositions that appear in Ψ.

- A self-reflective agent will be aware of what he is aware of. Semantically, this means that if $\varphi \in \mathcal{A}_i(s)$, then $A_i\varphi \in \mathcal{A}_i(s)$. This corresponds to the axiom $A_i\varphi \Rightarrow A_iA_i\varphi$.

- Similarly, an agent might know of which formulas he is or is not aware. Semantically, this means that if $(s, t) \in \mathcal{K}_i$, then $\mathcal{A}_i(s) = \mathcal{A}_i(t)$. This corresponds to the axioms $A_i\varphi \Rightarrow K_iA_i\varphi$ and $\neg A_i\varphi \Rightarrow K_i\neg A_i\varphi$. This restriction holds when the set of formulas that an agent is aware of is a function of his local state. It also holds when awareness is generated by a subset of primitive propositions, as discussed previously.

We now turn to examining the properties of knowledge in this logic. It is easy to see that the axiom system K_n is sound, since the semantics of K_i has not changed. Indeed, we can obtain a sound and complete axiomatization simply by adding the

axiom $X_i\varphi \Leftrightarrow (A_i\varphi \wedge K_i\varphi)$ to K_n (Exercise 9.40). These axioms, however, do not give us much insight into the properties of explicit knowledge.

In fact, despite the syntactic nature of the awareness operator, explicit knowledge retains many of the same properties as implicit knowledge, once we relativize to awareness. For example, corresponding to the Distribution Axiom

$$\big(K_i\varphi \wedge K_i(\varphi \Rightarrow \psi)\big) \Rightarrow K_i\psi$$

we have

$$\big(X_i\varphi \wedge X_i(\varphi \Rightarrow \psi) \wedge A_i\psi\big) \Rightarrow X_i\psi$$

(see Exercise 9.41). Thus, if you explicitly know φ and $\varphi \Rightarrow \psi$, then you will explicitly know ψ *provided* you are aware of ψ. Similarly, corresponding to the Knowledge Generalization Rule, we have

$$\text{from } \varphi \text{ infer } A_i\varphi \Rightarrow X_i\varphi$$

(see Exercise 9.41). That is, you explicitly know a valid formula if you are aware of it. Note the similarity between this rule and Theorem 9.4.2, which says that, in the impossible-worlds approach, knowledge of a tautology φ follows from knowledge of $complete(\varphi)$. In that setting, we can think of $K_i(p \vee \neg p)$ as saying that agent i is aware of p. If we take the view of awareness as being generated by a subset of primitive propositions, then $K_i(complete(\varphi))$ can be thought of saying that agent i is aware of φ. Thus, Theorem 9.4.2 can be viewed as saying that an agent knows a tautology if he is aware of it. In both cases, an agent must be aware of the relevant formula in order to know it explicitly.

As we saw earlier, we can capture certain properties of knowledge by imposing the appropriate conditions on the $\mathcal{K}_i$ relations. For example, if we assume that $\mathcal{K}_i$ is reflexive, then, as before, we obtain the axiom $X_i\varphi \Rightarrow \varphi$, since $X_i\varphi \Rightarrow K_i\varphi$ is valid, and reflexivity of $\mathcal{K}_i$ entails that $K_i\varphi \Rightarrow \varphi$ is valid. It may be tempting to think that if $\mathcal{K}_i$ is an equivalence relation, then we obtain the introspection axioms $X_i\varphi \Rightarrow X_iX_i\varphi$ and $\neg X_i\varphi \Rightarrow X_i\neg X_i\varphi$. It is easy to verify, however, that this is not the case. In fact, even the obvious modification of the introspection axioms, where an agent must be aware of a formula before she explicitly knows it, fails to hold:

$$(X_i\varphi \wedge A_iX_i\varphi) \Rightarrow X_iX_i\varphi$$
$$\big(\neg X_i\varphi \wedge A_i(\neg X_i\varphi)\big) \Rightarrow X_i\neg X_i\varphi$$

(see Exercise 9.42). The reason for this failure is the independence of the the awareness operator and the possibility relation; an agent may be aware of different formulas in states that she considers to be equivalent. It may be reasonable to assume that if an agent cannot distinguish between two states, then she is aware of the same formulas in both states, i.e., $(s, t) \in \mathcal{K}_i$ implies $\mathcal{A}_i(s) = \mathcal{A}_i(t)$. Intuitively, this means that the agent knows of which formulas she is aware. If this assumption holds and if $\mathcal{K}_i$ is an equivalence relation, then the modified introspection properties mentioned earlier hold (see Exercise 9.42). The phenomenon described by these axioms is similar to the phenomenon of the previous section, where knowledge of completeness or coherence was required. The first of the two axioms suggests how, as in the quotation from de Chardin at the beginning of the chapter, an animal may know, but not know that it knows: it might not be aware of its knowledge. The second axiom suggests how someone can fail to be conscious of his ignorance. By contrast, the Spinoza quotation suggests that people are aware of their knowledge. Of course, we can construct axioms analogous to these even if we do not assume that an agent knows what formulas she is aware of, although they are not quite as elegant (see Exercise 9.42).

As we observed earlier, if we are reasoning about a computer program that will never compute the truth of a formula unless it has computed the truth of all its subformulas, then awareness is closed under subformulas: if $\varphi \in \mathcal{A}_i(s)$ and ψ is a subformula of φ, then $\psi \in \mathcal{A}_i(s)$. Taking awareness to be closed under subformulas has some interesting consequences. First note that this property can be captured axiomatically by the following axioms:

$$A_i(\neg\varphi) \Rightarrow A_i\varphi$$
$$A_i(\varphi \wedge \psi) \Rightarrow (A_i\varphi \wedge A_i\psi)$$
$$A_i(X_j\varphi) \Rightarrow A_i\varphi$$
$$A_i(K_j\varphi) \Rightarrow A_i\varphi$$
$$A_i(A_j\varphi) \Rightarrow A_i\varphi.$$

(By changing $\Rightarrow$ to $\Leftrightarrow$ in these axioms, we can capture a notion of awareness generated by a set of primitive propositions; see Exercise 9.43.)

Although agents still do not explicitly know all valid formulas if awareness is closed under subformulas, an agent's knowledge is then closed under material implication; i.e., $X_i\varphi \wedge X_i(\varphi \Rightarrow \psi) \Rightarrow X_i\psi$ is then valid (see Exercise 9.44). Thus, the seemingly innocuous assumption that awareness is closed under subformulas has a rather powerful impact on the properties of explicit knowledge. Certainly this assumption is inappropriate for resource-bounded notions of awareness, where awareness of φ corresponds to being able to compute the truth of φ. As we remarked, it may be easy to see that $\varphi \vee \neg\varphi$ is a tautology without having to compute whether

either φ or $\neg\varphi$ follows from some information. Nevertheless, this observation shows that there are some natural interpretations of awareness and explicit knowledge (for example, an interpretation of awareness that is closed under subformulas and an interpretation of explicit knowledge that is not closed under material implication) that cannot be simultaneously captured in this framework.

What about the computational complexity of the validity problem? Clearly the addition of A_i and X_i cannot decrease the complexity. It turns out that this addition does not increase the complexity; the validity problem is still *PSPACE*-complete.

In summary, the awareness approach is very flexible and general. In a natural and appealing way, it can be used to demonstrate why various types of logical omniscience fail, and to give assumptions on what the agent must be aware of for these various types of logical omniscience to hold. It gains this flexibility through the use of a syntactic awareness operator. While at first this may seem to put us right back into the syntactic approach of Section 9.2, by isolating the syntactic component, we have more structure to study, while maintaining our intuition about knowledge being truth in all possible worlds.

This observation suggests that we focus on natural notions of awareness. We considered some notions already in this section. In Chapter 10, we describe a computational model of knowledge, which can be viewed as using a computational notion of awareness.

It is interesting to relate the awareness approach to the impossible-worlds approach. Both mix syntax and semantics, but in a very different way. In the impossible-worlds approach, knowledge depends on impossible worlds, where truth is defined syntactically. In the awareness approach, knowledge depends on awareness, which is defined syntactically. It turns out that in some sense the two approaches are equivalent: every impossible-worlds structure can be represented by an awareness structure and vice versa (see Exercise 9.45); thus, both approaches can be used to model the same situations.

9.6 Local Reasoning

An important difference between an idealized model of knowledge (such as a Kripke structure) and the knowledge of people in the real world is that in the real world people have inconsistent knowledge. That is, they may believe both φ and $\neg\varphi$ for some formula φ; this may happen when an agent believes both φ and $\neg\psi$ without realizing that φ and ψ are logically equivalent. We already have tools to model inconsistent knowledge: it is possible for an agent to believe both φ and $\neg\varphi$ in

a standard Kripke structure. Standard Kripke structures, however, provide a poor model for inconsistent knowledge. It is easy to see that the only way that an agent i in a standard Kripke structure $(S, \pi, \mathcal{K}_1, \ldots, \mathcal{K}_n)$ can have inconsistent knowledge in a state s is for $\mathcal{K}_i(s) = \{t \mid (s,t) \in \mathcal{K}_i\}$ to be empty, which implies that in state s, agent i knows *every* formula. Some of the approaches described earlier in this chapter can also be used to model inconsistent knowledge. For example, in an awareness structure $(S, \pi, \mathcal{K}_1, \ldots, \mathcal{K}_n, \mathcal{A}_1, \ldots, \mathcal{A}_n)$, it is possible for agent i to have contradictory explicit knowledge in state s without having explicit knowledge of every formula: this can be modeled by again letting $\mathcal{K}_i(s)$ be empty and taking $\mathcal{A}_i(s)$ to consist precisely of those formulas of which agent i has explicit knowledge.

In this section we describe another approach, in which inconsistent knowledge arises in a very natural way. Since this approach seems to be especially interesting when the agents are people, we describe the results in this section in these terms.

One reason that people have inconsistent knowledge is that knowledge tends to depend on an agent's frame of mind. We can view an agent as a society of minds, each with its own knowledge. The members of the society may have contradictory knowledge (or, perhaps better, beliefs). For example, in one frame of mind, a politician might believe in the importance of a balanced budget. In another frame of mind, however, he might believe it is necessary to greatly increase spending. This phenomenon seems to occur even in science. For example, the two great theories physicists reason with are the theory of quantum phenomena and the general theory of relativity. Some physicists work with both theories, even though they believe that the two theories might well be incompatible!

In Kripke structures, agents can be said to have a single frame of mind. We viewed $\mathcal{K}_i(s)$ as the set of states that agent i thinks possible in state s. In our next approach, there is not necessarily one set of states that an agent thinks possible, but rather a number of sets, each one corresponding to the knowledge of a different member of the society of minds. We can view each of these sets as representing the worlds the agent thinks possible in a given frame of mind, when he is focusing on a certain set of issues. This models agents with many frames of mind.

More formally, a *local-reasoning structure* is a tuple $M = (S, \pi, \mathcal{C}_1, \ldots, \mathcal{C}_n)$ where S is a set of states, $\pi(s)$ is a truth assignment to the primitive propositions for each state $s \in S$, and $\mathcal{C}_i(s)$ is a nonempty set of subsets of S. Intuitively, if $\mathcal{C}_i(s) = \{T_1, \ldots, T_k\}$, then in state s agent i sometimes (depending perhaps on his frame of mind or the issues on which he is focusing) considers the set of possible states to be precisely T_1, sometimes he considers the set of possible states to be precisely T_2, etc. Or, taking a more schizophrenic point of view, we could view each

f these sets as representing precisely the worlds that some member of the society in gent i's mind thinks possible.

We now interpret $K_i\varphi$ as "agent i knows φ in *some* frame of mind"; i.e., some nember of the society of minds making up agent i at s knows φ. Note that although ve are using the same symbol K_i, this notion is quite different from the notions of nowledge discussed earlier in this chapter. This form of knowledge could be called *ocal knowledge*, since it is local to one of the members of the society. The semantics or primitive propositions, conjunctions, and negations is just as for standard Kripke tructures. The semantics for knowledge, however, has changed:

$(M, s) \models K_i\varphi$ iff there is *some* $T \in \mathcal{C}_i(s)$ such that $(M, t) \models \varphi$ for all $t \in T$.

There is a stronger notion of knowledge where we would say that i knows φ if φ s known in *all* of i's frames of mind. Under the society of minds viewpoint, our notion of K_i corresponds to "some member (of agent i's society) knows," whereas his stronger notion corresponds to "all members know." We can get an even stronger notion by having i know φ only if φ is common knowledge among i's frames of mind. Going in the other direction, towards weaker notions of knowledge, there is a notion of distributed knowledge (among the frames of mind of agent i) analogous to that considered in Chapter 2. We do not pursue these directions here; Exercise 9.46 deals with the notion of distributed knowledge among the frames of mind.

Note that an agent may hold inconsistent knowledge in a local-reasoning structure: $K_i p \wedge K_i \neg p$ is satisfiable, since in one frame of mind agent i might know p, while in another he might know $\neg p$. In fact, $K_i(\mathit{false})$ is even possible: this will be true at state s if one of the sets in $\mathcal{C}_i(s)$ is the empty set. There is quite a difference between having inconsistent knowledge (that is, $K_i\varphi \wedge K_i\neg\varphi$) and knowing a contradiction (that is, $K_i(\varphi \wedge \neg\varphi)$). In the approach of this section, these are not equivalent. One can imagine a situation where contradictory statements φ and $\neg\varphi$ can both be known: this might correspond to having received contradictory information. It is harder to imagine knowing a contradictory statement $\varphi \wedge \neg\varphi$. Knowing contradictory statements can be forbidden (while still allowing the possibility of having inconsistent knowledge) by simply requiring that each set in each $\mathcal{C}_i(s)$ be nonempty.

If $M = (S, \pi, \mathcal{C}_1, \ldots, \mathcal{C}_n)$ is a local-reasoning structure, and if $\mathcal{C}_i(s)$ is a singleton set for each state s, say $\mathcal{C}_i(s) = \{T_i^s\}$, then M is equivalent to a Kripke structure $(S, \pi, \mathcal{K}_1, \ldots, \mathcal{K}_n)$, where $(s, t) \in \mathcal{K}_i$ exactly if $t \in T_i^s$ (see Exercise 9.47).

Clearly, there is a formal similarity between local-reasoning structures and MS structures, though the philosophy and the semantics are quite different. It is instructive to compare the two approaches. The Montague-Scott approach is more general than

the local reasoning approach. In fact, if we were to take a local-reasoning structure $M = (S, \pi, \mathcal{C}_1, \ldots, \mathcal{C}_n)$, and let $\mathcal{C}_i'$ be the set of all supersets of members of $\mathcal{C}_i$, then the MS structure $M' = (S, \pi, \mathcal{C}_1', \ldots, \mathcal{C}_n')$ is equivalent to the local-reasoning structure M, in the sense that $(M, s) \models \varphi$ iff $(M', s) \models \varphi$ (Exercise 9.48). Despite this formal embedding of local-reasoning structures in MS structures, we do not view the former as a special case of the latter. As we said earlier, the philosophy behind them is quite different. In Montague-Scott semantics, $\mathcal{C}_i(s)$ represents a set of propositions believed by i, while in local reasoning semantics $\mathcal{C}_i(s)$ represents the knowledge of each of the members of the society of minds. Thus the former is a model that explicitly represents knowledge, while the latter is a model for local reasoning.

What about logical omniscience? We already noted that closure under conjunction fails in the local reasoning semantics, since knowing φ and knowing $\neg\varphi$ is not equivalent to knowing $\varphi \wedge \neg\varphi$. It is easy to see that knowledge is not closed under material implication, but for different reasons than for the logics of the previous sections. The formula $K_i p \wedge K_i(p \Rightarrow q) \wedge \neg K_i q$ is satisfiable simply because in one frame of mind agent i might know p, in another he might know $p \Rightarrow q$, but he might never be in a frame of mind where he puts these facts together to conclude q. (See Exercise 9.49 to see how to guarantee closure under material implication.) We do have other types of omniscience: for example, in this approach, there is knowledge of valid formulas and closure under logical implication (Exercise 9.50). In fact, these two types of omniscience form the basis for a sound and complete axiomatization:

Theorem 9.6.1 *The following is a sound and complete axiomatization for validity with respect to local-reasoning structures:*

A1. All instances of tautologies of propositional logic

R1. From φ and $\varphi \Rightarrow \psi$ infer ψ (modus ponens)

R2. From φ infer $K_i\varphi$ (Knowledge of valid formulas)

R3. From $\varphi \Rightarrow \psi$ infer $K_i\varphi \Rightarrow K_i\psi$ (Closure under valid implication)

Proof See Exercise 9.51. ∎

The computational complexity of the satisfiability problem is only *NP*-complete, even if there are many agents. This contrasts with the complexity of the satisfiability problem for K_n, which is *PSPACE*-complete. This is essentially the phenomenon that we saw in Section 9.2.2, where in the absence of closure under conjunction the complexity of the satisfiability problem in MS structures is only *NP*-complete.

Theorem 9.6.2 *The satisfiability problem with respect to local-reasoning structures is NP-complete.*

Proof See Exercise 9.52. ∎

Just as we can impose conditions on the $\mathcal{K}_i$'s to capture various properties of knowledge, we can similarly impose conditions on the $\mathcal{C}_i$'s. We already noted that knowing contradictory statements can be forbidden (while still allowing the possibility of having inconsistent knowledge) by simply requiring that each set in each $\mathcal{C}_i(s)$ be nonempty (this is the analogue of the seriality condition of Section 3.1). We also gave a condition in Exercise 9.49 that guarantees closure under material implication. We now mention some other properties of knowledge that can be guaranteed by appropriate assumptions on the $\mathcal{C}_i$'s. By assuming that s is a member of every $T \in \mathcal{C}_i(s)$, we make $K_i\varphi \Rightarrow \varphi$ valid (this is the analogue to the reflexivity condition of Section 3.1). In Kripke structures, we capture positive and negative introspection by requiring the $\mathcal{K}_i$'s to be transitive and Euclidean, respectively. Here we can capture positive introspection by requiring that if $T \in \mathcal{C}_i(s)$ and $t \in T$, then $T \in \mathcal{C}_i(t)$. Intuitively, this says that in each frame of mind an agent considers it possible that he is in that frame of mind. We can capture negative introspection by requiring that if $T \in \mathcal{C}_i(s)$ and $t \in T$, then $\mathcal{C}_i(t) \subseteq \mathcal{C}_i(s)$. Intuitively, this says that in each frame of mind, the agent considers possible only the actual frames of mind. We note that these conditions are sufficient but not necessary. Exercises 9.53 and 9.54 deal with the conditions that we need to impose on the $\mathcal{C}_i$'s to capture various properties of knowledge.

A particularly interesting special case we can capture is one where in each frame of mind, an agent refuses to admit that he may occasionally be in another frame of mind. (This phenomenon can certainly be observed with people!) Semantically, we can capture this by requiring that if $T \in \mathcal{C}_i(s)$ and $s' \in T$, then $\mathcal{C}_i(s')$ is the singleton set $\{T\}$. This says that if an agent has a frame of mind T, then in every state in this frame of mind, he thinks that his only possible frame of mind is T. We call such agents *narrow-minded agents*.

A narrow-minded agent will believe he is consistent (even if he is not), since in a given frame of mind he refuses to recognize that he may have other frames of mind. Thus, $K_i(\neg(K_i\varphi \wedge K_i\neg\varphi))$ is valid in this case, even though $K_i\varphi \wedge K_i\neg\varphi$ is consistent (Exercise 9.55). Moreover, because an agent can do perfect reasoning within a given frame of mind, a narrow-minded agent will also believe that he is a perfect reasoner. Thus $K_i\big((K_i\varphi \wedge K_i(\varphi \Rightarrow \psi)) \Rightarrow K_i\psi\big)$ is a valid formula in all local-reasoning structures with narrow-minded agents (Exercise 9.55).

9.7 Concluding Remarks

The motivation behind this chapter is the observation that the semantics of knowledge in Kripke structures presented in Chapters 2 and 3, while adequate (and very useful!) for many applications, simply does not work for all applications. In particular, logical omniscience of agents, which is inherent in the standard possible-worlds approach, is in many cases inappropriate.

Just as we do not feel that there is one right, true definition of knowledge that captures all the nuances of the use of the word in English, we also do not feel that there is a single semantic approach to deal with the logical-omniscience problem. Thus, in this chapter we suggested a number of different approaches to avoiding or alleviating the logical-omniscience problem. With the exception of the explicit-representation approach (using either syntactic structures or MS structures), all of our approaches try to maintain the flavor of the possible-worlds approach, with knowledge defined as truth in all possible worlds. Nevertheless, they embody quite different intuitions. The nonstandard approach concedes that agents do not know all the logical consequences of their knowledge, at least if we consider all the logical consequences in standard logic. The hope is that by moving to a nonstandard logic, the fact that an agent's knowledge is closed under logical consequence will become more palatable. The impossible-worlds approach, while formally quite similar to the nonstandard approach, takes the point of view that, although we, the modelers, may know that the world satisfies the laws of standard logic, the agent may be confused, and consider "impossible" worlds possible. The awareness approach adds awareness as another component of knowledge, contending that one cannot explicitly know a fact unless one is aware of it. Finally, the local-reasoning approach tries to capture the intuition of a mind as a society of agents, each with its own (possibly inconsistent) beliefs.

One issue that we did not explore in this chapter is that of hybrid approaches, which combine features from several of the approaches discussed. We also did not address the interaction between knowledge and time. Combining several approaches and adding time to the models can greatly increase the complexity of the situations that can be captured. To see how the extra expressive power gained by adding time can be used, consider how people deal with inconsistencies. It has frequently been observed that people do not like inconsistencies. Yet occasionally they become aware that their beliefs are inconsistent. When this happens, people tend to modify their beliefs in order to make them consistent again. In a system with awareness, local reasoning, and time, this can be captured with the following axiom:

$$\big(X_i\varphi \wedge X_i\neg\varphi \wedge A_i(X_i\varphi \wedge X_i\neg\varphi)\big) \Rightarrow \bigcirc\big(\neg(X_i\varphi \wedge X_i\neg\varphi)\big).$$

This axiom says that if agent i has an inconsistent belief of which he is aware, then at the next step he will modify his belief so that it is no longer inconsistent. See Exercise 9.56 for a further discussion of adding time.

Ultimately, the choice of the approach used depends on the application. Certainly one criterion for an adequate approach is that it be expressive. As is shown in Exercises 9.45 and 9.48, there is a sense in which the syntactic approach, the impossible-worlds approach (in its full generality), and the awareness approach are all equally expressive, and more expressive than the other approaches we have considered. Nevertheless, while whatever approach we use must be expressive enough to describe the relevant details of the application being modeled, the "most expressive" approach is not always the one that does so in the most useful or most natural way. We expect that many applications can be usefully represented using the techniques we have presented, but this is an empirical question that deserves further study.

Exercises

9.1 Show that if $\mathcal{M}$ is any subclass of $\mathcal{M}_n$, then we have full logical omniscience with respect to $\mathcal{M}$.

9.2 This exercise considers some relations among various cases of omniscience.

(a) Assume that whenever φ logically implies ψ, then the formula $\varphi \Rightarrow \psi$ is valid. Show that closure under material implication and knowledge of valid formulas implies closure under logical implication.

(b) Assume that $\varphi \wedge \psi$ logically implies both φ and ψ, and that φ logically implies ψ iff φ is logically equivalent to $\varphi \wedge \psi$. Show that closure under logical implication is equivalent to the combination of closure under logical equivalence and the opposite direction of closure under conjunction (if agent i knows $\varphi \wedge \psi$, then agent i knows both φ and ψ).

9.3 Discuss various conditions that can be imposed on standard syntactic assignments in order to make the positive and negative introspection axioms valid in syntactic structures that satisfy these conditions.

9.4 Let $M = (S, \pi, \mathcal{K}_1, \ldots, \mathcal{K}_n)$ be a Kripke structure. Let M' be the MS structure $(S, \pi, \mathcal{C}_1, \ldots, \mathcal{C}_n)$, where $\mathcal{C}_i(s)$ is the set of all supersets of $\mathcal{K}_i(s)$. Show that $(M, s) \models \varphi$ iff $(M', s) \models \varphi$, for each formula φ.

9.5 Show that the only form of logical omniscience that holds in MS structures is closure under logical equivalence.

** **9.6** Prove Theorem 9.2.2. (Hint: use the maximal consistent set construction as in Chapter 3.)

** **9.7** Fill in the details of the proof of Theorem 9.2.3.

* **9.8** Consider the following possible axioms:

E1. $\neg K_i(\mathit{false})$

E2. $K_i(\mathit{true})$

E3. $K_i(\varphi \wedge \psi) \Rightarrow K_i\varphi$

E4. $K_i\varphi \wedge K_i\psi \Rightarrow K_i(\varphi \wedge \psi)$

E5. $K_i\varphi \Rightarrow K_iK_i\varphi$

E6. $\neg K_i\varphi \Rightarrow K_i\neg K_i\varphi$

E7. $K_i\varphi \Rightarrow \varphi$

Now consider the following conditions on $\mathcal{C}_i$:

C1. $\emptyset \notin \mathcal{C}_i(s)$

C2. $S \in \mathcal{C}_i(s)$

C3. If $T \in \mathcal{C}_i(s)$ and $T \subseteq U$, then $U \in \mathcal{C}_i(s)$

C4. If $T \in \mathcal{C}_i(s)$ and $U \in \mathcal{C}_i(s)$, then $T \cap U \in \mathcal{C}_i(s)$

C5. If $T \in \mathcal{C}_i(s)$ then $\{t \mid T \in \mathcal{C}_i(t)\} \in \mathcal{C}_i(s)$

C6. If $T \notin \mathcal{C}_i(s)$ then $\{t \mid T \notin \mathcal{C}_i(t)\} \in \mathcal{C}_i(s)$

C7. If $T \in \mathcal{C}_i(s)$, then $s \in T$

Define an *MS frame* to be a tuple $(S, \mathcal{C}_1, \ldots, \mathcal{C}_n)$ where S is a set (whose members are called states), and $\mathcal{C}_i(s)$ is a set of subsets of S. We say that the MS structure $(S, \pi, \mathcal{C}_1, \ldots, \mathcal{C}_n)$ is *based on* the MS frame $(S, \mathcal{C}_1, \ldots, \mathcal{C}_n)$. Note that the conditions C1–C7 are really conditions on frames.

Prove that for $1 \leq k \leq 7$, an MS frame N satisfies Ck if and only if every MS structure based on N satisfies Ek at every state. (Hint: the "if" direction may require the use of two primitive propositions.)