A Model-Theoretic Analysis of Knowledge*

Ronald Fagin Joseph Y. Halpern
Moshe Y. Vardi†

IBM Research
Almaden Research Center
650 Harry Road
San Jose, California 95120-6099

Chuangtse and Hueitse had strolled on to the bridge over the Hao, when the former observed, "See how the small fish are darting about! That is the happiness of the fish." "You are not a fish yourself," said Hueitse. "How can you know the happiness of the fish?" "And you not being I," retorted Chuangtse, "how can you know that I do not know?"

Chuangtse, c. 300 BC

Abstract

Understanding knowledge is a fundamental issue in many disciplines. In computer science, knowledge arises not only in the obvious contexts (such as knowledge-based systems), but also in distributed systems (where the goal is to have each processor "know" something, as in agreement protocols). A general semantic model of knowledge is introduced, to allow reasoning about statements such as "He knows that I know whether or not she knows whether or not it is raining." This approach more naturally models a state of knowledge than previous proposals (including Kripke structures). Using this notion of model, a model theory for knowledge is developed. This theory enables one to interpret the notion of a "finite amount of information".

*A preliminary version of this paper appeared in *Proc. 25th IEEE Symp. on Foundations of Computer Science*, 1984, pp. 268–278. This version is essentially identical to the version that appears in *Journal of the ACM* **38**:2, 1991, pp. 382–428.

†Part of the research reported here was performed while this author was visiting the Weizmann Institute of Science, Rehovot, Israel

1 Introduction

Epistemology, the theory of knowledge, has been a subject of philosophical investigation for millennia. Reasoning about knowledge and knowledge representation has also been an issue of concern in Artificial Intelligence for over two decades (cf. [MH, Mo1, BS]). More recently, researchers have realized that these issues also play a crucial role in other subfields of computer science, including cryptography, distributed computation, and database theory, as well as in mathematical economics (cf. [Ha1, Ha2]).

For example, a distributed protocol is often best analyzed in terms of the states of knowledge of the processors involved. In a protocol such as Byzantine Agreement [PSL, DS], it is essential that this analysis includes not only what a processor knows to be true about the world, but also its knowledge about what the other processors know. Such reasoning can, however, get very complicated. As Clark and Marshall point out [CM], while it may be somewhat difficult to keep straight a pipeline of gossip such as "Dean knows that Nixon knows that Haldeman knows that Magruder knows about the Watergate break-in", making sense out of "Dean doesn't know whether Nixon knows that Dean knows that Nixon knows about the Watergate break-in" is much harder. Yet this latter sentence precisely captures the type of reasoning that goes on in proving lower bounds for Byzantine Agreement [DS].

The need to formally model this type of reasoning is our motivation for constructing a semantic model for knowledge. The first attempt to do so was made by Hintikka [Hi], using essentially the notion of *possible worlds*. Hintikka's idea was that someone knows φ exactly if φ is true in all the worlds he thinks are possible. Possible-world semantics has been formalized (cf. [Sa]) using *Kripke structures* [Kr]. In a Kripke structure for knowledge, the "possible worlds" can be viewed as nodes on a graph that are joined by edges of various colors, one corresponding to each "knower" or "agent". Two possible worlds are joined by an edge for agent i exactly if they are indistinguishable as far as agent i is concerned.

There are situations where Kripke structures clearly model the state of knowledge. For example, assume that there is a set of processors, each with a set of clearly defined local states. We then define a Kripke structure whose states consist of the *global states* (which describe the local states of each of the processors), where two global states are indistinguishable to a processor if it has the same local state in both. This is the *situated-automata* approach, where knowledge is ascribed on the basis of the information carried by the state of a machine [Ro]. This approach has been used in a number of papers on distributed systems, including [HM1, DM, FHV, MT, RK]. However, there are situations where it is not clear how to use Kripke structures to model directly a state of knowledge.

Example 1.1: Consider a system with two communicating agents where message transmission is not guaranteed. Suppose two messages have been exchanged: a message from agent 1 to agent 2 saying p (think of p as being "the value is 3"), followed by an acknowledgment from agent 2 that is received by agent 1. Thus, agent 1 knows p, agent 2 knows that agent 1 knows p, agent 1 knows that agent 2 knows that agent 1 knows p, and this, in some sense, is all that is known. While it is easy to construct Kripke structures where the formulas K_1p, K_2K_1p, and $K_1K_2K_1p$ are all true (where $K_i\varphi$ is read "i knows φ"), it is not the least bit obvious which one captures precisely this simple situation, or even if there is one. (It follows from our results in Sections 3 and 4 there indeed is one, but that there is none with only finitely many nodes.) ∎

The difficulty in using Kripke structures to model directly knowledge states also sheds doubt on their adequacy as semantic models for knowledge. To get around this difficulty, various researchers have tried to characterize a state of knowledge syntactically, by the set of formulas that are true of this state (cf. [Hi, MSHI, Mo2]). This method, however, requires infinitely many formulas to characterize a state of knowledge, and still begs the question of what a model of a state of knowledge is. A model to us is a description of the world, not a collection of formulas. Describing a state by the formulas that are true in it seems to avoid the issue of modelling altogether.

In this paper we introduce *knowledge structures*, which are intended to model states of knowledge. We also use the idea of possible worlds, but in a somewhat different way than in Kripke structures. Roughly speaking, we proceed inductively by constructing worlds of each *depth*. A depth 0 world is a description of reality (in the propositional case, a truth assignment to all the primitive propositions); a depth 1 world consists of a set of depth 0 worlds for each agent, corresponding to the worlds that the agent thinks are possible; a depth 2 world consists roughly of a set of possible depth 1 worlds for each agent, etc.

Having modeled knowledge states, we can go back and examine Kripke structures. It turns out that we can now justify the use of Kripke structures as models for collections of knowledge states. More precisely, to every node in a Kripke structure there corresponds a knowledge structure where the same formulas are true, and conversely, for every knowledge structure we can build a Kripke structure one of whose nodes will satisfy the same formulas as the knowledge structure. This correspondence between knowledge structures and Kripke structures enables us to immediately apply to knowledge structures results concerning complete axiomatizations and decision procedures that have already been proved for Kripke structures (cf. [Sa, HM2]).

Although the same axioms characterize knowledge structures and Kripke structures, knowledge structures are a much more flexible tool for examining two concepts that seem to us fundamental - *finite information* and *common knowledge* - and their interaction. (A fact p is common knowledge if everyone knows that everyone knows that everyone knows ... that p. For a discussion of the significance of common knowledge for distributed systems, see [HM1].) We study two model-theoretic constructions, *no-information* and *least-information extensions*, that capture the notion of finite information, and finite information in the presence of common knowledge. An interesting corollary of this investigation is that finite Kripke structures cannot model lack of common knowledge.

Approaches similar to ours have been taken by van Emde Boas et al. [VSG] and by

Mertens and Zamir [MZ]. In [VSG] an epistemic model is used to analyze the Conway Paradox. Like ours, that model captures an infinite hierarchy of knowledge levels, but it does not have the expressive power of knowledge structures. In [MZ] the framework is Bayesian; a world is not just possible or impossible, but it has a probability associated with it. Mertens and Zamir's *infinite hierarchy of beliefs* is the analogue of our knowledge structures in a Bayesian setting. We will have more comments later about the relationship between these works and ours.

The rest of this paper is organized as follows. In the next section we formally describe knowledge structures and show how to use them to give semantics to formulas involving knowledge. In Section 3 we describe the correspondence between knowledge structures and Kripke structures, and show how we exploit this correspondence. In Section 4 we show how to model finite information, and give some tight bounds for what can be known with finite information. These results imply the surprisingly subtle fact that in many practical situations, there can be no nontrivial common knowledge. In this section we also show that finite Kripke structures cannot in general model finite information. In Section 5 we deal with common and joint knowledge. The theme in this section is that common and joint knowledge involve knowledge of transfinite depth, and we develop appropriate tools to deal with it. In Section 6 we compare our approach with the Bayesian approach to modelling knowledge and we comment on the flexibility and utility of knowledge structures, by showing how they can be extended to deal with belief and time. We conclude with some remarks in Section 7.

2 Knowledge structures

In this section we define *knowledge structures*, each of which models a state of knowledge. We assume a finite set of agents. The first step in designing a model of knowledge is to decide what the properties of knowledge should be. The nature of knowledge and its properties has been a matter of great dispute among philosophers. Rather than attempting to resolve these disputes here, we concentrate on one set of properties that seems natural, and mention later (in Section 6) how to modify the model to capture various others.

We take it to be part of the definition of knowledge that anything that someone knows is true. While someone may *believe* false things, it is impossible to have false knowledge. The motivation for the other properties of knowledge that we assume comes from considering a system of idealized rational agents, in which it is common knowledge that each agent is capable of perfect introspection and logical reasoning. In such a system, an agent knows exactly what he does and does not know, and knows also all the logical consequences of his knowledge. Finally, he knows that these properties hold for all the other agents' knowledge. These properties are essentially the axioms that characterize our notion of knowledge. Thus, our knowledge structures will be defined in such a way that they satisfy the following axioms (recall that $K_i\varphi$ means "agent i knows φ"):

1. All substitution instances of propositional tautologies.

2. $K_i\varphi \Rightarrow \varphi$ ("Whatever agent i knows is true").

3. $K_i\varphi \Rightarrow K_iK_i\varphi$ ("Agent i knows what he knows").

4. $\neg K_i\varphi \Rightarrow K_i\neg K_i\varphi$ ("Agent i knows what he does not know").

5. $K_i\varphi_1 \wedge K_i(\varphi_1 \Rightarrow \varphi_2) \Rightarrow K_i\varphi_2$ ("What agent i knows is closed under implication").

These axioms were first discussed by Hintikka [Hi]. The axioms, along with the inference rules of modus ponens ("from φ_1 and $\varphi_1 \Rightarrow \varphi_2$ infer φ_2") and knowledge generalization ("from φ infer $K_i\varphi$") imply that the agents are very wise: each knows all tautologies and all of the consequences of his knowledge, and each knows that all of the other agents are equally wise. It is well-known that Axiom 3 can be derived from the other axioms and inference rules [HC]. It is convenient to refer sometimes to Axiom 3 as describing *positive introspection*, and to Axiom 4 as describing *negative introspection*.

Before we formally define knowledge structures, let us discuss them informally. Assume first that there is only one agent. In this case, a knowledge structure consists of two parts. The first part describes "reality". For simplicity, in this paper we take reality to be a truth assignment to a fixed set of primitive propositions. The second part of a knowledge structure describes a set of "possible worlds", each of which is a truth assignment that the agent thinks is possible.

Example 2.1: Assume that p, q, and r are the primitive propositions, and that "reality" is the truth assignment $p\bar{q}r$, which means that p is true, q is false, and r is true. Assume that the agent knows that exactly one of p, q, or r is false, but that he does not know which. Then his set of "possible worlds" is $\{\bar{p}qr,\ p\bar{q}r,\ pq\bar{r}\}$. ∎

When there are two or more agents, then the situation becomes much more complex. Not only can agents have knowledge about reality, but they can also have knowledge about each other's knowledge.

Example 2.2: Assume there are two agents, Alice and Bob, and that there is only one primitive proposition p. At the "0th level" ("reality"), assume that p is true. The 1st level tells each agent's knowledge about reality. For example, Alice's knowledge at the 1st level could be "I (Alice) don't know whether p is true or false", and Bob's could be "I (Bob) know that p is true". The 2nd level tells each agent's knowledge about the other agent's knowledge about reality. For example, Alice's knowledge at the 2nd level could be "I know that Bob knows whether p is true or false", and Bob's could be "I don't know whether Alice knows p". Thus, Alice knows that either p is true and Bob knows this, or else p is false and Bob knows this. At the 3rd level, Alice's knowledge could be "I know that Bob does not know whether I know about p". This can continue for arbitrarily many levels. ∎

We now give the formal definition of a knowledge structure, and then explain more of the intuition underlying it. We assume a fixed finite set of primitive propositions, and a fixed finite set $\mathcal{P}$ of agents. A *0th-order knowledge assignment*, f_0, is a truth assignment to the primitive propositions. We call $\langle f_0 \rangle$ a *1-ary world* (since its "length" is 1). Intuitively, a 1-ary world is a description of reality. Assume inductively that k-ary worlds (or k-worlds, for short) have been defined. Let W_k be the set of all k-worlds. A *kth-order knowledge assignment* is a function $f_k : \mathcal{P} \to 2 \sup W_k$. Intuitively, f_k associates with each agent a set of "possible k-worlds"; the worlds in $f_k(i)$ are "possible" for agent i and the worlds in $W_k - f_k(i)$ are "impossible" for agent i. A $(k+1)$*-sequence of knowledge assignments* is a sequence $\langle f_0, \ldots, f_k \rangle$, where f_i is an ith-order knowledge assignment. A $(k+1)$*-world* is a $(k+1)$-sequence of knowledge assignments that satisfy certain semantic restrictions, which we shall list shortly. These restrictions enforce the properties of knowledge mentioned above. An infinite sequence $\langle f_0, f_1, f_2, \ldots \rangle$ is called a *knowledge structure* if each *prefix* $\langle f_0, \ldots, f_{k-1} \rangle$ is a k-world for each k. Thus, a k-world describes knowledge of depth $k-1$, and a knowledge structure describes knowledge of arbitrary depth.

Example 2.3: Before we list the restrictions on f_k, let us reconsider Example 2.2. In that example, f_0 is the truth assignment that makes p true. Also, $f_1(\text{Alice}) = \{p, \overline{p}\}$ (where by p (respectively, $\overline{p}$) we mean the 1-world $\langle f_0 \rangle$ (respectively, $\langle f_0' \rangle$), where f_0 (respectively, $\langle f_0' \rangle$) is the truth assignment that makes p true (respectively, false)), and $f_1(\text{Bob}) = \{p\}$. Saying $f_1(\text{Alice}) = \{p, \overline{p}\}$ means that Alice does not know whether p is true or false. We can write the 2-world $\langle f_0, f_1 \rangle$ as

$$\langle p,\ (\text{Alice} \mapsto \{p, \overline{p}\}, \text{Bob} \mapsto \{p\}) \rangle.$$

Let us denote this 2-world by w_1. Let w_2 be the 2-world

$$\langle \overline{p},\ (\text{Alice} \mapsto \{p, \overline{p}\}, \text{Bob} \mapsto \{\overline{p}\}) \rangle,$$

and let w_3 be

$$\langle p,\ (\text{Alice} \mapsto \{p\}, \text{Bob} \mapsto \{p\}) \rangle.$$

In Example 2.2, $f_2(\text{Alice}) = \{w_1, w_2\}$, since Alice thinks both w_1 (where p is true and Bob knows this) and w_2 (where p is false and Bob knows this) are possible worlds. Similarly, $f_2(\text{Bob}) = \{w_1, w_3\}$, since Bob thinks both w_1 (where p is true and Alice does not know it) and w_3 (where p is true and Alice knows this) are possible worlds. ∎

A $(k+1)$-world $\langle f_0, \ldots, f_k \rangle$ must satisfy the following restrictions for each agent i:

(K1) Correctness: $\langle f_0, \ldots, f_{k-1} \rangle \in f_k(i)$, if $k \geq 1$ ("The real k-world is one of the possibilities, for each agent"). In our example, we see that indeed $p \in f_1(\text{Alice})$ and $p \in f_1(\text{Bob})$. Furthermore, $w_1 \in f_2(\text{Alice})$ and $w_1 \in f_2(\text{Bob})$, where we recall that w_1 is the "real" 2-world $\langle f_0, f_1 \rangle$. Intuitively, this condition says that knowledge is always correct (unlike belief, which can be incorrect).

(K2) Introspection: If $\langle g_0, \ldots, g_{k-1} \rangle \in f_k(i)$, and $k > 1$, then $g_{k-1}(i) = f_{k-1}(i)$ ("Agent i knows exactly what he knows"). Let us consider our example. Alice thinks there are two possible 2-worlds, namely w_1 and w_2, since $f_2(\text{Alice}) = \{w_1, w_2\}$. If we write w_2 as $\langle g_0, g_1 \rangle$, then indeed $g_1(\text{Alice}) = \{p, \overline{p}\} = f_1(\text{Alice})$, as required. Intuitively, although Alice has doubts about Bob's knowledge, she has no doubts about her own knowledge. Thus, in all 2-worlds she considers possible, her knowledge is identical, namely, she does not know whether p is true or false. This condition implies that our agents are introspective about their knowledge.

(K3) Extension: $\langle g_0, \ldots, g_{k-2} \rangle \in f_{k-1}(i)$ iff there is a $(k-1)$st-order knowledge assignment g_{k-1} such that $\langle g_0, \ldots, g_{k-2}, g_{k-1} \rangle \in f_k(i)$, if $k > 1$ ("i's higher-order knowledge is an extension of i's lower-order knowledge"). In our example, since Alice thinks either p or $\overline{p}$ is possible, there is some 2-world she thinks possible (namely, w_1) in which p is true, and there is some 2-world she thinks possible (namely, w_2) in which p is false. Conversely, because she thinks w_1 and w_2 are both possible, it follows that she thinks either p or $\overline{p}$ is possible. Intuitively, this condition says that the different levels of knowledge describing a knowledge world are consistent with each other.

We note that (K1) implies that if $\langle f_0, \ldots, f_k \rangle$ is a $(k+1)$-world, then $\langle f_0, \ldots, f_j \rangle$ is a $(j+1)$-world, for all j such that $0 \leq j \leq k$. We also note that our three restrictions imply an apparent strengthening of (K2): namely, if $\langle g_0, \ldots, g_{k-1} \rangle \in f_k(i)$, and $k > 1$, then $g_j(i) = f_j(i)$ if $1 \leq j < k$. Similarly, our conditions imply that the "compatibility" between f_k and f_{k-1} as expressed by (K3) implies that the same compatibility holds between f_k and f_j if $0 < j < k$. Thus, i's higher-level knowledge determines his lower-level knowledge (that is, $f_k(i)$ determines $f_j(i)$ if $0 < j < k$). So, higher-order knowledge refines (that is, adds more detail to) lower-level knowledge.

We have phrased (K3) as a necessary and sufficient condition, but it is easy to see that one direction actually follows from (K1) and (K2). Suppose that $\langle g_0, \ldots, g_{k-2}, g_{k-1} \rangle \in f_k(i)$. Then, by (K2), $g_{k-1}(i) = f_{k-1}(i)$. By (K1), $\langle g_0, \ldots, g_{k-2} \rangle \in g_{k-1}(i)$. It follows that $\langle g_0, \ldots, g_{k-2} \rangle \in f_{k-1}(i)$. From now on, whenever we have to verify that (K3) holds, we will check only the nontrivial direction.

It is not obvious that every world is the prefix of some knowledge structure. McCarthy [Mc] posed essentially this question as an open problem in 1975 (in a different framework, of course). In fact, it may not be obvious to the reader that there are any knowledge structures at all. As we shall see in Section 4, the answer to McCarthy's question is positive.

There are tempting ways to "simplify" knowledge structures. It turns out that the alternative definitions are not expressive enough to model the full range of possibilities that knowledge structures can model. For example, one may want to define a kth-order knowledge assignment as an assignment to each agent of a set of $(k-1)$th-order knowledge assignments (instead of a set of k-worlds). This in fact is the approach taken by van Emde Boas et al. [VSG]. Unfortunately, with this definition we cannot describe the state of knowledge where Alice knows that either p is true and Bob knows it or p is false and Bob does not know it. Essentially, the simpler approach cannot model knowledge about relationships between knowledge and reality, and, more generally, it cannot model knowledge about relationships between different levels of knowledge.

Let $\mathbf{f}$ be the knowledge structure $\langle f_0, f_1, \ldots\rangle$. Define i's *view* of $\mathbf{f}$, denoted $\pi_i(\mathbf{f})$, to be the sequence $\langle f_1(i), f_2(i), \ldots\rangle$. If $\mathbf{f}$ and $\mathbf{f}'$ are knowledge structures, we say that $\mathbf{f}$ and $\mathbf{f}'$ are *i-equivalent*, written $\mathbf{f} \sim_i \mathbf{f}'$, if $\pi_i(\mathbf{f}) = \pi_i(\mathbf{f}')$. Thus, $\mathbf{f}$ and $\mathbf{f}'$ are i-equivalent if agent i cannot distinguish between them.

At this point we can imagine two notions of what it means for agent i to think that a k-world w is possible. The first is the one we have been implicitly using up to now: agent i *thinks* w *is possible in* a knowledge structure $\mathbf{f} = \langle f_0, f_1, \ldots\rangle$ if $w \in f_k(i)$. We say that agent i *thinks* w *is conceivable in* $\mathbf{f}$ if w is a prefix of some knowledge structure $\mathbf{f}'$ such that $\mathbf{f} \sim_i \mathbf{f}'$; i.e. w is the prefix of a knowledge structure that agent i cannot distinguish from $\mathbf{f}$. The following theorem assures us that the two notions of "possible world" are identical.

Theorem 2.4: *Agent i thinks that w is possible in $\mathbf{f}$ iff agent i thinks that w is conceivable in $\mathbf{f}$.*

Proof: Assume first that agent i thinks w is conceivable in $\mathbf{f}$, so $w = \langle f'_0, \ldots, f'_{k-1}\rangle$ is the prefix of a knowledge structure $\mathbf{f}' = \langle f'_0, f'_1, \ldots, \rangle$, where $\mathbf{f} \sim_i \mathbf{f}'$. In particular, we have $f_k(i) = f'_k(i)$. By (K1), $\langle f'_0, \ldots, f'_{k-1}\rangle \in f'_k(i)$. Hence, $w \in f_k(i)$, so agent i thinks w is possible in $\mathbf{f}$.

Conversely, suppose agent i thinks w is possible in $\mathbf{f}$, so that $w = \langle f'_0, \ldots, f'_{k-1}\rangle \in f_k(i)$. By (K2), $f'_{k-1}(i) = f_{k-1}(i)$. As we commented earlier, it follows easily that $f'_j(i) = f_j(i)$ if $1 \le j \le k-1$. Since $\langle f'_0, \ldots, f'_{k-1}\rangle \in f_k(i)$, it follows from (K3) that there is some f'_k such that $\langle f'_0, \ldots, f'_{k-1}, f'_k\rangle \in f_{k+1}(i)$. By (K2), $f'_k(i) = f_k(i)$. Similarly, we can find $f'_{k+1}, f'_{k+2}, \ldots$ such that $\mathbf{f} \sim_i \mathbf{f}' = \langle f'_0, f'_1, \ldots\rangle$. Since w is a prefix of $\mathbf{f}'$, it follows that w is conceivable in $\mathbf{f}$. ∎

The set of *formulas* is the smallest set that contains the primitive propositions, is closed under the Boolean connectives $\neg$ and $\wedge$, and contains $K_i\varphi$ if it contains φ (in Section 5, we discuss richer languages). Boolean connectives such as $\vee$ and $\Rightarrow$ are defined as usual. We now define the *depth* of a formula φ, denoted $\text{depth}(\varphi)$.

1. $\text{depth}(p) = 0$ if p is a primitive proposition;
2. $\text{depth}(\neg\varphi) = \text{depth}(\varphi)$;
3. $\text{depth}(\varphi_1 \wedge \varphi_2) = \max(\text{depth}(\varphi_1), \text{depth}(\varphi_2))$;
4. $\text{depth}(K_i\varphi) = \text{depth}(\varphi) + 1$.

We are almost ready to define what it means for a knowledge structure to *satisfy* a formula. We begin by defining what it means for an $(r+1)$-world $\langle f_0, \ldots, f_r \rangle$ to satisfy formula φ, written $\langle f_0, \ldots, f_r \rangle \models \varphi$, if $r \geq \mathrm{depth}(\varphi)$.

1. $\langle f_0, \ldots, f_r \rangle \models p$, where p is a primitive proposition, if p is true under the truth assignment f_0.

2. $\langle f_0, \ldots, f_r \rangle \models \neg\varphi$ if $\langle f_0, \ldots, f_r \rangle \not\models \varphi$.

3. $\langle f_0, \ldots, f_r \rangle \models \varphi_1 \wedge \varphi_2$ if $\langle f_0, \ldots, f_r \rangle \models \varphi_1$ and $\langle f_0, \ldots, f_r \rangle \models \varphi_2$.

4. $\langle f_0, \ldots, f_r \rangle \models K_i\varphi$ if $\langle g_0, \ldots, g_{r-1} \rangle \models \varphi$ for each $\langle g_0, \ldots, g_{r-1} \rangle \in f_r(i)$.

Let us reconsider Example 2.2. Let w_1 and w_2 be, as before, the two 2-worlds that Alice considers possible. Then $w_1 \models K_{Bob}p$, since according to w_1, the only 1-world Bob considers possible is $\langle p \rangle$. Similarly, $w_2 \models K_{Bob}\neg p$. Hence, both w_1 and w_2 satisfy $(K_{Bob}p \vee K_{Bob}\neg p)$. Since both of the 2-worlds that Alice considers possible satisfy $(K_{Bob}p \vee K_{Bob}\neg p)$, it follows that in our example $\langle f_0, f_1, f_2 \rangle \models K_{Alice}(K_{Bob}p \vee K_{Bob}\neg p)$.

The next lemma says that to determine whether a formula of depth k is satisfied by a world, we need only consider the $(k+1)$-ary prefix of the world.

Lemma 2.5: *Assume that depth(φ) $= k$ and $r \geq k$. Then $\langle f_0, \ldots, f_r \rangle \models \varphi$ iff $\langle f_0, \ldots, f_k \rangle \models \varphi$.*

Proof: The proof is by induction on formulas. The only nontrivial case is when φ is of the form $K_i\psi$, where we assume inductively that the lemma holds when φ is ψ. Assume that $\langle f_0, \ldots, f_r \rangle \models K_i\psi$, and that $\mathrm{depth}(K_i\psi) = k \leq r$. Let $\langle g_0, \ldots, g_{k-1} \rangle$ be an arbitrary member of $f_k(i)$. It follows from (K3) that there exist $g_k, \ldots, g_{r-1}$ such that $\langle g_0, \ldots, g_{k-1}, \ldots, g_{r-1} \rangle \in f_r(i)$. Since $\langle f_0, \ldots, f_r \rangle \models K_i\psi$, it follows by definition that $\langle g_0, \ldots, g_{r-1} \rangle \models \psi$. So, by inductive assumption, $\langle g_0, \ldots, g_{k-1} \rangle \models \psi$. Thus every member of $f_k(i)$ satisfies ψ, and so $\langle f_0, \ldots, f_k \rangle \models K_i\psi$, as desired. The proof of the converse is similar. ∎

We say that the knowledge structure $\mathbf{f} = \langle f_0, f_1, \ldots \rangle$ satisfies φ, written $\mathbf{f} \models \varphi$, if $\langle f_0, \ldots, f_k \rangle \models \varphi$, where $k = \mathrm{depth}(\varphi)$. This is a reasonable definition, since if $w = \langle f_0, \ldots, f_r \rangle$ is an arbitrary prefix of $\mathbf{f}$ such that $r \geq k$, then it then follows from Lemma 2.5 that $\mathbf{f} \models \varphi$ iff $w \models \varphi$. We say that φ is *satisfiable* if it is satisfied in some knowledge structure, and *valid* if it is satisfied in every knowledge structure.

Proposition 2.6: *All of the axioms are valid.*

Proof: (K1) (which says $\langle f_0, \ldots, f_{k-1} \rangle \in f_k(i)$, if $k \geq 1$) causes the axiom $K_i\varphi \Rightarrow \varphi$ to be valid. (K2) (which says that if $\langle g_0, \ldots, g_{k-1} \rangle \in f_k(i)$, and $k > 1$, then $g_{k-1}(i) = f_{k-1}(i)$) can be viewed as a combination of two restrictions, one with $g_{k-1}(i) \subseteq f_{k-1}(i)$, and one with $g_{k-1}(i) \supseteq f_{k-1}(i)$. The former restriction causes the axiom $K_i\varphi \Rightarrow K_iK_i\varphi$ to be valid, and the latter causes the axiom $\neg K_i\varphi \Rightarrow K_i\neg K_i\varphi$ to be valid. The remaining simple details are left to the reader. ∎

Theorem 2.7: $\mathbf{f} \models K_i\varphi$ *iff* $\mathbf{g} \models \varphi$ *whenever* $\mathbf{f} \sim_i \mathbf{g}$.

Proof: Assume that $\mathrm{depth}(K_i\varphi) = k$. We first show that if $\mathbf{f} \models K_i\varphi$ and $\mathbf{f} \sim_i \mathbf{g}$, then $\mathbf{g} \models \varphi$. Let $\mathbf{f}$ be $\langle f_0, f_1, \ldots\rangle$, and let w be the k-world that is a prefix of $\mathbf{g}$. By Theorem 2.4, we know that $w \in f_k(i)$. Since $\mathbf{f} \models K_i\varphi$, it follows by definition that every member of $f_k(i)$ (and in particular, w) satisfies φ. Since $w \models \varphi$, it follows by definition that $\mathbf{g} \models \varphi$, as desired.

Conversely, assume that $\mathbf{g} \models \varphi$ whenever $\mathbf{f} \sim_i \mathbf{g}$. To show that $\mathbf{f} \models K_i\varphi$, we must show that $w \models \varphi$ for each $w \in f_k(i)$. Assume that $w \in f_k(i)$. By Theorem 2.4, w is a prefix of some $\mathbf{g}$ such that $\mathbf{f} \sim_i \mathbf{g}$. By assumption, $\mathbf{g} \models \varphi$, and so by definition $w \models \varphi$. ∎

Thus, agent i knows φ precisely if φ holds in every knowledge structure that i thinks possible. Theorem 2.7 is a powerful tool. It shows the equivalence of two distinct notions of truth. The first notion of truth, which we can call "internal truth", says that $K_i\varphi$ is true if φ is true in every k-world that i thinks is possible (where $\mathrm{depth}(K_i\varphi) = k$). These k-worlds are obtained by "looking inside" the knowledge structure (at level k). Thus, internal truth is a finitistic notion. The other notion of truth, which we can call "external truth", says that $K_i\varphi$ is true if φ is true in each of the knowledge structures that i thinks is possible, of which there can be uncountably many.

3 Knowledge structures and Kripke structures

Previous attempts (cf. [Sa, MSHI, Mo1]) to provide a semantic foundation for reasoning about knowledge have made use of *Kripke structures* [Kr].

Suppose we have agents $1, \ldots, n$. The corresponding Kripke structure M is a tuple $(S, \pi, \mathcal{K}_1, \ldots, \mathcal{K}_n)$, where S is a set of *states*, $\pi(s)$ is a truth assignment to the primitive propositions for each state $s \in S$, and $\mathcal{K}_i$ is an equivalence relation on S (i.e., a reflexive, symmetric, and transitive binary relation on S). Intuitively, $(s,t) \in \mathcal{K}_i$ iff s and t are indistinguishable as far as agent i's knowledge is concerned. We now define what it means for a formula φ to be *satisfied* at a state s of M, written $M, s \models \varphi$.

1. $M, s \models p$, where p is a primitive proposition, if p is true under the truth assignment $\pi(s)$.
2. $M, s \models \neg\varphi$ if $M, s \not\models \varphi$.
3. $M, s \models \varphi_1 \wedge \varphi_2$ if $M, s \models \varphi_1$ and $M, s \models \varphi_2$.
4. $M, s \models K_i\varphi$ if $M, t \models \varphi$ for all t such that $(s,t) \in \mathcal{K}_i$.

It is not hard to show that with Kripke semantics, the modality K_i also has all the properties discussed in the previous section (see [HM2, Sa] for more details). The reflexivity of $\mathcal{K}_i$ gives us $K_i\varphi \Rightarrow \varphi$, transitivity gives us $K_i\varphi \Rightarrow K_iK_i\varphi$, and symmetry and transitivity together give us $\neg K_i\varphi \Rightarrow K_i\neg K_i\varphi$.

Even though Kripke structures give K_i the desired properties, it is not clear that they actually capture our intuition about knowledge. In particular, it is not clear what state of knowledge corresponds to a state in a Kripke structure. The following theorem clarifies this issue by providing an exact correspondence between knowledge structures and states in Kripke structures.

We say that a state s of Kripke structure M is *equivalent* to knowledge structure $\mathbf{f}$ if $M, s \models \varphi$ iff $\mathbf{f} \models \varphi$, for every formula φ. That is, s and $\mathbf{f}$ are equivalent if they satisfy the same formulas.

Theorem 3.1: *To every Kripke structure M and state s in M, there corresponds a knowledge structure $\mathbf{f}_{M,s}$ such that s is equivalent to $\mathbf{f}_{M,s}$. Conversely, there is a Kripke structure M_{know} such that for every knowledge structure $\mathbf{f}$ there is a state $s_{\mathbf{f}}$ in M_{know} such that $\mathbf{f}$ is equivalent to $s_{\mathbf{f}}$.*

Proof: Suppose M is a Kripke structure. For every state s in M we construct a knowledge structure $\mathbf{f}_{M,s} = \langle f_0^s, f_1^s, \ldots \rangle$. First, f_0^s is just the truth assignment $\pi(s)$. Suppose we have constructed $f_0^s, f_1^s, \ldots, f_k^s$ for each state s in M. Then $f_{k+1}^s(i) = \{\langle f_0^t, \ldots, f_k^t \rangle : (s,t) \in \mathcal{K}_i\}$. We leave it to the reader to check that $M_{know}, s \models \varphi$ iff $\mathbf{f}_{M,s} \models \varphi$.

For the converse, let $M_{know} = (S_{know}, \pi, \mathcal{K}_1, \ldots, \mathcal{K}_n)$, where S_{know} consists of *all* the knowledge structures, $\pi(\mathbf{f}) = f_0$, and $(\mathbf{f}, \mathbf{g}) \in \mathcal{K}_i$ iff $\mathbf{f} \sim_i \mathbf{g}$. Now using Theorem 2.7, we can show $M_{know}, \mathbf{f} \models \varphi$ iff $\mathbf{f} \models \varphi$. ∎

In [Sa] it is shown that the axioms of the previous section, with modus ponens and knowledge generalization as the rules of inference, give a complete axiomatization for the Kripke structure semantics of knowledge, while in [HM2] it is shown, again with respect to Kripke structure semantics, that the question of deciding if a formula is satisfiable is PSPACE-complete (provided there are at least two agents). From Theorem 3.1 it follows that these results also apply to the knowledge structure semantics, so we get:

Corollary 3.2: *The axioms of the previous section, together with modus ponens and knowledge generalization as the rules of inference, give a complete axiomatization for knowledge structures.*

Corollary 3.3: *The problem of deciding if a formula is satisfiable is PSPACE-complete (provided there are at least two agents).*

We note that in [FV1], it is shown how to obtain an elegant, constructive proof of Corollary 3.2, by working only with knowledge structures and not making use of the completeness theorem for Kripke structures.

Theorem 3.1 shows that knowledge structures and Kripke structures have the same theory, but its implications are deeper. It shows that knowledge structures and Kripke structures complement each other in modelling knowledge: knowledge structures model states of knowledge, and Kripke structures model collections of knowledge states.

4 Modelling Finite Information

4.1 The no-information extension

A knowledge structure fully describes a state of knowledge; i.e., it describes arbitrarily deep levels of knowledge. In reality, however, agents have only a finite amount of information. In this section we study the knowledge states that arise from finite amounts of information. Put differently, we study knowledge structures that have finite descriptions.

Consider a variant of Example 1.1, where we have a system with three communicating agents, Alice, Bob, and Charlie. Assume that Bob has sent no messages and has received only one message, a message from Alice saying "p" (i.e. p is true). For ease of exposition, let us also assume that p is the only primitive proposition. Intuitively, all that Bob knows at this point is that Alice knows p. But what state of knowledge does this correspond to?

The answer to this question depends in part on the underlying model of knowledge acquisition (cf. [FV2, FHV]). For example, is it possible as far as Bob is concerned that Charlie knows that Bob knows p, even though Bob never sent any messages? The answer may be yes if each agent stores the information he has about primitive propositions and about the information he has received from other agents in a database, and if databases are insecure, so that agents can read each other's databases (then Charlie can find out what information Bob has received, without receiving any messages from other agents). It may also be yes if messages are *guaranteed* to arrive in one round of communication, for in that case, for all Bob knows, Alice may have sent Charlie a message (in the first round) saying that she would also send Bob the message "p" in that round. On the other hand, if message communication is *not* guaranteed and databases are secure, then Bob knows that Charlie does *not* know that Bob knows p. Is it possible that Alice knows that Bob knows that Alice knows p? Again, the answer depends in part on whether communication is guaranteed. If there is a chance that messages may not arrive, then it is not possible for Alice to have such depth 3 knowledge at the end of the first round.

In order to characterize Bob's knowledge state, we will first consider the most "permissive" situation, where we assume that agents have no knowledge about how other agents acquire information. Thus, agents should allow for all possibilities that are consistent with the information they have. Since Bob has received a message from Alice saying p is true (and we assume that messages are honest), Bob knows that Alice knows p. Of course, Bob also knows that he himself knows p, but he has no idea whether Charlie knows p. Thus there are two 2-worlds that Bob thinks are possible:

$$\langle p, (\text{Alice} \mapsto \{p\}, \text{Bob} \mapsto \{p\}, \text{Charlie} \mapsto \{p\})\rangle$$

and

$$\langle p, (\text{Alice} \mapsto \{p\}, \text{Bob} \mapsto \{p\}, \text{Charlie} \mapsto \{p, \bar{p}\})\rangle$$

.

Let W be the set consisting of these two 2-worlds. What 3-worlds does Bob think possible? Intuitively, Bob should consider a 3-world $w = \langle g_0, g_1, g_2 \rangle$ possible if it is consistent with Bob's information, that is, if $g_2(\text{Bob}) = W$. Let W' be the set of 3-worlds that satisfy this condition. These are the 3-worlds that Bob considers possible. This idea extends. The set of 4-worlds that Bob thinks are possible consists of all the 4-worlds $\langle g_0, g_1, g_2, g_3 \rangle$ such that $g_3(\text{Bob}) = W'$. We shall generalize this idea shortly when we define the no-information extension.

Let W be a set of $(k-1)$-worlds, and i an agent. Define the *i-extension of W* to be the set of k-worlds given by $\{\langle g_0, \ldots, g_{k-1} \rangle : g_{k-1}(i) = W\}$. Intuitively, this is the set of k-worlds w such that in w, agent i considers W as the set of possible $(k-1)$-worlds.

Definition 4.1: Let $w = \langle f_0, \ldots, f_k \rangle$ be a $(k+1)$-world. The *one-step no-information extension w^+ of w* is the $(k+2)$-world $\langle f_0, \ldots, f_k, f_{k+1} \rangle$, where $f_{k+1}(i)$ is the i-extension of $f_k(i)$. Thus, $f_{k+1}(i) = \{\langle g_0, \ldots, g_k \rangle : g_k(i) = f_k(i)\}$. ∎

In the above definition (and later), we use the convention that $f_0(i)$ is the empty set for each agent i. Hence, $\langle f_0 \rangle^+ = \langle f_0, f_1 \rangle$, where $f_1(i)$ is the set of all truth assignments, for each agent i. Intuitively, the one-step no-information extension $\langle f_0, \ldots, f_k, f_{k+1} \rangle$ of $\langle f_0, \ldots, f_k \rangle$ describes what each agent knows at depth $k+1$, assuming that "all that each agent i knows" is already described by $f_k(i)$ and given the underlying "permissive" model described above. Thus, $f_{k+1}(i)$ is the set of all k-worlds that are compatible with i's lower-depth knowledge.

We can relate this definition to the notion of i-equivalence defined in the previous section as follows. Let $w = \langle f_0, ..., f_{k-1} \rangle$ and $w' = \langle f'_0, ..., f'_{k-1} \rangle$ be k-worlds. By analogy with our definition of i-equivalence for knowledge structures, let us say that the worlds w and w' are *i-equivalent* (written $w \sim_i w'$), if $f_j(i) = f'_j(i)$ for $0 < j < k$. Then (as noted in Section 2), $w \sim_i w'$ iff $f_{k-1}(i) = f'_{k-1}(i)$. So, $f_{k+1}(i)$ is the i-extension of $f_k(i)$ if $f_{k+1}(i) = \{\langle g_0, \ldots, g_k \rangle : \langle g_0, \ldots, g_k \rangle \sim_i \langle f_0, \ldots, f_k \rangle\}$.

The intention of the one-step no-information extension is that w^+ describes the knowledge of the agents one level higher than the description of their knowledge in w, if w completely describes the information they have. However, *a priori*, it is not clear that w^+ is even a world, since it may not satisfy the three restrictions described in Section 2. Before removing this doubt we need some auxiliary machinery.

The following lemma, whose proof is left to the reader, shows that knowledge assignments can be "mixed".

Lemma 4.2: *Let $\langle f_0, \ldots, f_k \rangle$ and $\langle g_0, \ldots, g_k \rangle$ be $(k+1)$-worlds such that $\langle f_0, \ldots, f_{k-1} \rangle \sim_i \langle g_0, \ldots, g_{k-1} \rangle$. Let h_k be a kth-order knowledge assignment such that $h_k(i) = f_k(i)$ and $h_k(j) = g_k(j)$ for $j \neq i$. Then $\langle g_0, \ldots, g_{k-1}, h_k \rangle$ is a $(k+1)$-world.*

Let $w = \langle f_0, \ldots, f_k \rangle$ be a $(k+1)$-world and let $\langle g_0, \ldots, g_{k-1} \rangle \in f_k(i)$. The *$i$-matching extension* of $\langle g_0, \ldots, g_{k-1} \rangle$ with respect to w is the sequence $\langle g_0, \ldots, g_{k-1}, g_k \rangle$, where $g_k(i) = f_k(i)$ and $g_k(j)$ is the j-extension of $g_{k-1}(j)$ for $j \neq i$.

Lemma 4.3:

1. *Let* $\langle f_0, \ldots, f_{k-1}\rangle$ *be a* k*-world. Then the one-step no-information extension is a* $(k+1)$*-world.*

2. *Let* $w = \langle f_0, \ldots, f_k\rangle$ *be a* $(k+1)$*-world and let* $\langle g_0, \ldots, g_{k-1}\rangle \in f_k(i)$*. Then the* i*-matching extension of* $\langle g_0, \ldots, g_{k-1}\rangle$ *with respect to* w *is a* $(k+1)$*-world.*

Proof: We prove parts (1) and (2) simultaneously by induction on k. To prove part (1) we consider a k-world $w = \langle f_0, \ldots, f_{k-1}\rangle$. Let $\langle f_0, \ldots, f_k\rangle$ be the one-step no-information extension of w. We have to show that $\langle f_0, \ldots, f_k\rangle$ satisfies the restrictions (K1), (K2), and (K3). The case $k = 1$ is immediate. For the inductive step, again the fact that (K1) and (K2) hold follows immediately from the definition of the one-step no-information extension. For (K3), suppose $\langle g_0, \ldots, g_{k-2}\rangle \in f_{k-1}(i)$. Let w' be the i-matching extension of $\langle g_0, \ldots, g_{k-2}\rangle$ with respect to w. By the induction hypothesis, w' is a k-world. By construction $w \sim_i w'$, so by the definition of the one-step no-information extension, $w' \in f_k(i)$. Thus, every element of $f_k(i)$ has an extension in $f_{k+1}(i)$, and (K3) holds.

To prove part (2) we consider first the one-step no-information extension $\langle g_0, \ldots, g_k\rangle$ of $\langle g_0, \ldots, g_{k-1}\rangle$. By the induction hypothesis, $\langle g_0, \ldots, g_k\rangle$ is a $(k+1)$-world. Since $\langle g_0, \ldots, g_{k-1}\rangle \in f_k(i)$, we have that $\langle f_0, \ldots, f_{k-1}\rangle \sim_i \langle g_0, \ldots, g_{k-1}\rangle$. The claim now follows by Lemma 4.2. ∎

Part (1) of Lemma 4.3 tells us that indeed, the one-step no-information extension of a world is a world.

We now develop some machinery that justifies the name "no-information extension". We have defined what it means for an $(r+1)$-world $\langle f_0, \ldots, f_r\rangle$ to satisfy formula φ, written $\langle f_0, \ldots, f_r\rangle \models \varphi$, if $r \geq \text{depth}(\varphi)$. We now give an extension of this definition to formulas of greater depth. Let us say that a world w *must eventually satisfy* φ, written $w \models^+ \varphi$, if for every knowledge structure $\mathbf{f}$ with w as a prefix, $\mathbf{f} \models \varphi$. For example, if the primitive proposition p is true under the truth assignment f_0, then $\langle f_0\rangle \models^+ \neg K_1 \neg p$, since if p is true, then it is not possible for an agent to know $\neg p$. Note that we cannot replace $\models^+$ by $\models$ (in other words, it is not the case that $\langle f_0\rangle \models \neg K_1\neg p$), since the depth of the formula $\neg K_1 \neg p$ is too big.

Say that a set Σ of formulas *logically implies* the formula σ, written $\Sigma \models \sigma$, if every knowledge structure that satisfies every formula of Σ also satisfies σ. That is, Σ logically σ if there is no "counterexample" knowledge structure that satisfies every formula of Σ but not σ.

The next proposition justifies the name "no-information extension", by characterizing when an agent i knows a formula φ of depth at most $k-1$ in the one-step no-information extension of a k-ary world $w = \langle f_0, \ldots, f_{k-1}\rangle$. The proposition says that this happens precisely when the truth of the formula $K_i\varphi$ is already guaranteed by w anyway. There are two ways that we make precise "the truth of the formula $K_i\varphi$ is already guaranteed by w anyway". In the first sense (part (2) of Proposition 4.4 below), $w \models^{+} K_i\varphi$; that is, w must eventually satisfy $K_i\varphi$, as defined above. The second sense (part (3) of Proposition 4.4) says that knowledge (and lack of knowledge) of agent i, as described by w, is sufficient to logically imply $K_i\varphi$.

Proposition 4.4: *Assume that i is an agent, w is a k-world, and φ is a formula of depth at most $k-1$. The following are equivalent.*

1. $w^{+} \models K_i\varphi$
2. $w \models^{+} K_i\varphi$
3. *Let Σ be the set of all formulas γ of depth at most $k-1$ of the form $K_i\psi$ or $\neg K_i\psi$ that are satisfied by w. Then $\Sigma \models K_i\varphi$.*[1]

Proof: It is immediate that (2) implies (1), since w is a prefix of w^{+}. We now show that (3) implies (2). Assume that (3) holds. Let $\mathbf{f}$ be a knowledge structure with w as a prefix. We must show that $\mathbf{f} \models K_i\varphi$. Since w satisfies Σ, so does $\mathbf{f}$. Since $\Sigma \models K_i\varphi$, it follows that $\mathbf{f} \models K_i\varphi$, as desired.

We now show that (1) implies (3). Assume that (1) holds. Let $w^{+} = \langle f_0, \ldots, f_k\rangle$. Let Σ be as in (3), and let $\mathbf{f}' = \langle f_0', f_1', \ldots\rangle$ be a knowledge structure that satisfies Σ. We must show that $\mathbf{f}' \models K_i\varphi$. That is, let $v = \langle g_0, \ldots, g_{k-1}\rangle$ be an arbitrary member of $f_k'(i)$; we must show that $v \models \varphi$.

Since $\mathbf{f}'$ satisfies Σ, so does its k-ary prefix $\langle f_0', \ldots, f_{k-1}'\rangle$. It therefore follows from Lemma A.1 of the appendix that $f_{k-1}(i) = f_{k-1}'(i)$. Since $v = \langle g_0, \ldots, g_{k-1}\rangle \in f_k'(i)$, it follows from (K2) that $g_{k-1}(i) = f_{k-1}'(i)$. Hence, $g_{k-1}(i) = f_{k-1}(i)$. Therefore, by definition of the one-step no-information extension, $v \in f_k(i)$. Since $w^{+} = \langle f_0, \ldots, f_k\rangle \models K_i\varphi$, it follows that $v \models \varphi$. This was to be shown. ∎

We now define the no-information extension of a world w to be the result of repeatedly taking one-step no-information extensions. Formally, the *no-information extension* w^* of w is the sequence $\langle f_0, \ldots, f_k, f_{k+1}, \ldots\rangle$, where $\langle f_0, \ldots, f_{m+1}\rangle$ is the one-step no-information extension of $\langle f_0, \ldots, f_m\rangle$ for each $m \geq k$. Intuitively, the no-information extension w^* is a knowledge structure that describes the knowledge of the agents, if w completely describes the information they have.

We might hope that an analogous proposition to Proposition 4.4 would hold for the (full) no-information extension. However, this is not the case. To understand why, let us denote the *two-step no-information extension* $(w^{+})^{+}$ of w by w^{++}. If we replace w^{+} everywhere in Proposition 4.4 by w^{++}, and let φ be of depth k, then the proposition no longer holds. For example, let p be a primitive proposition that is true under the truth assignment f_0. Then $\langle f_0\rangle^{+} \models \neg K_1 p$, so by negative introspection, $\langle f_0\rangle^{++} \models K_1\neg K_1 p$.

[1] Note that if $k=1$, then there are no formulas γ of depth 0 of the form $K_i\psi$ or $\neg K_i\psi$. Thus, if $k=1$ then Σ is the empty set.

Therefore, if φ is $\neg K_1 p$, then $\langle f_0\rangle^{++} \models K_i\varphi$, although $\langle f_0\rangle \not\models^{+} K_i\varphi$ and $\Sigma \not\models K_i\varphi$, where Σ is as in part (3) of Proposition 4.4. What is happening is that negative knowledge at one level induces positive knowledge at the next level. This would not happen if we modified the definition of knowledge structures by eliminating negative introspection, as is done in [Va1].

Let $w = \langle f_0, \ldots, f_k\rangle$ be a $(k+1)$-world, and let $w^+ = \langle f_0, \ldots, f_k, f_{k+1}\rangle$ be the one-step no-information extension. Note that for each $(k+2)$-world $\langle f_0, \ldots, f_k, f'_{k+1}\rangle$ that extends w, we have $f'_{k+1}(i) \subseteq f_{k+1}(i)$ for each agent i (this follows from the restriction that for every $\langle g_0, \ldots, g_k\rangle \in f'_{k+1}(i)$, necessarily $g_k(i) = f_k(i)$). Thus, the one-step no-information extension can be characterized by the fact that $f_{k+1}(i)$ is maximal for each i. This explains why "no extra knowledge" is added in taking the one-step no-information extension, since the more possible worlds there are, the less knowledge there is. However, now let $\langle f_0, \ldots, f_k, f_{k+1}, f_{k+2}\rangle$ be the two-step no-information extension w^{++}. It is *not* the case that $f_{k+2}(i)$ is maximal for each i, among all two-step extensions $\langle f_0, \ldots, f_k, f'_{k+1}, f'_{k+2}\rangle$ of w. This is because if $f'_{k+1}(i) \neq f_{k+1}(i)$, then $f'_{k+2}(i)$ and $f_{k+2}(i)$ are incomparable (and in fact, disjoint from each other), since every $\langle g_0, \ldots, g_{k+1}\rangle \in f'_{k+2}$ has $g_{k+1}(i) = f'_{k+1}(i)$, whereas every $\langle g_0, \ldots, g_{k+1}\rangle \in f_{k+2}$ has $g_{k+1}(i) = f_{k+1}(i)$. Again, the point is that lack of knowledge at one level induces knowledge at the next level, by negative introspection.

The next theorem follows immediately from part (1) of Lemma 4.3.

Theorem 4.5: *For all worlds w, the no-information extension w^* is a knowledge structure.*

Corollary 4.6: *Every world is the prefix of a knowledge structure.*

We note that in fact, it is not hard to show that every world is the prefix of uncountably many distinct knowledge structures. Corollary 4.6 answers McCarthy's question (see Section 2) positively.

We shall investigate some properties of the no-information extension in the next section.

4.2 On the Presence of Common Knowledge

Recall that state s of Kripke structure M is *equivalent* to knowledge structure $\mathbf{f}$ if $M, s \models \varphi$ iff $\mathbf{f} \models \varphi$, for every formula φ. Since a no-information extension captures what is perhaps the most natural notion of finite information, one might hope that for each no-information extension w^*, there would be a finite Kripke structure (i.e., one with finitely many states), one of whose states is equivalent to w^*. However, this is not true. In fact, it is very far from the truth: for *no* no-information extension is there such a finite Kripke structure. To understand why, we must first consider the notion of *common knowledge*.

Assume that the agents are 1,...,n. Let $E\varphi$ be a shorthand for $K_1\varphi \wedge \ldots \wedge K_n\varphi$, that is, "Everyone knows φ". Let $E \sup 0\varphi$ be φ, and let $E \sup j\varphi$ abbreviate $EE \sup j-1\varphi$ for $j \geq 1$. We say that the formula φ is *common knowledge in knowledge structure* **f** if $\mathbf{f} \models E \sup j\varphi$ for every $j > 0$. Similarly, we say that the formula φ is *common knowledge in state s of Kripke structure M* if $M, s \models E \sup j\varphi$ for every $j > 0$.[2]

As we shall show, there is *never* any nontrivial common knowledge in a no-information extension, as long as there are at least two agents. In fact, we shall show (in Corollary 4.13 below) that φ is common knowledge in w^* iff φ is valid. We now show that the situation is completely different for finite Kripke structures. We first need some preliminary definitions, which will also be useful for some of the theorems we prove later.

Definition 4.7: Let $\mathbf{p} = i_1 \ldots i_s$ be a finite string of agents. The *length* of $\mathbf{p}$ is s. The *reverse* of $\mathbf{p}$, written $\mathbf{p} \sup R$, is $i_s \ldots i_1$. If φ is a formula, then let $K_{\mathbf{p}}\varphi$ be an abbreviation for the formula $K_{i_1} \ldots K_{i_s}\varphi$. If $\mathbf{p}$ is the empty string, then $K_{\mathbf{p}}\varphi$ is taken to be φ. ∎

Definition 4.8: Let $M = (S, \pi, \mathcal{K}_1, \ldots, \mathcal{K}_n)$ be a Kripke structure. We say that there is a *path of length k* between two states s and t in S if there is a sequence $u_0, \ldots, u_k$ of states in S such that $s = u_0$, $t = u_k$ and for all $0 \leq i \leq k-1$ we have that $(u_i, u_{i+1}) \in \mathcal{K}_j$ for some $1 \leq j \leq n$. If there is such a path, then we say that the two states are *connected*. The *distance* between s and t is the length of the shortest path between s and t if such a path exists, and is undefined otherwise. ∎

Theorem 4.9: *Assume that there are at least two agents. Then for each finite Kripke structure M, there is some nonvalid formula φ that is common knowledge in every state of M.*

Proof: Let $M = (S, \pi, \mathcal{K}_1, \ldots, \mathcal{K}_n)$, $n \geq 2$. Let δ be the maximal distance between any two connected states in S, and let p be a primitive proposition.

Let ψ be the formula $E \sup \delta p \Rightarrow E \sup \delta + 1p$. We claim that ψ is not valid. By way of proof, consider the Kripke structure $M' = (S', \pi', \mathcal{K}'_1, \ldots, \mathcal{K}'_n)$, where $S' = \{0, \ldots, \delta, \delta + 1\}$, $\pi'(i)$ makes p true iff $0 \leq i \leq \delta$, for $j \neq 1$ and $j \neq 2$ the relation $\mathcal{K}'_j$ is the trivial equivalence relation $\{(i,i) : 0 \leq i \leq \delta+1\}$, $\mathcal{K}'_1$ is the reflexive closure of $\{(i, i+1) : 0 \leq i \leq \delta$ and i is even$\}$, and $\mathcal{K}'_2$ is the reflexive closure of $\{(i, i+1) : 0 \leq i \leq \delta$ and i is odd$\}$. We leave it to the reader to verify that $M', 0 \not\models \psi$. (Note, however, that ψ is valid if there is only one agent.)

We now claim that ψ is true in every state of S. Suppose that $E \sup \delta\varphi$ is true in a state s in S. Then φ must be true in all states whose distance from s is less or equal to δ. By the definition of δ, it follows that φ must be true in all states that are connected to s, and consequently $E \sup \delta + 1\varphi$ is true in s. Since ψ is true in all states of M, it follows that ψ is common knowledge in every state of M. ∎

[2] There are other interpretations of the notion of common knowledge. See [Ba].

Remark 4.10: Interestingly, the theorem does not hold if there is only one agent. The intuitive reason is that in that case there are only finitely many distinct knowledge states (given our assumption of a finite set of primitive propositions) and each one of them may hold in one of the connected components of M. A weaker version of the theorem is still true, however. For each finite Kripke structure M and each state s of M, there is some nonvalid formula φ that is common knowledge in s. ∎

We now show that agents have arbitrarily deep nontrivial knowledge in a no-information extension. It is instructive to consider first the "simplest" no-information extension. Assume that there is only one primitive proposition p, and that there are only two agents, Alice and Bob. As before, for convenience let us for now denote simply by p the truth assignment that makes p true. Intuitively, the no-information extension $\langle p\rangle^*$ is the knowledge structure where p is true, and where Alice and Bob have no information. Assume that $\langle p\rangle^* \models K_{Alice}\varphi$. We might conjecture that φ must then be valid, since, after all, Alice has "no information" in $\langle p\rangle^*$. This, however, is not the case. For, since Alice does not know p, she knows that she does not know p. That is, $\langle p\rangle^* \models K_{Alice}\neg K_{Alice}p$, although $\neg K_{Alice}p$ is not valid. What if $\langle p\rangle^* \models K_{Alice}K_{Bob}\varphi$? We are certainly tempted to conjecture that φ must then be valid. After all, in $\langle p\rangle^*$, Alice "has no information" about Bob. Once again, our intuition is incorrect. For, since Alice knows that the formula $K_{Alice}p$ is false, she also knows that Bob cannot know this formula (because anything that Bob knows must be true); hence, Alice knows that $\neg K_{Bob}K_{Alice}p$ is true. Since Alice knows that Bob is introspective, she knows that if Bob does not know something, then he knows that he does not know it. Thus, $\langle p\rangle^* \models K_{Alice}K_{Bob}\neg K_{Bob}K_{Alice}p$. Hence, if φ is the (non-valid) formula $\neg K_{Bob}K_{Alice}p$, then $\langle p\rangle^* \models K_{Alice}K_{Bob}\varphi$, contrary to the tempting conjecture. In fact, it follows from the proof of Theorem 4.11 below that this generalizes, so that if $\mathbf{q} \in \{Alice, Bob\}^*$, then $\langle p\rangle^* \models K_{\mathbf{q}}\neg K_{(\mathbf{q}\sup R)}p$. Thus, in the no-information extension $\langle p\rangle^*$, where Alice and Bob have "no information" about each other, they nevertheless have arbitrarily deep nontrivial knowledge! Moreover, by taking φ to be the disjunction of $\neg K_{(\mathbf{q}\sup R)}p$ over all $\mathbf{q}$ of length k, it follows that $\langle p\rangle^* \models E\sup k\varphi$. More generally, we have the following theorem, which we shall show to be the best possible.

Theorem 4.11: *For every k, there is a k-world w such that for every l, there is a nonvalid formula φ of depth l where $w^* \models E\sup l+k-1\varphi$.*

Proof: We assume for convenience that there is exactly one primitive proposition p. Let w be the k-world $\langle f_0,\ldots,f_{k-1}\rangle$ where f_0 is the truth assignment that makes p true, and where $f_j(i) = \{\langle f_0,\ldots,f_{j-1}\rangle\}$ if $1 \le j < k$, for each agent i. Thus, $f_j(i)$ is a singleton set for $1 \le j < k$ and for each agent i. For each $\mathbf{r} \in \mathcal{P}^*$ of length $k-1$, it is easy to see that $w \models K_{\mathbf{r}}p$. Let $\mathbf{q} = \mathbf{rs}$ be an arbitrary member of $\mathcal{P}^*$ of length $l+k-1$, where $\mathbf{r}$ is of length $k-1$ and $\mathbf{s}$ is of length l. Let $\psi_{\mathbf{s}}$ be the formula $\neg K_{(\mathbf{s}\sup R)}\neg p$. We shall now show that $w^* \models K_{\mathbf{q}}\psi_{\mathbf{s}}$.

We first show, by induction on l, that for every $\mathbf{p} \in \mathcal{P}^*$ of length l, the formula p implies $K_{\mathbf{p}}\neg K_{(\mathbf{p}\sup R)}\neg p$. The base case ($l = 0$) is immediate. Assume inductively that

p implies $K_{\mathbf{p}}\neg K_{(\mathbf{p}\sup R)}\neg p$; we shall show that p implies $K_{\mathbf{p}i}\neg K_{i(\mathbf{p}\sup R)}\neg p$, where i is an arbitrary agent. It is convenient to give names for reference to the following two simple facts.

Fact 1. α implies $\neg K_i \neg\alpha$.

Fact 2. If α implies β, then $K_{\mathbf{t}}\alpha$ implies $K_{\mathbf{t}}\beta$, for $\mathbf{t} \in \mathcal{P}^*$.

By Fact 1, where we let α be the formula $\neg K_{(\mathbf{p}\sup R)}\neg p$, we know that $\neg K_{(\mathbf{p}\sup R)}\neg p$ implies $\neg K_i K_{(\mathbf{p}\sup R)}\neg p$. Let γ be $K_{(\mathbf{p}\sup R)}\neg p$. By the axioms, we know that $\neg K_i\gamma$ implies $K_i\neg K_i\gamma$. So, $\neg K_{(\mathbf{p}\sup R)}\neg p$ implies $K_i\neg K_i K_{(\mathbf{p}\sup R)}\neg p$. By using Fact 2, we therefore infer that $K_{\mathbf{p}}\neg K_{(\mathbf{p}\sup R)}\neg p$ implies $K_{\mathbf{p}}K_i\neg K_i K_{(\mathbf{p}\sup R)}\neg p$. That is, $K_{\mathbf{p}}\neg K_{(\mathbf{p}\sup R)}\neg p$ implies $K_{\mathbf{p}i}\neg K_{i(\mathbf{p}\sup R)}\neg p$. Since by inductive assumption p implies $K_{\mathbf{p}}\neg K_{(\mathbf{p}\sup R)}\neg p$, it follows that p implies $K_{\mathbf{p}i}\neg K_{i(\mathbf{p}\sup R)}\neg p$. This completes the induction step.

By what we just showed, p implies $K_{\mathbf{s}}\neg K_{(\mathbf{s}\sup R)}\neg p$. So by Fact 2, we know that $K_{\mathbf{r}}p$ implies $K_{\mathbf{r}}K_{\mathbf{s}}\neg K_{(\mathbf{s}\sup R)}\neg p$. But $K_{\mathbf{r}}K_{\mathbf{s}}\neg K_{(\mathbf{s}\sup R)}\neg p$ is just $K_{\mathbf{q}}\psi_{\mathbf{s}}$. Hence, $K_{\mathbf{r}}p$ implies $K_{\mathbf{q}}\psi_{\mathbf{s}}$. Since (a) $w \models K_{\mathbf{r}}p$, (b) the k-ary prefix of w^* is w, and (c) the formula $K_{\mathbf{r}}p$ is of depth $k-1$, it follows that $w^* \models K_{\mathbf{r}}p$. Therefore, from what we just showed, $w^* \models K_{\mathbf{q}}\psi_{\mathbf{s}}$.

Let φ be $\bigvee\{\psi_{\mathbf{s}} : \mathbf{s} \in \mathcal{P}^* \text{ and } \mathbf{s} \text{ is of length } l\}$. Let $\mathbf{q} = \mathbf{rs}$ be an arbitrary member of $\mathcal{P}^*$ of length $l+k-1$, where $\mathbf{r}$ is of length $k-1$ and $\mathbf{s}$ is of length l. Since, as $w^* \models K_{\mathbf{q}}\psi_{\mathbf{s}}$, it follows that $w^* \models K_{\mathbf{q}}\varphi$. Since $\mathbf{q}$ was arbitrary, it follows that $w^* \models E\sup l+k-1\varphi$.

Finally, it remains to show that φ is not valid. Let $\mathbf{f}$ be the knowledge structure $\langle f_0, f_1, \ldots\rangle$ where f_0 is the truth assignment that makes p false, and where $f_j(i) = \{\langle f_0, \ldots, f_{j-1}\rangle\}$ for $j \geq 1$ and for each agent i. Thus, $f_j(i)$ is a singleton set for $j \geq 1$ and for each agent i. It is easy to see that $w \models K_{\mathbf{r}}\neg p$, for each $\mathbf{r} \in \mathcal{P}^*$. Thus $\mathbf{f} \models \neg\varphi$, so φ is not valid. ∎

Recall that we claimed above that there is no nontrivial common knowledge in a no-information extension. We might hope to prove this by showing that if w is a k-world and $w \models E\sup j\varphi$ for some sufficiently large j (say $j > k+1$), then φ is necessarily valid. However, Theorem 4.11 shows that this approach will not work. We must take a more sophisticated approach and consider also the depth of φ, as in the following result, whose proof appears in the appendix.

Theorem 4.12: *Assume that there are at least two agents, φ is a formula of depth r, and w is a k-world. If $w^* \models E\sup r+k\varphi$, then φ is valid.*

Note that the result of Theorem 4.12 is tight, since Theorem 4.11 shows that we cannot replace "$r+k$" in 4.12 by "$r+k-1$". And of course it now follows immediately that there is no nontrivial common knowledge in a no-information extension.

Corollary 4.13: *Assume that there are at least two agents, and w is a world. Then φ is common knowledge in w^* iff φ is valid.*

Corollary 4.13 contrasts with the situation for finite Kripke structures (Theorem 4.9). In fact, the following theorem is a simple consequence of Theorem 4.9 and Corollary 4.13.

Theorem 4.14: *Assume that there are at least two agents and that w^* is a no-information extension. Then there is no state s of a finite Kripke structure M that is equivalent to w^*. That is, if s is a state of a finite Kripke structure M, then there is a formula φ such that $M, s \models \varphi$ but $w^* \models \neg\varphi$.*

Proof: Let ψ be a nonvalid formula which is common knowledge in every state of M. Such a formula ψ exists, by Theorem 4.9. By Corollary 4.13, there is k such that $w^* \models \neg E \sup k\psi$. However, since ψ is common knowledge in every state of M, we know that $M, s \models E \sup k\psi$. So, if we let φ be $E \sup k\psi$, then the theorem follows. ∎

Thus, even to model Example 1.1 requires infinitely many states if we use Kripke structures.

4.3 The least-information extension

When defining the no-information extension of a world w we assumed that agents consider possible every world that is compatible with w. The justification is that no assumption should be made about the underlying mode of knowledge acquisition. In practice, however, agents usually do have information about how knowledge is acquired. Furthermore, this information is often common knowledge. For example, it may be common knowledge that the only way new knowledge is acquired is via message passing, and that communication proceeds in synchronous rounds. In this case the no-information extension is inappropriate, since it does not capture this common knowledge (thus, it is common knowledge that after, say, one round, Alice does not know that Bob knows that Alice knows the primitive proposition p). As another example, if it is common knowledge that each agent stores his information about primitive propositions and about the information he has received from other agents in a database, and it is common knowledge that databases are insecure, then again the no-information extension is inappropriate. For, if Alice has peeked at Bob's database and thereby knows that Bob knows p, then Alice does not consider it possible that Bob knows that Charlie does not know that Bob knows p, since it is common knowledge that Charlie could have peeked at Bob's database also.

We now consider a generalization of the no-information extension; in particular, we construct the *least-information extension*, which is designed to capture the idea of "all you know, given some common knowledge". In order to explain our construction, we first need to investigate the notion of common knowledge a little more.

Definition 4.15: A world w *appears*$_0$ in a world u if w is a prefix of u; w *appears*$_{j+1}$ in u if w appears$_j$ in u or some world $\langle g_0, \ldots, g_k\rangle$ appears$_j$ in u and $w \in g_k(i)$ for some agent i. A world w *appears in world* u if w appears$_j$ in u for some j. A world w *appears in* *knowledge structure* $\mathbf{f}$ if w appears in some prefix of $\mathbf{f}$. Let $worlds_k(\mathbf{f})$ (resp. $worlds_k(w)$) be the set of k-worlds appearing in $\mathbf{f}$ (resp. w).

∎

Lemma 4.16: *Suppose $depth(\varphi) = k$. Then φ is common knowledge in $\mathbf{f}$ iff φ is true in all the worlds in $worlds_{k+1}(\mathbf{f})$.*

Proof: It is easy to show by induction on m that $\mathbf{f} \models E \sup m\varphi$ iff φ is true in all the $(k+1)$-worlds that appear_m in $\mathbf{f}$. ∎

Using the intuition brought out by this lemma, we can now describe our construction of the least-information extension. More precisely, given a set $\mathcal{C}$ of k-worlds (for example, these worlds can be all the k-worlds satisfying a formula that is commonly known to be true) and a world $w \in \mathcal{C}$, we construct the *least-information extension of w with respect to $\mathcal{C}$*. The idea is to build a knowledge structure where *$\mathcal{C}$ is common knowledge*; i.e., the only k-worlds that appear are those in $\mathcal{C}$. The construction is completely analogous to that of the no-information extension, except that everything is relativized to $\mathcal{C}$.

Let $\mathcal{C}$ be a set of k-worlds, let W be a set of $(m-1)$-worlds, and i an agent. Define the *i-extension of W with respect to $\mathcal{C}$* to be the set of m-worlds given by $\{\langle g_0, \ldots, g_{m-1}\rangle : g_{m-1}(i) = W \text{ and } worlds_k(\langle g_0, \ldots, g_{m-1}\rangle) \subseteq \mathcal{C}\}$.

Intuitively, this is the set of m-worlds w such that in w, agent i considers W as the set of possible $(m-1)$-worlds, subject to the restriction that it is common knowledge that the only possible k-worlds are those in $\mathcal{C}$.

Definition 4.17: Let $\mathcal{C}$ be a set of k-worlds, and let $w = \langle f_0, \ldots, f_m\rangle$ be an $(m+1)$-world. The *one-step least-information extension of w with respect to $\mathcal{C}$*, written $(w, \mathcal{C})^+$, is the $(m+2)$-world $\langle f_0, \ldots, f_m, f_{m+1}\rangle$, where $f_{m+1}(i)$ is the i-extension of $f_m(i)$ with respect to $\mathcal{C}$. Thus, $f_{m+1}(i) = \{\langle g_0, \ldots, g_m\rangle : g_m(i) = f_m(i) \text{ and } worlds_k(\langle g_0, \ldots, g_m\rangle) \subseteq \mathcal{C}\}$. ∎

Intuitively, the one-step least-information extension $\langle f_0, \ldots, f_m, f_{m+1}\rangle$ of $\langle f_0, \ldots, f_m\rangle$ describes what each agent knows at depth $m+1$, assuming that "all that each agent i knows" is already described by $f_m(i)$ and the fact that it is common knowledge that $\mathcal{C}$ is the set of possible k-worlds. Thus, $f_{m+1}(i)$ is the set of all m-worlds that are compatible with i's lower-depth knowledge, subject to the constraint that it is common knowledge that $\mathcal{C}$ is the set of possible k-worlds. Note that the one-step no-information extension is a special case of the one-step least-information extension, where we take $\mathcal{C}$ to be the set of *all* k-worlds. As in the case of the one-step no-information extension, it is not *a priori* clear that $(w, \mathcal{C})^+$ is a world. In fact, in general it is not. We shall investigate this issue shortly.

The next proposition is analogous to Proposition 4.4.

Proposition 4.18: *Assume that $\mathcal{C}$ is a set of k-worlds, i is an agent, w is a k-world, and φ is a formula of depth at most $k-1$. Assume that $(w, \mathcal{C})^+$ is a world. The following are equivalent.*

1. $(w, \mathcal{C})^+ \models K_i\varphi$

2. $w \models^{+,\mathcal{C}} K_i\varphi$

3. Let Σ *be the set of all formulas* γ *of depth at most* $k-1$ *of the form* $K_i\psi$ *or* $\neg K_i\psi$ *that are satisfied by* w*. Let* Γ *be the set of all formulas* $E^r\psi$*, where* ψ *is a depth* $(k-1)$ *formula that is satisfied by every member of* $\mathcal{C}$*. Then* $\Sigma \cup \Gamma \models K_i\varphi$*.*

Proof: The proof is very similar to that of Proposition 4.4. We also make use of the fact that if a knowledge structure $\mathbf{f}$ satisfies Γ, then $worlds_k(\mathbf{f}) \subseteq \mathcal{C}$. Details are omitted. ∎

In part (3) of Proposition 4.18, the set Γ of formulas say that every depth $k-1$ formula that is satisfied by every member of $\mathcal{C}$ is common knowledge. In the next section, we shall enrich our language so that "ψ is common knowledge" can be expressed in the language (by $C\psi$). For each ψ as described in Proposition 4.18, we could then, of course, replace the set of formulas $E^r\psi$ by the single formula $C\psi$. So, part (3) of Proposition 4.18 says that knowledge (and lack of knowledge) of agent i, as described by w, along with the fact that it is common knowledge that $\mathcal{C}$ is the set of possible k-worlds, is sufficient to logically imply $K_i\varphi$.

Just as we did with the no-information extension, we define the *least-information extension* by taking one-step least-information extensions repeatedly. Formally, if $w = \langle f_0, \ldots, f_{k-1}\rangle \in \mathcal{C}$, then the *least-information extension* of w with respect to $\mathcal{C}$, written $(w, \mathcal{C})^*$, is the sequence $\langle f_0, \ldots, f_{k-1}, f_k, \ldots\rangle$, where $\langle f_0, \ldots, f_{m+1}\rangle$ is the one-step least-information extension of $\langle f_0, \ldots, f_m\rangle$ with respect to $\mathcal{C}$, for each $m \geq k-1$.

As with one-step extensions, the no-information extension is a special case of the least-information extension, where we take $\mathcal{C}$ to be the set of *all* k-worlds.

As we remarked earlier, $(w, \mathcal{C})^+$ is not necessarily a world, so of course $(w, \mathcal{C})^*$ is not necessarily a knowledge structure. As we shall see, there may not even be a knowledge structure where $\mathcal{C}$ is common knowledge. In order to characterize when least-information extensions exist, we need a few technical definitions.

Definition 4.19: Let $\mathcal{C}$ be a set of k-worlds, and let i be an agent. $\mathcal{C}$ is *i-closed* if either (a) $k = 1$ or (b) $k > 1$ and whenever $\langle f_0, ..., f_{k-1}\rangle \in \mathcal{C}$ and $\langle g_0, ..., g_{k-2}\rangle \in f_{k-1}(i)$ then there is g_{k-1} such that $\langle g_0, ..., g_{k-2}, g_{k-1}\rangle \in \mathcal{C}$ and $g_{k-1}(i) = f_{k-1}(i)$. $\mathcal{C}$ is *closed* if it is i-closed for each agent i. We say w' is *reachable* from w in $\mathcal{C}$ if there is a sequence $w_0, \ldots, w_k$ of worlds in $\mathcal{C}$ such that $w = w_0$, $w' = w_k$, and for all $j < k$, we have $w_j \sim_i w_{j+1}$ for some agent i. In this case we say that w' is *distance k from w*. Let $reach(w, \mathcal{C})$ be the set of worlds reachable from w in $\mathcal{C}$. ∎

The intuition behind closedness is that a set $\mathcal{C}$ of worlds is closed if all the worlds that are considered possible in worlds in $\mathcal{C}$ are themselves in $\mathcal{C}$. Thus, if agent i knows that only worlds in $\mathcal{C}$ are possible, and

1. $\langle f_0, ..., f_{k-1}\rangle \in \mathcal{C}$, and

2. there is no g_{k-1} such that $\langle g_0, ..., g_{k-2}, g_{k-1}\rangle \in \mathcal{C}$ and $g_{k-1}(i) = f_{k-1}(i)$,

then we cannot have $\langle g_0, ..., g_{k-2}\rangle \in f_{k-1}(i)$. The intuition behind reachability is that if w and w' are in $\mathcal{C}$ and $w \sim_i w'$, then w' is possible for agent i from w. Thus, the worlds reachable from w in $\mathcal{C}$ are the worlds that are in some sense considered possible in w.

To further motivate the above notions we present the following proposition.

Proposition 4.20: *If* $\mathbf{f} = \langle f_0, f_1, \ldots\rangle$ *is a knowledge structure, then for all* $k \geq 0$*,*

1. *the set* $worlds_k(\mathbf{f})$ *is closed, and*

2. $worlds_k(\mathbf{f}) = reach(\langle f_0, \ldots, f_{k-1}\rangle, worlds_k(\mathbf{f}))$*.*

Proof: For (1), first note that if a world $\langle g_0, \ldots, g_{k-1}\rangle$ appears in $\mathbf{f}$, then some extension $\langle g_0, \ldots, g_{k-1}, g_k\rangle$ also appears in $\mathbf{f}$. This is proved for appears$_j$ by an easy induction on j, using (K3); we leave details to the reader. Now fix k, and suppose that $\langle g_0, \ldots, g_{k-1}\rangle \in worlds_k(\mathbf{f})$ and $\langle h_0, \ldots, h_{k-2}\rangle \in g_{k-1}(i)$. From the observation above it follows that there is some extension $\langle g_0, \ldots, g_k\rangle$ that appears in $\mathbf{f}$. By (K3) again, there is some h_{k-1} such that $\langle h_0, \ldots, h_{k-2}, h_{k-1}\rangle \in g_k(i)$. By definition, $\langle h_0, \ldots, h_{k-1}\rangle \in worlds_k(\mathbf{f})$ and by (K2), $g_{k-1}(i) = h_{k-1}(i)$. Thus $worlds_k(\mathbf{f})$ is i-closed. Since i was arbitrary, $worlds_k(\mathbf{f})$ is closed.

For (2), we prove by a straightforward induction on j that for all m, if $\langle g_0, \ldots, g_{m-1}\rangle$ appears$_j$ in $\mathbf{f}$, then $\langle g_0, \ldots, g_{m-1}\rangle \in reach(\langle f_0, \ldots, f_{m-1}\rangle, worlds_m(\mathbf{f}))$. If $j = 0$ the result is immediate. For $j \geq 1$, suppose that there is some $\langle h_0, \ldots, h_m\rangle$ that appears$_{j-1}$ in $\mathbf{f}$ such that $\langle g_0, \ldots, g_{m-1}\rangle \in h_m(i)$ for some agent i. By the induction hypothesis, $\langle h_0, \ldots, h_m\rangle \in reach(\langle f_0, \ldots, f_m\rangle, worlds_{m+1}(\mathbf{f}))$. It is easy to see that we must also have $\langle h_0, \ldots, h_{m-1}\rangle \in reach(\langle f_0, \ldots, f_{m-1}\rangle, worlds_m(\mathbf{f}))$. And since $\langle g_0, \ldots, g_{m-1}\rangle \sim_i \langle h_0, \ldots, h_{m-1}\rangle$, it follows that $\langle g_0, \ldots, g_{m-1}\rangle \in reach(\langle f_0, \ldots, f_{m-1}\rangle, worlds_m(\mathbf{f}))$. ∎

The next theorem, which is proven in the appendix, characterizes when $(w, \mathcal{C})^*$ is a knowledge structure.

Theorem 4.21: $(w, \mathcal{C})^*$ *is a knowledge structure iff* $reach(w, \mathcal{C})$ *is a closed set.*

Suppose $\mathcal{C}$ is a set of k-worlds and $w \in \mathcal{C}$. The least-information extension $(w, \mathcal{C})^*$ (provided it exists) describes a knowledge structure where it is common knowledge that the only k-worlds that appear are those in $\mathcal{C}$. But the k-worlds that appear might be a proper subset $\mathcal{C}' \subset \mathcal{C}$. In such a case, it would be common knowledge that the only k-worlds that appear are those in $\mathcal{C}'$. When is it the case that *all* the worlds of $\mathcal{C}$ are considered possible? That is, under what conditions do all the worlds in $\mathcal{C}$ appear in $(w, \mathcal{C})^*$? We get a clue to the answer from Proposition 4.20, from which it follows easily that if $\mathcal{C} = worlds_k(\mathbf{f})$ and w is the k-ary prefix of $\mathbf{f}$, then $reach(w, \mathcal{C})$ is closed and $\mathcal{C} = reach(w, \mathcal{C})$. The next theorem, whose proof appears in the appendix, says that these two conditions on w and $\mathcal{C}$ characterize when $(w, \mathcal{C})^*$ is a knowledge structure where all the worlds of $\mathcal{C}$ appear.

Theorem 4.22: *$(w,\mathcal{C})^*$ is a knowledge structure where all the worlds of $\mathcal{C}$ appear iff $\mathcal{C}$ is closed and $\mathcal{C} = reach(w,\mathcal{C})$.*

We remark that from now on, whenever we write $(w,\mathcal{C})^*$, we will always assume that $\mathcal{C}$ is a set of k-worlds for some k, and $reach(w,\mathcal{C})$ is closed, so that in fact $(w,\mathcal{C})^*$ is a knowledge structure.

We shall shortly give a proposition that justifies the name "least-information extension", just as the analogous Proposition 4.4 justifies the name "no-information extension". We first need a definition analogous to that of $\models^+$. Recall that a world w *must eventually satisfy* φ, written $w \models^+ \varphi$, if for every knowledge structure $\mathbf{f}$ with w as a prefix, $\mathbf{f} \models \varphi$. Let $\mathcal{C}$ be a set of k-worlds. We say that a world w *must eventually satisfy* φ*, if* $\mathcal{C}$ *is common knowledge*, written $w \models^{+,\mathcal{C}} \varphi$, if for every knowledge structure $\mathbf{f}$ with w as a prefix such that $worlds_k(\mathbf{f}) \subseteq \mathcal{C}$, we have $\mathbf{f} \models \varphi$.

There is a natural sense in which we can view $(w,\mathcal{C})^*$ as a "finite model", since it is a finite description of an (infinite) knowledge structure. It is also natural to view a pair (M,s), where M is a finite Kripke structure and s is a state of M, as a "finite model". As we now show, the former class of "finite models" is richer than the latter class: every state in a finite Kripke structure is equivalent to some least-information extension. However, by Theorem 4.14, the converse does not hold, since a no-information extension is a special case of a least-information extension and is not equivalent to any state in a finite Kripke model.

If M is a Kripke structure and s is a state of M, then let $\mathbf{f}_{M,s}$ be the knowledge structure constructed in the proof of Theorem 3.1. (Recall that $\mathbf{f}_{M,s}$ is equivalent to the state s of M, that is, $M,s \models \varphi$ iff $\mathbf{f}_{M,s} \models \varphi$, for every formula φ.)

Theorem 4.23: *If M is a finite Kripke structure and s is a state of M, then $f_{M,s}$ is a least-information extension; i.e., there exists a set $\mathcal{C}$ of worlds and a world $w \in \mathcal{C}$ such that $f_{M,s} = (w,\mathcal{C})^*$.*

Proof: Let M be the Kripke structure $(S,\pi,\mathcal{K}_1,\ldots,\mathcal{K}_n)$, where S, the set of states, is finite. Throughout this proof, we shall denote $\mathbf{f}_{M,s}$ by $\mathbf{f}_s$, and write $\mathbf{f}_s$ as $\langle s_0, s_1, \ldots\rangle$, for each $s \in S$. Recall from the construction of Theorem 3.1 that $s_k(i) = \{\langle t_0,\ldots,t_{k-1}\rangle : (s,t) \in \mathcal{K}_i\}$ for $k > 0$. Choose N such that if $s,t \in S$ and $\mathbf{f}_s \neq \mathbf{f}_t$, then $s_{N-1} \neq t_{N-1}$. Since S is finite, there is some finite N with this property. In fact, we can simply take N to be the maximum of $N_{s,t}+1$ for all states s,t in S, where $N_{s,t}$ is the least m such that $s_m \neq t_m$ if $\mathbf{f}_s \neq \mathbf{f}_t$, and 0 if $\mathbf{f}_s = \mathbf{f}_t$. Let $\mathcal{C} = \{\langle s_0,\ldots,s_N\rangle : s \in S\}$. It is easy to see from the construction of $\mathbf{f}_s$ that the $(N+1)$-worlds that appear in $\mathbf{f}_s$ are precisely those in $reach(\langle s_0,\ldots,s_N\rangle,\mathcal{C})$, and by Proposition 4.20, these worlds form a closed set. Thus $(\langle s_0,\ldots,s_N\rangle,\mathcal{C})^*$ is a knowledge structure by Theorem 4.21. We now show that for all $s \in S$, we have $\mathbf{f}_s = (\langle s_0,\ldots,s_N\rangle,\mathcal{C})^*$.

In fact, we prove the following claim: Suppose $s \in S$, $N' \geq N$, and $w = \langle s'_0,\ldots,s'_{N'}\rangle$ is such that (a) $worlds_{N+1}(w) \subseteq \mathcal{C}$ and (b) $s'_i = s_i$ for $i \leq N$. Then $s'_i = s_i$ for all $i \leq N'$.

We prove the claim by showing, by induction on m, that $s_m = s'_m$, for $0 \le m \le N'$. For $m \le N$ this follows by assumption. For $m > N$, suppose $\langle t'_0, \ldots, t'_{m-1} \rangle \in s'_m(i)$. Now $\langle t'_0, \ldots, t'_{m-2} \rangle \in s'_{m-1}(i)$ by (K3), and $s'_{m-1}(i) = s_{m-1}(i)$ by the inductive hypothesis. Thus, $\langle t'_0, \ldots, t'_{m-2} \rangle \in s_{m-1}(i)$, so for some t such that $(s,t) \in \mathcal{K}_i$, we have $t'_l = t_l$ for $0 \le l \le m-2$. Moreover, since the prefix $\langle t'_0, \ldots, t'_N \rangle$ must be in $\mathcal{C}$ by assumption (a), it follows that for some $u \in S$, we have $t'_l = u_l$ for $0 \le l \le N$. Now using the induction hypothesis, we have that $t'_l = u_l$ for $0 \le l \le m-1$. Since $t_l = t'_l$ for $0 \le l \le m-2$, it follows that $t_l = u_l$ for $0 \le l \le m-2$, and, since $m-2 \ge N-1$, by choice of N we have $u_{m-1} = t_{m-1}$. Thus, $t'_l = t_l$ for $0 \le l \le m-1$. Since $\langle t_0, \ldots, t_{m-1} \rangle \in s_m(i)$ by definition of $s_m(i)$, we have that $\langle t'_0, \ldots, t'_{m-1} \rangle \in s_m(i)$. Thus, $s'_m(i) \subseteq s_m(i)$. For the converse, suppose that $\langle t_0, \ldots, t_{m-1} \rangle \in s_m(i)$. Thus by (K3), $\langle t_0, \ldots, t_{m-2} \rangle \in s_{m-1}(i)$, so by the inductive hypothesis, $\langle t_0, \ldots, t_{m-2} \rangle \in s'_{m-1}(i)$. By (K3), there exists some t'_{m-1} such that $\langle t_0, \ldots, t_{m-2}, t'_{m-1} \rangle \in s'_m(i)$. Now if $m-1 > N$, it immediately follows from the induction hypothesis that we must have $t'_{m-1} = t_{m-1}$. If $m-1 = N$, then by assumption (a) it follows that $\langle t_0, \ldots, t_{m-2}, t'_{m-1} \rangle \in \mathcal{C}$. By the definition of $\mathcal{C}$, we also have that $\langle t_0, \ldots, t_{m-2}, t_{m-1} \rangle \in \mathcal{C}$. By choice of N, we must have $t'_{m-1} = t_{m-1}$. In either case, we get $s_m(i) \subseteq s'_m(i)$. Thus it follows that $s_m = s'_m$, and we are done with the proof of the claim.

Now taking $(\langle s_0, \ldots, s_N \rangle, \mathcal{C})^* = \langle s'_0, s'_1, \ldots \rangle$, an easy induction on m using this claim shows that $s_k = s'_k$ for all k, so that $\mathbf{f}_s = (\langle s_0, \ldots, s_N \rangle, \mathcal{C})^*$ as desired. ∎

Remark 4.24: The above proof does not give us any information about the length of the worlds in $\mathcal{C}$. We now show that we can bound this length.

We say that a state s is *equivalent to t at the jth stage with respect to agent i*, denoted $s \sim_{i,j} t$, if $s_j(i) = t_j(i)$. It is easy to see that if $s \sim_{i,j} t$, then $s \sim_{i,j-1} t$.

Suppose now that for some $j \ge 0$ we have that $s \sim_{i,j} t$ iff $s \sim_{i,j-1} t$ for all states $s, t \in S$ and for all agents $i \in \mathcal{P}$. If that is the case, then we say that $j-1$ is *stable*. We claim that if $j-1$ is stable, then j is also stable. As we already observed, if $s \sim_{i,j+1} t$ then $s \sim_{i,j} t$, so it remains to prove the other direction.

Assume $s \sim_{i,j} t$. We first prove that $s_{j+1}(i) \subseteq t_{j+1}(i)$. Let $u \in S$ be such that $(s,u) \in \mathcal{K}_i$, so $\langle u_0, \ldots, u_j \rangle \in s_{j+1}(i)$. Since $\langle u_0, \ldots, u_{j-1} \rangle \in s_j(i) = t_j(i)$, there must be some $v \in S$ such that $(t,v) \in \mathcal{K}_i$ and $\langle u_0, \ldots, u_{j-1} \rangle = \langle v_0, \ldots, v_{j-1} \rangle$. By assumption it follows that $\langle u_0, \ldots, u_j \rangle = \langle v_0, \ldots, v_j \rangle$, so $\langle u_0, \ldots, u_j \rangle \in t_{j+1}(i)$, as desired. Analogously we can show that $t_{j+1}(i) \subseteq s_{j+1}(i)$. Thus $s \sim_{i,j+1} t$.

Since the relations $\sim_{i,j}$ are decreasing as a function of j, there is some $j_0 \ge 0$ such that all $j \ge j_0$ are stable. Assume that j_0 is minimal with respect to that property. Thus, if $0 \le j < j_0$, then for some agent i, the equivalence relation $\sim_{i,j}$ strictly refines $\sim_{i,j-1}$. Let m be the number of states in S, and let n be the number of agents. An equivalence relation on S can be strictly refined at most $m-1$ times. Thus, $j_0 \le mn$. That is, if we take N to be $mn+1$, and if $s, t \in S$ and $f_{\mathbf{s}} \ne f_{\mathbf{t}}$, then $s_{N-1} \ne t_{N-1}$. Consequently, $\mathcal{C}$ can be taken to be a set of $(mn+2)$-worlds.

We note that if $s \sim_{i,j_0} t$ for all agents $i \in \mathcal{P}$, then $\mathbf{f}_s = \mathbf{f}_t$. This notion of equivalence between states in Kripke structures is closely related to the notions of equivalence [Ho] and bisimulation [Pa] between states in finite-state automata. ∎

Example 4.25: We now consider a very interesting situation where both no-information extensions and least-information extensions enter the picture. Suppose we have three agents, Alice, Bob, and Charlie, and suppose that there are two primitive propositions, p and q. All the agents observe "reality" (so that they get some information about p and q), but do not communicate, and intuitively have "no information about each other's knowledge". Moreover, it is common knowledge that this is the case. Further, suppose that it is the case that p and q are both true, but Alice just knows that p is true and has no knowledge of q, Bob knows that either p is true or q is true, while Charlie knows that both p and q are true. What knowledge structure $\mathbf{f}$ describes this situation? Clearly, its 2-ary prefix is $\langle f_0, f_1 \rangle$, where $f_0 = pq$, $f_1(Alice) = \{pq, p\bar{q}\}$, $f_1(Bob) = \{pq, p\bar{q}, \bar{p}q\}$, and $f_1(Charlie) = \{pq\}$. Now $f_2(Alice)$ is clearly the Alice-extension of $f_1(Alice)$; Alice considers any 2-world consistent with her own information possible. Similarly, we see that $\langle f_0, f_1, f_2 \rangle$ is the one-step no-information extension of $\langle f_0, f_1 \rangle$. What about f_3? Should we continue taking one-step no-information extensions? The answer is no, since it is common knowledge that "no one has any knowledge about anyone else's knowledge", so it is also common knowledge that the only 2-worlds possible are one-step no-information extensions! Let $\mathcal{C}$ be the set of all 2-worlds that are one-step no-information extensions. Then $\mathbf{f} = (\langle f_0, f_1, f_2 \rangle, \mathcal{C})^*$. ∎

This example can be generalized. Consider a situation where knowledge is acquired by unreliable synchronous communication. Intuitively, before the first round of communication we can paradoxically say that it is common knowledge that "nobody has nontrivial knowledge of depth greater than 1". (Note that if communication is reliable, then common knowledge about reality can be achieved in one round of communication.) Similarly, after r rounds of communication we can say that it is common knowledge that "nobody has nontrivial knowledge of depth greater than $r+1$". Suppose that $\mathbf{f}$ describes the state of knowledge after r rounds of communication, where $\mathbf{f} = \langle f_0, f_1, \ldots \rangle$. Then essentially the same reasoning as that above shows that $f_{r+2}(i)$ is the i-extension of $f_{r+1}(i)$, and if $\mathcal{C}$ is the set of all $(r+3)$-ary one-step no-information extensions, then $\mathbf{f} = (\langle f_0, \ldots, f_{r+2} \rangle, \mathcal{C})^*$. The knowledge structure $\mathbf{f}$ can loosely be described as "the least-information extension of a one-step no-information extension".

It turns out that this situation can also be captured by a finite Kripke model. Let $1, \ldots, n$ be the agents. Let M be the Kripke structure $(S, \pi, \mathcal{K}_1, \ldots, \mathcal{K}_n)$, where S is the set of all k-worlds, where $\pi(\langle f_0, \ldots, f_{k-1} \rangle) = f_0$, and where $\mathcal{K}_i = \{(w, w') : w \sim_i w'\}$ for each agent i. This construction is analogous to that of Theorem 3.1, except there we took S to consist of all knowledge structures, rather than all the k-worlds. Of course, in this case M is a finite Kripke structure, since there are only a finite number of k-worlds. By Theorem 4.23, we know that $\mathbf{f}_{M,s}$ is a least-information extension for every state s

of M. In fact, it turns out that $\mathbf{f}_{M,s} = (\langle s_0, \ldots, s_k \rangle, \mathcal{C})^*$, where $s_k(i)$ is the i-extension of $s_{k-1}(i)$ for each agent i and $\mathcal{C}$ consists of all the $(k+1)$-ary one-step no-information extensions. The details of the proof of this fact are straightforward and left to the reader.

4.4 Summary

We now summarize our results on modelling finite information. First, we introduce the no-information extension w^* of a world w, which intuitively represents the state of knowledge if all of the information of the agents (about reality and about each other) is already given by the world w. We show that the no-information extension is indeed a knowledge structure. In particular, this shows that every world is the prefix of some knowledge structure. Although on the face of it, both a finite Kripke structure (along with a state of the Kripke structure) and a no-information extension can be considered as "finite models", we show that there is an important difference, when there are at least two agents. In every finite Kripke structure, some nonvalid formula is common knowledge in every state. However, in each no-information extension, the only formulas that are common knowledge are valid formulas.

We then define the least-information extension, which is a generalization of the no-information extension that allows certain common knowledge. Let $\mathcal{C}$ be a set of k-ary worlds, and let w be a world in $\mathcal{C}$. Intuitively, the least-information extension $(w, \mathcal{C})^*$ represents the state of knowledge if all of the information of the agents (about reality and about each other) is already given by the world w, subject to the constraint that it is common knowledge that the only possible k-worlds are those in $\mathcal{C}$. Unlike the no-information extension, the least-information extension is not always a knowledge structure. We characterize when the least-information extension $(w, \mathcal{C})^*$ is a knowledge structure. We also characterize when it represents a situation where it is common knowledge that in fact the possible k-worlds are *precisely* those in $\mathcal{C}$ (rather than possibly a proper subset of $\mathcal{C}$). We show that least-information extensions are the most general notion of "finite model" of those we have discussed, in that not only is every no-information extension a least-information extension, but also the knowledge structure that represents the information at a state of a finite Kripke structure is also a least-information extension.

5 Modelling Common and Joint Knowledge

Convention: We use lower-case Roman letters such as i, j, etc. to range over natural numbers, and we use lower-case Greek letters such as δ, λ, etc. to range over ordinals.

5.1 Extended syntax and semantics

In Section 4, we mentioned the important concept of *common knowledge*. Nevertheless, common knowledge was defined as a metalogical concept, and we could not express it directly in our logic. Thus, it is natural to extend our logic and add to it the notion of common knowledge. That is, if φ is a formula, then we would also like $C\varphi$ ("φ is common knowledge") to be a formula, so that we can allow formulas with C "inside". Another important notion that we would like to add to our logic is that of *joint knowledge*. A fact φ is joint knowledge of a group S if "everybody in S knows that everybody in S knows ... φ". Common knowledge is, of course, a special case of joint knowledge, where S is the set of all agents. Joint knowledge is important in situations where some agents are reasoning about the knowledge shared by certain groups of agents (see for example [DM]). Thus, we extend our language by adding a new modality C_S, for each nonempty set S of agents. Let $E_S\varphi$ be an abbreviation for $\bigwedge_{i\in S} K_i\varphi$ (i.e., "everyone in S knows φ"). Furthermore, let $E_S\,\mathrm{sup}\,0\varphi$ denote φ, and let $E_S\,\mathrm{sup}\,i\varphi$ denote $E_S E_S\,\mathrm{sup}\,i-1\varphi$. Then $C_S\varphi$ is intended to mean $\bigwedge_{i>0} E_S\,\mathrm{sup}\,i\varphi$.

We now want to give semantics to the extended language. The semantics defined in Section 2 depended on the notion of *depth* of formulas. Since a joint knowledge formula is intended to be equivalent to an infinite conjunction of formulas of increasing depth, it seems that the depth of a joint knowledge formula should be infinite. This motivates using ordinals to define depth of formulas. More formally, we define the depth of formulas in the extended logic as follows.

1. $\mathrm{depth}(p) = 0$ if p is a primitive proposition;
2. $\mathrm{depth}(\neg\varphi) = \mathrm{depth}(\varphi)$;
3. $\mathrm{depth}(\varphi_1 \wedge \varphi_2) = max\,(\mathrm{depth}(\varphi_1), \mathrm{depth}(\varphi_2))$;
4. $\mathrm{depth}(K_i\varphi) = \mathrm{depth}(\varphi) + 1$.
5. $\mathrm{depth}(C_S\varphi) = min\{\lambda{:}\lambda \geq \mathrm{depth}(\varphi) + i \text{ for all } i \geq 0\}$.

In other words, $\mathrm{depth}(C_S\varphi)$ is the first limit ordinal greater than $\mathrm{depth}(\varphi)$. For example, if p is a primitive proposition then the depth of $K_1K_2\neg C_{\{3,4\}}p$ is $\omega + 2$ and the depth of $C_{\{1,2\}}\neg C_{\{3,4\}}p$ is $\omega \times 2$.

It is easy to verify the following proposition.

Proposition 5.1: *For all extended formulas φ, we have depth$(\varphi) < \omega\,\mathrm{sup}\,2$.*

To define the semantics of formulas of infinite depth we need to define worlds of infinite length, that are indexed by ordinals rather than only by natural numbers. The definition is a natural extension of the definitions in Section 2. Instead of defining k-ary worlds for every *natural number* k, we define λ-ary worlds for every *ordinal* λ. A *0th-order knowledge assignment* f_0 is a truth assignment to the primitive propositions. We call $\langle f_0\rangle$ a 1-*ary world*. Let W_λ be the set of all λ-ary worlds. A *λth-order knowledge assignment* is a function $f_\lambda : \mathcal{P} \to 2\,\mathrm{sup}\,W_\lambda$. A λ-sequence of knowledge assignments is a sequence $\langle f_0, f_1, \ldots\rangle$ of length λ, where f_i is an ith-order knowledge assignment. A *λ-ary world* (or λ-world, for short) $\mathbf{f}$ is a λ-sequence of knowledge assignments satisfying certain restrictions. For example, an $(\omega+1)$-world is of the form $\langle f_0, f_1, ..., f_\omega\rangle$, where $f_\omega(i)$ is a set of ω-worlds and certain other restrictions are satisfied. If $\kappa \leq \lambda$, then the *κ-prefix* of $\mathbf{f}$, denoted $\mathbf{f}_{<\kappa}$, is the κ-sequence that is the restriction of $\mathbf{f}$ to κ.

We now describe the restrictions that a $(\lambda+1)$-world $\mathbf{f}$ has to satisfy for each agent i.

(K1′) $\mathbf{f}_{<\lambda} \in f_\lambda(i)$.

(K2′) If $\mathbf{g} \in f_\lambda(i)$, and $\lambda > 1$, then $g_\kappa(i) = f_\kappa(i)$ for all $\kappa < \lambda$.

(K3′) Let $0 < \kappa < \lambda$. Then $\mathbf{g} \in f_\kappa(i)$ iff there is some $\mathbf{h} \in f_\lambda(i)$ such that $\mathbf{g} = \mathbf{h}_{<\kappa}$.

We note that it follows from (K1′) and (K2′) that if $\mathbf{h} \in f_\lambda(i)$ and $0 < \kappa < \lambda$, then $\mathbf{h}_{<\kappa} \in f_\kappa(i)$. Thus, only the other direction of (K3′) is nontrivial. We also note that it follows from (K1′) that if $\mathbf{f}$ is a λ-world, then $\mathbf{f}_{<\kappa}$ is a κ-world for all $\kappa \leq \lambda$.

Clearly, (K1′)-(K3′) generalize restrictions (K1)-(K3). It is easy to see that knowledge structures are simply ω-worlds.

We can now define what it means for a λ-world $\mathbf{f}$ to satisfy a formula φ of depth κ, written $\mathbf{f} \models \varphi$. We first define a binary relation $\preceq$ on ordinals. We say that $\kappa \preceq \lambda$ (or $\lambda \succeq \kappa$) if either κ is a successor ordinal and $\kappa < \lambda$, or κ is a limit ordinal and $\kappa \geq \lambda$ (in other words, whether κ is a successor ordinal or a limit ordinal, $\lambda > \mu + 1$ for all $\mu < \kappa$). The relation $\models$ is defined between $\mathbf{f}$ and φ iff $\kappa \preceq \lambda$.

1. $\mathbf{f} \models p$, where p is a primitive proposition, if p is true under the truth assignment f_0.
2. $\mathbf{f} \models \neg\psi$ if $\mathbf{f} \not\models \psi$.
3. $\mathbf{f} \models \varphi_1 \wedge \varphi_2$ if $\mathbf{f} \models \varphi_1$ and $\mathbf{f} \models \varphi_2$.
4. If λ is a successor ordinal, then $\mathbf{f} \models K_i\psi$ if $\mathbf{g} \models \psi$ for each $\mathbf{g} \in f_{\lambda-1}(i)$.
5. If λ is a limit ordinal, then $\mathbf{f} \models K_i\psi$ if $\mathbf{f}_{<\kappa+1} \models K_i\psi$.
6. $\mathbf{f} \models C_S\psi$ if $\mathbf{f} \models E_S \sup i\psi$ for all $i > 0$.

It is easy to see that the definitions of Section 2 are a special case of the definitions here. In Section 2, however, we defined satisfaction with respect to structures as a total relation, where here we leave satisfaction as a partial relation between worlds and formulas.

The following lemma, which is analogous to Lemma 2.5, indicates the robustness of the definitions above.

Lemma 5.2: *Let $\mathbf{f}$ be a λ-world, and let φ be a formula such that depth(φ) $= \kappa$ and $\lambda > \kappa$. Then $\mathbf{f} \models \varphi$ iff $\mathbf{f}_{<\kappa+1} \models \varphi$. Furthermore, if κ is a limit ordinal, then $\mathbf{f} \models \varphi$ iff $\mathbf{f}_{<\kappa} \models \varphi$.*

Proof: The proof is by simultaneous induction on formulas and worlds. The nontrivial cases in the induction on formulas are when φ is of the form $K_i\psi$ or $C_S\psi$, where we assume inductively that the lemma holds for ψ.

Consider first the case that φ is of the form $K_i\psi$. Assume that λ is a successor ordinal. Suppose that $\mathbf{f} \models K_i\psi$. Let $\mathbf{g}$ be an arbitrary member of $f_\kappa(i)$. It follows from (K3′) that $\mathbf{g} = \mathbf{h}_{<\kappa}$ for some $\mathbf{h} \in f_{\lambda-1}(i)$. Since $\mathbf{f} \models K_i\psi$, it follows by definition that $\mathbf{h} \models \psi$. By inductive assumption, $\mathbf{g} \models \psi$. Thus, every member of $f_\kappa(i)$ satisfies ψ, and so $\mathbf{f}_{<\kappa+1} \models \varphi$. The proof of the converse is similar.

Assume now that λ is a limit ordinal. Then $\mathbf{f} \models K_i\psi$ iff $\mathbf{f}_{<\kappa+1} \models K_i\psi$, and the claim for this case holds by the induction hypothesis.

Consider now the case that φ is of the form $C_S\psi$. Suppose that $\mathbf{f} \models C_S\psi$. Then $\mathbf{f} \models E_S \sup i\psi$ for all $i \geq 0$. Let $\mu = \mathrm{depth}(\psi)$. By the induction hypothesis, $\mathbf{f}_{<\mu+i+1} \models E_S \sup i\psi$ for all $i \geq 0$. Again by the induction hypothesis, $\mathbf{f}_{<\nu} \models E_S \sup i\psi$ for all $i \geq 0$, whenever $\nu \geq \mu + i$ for all $i \geq 0$. In particular, $\mathbf{f}_{<\kappa} \models \varphi$. Again, the proof of the converse is similar. ∎

We now describe some axioms for joint knowledge. These are generalization of axioms for common knowledge due to Lehmann [Le] and Milgrom [Mi].

1. $C_S\varphi \Rightarrow \varphi$
2. $C_S\varphi \Rightarrow C_SC_S\varphi$
3. $\neg C_S\varphi \Rightarrow C_S\neg C_S\varphi$
4. $C_S\varphi_1 \wedge C_S(\varphi_1 \Rightarrow \varphi_2) \Rightarrow C_S\varphi_2$
5. $C_{\{i\}}\varphi \equiv K_i\varphi$
6. $C_S\varphi \Rightarrow C_T\varphi$ if $T \subseteq S$
7. $C_S(\varphi \Rightarrow E_S\varphi) \Rightarrow (\varphi \Rightarrow C_S\varphi)$.

Axioms 1-4 are analogous to axioms 2-5 for knowledge. They say that joint knowledge is correct, introspective, and closed under implication. Axiom 5 deals with the degenerate case of a single agent. Axiom 6 says that joint knowledge is inherited by subsets, and Axiom 7 describes how joint knowledge is built as a fixpoint of knowledge.

We want to show that the axioms are valid. Since satisfaction is a partial relation between worlds and formulas, we have to redefine the notion of *validity*. We say that a formula φ is valid if it is satisfied by every world for which the satisfaction relation is defined with respect to that formula. An axiom is valid if all its instances are valid.

Proposition 5.3: *All the axioms above are valid.*

Proof: We demonstrate the validity of Axiom 7. The other axioms are left to the reader. Let $\mathbf{f}$ be a λ-world, and let φ be a formula of depth κ. Assume that $\mathbf{f} \models C_S(\varphi \Rightarrow E_S\varphi)$ and $\mathbf{f} \models \varphi$. By definition we have that $\mathbf{f} \models E_S \sup i(\varphi \Rightarrow E_S\varphi)$, for all $i > 0$. By the knowledge axioms, it follows that $\mathbf{f} \models E_S \sup i\varphi \Rightarrow E_S \sup i + 1\varphi$, for all $i > 0$. We now show, by induction on i, that $\mathbf{f} \models E_S \sup i\varphi$, for all $i > 0$. For $i = 1$, we have that $\mathbf{f} \models E_S\varphi$, since $\mathbf{f} \models \varphi$ and $\mathbf{f} \models C_S(\varphi \Rightarrow E_S\varphi)$. Assume now that $\mathbf{f} \models E_S \sup i\varphi$. Since $\mathbf{f} \models E_S \sup i \Rightarrow E_S \sup i + 1\varphi$, it follows that $\mathbf{f} \models E_S \sup i + 1\varphi$. Thus, $\mathbf{f} \models C_S\varphi$. ∎

In [FV1] it is shown that the above axiomatization for knowledge and joint knowledge together with modus ponens and joint knowledge generalization ("from φ infer $C_S\varphi$") is indeed complete. Another axiomatization is given in [HM2].

5.2 Model-theoretic constructions

We now extend the machinery developed in the previous sections. Our goal is to prove the equivalence of "internal" truth and "external" truth, in analogy with Theorem 2.7 This will then enable us to relate knowledge worlds and Kripke structures as in Theorem 3.1.

Let $\mathbf{f}$ and $\mathbf{g}$ be λ-worlds. We say that $\mathbf{f}$ and $\mathbf{g}$ are *i-equivalent*, written $\mathbf{f} \sim_i \mathbf{g}$, if $f_\kappa(i) = g_\kappa(i)$ for all κ such that $0 \leq \kappa < \lambda$. We call $\{\mathbf{g} : \mathbf{f} \sim_i \mathbf{g}\}$ the *i-equivalence class of* $\mathbf{f}$.

We now generalize the no-information extension.

Definition 5.4: Let $\mathbf{f}$ be a λ-world. Let $\mu \geq \lambda$. The *μ-no-information extension* of $\mathbf{f}$, denoted $\mathbf{f} \sup \mu$, is a μ-sequence of knowledge assignments defined as follows:

1. $\mathbf{f} \sup \lambda = \mathbf{f}$.
2. If $\mu > \lambda$ is a successor ordinal, then $\mathbf{f}_{<\mu-1} \sup \mu = \mathbf{f} \sup \mu - 1$ and $f_{\mu-1} \sup \mu(i)$ is the i-equivalence class of $\mathbf{f} \sup \mu - 1$ for all agents i.
3. If μ is a limit ordinal, then for each $\nu < \mu$, we have $\mathbf{f}_\nu \sup \mu = \mathbf{f}_\nu \sup \nu + 1$. ∎

It is easy to see that when $\lambda < \omega$ (so that $\mathbf{f}$ is just a world of finite length), then the ω-no-information extension of $\mathbf{f}$ is the same as the no-information extension as we defined it in Section 4.

We need to prove that the no-information extension yields knowledge worlds. We first need the analogue of Lemma 4.2.

Lemma 5.5: *Let $\mathbf{f}$ and $\mathbf{g}$ be λ-world for some successor ordinal λ such that $\mathbf{f}_{<\lambda-1} \sim_i \mathbf{g}_{<\lambda-1}$. Let h_λ be a λth-order knowledge assignment such that $h_\lambda(i) = f_\lambda(i)$ and $h_\lambda(j) = g_\lambda(j)$ for $j \neq i$. Then $\langle g_0, \ldots, g_{\lambda-1}, h_\lambda \rangle$ is a $(\lambda + 1)$-world.*

We also need a generalization of the matching extension. Let $\mathbf{f}$ be a λ-world, where λ is a successor ordinal. Let $\mathbf{g} \in f_{\lambda-1}(i)$. The *$i-matching\,extension$* of $\mathbf{g}$ with respect to $\mathbf{f}$ is the λ-sequence $\mathbf{h}$ of knowledge assignments, where $\mathbf{h}_{<\lambda-1} = \mathbf{g}$, $h_{\lambda-1}(i) = f_{\lambda-1}(i)$, and $h_{\lambda-1}(j)$ is the j-equivalence class of $\mathbf{g}$ for $j \neq i$. We now prove the analogue of Lemma 4.3.

Theorem 5.6: *Let $\mathbf{f}$ be a λ-world.*

1. *Let $\mu \geq \lambda$. Then $\mathbf{f}\sup\mu$ is a μ-world.*
2. *If λ is a successor ordinal, and $\mathbf{g} \in f_{\lambda-1}(i)$, then the i-matching extension of $\mathbf{g}$ with respect to $\mathbf{f}$ is a λ-world.*

Proof: The proof is by induction on λ.

To prove (1), it suffices to prove that if $\mathbf{f}$ is a λ-world, then $\mathbf{f}\sup\lambda+1$ satisfies the restrictions (K1′), (K2′), and (K3′). The fact that (K1′) and (K2′) hold is immediate from the definition. We prove that (K3′) holds by induction on λ.

If $\lambda = 1$, then the claim holds vacuously. For the inductive step, suppose that $0 < \kappa < \lambda$ and $\mathbf{g} \in f_\kappa(i)$. We will construct a λ-world $\mathbf{h}$ such that $\mathbf{g} = \mathbf{h}_{<\kappa}$ and $\mathbf{h} \sim_i \mathbf{f}$. Thus, $\mathbf{h} \in f_\lambda(i)$. We inductively describe $\mathbf{h}_{<\mu}$ for $\kappa \leq \mu \leq \lambda$, where $\mathbf{h}_{<\mu} \sim_i \mathbf{f}_{<\mu}$.

For the basis of the induction, we take $\mathbf{h}_{<\kappa}$ to be $\mathbf{g}$. Assume now that $\mu \leq \lambda$ and $\mathbf{h}_{<\nu}$ has been defined for all $\nu < \mu$. If μ is a limit ordinal, then $\mathbf{h}_{<\mu}$ is already defined, so suppose that μ is a successor ordinal. Then we have $\mathbf{h}_{<\mu-1} \sim_i \mathbf{f}_{<\mu-1}$. Let $\mathbf{h}_{<\mu}$ be the i-matching extension of $\mathbf{h}_{<\mu-1}$ with respect to $\mathbf{f}_{<\mu}$. By the induction hypothesis, $\mathbf{h}_{<\mu}$ is a μ-world, and clearly $\mathbf{h}_{<\mu} \sim_i \mathbf{f}_{<\mu}$. This completes the proof that (K3′) holds.

To prove part (2) we consider first the λ-sequence $\mathbf{g}\sup\lambda$, which by the induction hypothesis is a λ-world. Since $\mathbf{g} \in f_{\lambda-1}(i)$, we have that $\mathbf{g} \sim_i \mathbf{f}_{<\lambda-1}$. The claim now follows by Lemma 5.5. ∎

A consequence of the theorem is that every λ-world can be extended to a μ-world, for any $\mu > \lambda$. This generalizes the result in Section 4 that every world can be extended to a structure. Furthermore, while proving the theorem we also proved another useful result.

Lemma 5.7: *Let $\mathbf{f}$ be a λ-world, and let $\mathbf{g} \in f_\kappa(i)$, where $\kappa < \lambda$. Then there is some λ-world $\mathbf{e}$ such that $\mathbf{g} = \mathbf{e}_{<\kappa}$ and $\mathbf{e} \sim_i \mathbf{f}$.*

Let S be a set of agents. We say that a λ-world $\mathbf{g}$ is *S-reachable* from a λ-world $\mathbf{f}$ if there is a sequence $\mathbf{f}_1, \ldots, \mathbf{f}_k$ such that $\mathbf{f} = \mathbf{f}_1$, $\mathbf{g} = \mathbf{f}_k$, and for each j such that $1 \leq j \leq k-1$ there is some agent $i \in S$ such that $\mathbf{f}_j \sim_i \mathbf{f}_{j+1}$. In this case we say that $\mathbf{g}$ is *S-distance k* from $\mathbf{g}$.

We can now prove the analogue of Theorem 2.7, which shows the equivalence of internal and external notions of truth.

Theorem 5.8: *Let* $\mathbf{f}$ *be a* λ*-world.*

1. $\mathbf{f} \models K_i\varphi$ *iff* $\mathbf{g} \models \varphi$ *whenever* $\mathbf{f} \sim_i \mathbf{g}$.
2. $\mathbf{f} \models C_S\varphi$ *iff* $\mathbf{g} \models \varphi$ *whenever* $\mathbf{g}$ *is* S*-reachable from* $\mathbf{f}$.

Proof:

1. Suppose first that λ is a successor ordinal.

   (a) Assume that $\mathbf{f} \models K_i\varphi$. By definition, $\mathbf{h} \models \varphi$ whenever $\mathbf{h} \in f_{\lambda-1}(i)$. Let $\mathbf{g}$ be such that $\mathbf{f} \sim_i \mathbf{g}$. By (K1′), $\mathbf{g}_{<\lambda-1} \in g_{\lambda-1}(i)$. But $g_{\lambda-1}(i) = f_{\lambda-1}(i)$. It follows that $\mathbf{g}_{<\lambda-1} \in f_{\lambda-1}(i)$, so $\mathbf{g}_{<\lambda-1} \models \varphi$. By Lemma 5.2, we have $\mathbf{g} \models \varphi$.

   (b) Assume that $\mathbf{g} \models \varphi$ whenever $\mathbf{f} \sim_i \mathbf{g}$. Let $\mathbf{h} \in f_{\lambda-1}(i)$. By Lemma 5.7, there is a λ-world $\mathbf{g}$ such that $\mathbf{h} = \mathbf{g}_{<\lambda-1}$ and $\mathbf{g} \sim_i \mathbf{f}$. Thus, $\mathbf{g} \models \varphi$. By Lemma 5.2, it follows that $\mathbf{h} \models \varphi$.

   Suppose now that λ is a limit ordinal, and assume that the claim has been proven for all smaller ordinals. Let $\kappa = \text{depth}(\varphi)$.

   (a) Assume that $\mathbf{f} \models K_i\varphi$. Then, by Lemma 5.2, $\mathbf{f}_{<\kappa+2} \models K_i\varphi$. Since $\kappa + 2 < \lambda$, we have that $\mathbf{h} \models \varphi$ for every $(\kappa+2)$-world $\mathbf{h} \sim_i \mathbf{f}_{<\kappa+2}$. Now assume $\mathbf{g} \sim_i \mathbf{f}$. It follows that $\mathbf{g}_{<\kappa+2} \sim_i \mathbf{f}_{<\kappa+2}$. Thus, by Lemma 5.2, it follows that $\mathbf{g} \models \varphi$.

   (b) Assume that $\mathbf{g} \models \varphi$ whenever $\mathbf{g} \sim_i \mathbf{f}$. Let $\mathbf{h} \in f_{\kappa+1}(i)$. By (K2′), $\mathbf{h} \sim_i \mathbf{f}_{<\kappa+1}$. By Lemma 5.7, there is a λ-world $\mathbf{g}$ such that $\mathbf{g} \sim_i \mathbf{f}$ and $\mathbf{h} = \mathbf{g}_{<\kappa+1}$. By assumption, $\mathbf{g} \models \varphi$, so $\mathbf{h} \models \varphi$, by Lemma 5.2. We have shown that $\mathbf{h} \models \varphi$ whenever $\mathbf{h} \in f_{\kappa+1}(i)$. It follows that $\mathbf{f}_{<\kappa+2} \models K_i\varphi$, and by Lemma 5.2, we have that $\mathbf{f} \models K_i\varphi$.

2. It is easy to prove, by induction on i, that $\mathbf{f} \models E_S \sup i\varphi$ iff $\mathbf{g} \models \varphi$ whenever $\mathbf{g}$ is S-distance i from $\mathbf{f}$. The claim follows, since $\mathbf{f} \models C_S\varphi$ iff $\mathbf{f} \models E_S \sup i\varphi$ for all $i \geq 0$. ∎

The next result is analogous to Theorem 3.1. Here we consider a state s of Kripke structure M to be *equivalent* to an ω^2-world if they satisfy the same extended formulas.

Corollary 5.9: *To every Kripke structure* M *and state* s *in* M*, there corresponds an* $\omega \sup 2$*-world* $\mathbf{f}_{M,s}$ *such that* s *is equivalent to* $\mathbf{f}_{M,s}$*. Conversely, there is a Kripke structure* M_{know} *such that for every* $\omega \sup 2$*-world* $\mathbf{f}$ *there is a state* $s_{\mathbf{f}}$ *in* M_{know} *such that* $\mathbf{f}$ *is equivalent to* $s_{\mathbf{f}}$.

Proof: The proof is analogous to the proof of Theorem 3.1. There are two differences. First, the $\omega \sup 2$-world $\mathbf{f}_{M,s}$ is constructed by transfinite induction. Second, Theorem 5.8 is used instead of Theorem 2.7. ∎

5.3 Was that necessary?

So far we have claimed that it is necessary to define infinitary worlds in order to give semantics to extended formulas. But is that really the case? We are now going to show some evidence to the contrary.

Let *KC-formulas* be formulas that use the modalities K_i and C (i.e., $C_{\mathcal{P}}$, where $\mathcal{P}$ is the set of all agents), but not C_S when S is a proper subset of $\mathcal{P}$. Let K-formulas be formulas that use only the K_i modalities.

The next theorem says that for KC-formulas, the extension beyond ω is redundant.

Theorem 5.10: *Let σ be a KC-formula. Let $\mathbf{f}$ and $\mathbf{g}$ be λ-worlds, such that $\mathbf{f}_{<\omega} = \mathbf{g}_{<\omega}$. Then $\mathbf{f} \models \sigma$ iff $\mathbf{g} \models \sigma$.*

Proof: See appendix. ∎

The above theorem indicates how we can define the semantics of arbitrary KC-formulas in ω-worlds. We denote this new satisfaction relation by the symbol $\Vdash$.

1. $\mathbf{f} \Vdash p$, where p is a primitive proposition, if p is true under the truth assignment f_0.
2. $\mathbf{f} \Vdash \neg\psi$ if it is not the case that $\mathbf{f} \Vdash \psi$.
3. $\mathbf{f} \Vdash \varphi_1 \wedge \varphi_2$ if $\mathbf{f} \Vdash \varphi_1$ and $\mathbf{f} \Vdash \varphi_2$.
4. $\mathbf{f} \Vdash K_i\psi$ if $\mathbf{g} \Vdash \psi$ whenever $\mathbf{g} \sim_i \mathbf{f}$.
5. $\mathbf{f} \Vdash C\psi$ if $\mathbf{f} \Vdash E \sup i\psi$ for all $i > 0$.

Theorem 5.11: *Let σ be a KC-formula of depth κ, and let $\mathbf{f}$ be a λ-world, $\kappa \preceq \lambda$. Then $\mathbf{f} \models \sigma$ iff $\mathbf{f}_{<\omega} \Vdash \sigma$.*

Proof: We prove the claim by transfinite induction on the depth of σ and an induction on the Boolean structure of σ. The nontrivial cases are where σ is either of the form $K_i\varphi$ or of the form $C\varphi$.

Consider the first case. Suppose that $\mathbf{f} \models K_i\varphi$. Let $\mathbf{h}$ be an ω-world such that $\mathbf{h} \sim_i \mathbf{f}_{<\omega}$. Consider $\mathbf{h} \sup \lambda$. By Theorem 5.10, we have $\mathbf{h} \sup \lambda \models K_i\varphi$, so we also have $\mathbf{h} \sup \lambda \models \varphi$, and by the induction hypothesis, $\mathbf{h} \Vdash \varphi$. Since $\mathbf{h}$ is an arbitrary ω-world such that $\mathbf{h} \sim_i \mathbf{f}_{<\omega}$, it follows that $\mathbf{f}_{<\omega} \Vdash K_i\varphi$.

Suppose that $\mathbf{f}_{<\omega} \Vdash K_i\varphi$. Let $\mathbf{g}$ be a λ-world such that $\mathbf{g} \sim_i \mathbf{f}$. So $\mathbf{g}_{<\omega} \Vdash \varphi$, by definition. By the induction hypothesis, $\mathbf{g} \models \varphi$. Since $\mathbf{g}$ is an arbitrary λ-world such that $\mathbf{g} \sim_i \mathbf{f}$, it follows, by Theorem 5.8, that $\mathbf{f} \models K_i\varphi$.

Consider the second case. Then $\mathbf{f} \models C\varphi$ iff $\mathbf{f} \models E \sup j\varphi$ for all $j \geq 0$ iff (by the induction hypothesis) $\mathbf{f}_{<\omega} \Vdash E \sup j\varphi$ for all $j \geq 0$ iff $\mathbf{f}_{<\omega} \Vdash C\varphi$. ∎

According to the above theorem $\models$ and $\Vdash$ are consistent with each other, so we need not distinguish between them. Note, however, that $\Vdash$ may be defined where $\models$ is undefined.

As a consequence of the above theorem we show that when dealing with satisfiability of KC-formulas, it is sufficient to consider least-information extensions.

Theorem 5.12: *Let φ be a satisfiable KC-formula. Then there is an ω-world $\mathbf{f}$ such that $\mathbf{f} \Vdash \varphi$ and $\mathbf{f}$ is a least-information extension.*

Proof: By [HM2], if φ is satisfiable, then there is a finite Kripke structure M and a state s in M such that $M, s \models \varphi$. By Corollary 5.9, $\mathbf{f}_{M,s} \models \varphi$. Let $\mathbf{g} = \mathbf{f}_{M,s}$. By Theorem 5.11, $\mathbf{g}_{<\omega} \Vdash \varphi$. But by Theorem 4.23, $\mathbf{g}_{<\omega}$ is a least-information extension. ∎

Since a least-information extension has a finite description, this theorem can be viewed as a "finite model" theorem: if a KC-formula is satisfiable, then it is satisfiable in a "finite" model.

We can now use KC-formulas to give another demonstration of the subtlety of Corollary 4.13. Corollary 4.13 says that if φ is a K-formula that is common knowledge in a no-information extension, then φ is valid. The theorem is false if φ is allowed to be a KC-formula. For example, if p is a primitive proposition, then we can show that the KC-formula $\neg Cp$, which is not valid, is common knowledge in every no-information extension. As a side remark, we note that this formula $\neg Cp$ can be viewed as an abbreviation for the infinite disjunction $\neg Ep \vee \neg EEp \vee \neg EEEp \vee \ldots$. It is interesting to note that although this infinite disjunction is common knowledge in every no-information extension, no finite part of it is common knowledge in any no-information extension.

The above theorems show that it is sufficient to consider ω-worlds when dealing with KC-formulas. The question then arises whether there is indeed a need to define "longer" worlds.

Example 5.13: Consider the following situation. Agents 1 and 2 are communicating about a fact p through an unreliable channel, one over which messages are not guaranteed to arrive. As shown in [HM1], under such conditions, arbitrarily deep knowledge is attainable, but common knowledge is not. More precisely, $E_{\{1,2\}} \sup kp$ is attainable for all $k \geq 1$, but $C_{\{1,2\}}p$ is not attainable. If agent 3 does not know how many rounds of successful communication have transpired, then $K_3 \neg C_{\{1,2\}}p$ holds and $\neg K_3 \neg E_{\{1,2\}} \sup kp$ holds for all $k \geq 1$.

We claim that we need an $(\omega+1)$-world to model this situation. That is, we claim that there is no ω-world $\mathbf{f}$ where $K_3 \neg C_{\{1,2\}}p$ holds and $\neg K_3 \neg E_{\{1,2\}} \sup kp$ holds for all $k \geq 1$. To make sense of what it means for a formula such as $K_3 \neg C_{\{1,2\}}p$ to hold in $\mathbf{f}$, we extend our definition of $\Vdash$ in the natural way to apply to such formulas. Thus, $\mathbf{f} \Vdash K_3 \neg C_{\{1,2\}}p$ iff for every ω-world $\mathbf{g}$ such that $\mathbf{f} \sim_3 \mathbf{g}$, necessarily $\mathbf{g} \Vdash \neg C_{\{1,2\}}p$ (which, because $\neg C_{\{1,2\}}p$ is of depth ω, means that $\mathbf{g} \models \neg C_{\{1,2\}}p$).

Assume now that $\mathbf{f} = \langle f_0, f_1, f_2, \ldots\rangle$ is an ω-world and $\mathbf{f} \models \neg K_3\neg E_{\{1,2\}} \sup kp$ for all $k \geq 1$; to prove our claim we must show that $\mathbf{f} \Vdash \neg K_3 \neg C_{\{1,2\}}p$.

We can assume without loss of generality that p is the only primitive proposition (otherwise, we can "restrict" $\mathbf{f}$ by "erasing" all of the primitive propositions other than p; the straightforward details are left to the reader). Let T be a tree with levels $0, 1, 2, \ldots$, where the kth level of the tree contains all of the members $\langle g_0, \ldots, g_k\rangle$ of $f_{k+1}(3)$ that satisfy $E_{\{1,2\}} \sup kp$, and where the parent of the $(k+1)$-ary world $\langle g_0, ..., g_k\rangle$ is its k-ary prefix $\langle g_0, ..., g_{k-1}\rangle$ if $k \geq 1$. Thus, the kth level contains all of the $(k+1)$-ary worlds that agent 3 consider possible and that satisfy $E_{\{1,2\}} \sup kp$. (We see that T is a tree rather than a forest, since there is only one member at level 0, namely $\langle g_0\rangle$, where g_0 is the truth assignment where p is true.) For each k, there is some world $\langle g_0, \ldots, g_k\rangle$ at level k, since $\mathbf{f} \models \neg K_3\neg E_{\{1,2\}} \sup kp$. Since $E_{\{1,2\}} \sup kp \Rightarrow E_{\{1,2\}} \sup rp$ for each $r \leq k$, it follows that if $\langle g_0, \ldots, g_k\rangle$ is in T, then so are all of its prefixes. Thus, there are arbitrarily long finite paths in the tree. The tree has finite fanout, since there are only a finite number of possible k-worlds for each k. By König's Infinity Lemma, T contains an infinite path. This infinite path corresponds to a knowledge structure $\mathbf{g} = \langle g_0, g_1, g_2, \ldots\rangle$. Since $\langle g_0, \ldots, g_k\rangle \in f_{k+1}(3)$, it follows by restriction (K2) on knowledge structures that $g_k(3) = f_k(3)$, for every k. So $\mathbf{f} \sim_3 \mathbf{g}$. Also $\mathbf{g} \models C_{\{1,2\}}p$, since $\mathbf{g} \models E_{\{1,2\}} \sup kp$ for each k. Thus, $\mathbf{f} \Vdash \neg K_3\neg C_{\{1,2\}}p$, as desired.

In our example we would like to be able to model a situation where the facts $\neg K_3\neg E^k_{\{1,2\}}p$, $k \geq 1$, and $K_3\neg C_{\{1,2\}}p$ all hold. It is easy to imagine another situation where the facts $\neg K_3\neg E^k_{\{1,2\}}p$, $k \geq 1$, and $\neg K_3\neg C_{\{1,2\}}p$ all hold. The crucial point is that the facts $\neg K_3\neg E^k_{\{1,2\}}p$, $k \geq 1$, should not determine whether $K_3\neg C_{\{1,2\}}p$ holds. If we restrict attention to ω-worlds then, as we showed, this independence fails. In an ω-world where the facts $\neg K_3\neg E_{\{1,2\}}{}^k p$, $k \geq 1$, are all true, the fact $K_3\neg C_{\{1,2\}}p$ is forced to be true as well. This can be explained as follows. It is straightforward to show that under our $\Vdash$ semantics, agent 3 considers an ω-world $\mathbf{g} = \langle g_0, g_1, g_2, \ldots\rangle$ possible precisely if agent 3 considers the k-ary prefix $\langle g_0, ..., g_{k-1}\rangle$ possible for every $k \geq 1$ (that is, $\langle g_0, ..., g_{k-1}\rangle \in f_k(3)$). Under the assumption that $\neg K_3\neg E^k_{\{1,2\}}p$ holds for every $k \geq 1$, our arguments above actually show that there is an ω-world $\mathbf{g}$ that satisfies $C_{\{1,2\}}p$ such that agent 3 considers every finite prefix of $\mathbf{g}$ possible. So under the $\Vdash$ semantics, agent 3 is forced to consider $\mathbf{g}$ possible. The whole point of having an ωth-order knowledge assignment f_ω is to be able to model the fact that agent 3 considers $\mathbf{g}$ impossible (via $\mathbf{g} \notin f_\omega(3)$) even though agent 3 considers every finite prefix of $\mathbf{g}$ possible. ∎

The above example suggests that our transfinite construction is indeed necessary. Intuitively, "long worlds" are needed to model "deep" knowledge. This intuition, however, needs to be sharpened, since KC-formulas can express deep knowledge but they do not require long models. The answer is that KC-formulas do not really express knowledge of depth greater than ω, since they are always equivalent to formulas of depth ω. On the other hand, a formula such as $K_3\neg C_{\{1,2\}}p$ is inherently of depth $\omega + 1$. The next theorem says that the situation in general is unlike that with KC-formulas φ, where we could decide the truth of φ by considering only the prefix of length ω. Specifically, the theorem says that there is no $\lambda < \omega$ such that for every formula φ, we can decide the truth of φ in a world w by considering only the prefix of w of length λ.

Theorem 5.14: *For every ordinal $1 \leq \lambda < \omega \sup 2$ there is a formula σ_λ and there are $(\lambda+1)$-worlds $\mathbf{f}$ and $\mathbf{g}$, such that $\mathbf{f}_{<\lambda} = \mathbf{g}_{<\lambda}$, $\mathbf{f} \models \sigma_\lambda$, and $\mathbf{g} \not\models \sigma_\lambda$.*

Proof: See appendix. ▮

We note that another approach to modelling joint knowledge is described in [FV1]. In that approach knowledge assignments assign sets of worlds to *sets* of agents, rather then individual agents. The advantage of that approach is that one does not need to consider λ-worlds for $\lambda > \omega$.

5.4 Summary

We now summarize our results on modelling common and joint knowledge. Common and joint knowledge are described by formulas of infinite depth (in fact, by formulas whose depth is given by ordinals up to ω^2). Therefore, to properly model states of knowledge where such formulas might hold, we need to consider not just worlds of length ω (that is, knowledge structures, as in the previous sections), but worlds of length up to ω^2. We show that as before, the truth of a formula is determined by a prefix of appropriate length of the world. We generalize the definition of no-information extension, and show that as before, the result is indeed a world. In particular, this shows that every λ-world is a prefix of some μ-world, whenever $\mu \geq \lambda$, which generalizes the result of Section 4 that says this when λ is finite and $\mu = \omega$. We show that as before, internal and external notions of truth coincide. As before, we use this to prove an equivalence between knowledge worlds and Kripke structures.

We show that if we restrict our attention to KC-formulas φ (those where there is no joint knowledge over proper subsets of the set of agents), then the truth of φ in a λ-world $\mathbf{f}$ is already determined by the ω-prefix of $\mathbf{f}$. We then show how to define the semantics of KC-formulas in ω-worlds directly. We also prove a finite model theorem, by showing that if a KC-formula is satisfiable, then it is satisfiable in an ω-world that is a least-information extension. Finally, we show the rather difficult technical result that when we allow general extended formulas (which may discuss joint knowledge among proper subsets of the set of agents), then we really need to allow "long worlds". Thus, there is no $\lambda < \omega^2$ such that for every formula φ, we can decide the truth of φ in a world $\mathbf{f}$ by considering only the prefix of $\mathbf{f}$ of length λ.

6 Extensions of the Approach

6.1 A Bayesian Approach

Economists have taken a *Bayesian* approach to modelling knowledge, where instead of having possible and impossible worlds we associate a probability distribution on worlds with each agent [Au, BD1]. In a non-Bayesian setting, an agent knows a fact p if p holds in all the worlds that the agent consider possible. In a Bayesian setting, an agent knows a fact p if the probability that p holds according to the agent's distribution is 1 [BD1]. (See also [FH, MS] for an approach that mixes Bayesian and non-Bayesian approaches).

Mertens and Zamir describe a Bayesian analogue to knowledge structures [MZ], which they call *infinite hierarchies of beliefs*. If X is a set, then let $\Delta(X)$ denote the space of probability distributions over X. Mertens and Zamir start with a set S called the *uncertainty space* (for technical reasons, this set is required to satisfy certain topological properties). Intuitively, S consists of all possible states of nature. A 0th-order Bayesian assignment f_0 is simply an element of S and $\langle f_0 \rangle$ is a Bayesian 1-world.

Assume inductively that the set X_k of Bayesian k-worlds have been defined. A kth-order Bayesian assignment is a function $f_k : \mathcal{P} \rightarrow \Delta(X_k)$. Intuitively, f_k associates with every agent a probability distribution on the set of Bayesian k-worlds. A $(k+1)$-*sequence of Bayesian assignments* is a sequence $\langle f_0, \ldots, f_k \rangle$, where f_i is an ith-order Bayesian assignment. A $(k+1)$-*world* is a $(k+1)$-sequence of Bayesian assignments that satisfy certain semantic restrictions, which we shall not list. An infinite sequence $\langle f_0, f_1, f_2, \ldots \rangle$ is called a *Bayesian knowledge structure* if each prefix $\langle f_0, \ldots, f_{k-1} \rangle$ is a Bayesian k-world for each k. The Bayesian approach has the interesting feature that there is no point in explicitly defining transfinite assignments, since these are already determined by the kth-order assignments (this result is implicit in Theorem 2.9 of [MZ]). We will come back to this point later.

The connection between Bayesian Kripke structures ([Au, BD1]) and Bayesian knowledge structures ([MZ]) has been studied in [BD2, MZ, TW]. The conclusion is that Bayesian Kripke structures and Bayesian knowledge structures have a relationship somewhat analogous to the one exhibited in this paper between Kripke structures and knowledge structure. Note that the results cannot be precisely analogous, since none of those paper has a notion of a logical language in which assertions about the structures can be made.

It may seem that the Bayesian approach is more expressive than the non-Bayesian approach. After all, in the Bayesian approach not only we distinguish between possible (having positive probability) and impossible (having probability zero) worlds, but we actually supply a degree of possibility to worlds. This is indeed the case with finite-order assignments over a finite uncertainty space. However, once we consider infinite uncertainty spaces or transfinite assignments, the picture gets more involved. Now, we cannot represent the fact that a world is impossible just by assigning it probability zero; in fact, probability is assigned to sets of worlds and not to individual worlds. Thus, in general the expressive power of knowledge assignments and Bayesian assignments is incomparable. Consider an $(\omega+1)$-world $\mathbf{f} = \langle f_0, f_1, \ldots, f_\omega \rangle$ in which $f_\omega(i)$ is uncountable, i.e., there are uncountably many ω-worlds that agent i considers possible (an example occurs in the $(\omega+1)$-no-information extension of a k-ary world, $k < \omega$, when there are at least two agents). For Bayesian assignments, we have made the convention that an agent knows a fact precisely if the probability of that fact (according to the agent's distribution) is 1. Hence, an agent considers a fact possible precisely if its probability is positive, since if its probability is 0, then the agent would know the negation of the fact. In the situation we are now considering, agent i considers an uncountable number of ω-worlds possible, and hence, under the Bayesian approach, agent i must assign to each of these ω-worlds a positive probability. However, it is well-known that it is impossible to assign positive probabilities to uncountably many disjoint events. Thus, this situation cannot be captured by the Bayesian approach.

Example 6.1: It is instructive to re-examine in a Bayesian setting the situation described in Example 5.13, where $\neg K_3 \neg E_{\{1,2\}} \sup kp$ holds for all $k \geq 1$. Probabilistically speaking, that means that the probability assigned by agent 3 to the events $E_{\{1,2\}} \sup kp$ is greater than 0 for all $k \geq 1$.

Let p_k be the probability assigned by agent 3 to $E_{\{1,2\}} \sup kp$ (and hence to the equivalent formula $\wedge_{i=1} \sup kE_{\{1,2\}} \sup ip$), for $k \geq 1$. Since $\wedge_{i=1} \sup \infty E_{\{1,2\}} \sup kp$ is equivalent to $C_{\{1,2\}}p$, the probability that agent 3 assigns to $C_{\{1,2\}}p$ is $\lim_{k\to\infty} p_k$ (this follows from the countable additivity of probability functions). Thus, $K_3\neg C_{\{1,2\}}p$ holds precisely when $\lim_{k\to\infty} p_k = 0$. So, in this case, we can see why it is not necessary to go beyond level ω: the probability (and hence the truth) of $K_3\neg C_{\{1,2\}}p$ is determined by probabilities at the finite levels. By contrast, as we discussed in Example 5.13, in the non-Bayesian setting we need to examine the ωth-order assignment f_ω to determine whether $K_3\neg C_{\{1,2\}}p$ holds.

The crucial point is that in the Bayesian setting, the probabilities assigned by agent 3 to the facts $E^k_{\{1,2\}}p$, $k \geq 1$, determine the probability he assigns to the fact $C_{\{1,2\}}p$. This lack of independence is a general phenomenon. Let A be a set of ω-worlds in the Bayesian setting, and let A_k be the set of all k-ary prefixes of members of A. The probability that agent 3 assigns to the set A is the limit (as $k \to \infty$) of the probability that agent 3 assigns to the set A_k. The probabilities at the finite levels completely determine the probabilities at level ω. As a consequence, we do not need ωth-order assignments in a Bayesian setting. By contrast, as we saw in Example 5.13, in our setting we need level ω, to provide additional information: if agent 3 considers every finite prefix of an ω-world $\mathbf{g}$ possible, then the ωth-order assignment f_ω tells us whether or not agent 3 considers $\mathbf{g}$ possible. ∎

The above example demonstrates that countable additivity of probability functions is the reason that transfinite assignments are redundant in the Bayesian approach. We note that countable additivity is crucial to the results in [BD2, MZ].

6.2 Further Extensions

Knowledge structures serve as a useful and robust tool for a deep investigation of various knowledge-related issues. While in this paper we focus on a particular variety of knowledge, the methodology we presented can be used to study other varieties.

For example, if we want to study *belief* rather than knowledge, then we replace the semantic restriction (K1) ("$\langle f_0, \ldots, f_{k-1}\rangle \in f_k(i)$, if $k \geq 1$") by "$f_k(i)$ is nonempty if $k \geq 1$" in our definition of knowledge structures, and we get *belief structures* (where it is possible to "believe" something that may not be true). We can also define *knowledge-belief structures* that deal with both knowledge and belief simultaneously, where there is a semantic restriction that implies that every known fact is also believed.

We can also incorporate time into knowledge structures, so that we can give semantics to a sentence such as "Alice knows that tomorrow Bob will know p" (or even to a sentence "Alice knows that tomorrow she (Alice) will know whether p is true or false"). The first step is to define a 0th-order assignment as a function f_0 from ω (which we take to represent time; 0 is today, 1 is tomorrow, etc.) to truth assignments to the primitive propositions. The second step is to define a kth-order assignment as a function $f_k : \mathcal{P} \times \omega \rightarrow 2\,\mathrm{sup}\, W_k$, where W_k here is the set of all k-worlds involving time. Our semantic restrictions (K1), (K2), and (K3) generalize naturally. One can also add other natural semantic restrictions; e.g., a restriction that says that each agent's knowledge increases monotonically with time. (Cf. [Le, HV] for the Kripke semantics of knowledge and time.)

The above examples suggest that the methodology described in this paper is quite general. This line of thought is pursued in [FV1], which investigates the applicability of the approach presented here to the modelling of other modal logics.

7 Concluding Remarks

In this paper we introduced a new semantic approach to modelling knowledge using knowledge structures. Although in a certain sense knowledge structures are equivalent to the well-known Kripke structures, they have a number of advantages over Kripke structures:

- Although there are situations where using Kripke structures is the appropriate approach to modelling knowledge (such as the situated-automata approach, where knowledge is ascribed on the basis of the information carried by the state of a machine), there are other situations where it is not clear how to use Kripke structures to model knowledge states. In such situations our approach offers a more intuitive approach to modelling knowledge.

- Our notion of a *no-information extension* models directly the notion of a "finite amount of information", where in particular there is no common knowledge. However, we show that no finite Kripke structure can capture this. By means of the *least-information extension*, we model the notion of a "finite amount of information in the presence of common knowledge".

- As shown in [FV1], by using knowledge structures, one can obtain proofs of decidability and compactness that are almost straightforward, and an elegant and constructive completeness proof.

On the other hand for some applications using Kripke structures is clearly the preferred approach. For example, the graph-theoretic nature of Kripke structures makes them the tool of choice when developing efficient decision procedures (cf. [HM2, La, Va2]).

In summary, knowledge structures are a new semantic representation for knowledge. Although they do not replace the widely-used Kripke structures, they do complement them: there are times when we can gain more insight by modelling knowledge with knowledge structures rather than with Kripke structures.

Acknowledgments. We thank Mike Fischer and Neil Immerman for stimulating discussions during the formative stages of this paper. We are grateful to Flaviu Cristian, Jakob Gonczarowski, Yuri Gurevich, Carl Hauser, Daniel Lehmann, Hector Levesque, Yoram Moses, Ray Strong, Johan van Benthem, and Ed Wimmers, who read a draft of this paper and made helpful comments. We also thank Eddie Dekel-Tabak, Jean-Francois Mertens, and John McCarthy for interesting discussions. Finally, the two referees deserve thank for many suggestions that helped to improve the quality of the paper. In particular, the equivalence of (1) and (3) in Proposition 4.4 was suggested by one of the referees.

A Appendix

In this appendix, we give the results and proofs we promised in the body of the paper.

A.1 A lemma used for Proposition 4.4

We begin with a lemma that was used in the proof of Proposition 4.4.

Lemma A.1: *Let w and w' be k-worlds that agree on all formulas of depth at most $k-1$ of the form $K_i\psi$ or $\neg K_i\psi$. Then $w \sim_i w'$.*

Proof: We prove this by induction on k. The case $k = 1$ is trivial. For the inductive step, assume that $w = \langle f_0, \ldots, f_{k-1}\rangle$ and $w' = \langle f'_0, \ldots, f'_{k-1}\rangle$ are as in the statement of the lemma. We must show that $f_{k-1}(i) = f'_{k-1}(i)$. If not, then without loss of generality, we can assume that there is some $(k-1)$-world v that is in $f_{k-1}(i)$ but not in $f'_{k-1}(i)$. Assume that $f'_{k-1}(i) = \{v'_1, \ldots, v'_r\}$. Thus, v is distinct from each of the v'_j's. Hence, for each j $(1 \le j \le r)$, either first component g_0 of v is distinct from the first component of v'_j, or else there is i such that $v \not\sim_i v'_j$. So by inductive hypothesis, there is a formula ψ_j of depth at most $k-2$ such that $v'_j \models \psi_j$ but $v \not\models \psi_j$. Let ψ be the formula $\psi_1 \vee \cdots \vee \psi_r$.

Then $w' \models K_i\psi$ but $w \not\models K_i\psi$. Since the formula $K_i\psi$ is of depth at most $k-1$, this contradicts our assumption. ∎

A.2 Proof of Theorem 4.12

Our next goal is to prove Theorem 4.12, which is as follows.

Theorem 4.12: *Assume that there are at least two agents, φ is a formula of depth r, and w a k-world. If $w^* \models E \sup r + k\varphi$, then φ is valid.*

We first need some preliminary concepts and results.

Lemma A.2: *Let $\langle f_0, \ldots, f_{k-1}\rangle$ and $\langle g_0, \ldots, g_{k-1}\rangle$ be k-worlds. If $f_{k-1}(i) = g_{k-1}(i)$, then $\langle f_0, \ldots, f_{k-1}\rangle^* \sim_i \langle g_0, \ldots, g_{k-1}\rangle^*$.*

Proof: Let $\langle f_0, \ldots, f_{k-1}\rangle^*$ be $\langle f_0, \ldots, f_{k-1}, f_k, f_{k+1}, \ldots\rangle$, and similarly for $\langle g_0, \ldots, g_{k-1}\rangle^*$. Since $f_{k-1}(i) = g_{k-1}(i)$, it follows, as noted earlier, that $f_j(i) = g_j(i)$ whenever $0 < j < k$. Assume inductively that we have shown that $f_r(i) = g_r(i)$ for some $r \geq k-1$. Then $f_{r+1}(i) = \{\langle h_0, \ldots, h_r\rangle : h_r(i) = f_r(i)\} = \{\langle h_0, \ldots, h_r\rangle : h_r(i) = g_r(i)\} = g_{r+1}(i)$. This completes the induction step. ∎

Definition A.3: If $\mathbf{p} = i_1, \ldots, i_s$ is a string of agents and if $\mathbf{f}$ and $\mathbf{f}'$ are knowledge structures, then we say that $\mathbf{f} \sim_{\mathbf{p}} \mathbf{f}'$ if there are knowledge structures $\mathbf{f}_1, \ldots, \mathbf{f}_{s+1}$ such that (a) $\mathbf{f}_1 = \mathbf{f}$, (b) $\mathbf{f}_{s+1} = \mathbf{f}'$, and (c) $\mathbf{f}_j \sim_{i_j} \mathbf{f}_{j+1}$ whenever $1 \leq j \leq s$. We may then say that there is a *path* between $\mathbf{f}$ and $\mathbf{f}'$. ∎

Note in particular that if $\mathbf{p}$ is the empty string, then $\mathbf{f} \sim_{\mathbf{p}} \mathbf{f}'$ precisely if $\mathbf{f} = \mathbf{f}'$. Note that for an arbitrary string $\mathbf{p}$ the relation $\sim_{\mathbf{p}}$ need not be an equivalence relation.

We shall take advantage of the following simple properties of $\sim_{\mathbf{p}}$, where $\mathbf{pq}$ is the concatenation of the strings $\mathbf{p}$ and $\mathbf{q}$.

(Reversibility): If $\mathbf{f} \sim_{\mathbf{p}} \mathbf{f}'$, then $\mathbf{f}' \sim_{\mathbf{p} \sup R} \mathbf{f}$.

(Transitivity): If $\mathbf{f} \sim_{\mathbf{p}} \mathbf{f}'$ and $\mathbf{f}' \sim_{\mathbf{q}} \mathbf{f}''$, then $\mathbf{f} \sim_{\mathbf{pq}} \mathbf{f}''$.

(Collapsibility): Let i be an agent. Then $\mathbf{f} \sim_{\mathbf{p}ii\mathbf{q}} \mathbf{f}'$ if and only if $\mathbf{f} \sim_{\mathbf{p}i\mathbf{q}} \mathbf{f}'$.

If $\mathbf{p} = i_1 \ldots i_s$ and no two consecutive agents in $\mathbf{p}$ are the same, (i.e., if $i_j \neq i_{j+1}$ for $1 \leq j < s$), then we say that $\mathbf{p}$ is *nonduplicating*. By collapsibility, we can use nonduplicating strings without loss of generality.

We are interested in our notion of $\sim_{\mathbf{p}}$ because of the following lemma, whose simple proof (by induction on the length of $\mathbf{p}$) is left to the reader.

Lemma A.4: *If $\mathbf{f} \models K_{\mathbf{p}}\varphi$ and $\mathbf{f} \sim_{\mathbf{p}} \mathbf{g}$, then $\mathbf{g} \models \varphi$.*

The next proposition implies that there is a path between every pair of no-information extensions, and gives us information as to the length of the path.

Proposition A.5: *Let w be a k-world, let w' be a k'-world, and let $\mathbf{p} \in \mathcal{P}^*$ be nonduplicating and of length $k + k' - 1$. Then $w^* \sim_{\mathbf{p}} w'^*$.*

Proof: We first prove the following special case.

> **Special case:** Let $\langle f_0, \ldots, f_{k-1} \rangle$ be a k-world, and let $\mathbf{p} \in \mathcal{P}^*$ be nonduplicating and of length k. Then $\langle f_0 \rangle^* \sim_{\mathbf{p}} \langle f_0, \ldots, f_{k-1} \rangle^*$.

We prove the special case by induction on k. The base case (k=1) is immediate. For the inductive step, let $\langle f_0, \ldots, f_k \rangle$ be a $(k+1)$-world, and let $\mathbf{p} = i_1 \ldots i_{k+1}$ be nonduplicating. By the inductive assumption, $\langle f_0 \rangle^* \sim_{i_1 \ldots i_k} \langle f_0, \ldots, f_{k-1} \rangle^*$. We must show that $\langle f_0 \rangle^* \sim_{i_1 \ldots i_{k+1}} \langle f_0, \ldots, f_k \rangle^*$. Define f'_k by letting $f'_k(j)$ be the j-equivalence class of $f_{k-1}(j)$ for each agent j. By Lemma 4.3, we know that $\langle f_0, \ldots, f_{k-1}, f'_k \rangle$ is a $(k+1)$-world. Let $\langle f_0, \ldots, f_{k-1}, g_k \rangle$ be the i_{k+1}-matching extension of $\langle f_0, \ldots, f_{k-1} \rangle$ with respect to f_k. Again, by Lemma 4.3, we know that $\langle f_0, \ldots, f_{k-1}, g_k \rangle$ is a $(k+1)$-world.

Since $g_k(i_k) = f'_k(i_k)$, it follows from Lemma A.2 that

$$\langle f_0, \ldots, f_{k-1}, f'_k \rangle^* \sim_{i_k} \langle f_0, \ldots, f_{k-1}, g_k \rangle^*.$$

Since $g_k(i_{k+1}) = f_k(i_{k+1})$, it follows from Lemma A.2 that

$$\langle f_0, \ldots, f_{k-1}, g_k \rangle^* \sim_{i_{k+1}} \langle f_0, \ldots, f_{k-1}, f_k \rangle^*.$$

Since $f'_k(j)$ is the j-equivalence class of $f_{k-1}(j)$ for each agent j, it follows that $\langle f_0, \ldots, f_{k-1}, f'_k \rangle^* = \langle f_0, \ldots, f_{k-1} \rangle^*$. Putting these last few observations together, we see that $\langle f_0, \ldots, f_{k-1} \rangle^* \sim_{i_k i_{k+1}} \langle f_0, \ldots, f_{k-1}, f_k \rangle^*$. Putting this fact together with our inductive assumption that $\langle f_0 \rangle^* \sim_{i_1 \ldots i_k} \langle f_0, \ldots, f_{k-1} \rangle^*$, it follows by transitivity that $\langle f_0 \rangle^* \sim_{(i_1 \ldots i_k)(i_k i_{k+1})} \langle f_0, \ldots, f_k \rangle^*$. By collapsibility, $\langle f_0 \rangle^* \sim_{i_1 \ldots i_{k+1}} \langle f_0, \ldots, f_k \rangle^*$, which was to be shown. This concludes the proof of the special case.

Let $w = \langle f_0, \ldots, f_{k-1} \rangle$ and $w' = \langle f'_0, \ldots, f'_{k-1} \rangle$. Let $\mathbf{p} = i_1 \ldots i_{k+k'-1}$. By the special case, $\langle f_0 \rangle^* \sim_{i_k \ldots i_1} w^*$ and $\langle f'_0 \rangle^* \sim_{i_k \ldots i_{k+k'-1}} w'^*$. By reversibility, $w^* \sim_{i_1 \ldots i_k} \langle f_0 \rangle^*$. By Lemma A.2, $\langle f_0 \rangle^* \sim_{i_k} \langle f'_0 \rangle^*$. By transitivity (applied twice), $w^* \sim_{(i_1 \ldots i_k) i_k (i_k \ldots i_{k+k'-1})} w'^*$. By collapsibility (applied twice), $w^* \sim_{\mathbf{p}} w'^*$, as desired. ∎

We can now prove Theorem 4.12.

Proof of Theorem 4.12. Assume that there are at least two agents, that φ is a formula of depth r, that w a k-world, and that $w^* \models E \sup r + k\varphi$. We must show that φ is valid. Assume not; we shall derive a contradiction.

Let $\mathbf{p} \in \mathcal{P}^*$ be nonduplicating and of length $r + k$ (there is obviously such a string $\mathbf{p}$, since there are at least two agents). Since $w^* \models E \sup r + k\varphi$, clearly $w^* \models K_{\mathbf{p}}\varphi$. Since φ is of depth r and not valid, there is an $(r+1)$-world $w' = \langle f_0, \ldots, f_r \rangle$ such that $w' \not\models \varphi$. Hence, $w'^* \not\models \varphi$. By Lemma A.5, $w^* \sim_{\mathbf{p}} w'^*$. Therefore, since $w^* \models K_{\mathbf{p}}\varphi$, it follows from Proposition A.4 that $w'^* \models \varphi$. This is a contradiction. ∎

A.3 Proof of Theorem 4.21

We now begin a development that will lead to the proof of Theorem 4.21, which is as follows:

Theorem 4.21: *$(w, \mathcal{C})^*$ is a knowledge structure iff $reach(w, \mathcal{C})$ is a closed set.*

In order to prove the theorem, we need a few preliminary lemmas.

Lemma A.6: *If $\mathcal{C}$ is a set of k-worlds, $w \in \mathcal{C}$, and $\mathcal{C}' = reach(w, \mathcal{C})$, then $(w, \mathcal{C})^* = (w, \mathcal{C}')^*$.*

Proof: Suppose $w = \langle f_0, \ldots, f_{k-1} \rangle$ and $(w, \mathcal{C})^* = \langle f_0, \ldots, f_{k-1}, f_k, f_{k+1}, \ldots \rangle$. We can prove, by induction on $m \geq k$, that $worlds_k(\langle f_0, \ldots, f_{m-1} \rangle) \subseteq reach(w, \mathcal{C})$; the proof is left to the reader. It now follows from the definition of the least-information extension that $(w, \mathcal{C})^* = (w, \mathcal{C}')^*$. ∎

Lemma A.7: *Let $\mathcal{C}$ be a closed set of k-worlds, $m \geq k - 1$, and $v = \langle g_0, \ldots, g_m \rangle$ be a world such that $worlds_k(v) \subseteq \mathcal{C}$. Define $g_{m+1}(i) = \{\langle h_0, \ldots, h_m \rangle : h_m(i) = g_m(i)$ and $worlds_k(\langle h_0, \ldots, h_m \rangle) \subseteq \mathcal{C}\}$ for each agent i. Then $\langle g_0, \ldots, g_{m+1} \rangle$ is a world.*

Proof: The fact that (K1) and (K2) hold is immediate from the definition. To see that (K3) holds, we proceed by induction on m. First suppose $m = k - 1 > 0$ and $\langle h_0, \ldots, h_{m-1} \rangle \in g_m(i)$. From the fact that $\mathcal{C}$ is closed and $v \in \mathcal{C}$ (since $v \in worlds_k(v) \subseteq \mathcal{C}$), it follows that there is some $\langle h_0, \ldots, h_{m-1}, h_m \rangle \in \mathcal{C}$ such that $h_m(i) = g_m(i)$. This world is in $g_{m+1}(i)$ by definition, so (K3) holds. Suppose now that $m > k - 1$ and $\langle h_0, \ldots, h_{m-1} \rangle \in g_m(i)$. Note that $worlds_k(\langle h_0, \ldots, h_{m-1} \rangle) \subseteq \mathcal{C}$ by assumption. Define $h_m(i) = g_m(i)$ and $h_m(j) = \{\langle h'_0, \ldots, h'_{m-1} \rangle : h'_{m-1}(j) = h_{m-1}(j)$ and $worlds_k(\langle h'_0, \ldots, h'_{m-1} \rangle) \subseteq \mathcal{C}\}$ for $j \neq i$. By the inductive hypothesis, $\langle h_0, \ldots, h_m \rangle$ is a world, and by construction, it is in $g_{m+1}(i)$. Thus there is some extension of $\langle h_0, \ldots, h_{m-1} \rangle$ in $g_{m+1}(i)$, so (K3) holds. ∎

Lemma A.8: *Suppose $\mathcal{C}$ is a set of k-worlds, $w, w' \in \mathcal{C}$ with $w \sim_i w'$, and $(w, \mathcal{C})^*$ is a knowledge structure. Then $(w', \mathcal{C})^*$ is a knowledge structure.*

Proof: Suppose $(w, \mathcal{C})^* = \langle f_0, f_1, \ldots \rangle$ and $(w', \mathcal{C})^* = \langle g_0, g_1, \ldots \rangle$. We prove by induction on m that $\langle g_0, \ldots, g_m \rangle$ is an $(m+1)$-world for all m. If $m < k$ then $\langle g_0, \ldots, g_m \rangle$ is a prefix of the world w', so the result is immediate. If $m \geq k$, it is easy to see that properties (K1) and (K2) hold from the construction. For (K3), suppose that $m > 1$ and $\langle h_0, \ldots, h_{m-2} \rangle \in g_{m-1}(j)$. By assumption, $\langle g_0, \ldots, g_{m-1} \rangle$ is a world. By the construction of least-information extensions, $g_{m-1}(i) = f_{m-1}(i)$ and $worlds_k(\langle g_0, \ldots, g_{m-1} \rangle) \subseteq \mathcal{C}$, so $\langle g_0, \ldots, g_{m-1} \rangle \in f_m(i)$. Since $(w, \mathcal{C})^*$ is a knowledge structure, there is some g'_m such that $\langle g_0, \ldots, g_{m-1}, g'_m \rangle \in f_{m+1}(i)$. By property (K3), there must be some h'_{m-1} such that

$\langle h_0, \ldots, h_{m-2}, h'_{m-1}\rangle \in g'_m(i)$. By construction again, we must have $h'_{m-1}(i) = g_{m-1}(i)$ and $worlds_k(\langle h_0, \ldots, h_{m-2}, h'_{m-1}\rangle) \subseteq \mathcal{C}$. Thus, $\langle h_0, \ldots, h_{m-2}, h'_{m-1}\rangle \in g_m(i)$, so (K3) holds. ∎

Proof of Theorem 4.21. First suppose that $(w, \mathcal{C})^*$ is a knowledge structure. Note that it follows (by an easy induction on distance using Lemma A.8), that $(w', \mathcal{C})^*$ is a knowledge structure for all $w' \in reach(w, \mathcal{C})$. We now show that $reach(w, \mathcal{C})$ is closed. Suppose $\langle g_0, \ldots, g_{k-1}\rangle \in reach(w, \mathcal{C})$ and $\langle h_0, \ldots, h_{k-2}\rangle \in g_{k-1}(i)$. We want to show that for some h'_{k-1}, we have $h'_{k-1}(i) = g_{k-1}(i)$ and $\langle h_0, \ldots, h_{k-2}, h'_{k-1}\rangle \in reach(w, \mathcal{C})$. Suppose $(\langle g_0, \ldots, g_{k-1}\rangle, \mathcal{C})^* = \langle g_0, \ldots, g_{k-1}, g_k, \ldots\rangle$. By property (K3), there is some h'_{k-1} such that $\langle h_0, \ldots, h_{k-2}, h'_{k-1}\rangle \in g_k(i)$. By the construction of least-information extensions, we must have $\langle h_0, \ldots, h_{k-2}, h'_{k-1}\rangle \in \mathcal{C}$ and $h'_{k-1}(i) = g_{k-1}(i)$. Thus $\langle h_0, \ldots, h_{k-2}, h'_{k-1}\rangle \in reach(w, \mathcal{C})$. This shows that $reach(w, \mathcal{C})$ is closed.

For the converse, suppose that $\mathcal{C}' = reach(w, \mathcal{C})$ is closed. By Lemma A.6, it follows that $(w, \mathcal{C}')^* = (w, \mathcal{C})^*$. Suppose that $(w, \mathcal{C}')^* = \langle f_0, f_1, \ldots\rangle$. Now a straightforward induction on m using Lemma A.7 shows that $\langle f_0, \ldots, f_m\rangle$ is an $(m+1)$-world. Thus $(w, \mathcal{C})^*$ is a knowledge structure. ∎

A.4 Proof of Theorem 4.22

In this subsection, we prove Theorem 4.22, which is as follows:

Theorem 4.22: *$(w, \mathcal{C})^*$ is a knowledge structure where all the worlds of $\mathcal{C}$ appear iff $\mathcal{C}$ is closed and $\mathcal{C} = reach(w, \mathcal{C})$.*

We begin with a lemma.

Lemma A.9: *Suppose $\mathcal{C}$ is a set of k-worlds, $w \in \mathcal{C}$, and $(w, \mathcal{C})^*$ is a knowledge structure. Then $worlds_k((w, \mathcal{C})^*) = reach(w, \mathcal{C})$; i.e., the k-worlds that appear in $(w, \mathcal{C})^*$ are precisely those that are reachable from w.*

Proof: Let $\mathcal{C}' = reach(w, \mathcal{C})$. By Lemma A.6, it follows that $worlds_k((w, \mathcal{C})^*) \subseteq \mathcal{C}'$. To get containment in the other direction, we show by induction on m that if w' is distance m from w, then w' appears in $(w, \mathcal{C})^*$. The result is trivial if $m = 0$. If $m > 0$, there exists $w'' \in \mathcal{C}$, where $w \sim_i w''$ for some agent i, and w' is distance $m - 1$ from w''. By Lemma A.8, $(w'', \mathcal{C})^*$ is a knowledge structure, and by induction hypothesis, w' appears in $(w'', \mathcal{C})^*$. Suppose $(w, \mathcal{C})^* = \langle f_0, f_1, \ldots\rangle$ and $(w'', \mathcal{C})^* = \langle g_0, g_1, \ldots\rangle$. Thus w' appears in $\langle g_0, \ldots, g_l\rangle$ for some l. Since $w \sim_i w''$, by the construction of least-information extensions, we must have $f_l(i) = g_l(i)$ and thus $\langle g_0, \ldots, g_l\rangle \in f_{l+1}(i)$. Therefore w' appears in $\langle f_0, \ldots, f_{l+1}\rangle$. ∎

We now prove Theorem 4.22. If $(w, \mathcal{C})^*$ is a knowledge structure where all the worlds in $\mathcal{C}$ appear, then by Theorem 4.21, $reach(w, \mathcal{C})$ is closed, and by Lemma A.9, $\mathcal{C} = reach(w, \mathcal{C})$. Conversely, if $\mathcal{C}$ is closed and $\mathcal{C} = reach(w, \mathcal{C})$, then $reach(w, \mathcal{C})$ is closed, so by Theorem 4.21, $(w,\mathcal{C})^*$ is a knowledge structure. By Lemma A.9, it also follows that all the worlds of $\mathcal{C}$ appear in $(w,\mathcal{C})^*$. ∎

A.5 Proof of Theorem 5.10

In this subsection, we prove Theorem 5.10. We first need a lemma.

Lemma A.10: *Let σ be a K-formula, and let $\mathbf{f}$ and $\mathbf{g}$ be ω-worlds such that $\mathbf{f} \sim_i \mathbf{g}$ for each i. Then $\mathbf{f} \models C\sigma$ iff $\mathbf{g} \models C\sigma$.*

Proof: Assume $\mathbf{f} \models C\sigma$. Then $\mathbf{f} \models E \sup j\sigma$ for all $j > 0$. It follows that $\mathbf{f} \models K_i E \sup j\sigma$ for all $j > 0$. But $\mathbf{f} \sim_i \mathbf{g}$, so $\mathbf{g} \models E \sup j\sigma$ for all $j > 0$. Thus, $\mathbf{g} \models C\sigma$. ∎

We can now restate and prove the theorem.

Theorem 5.10: *Let σ be a KC-formula. Let $\mathbf{f}$ and $\mathbf{g}$ be λ-worlds, such that $\mathbf{f}_{<\omega} = \mathbf{g}_{<\omega}$. Then $\mathbf{f} \models \sigma$ iff $\mathbf{g} \models \sigma$.*

Proof: By Lemma 5.2, to prove the theorem it suffices to show that every KC-formula is equivalent to a formula whose depth is less than or equal to ω. The proof uses the following valid axiom schemes:

(1) $K_i(\varphi_1 \wedge \ldots \wedge \varphi_k) \equiv ((K_i\varphi_1) \wedge \ldots \wedge (K_i\varphi_k))$

(2) $C(\varphi_1 \wedge \ldots \wedge \varphi_k) \equiv ((C\varphi_1) \wedge \ldots \wedge (C\varphi_k))$

(3) $K_iC\varphi \equiv C\varphi$

(4) $K_i\neg C\varphi \equiv \neg C\varphi$

(5) $CC\varphi \equiv C\varphi$

(6) $C\neg C\varphi \equiv \neg C\varphi$

(7) $K_i\varphi \vee K_i\psi \Rightarrow K_i(\varphi \vee \psi)$

(8) $C\varphi \vee C\psi \Rightarrow C(\varphi \vee \psi)$

(9) $K_i\varphi \Rightarrow \varphi$

(10) $C\varphi \Rightarrow \varphi$

We want to show that every KC-formula is equivalent to a formula where there is no C in the scope of another C or a K_i. The proof is by structural induction. By (1) and (2) it suffices to prove that the following axiom schemes are valid, where φ and ψ are K-formulas.

(a) $K_i(C\varphi \vee \psi) \equiv (C\varphi \vee K_i\psi)$

(b) $K_i(\neg C\varphi \vee \psi) \equiv (\neg C\varphi \vee K_i\psi)$

(c) $C(C\varphi \vee \psi) \equiv (C\varphi \vee C\psi)$

(d) $C(\neg C\varphi \vee \psi) \equiv (\neg C\varphi \vee C\psi)$

In (a)-(d), implication from right to left follows easily from (3)-(10), so we consider only implication from left to right. We first prove (a). By Lemma 5.2, it suffices to consider $(\omega+1)$-worlds. Let $\mathbf{h}$ be an $(\omega+1)$-world.

Suppose that $\mathbf{h} \models K_i(C\varphi \vee \psi)$. Then for every $\mathbf{e} \in f_\omega(i)$ we have $\mathbf{e} \models C\varphi \vee \psi$. There are two possibilities to consider. First, it is possible that for some $\mathbf{e} \in f_\omega(i)$, we have that $\mathbf{e} \models C\varphi$. Then, by Lemma A.10, for every $\mathbf{e} \in f_\omega(i)$ we have that $\mathbf{e} \models C\varphi$. Thus, $\mathbf{h} \models K_i C\varphi$. Consequently, by (3), $\mathbf{h} \models C\varphi \vee K_i\psi$. The other possibility is that for all $\mathbf{e} \in f_\omega(i)$ we have that $\mathbf{e} \models \neg C\varphi$. In that case, we have $\mathbf{h} \models K_i\psi$, so $\mathbf{h} \models C\varphi \vee K_i\psi$.

The proof of (b) is similar and left to the reader. We now prove (c). Let $\mathbf{h}$ be a world such that $\mathbf{h} \models C(C\varphi \vee \psi)$. Then $\mathbf{h} \models E \sup i(C\varphi \vee \psi)$ for all $i > 0$. It follows by (a) that $\mathbf{h} \models (C\varphi \vee E \sup i\psi)$ for all $i \geq 0$. There are now two possibilities to consider. First, it is possible that $\mathbf{h} \models C\varphi$, in which case clearly, $\mathbf{h} \models C\varphi \vee C\psi$. The other possibility is that $\mathbf{h} \not\models C\varphi$. In that case we have $\mathbf{h} \models E \sup i\psi$ for all $i \geq 0$, i.e., $\mathbf{h} \models C\psi$. It follows that $\mathbf{h} \models C\varphi \vee C\psi$.

The proof of (d) is similar and left to the reader. ∎

A.6 Proof of Theorem 5.14

In this subsection, we prove Theorem 5.14, which is as follows:

Theorem 5.14: *For every ordinal* $1 \leq \lambda < \omega \sup 2$ *there is a formula* σ_λ *and there are* $(\lambda+1)$*-worlds* $\mathbf{f}$ *and* $\mathbf{g}$*, such that* $\mathbf{f}_{<\lambda} = \mathbf{g}_{<\lambda}$*,* $\mathbf{f} \models \sigma_\lambda$*, and* $\mathbf{g} \not\models \sigma_\lambda$*.*

Proof: We first prove the claim for $1 \leq \lambda < \omega$. The case $\lambda = 1$ is easy and is left to the reader (one agent suffices). Consider the case where $1 < \lambda < \omega$. Here we need two agents, 1 and 2, and one primitive proposition p. Let f_0 make p true. For $1 \leq k \leq \lambda$, let $f_k(1) = f_k(2) = \{\mathbf{f}_{<k}\}$. It is easy to verify that $\mathbf{f}$ is a $(\lambda+1)$-world. Also, one can show by induction on k, $1 \leq k \leq \lambda$ that $\mathbf{f}_{<k+1} \models E \sup kp$. In particular, $\mathbf{f} \models E \sup \lambda p$. Let $\mathbf{g}$ be the 2-matching extension of $\mathbf{f}_{<\lambda}$ with respect to $\mathbf{f}$. That is, $\mathbf{g}_{<\lambda} = \mathbf{f}_{<\lambda}$, $g_\lambda(2) = \{\mathbf{g}_{<\lambda}\}$, and $g_\lambda(1)$ is the 1-equivalence class of $\mathbf{g}_{<\lambda}$. We now show that $\mathbf{g} \not\models E \sup \lambda p$. The proof is by induction on λ. For $\lambda = 1$, we have $g_1(1) = \{p, \overline{p}\}$, so $\mathbf{g} \not\models K_1 p$, and consequently, $\mathbf{g} \not\models Ep$. For $\lambda > 1$, let $\mathbf{g}'$ be the 1-matching extension of $\mathbf{g}_{<\lambda-1}$ with respect to $\mathbf{g}_{<\lambda}$; that is, $\mathbf{g}'_{<\lambda-1} = \mathbf{g}_{<\lambda-1}$, $g'_{\lambda-1}(1) = \{\mathbf{g}_{<\lambda-1}\}$, and $g'_{\lambda-1}(2)$ is the 2-equivalence class of $\mathbf{g}_{<\lambda-1}$. By the inductive hypothesis and a symmetry argument we have that $\mathbf{g}' \not\models E \sup \lambda - 1p$. But $\mathbf{g}' \sim_1 \mathbf{g}_{<\lambda}$, so $\mathbf{g} \not\models K_1 E \sup \lambda - 1p$. Consequently, $\mathbf{g} \not\models E \sup \lambda p$.

We now prove the claim for $\omega \leq \lambda < \omega \sup 2$. Here we use three agents, 1, 2, and 3, and one primitive proposition p (note that three agents are necessary by Theorem 5.10). Let $\mu < \omega \sup 2$. Note that $\mu = \omega \times k + l$ for some $k, l \geq 0$. We define the classes U_μ and V_μ of μ-worlds. As we shall see later, the worlds $\mathbf{f}$ and $\mathbf{g}$ whose existence is claimed by the theorem will be members of U_μ and V_μ, correspondingly. Worlds in U_μ are constructed in such a manner as to prevent agents 2 and 3 from having joint knowledge, and to make sure that agent 1 knows it. In worlds in V_μ agents 2 and 3 do have joint knowledge of agent 1's knowledge.

The construction is by induction on μ. The class U_1 contains the single 1-world $\langle h_0 \rangle$, where h_0 makes p true. Let $V_1 = U_1$.

A 2-world $\mathbf{h}$ is in U_2 if $\mathbf{h}_{<1} \in U_1$ and $h_1(1) = \{\mathbf{h}_{<1}\}$. An l-world $\mathbf{h}$ is in U_l for $l > 2$ if $\mathbf{h}_{<2}$ is in U_2, and either $h_{l-1}(2)$ is the 2-equivalence class of $\mathbf{h}_{<l-1}$ or $h_{l-1}(3)$ is the 3-equivalence class of $\mathbf{h}_{<l-1}$. An ω-world $\mathbf{h}$ is in U_ω if $h_l(1)$ is the 1-equivalence class of $\mathbf{h}_{<l}$ for all $l > 1$, and $\mathbf{h}_{<l}$ is in U_l for some $l > 2$. An l-world $\mathbf{h}$ is in V_l for $l > 1$ if $\mathbf{h}_{<2}$ is in U_2, and for all m such that $1 \leq m < l$ we have that $h_m(2) = \{\mathbf{e} : \mathbf{e} \sim_2 \mathbf{h}_{<m} \text{ and } \mathbf{e} \in V_m\}$, and $h_m(3) = \{\mathbf{e} : \mathbf{e} \sim_3 \mathbf{h}_{<m} \text{ and } \mathbf{e} \in V_m\}$. An ω-world $\mathbf{h}$ is in V_ω if $h_l(1)$ is the 1-equivalence class of $\mathbf{h}_{<l}$ for all $l > 1$, and $\mathbf{h}_{<l}$ is in V_l for all $l > 1$.

Inductively, let $\mu < \omega \sup 2$ be a limit ordinal. A world $\mathbf{h}$ is in $U_{\mu+1}$ if $\mathbf{h}_{<\mu}$ is in U_μ and $h_\mu(1) = \{\mathbf{e} : \mathbf{e} \sim_1 \mathbf{h}_{<\mu} \text{ and } \mathbf{e} \in U_\mu\}$. A $(\mu + l)$-world $\mathbf{h}$ is in $U_{\mu+l}$ for $l > 1$ if $\mathbf{h}_{<\mu+1}$ is in $U_{\mu+1}$ and either $h_{\mu+l-1}(2)$ is the 2-equivalence class of $\mathbf{h}_{<\mu+l-1}$ or $h_{\mu+l-1}(3)$ is the 3-equivalence class of $\mathbf{h}_{<\mu+l-1}$. A $(\mu + \omega)$-world $\mathbf{h}$ is in $U_{\mu+\omega}$ if $h_{\mu+l}(1)$ is the 1-equivalence class of $\mathbf{h}_{<\mu+l}$ for all $l > 0$, and $\mathbf{h}_{<\mu+l}$ is in $U_{\mu+l}$ for some $l > 1$. A $(\mu + l)$-world $\mathbf{h}$ is in $V_{\mu+l}$ for $l > 0$ if $\mathbf{h}_{<\mu+1}$ is in $U_{\mu+1}$, $h_\mu(2) = \{\mathbf{e} : \mathbf{e} \sim_2 \mathbf{h}_{<\mu} \text{ and } \mathbf{e} \in U_\mu\}$, $h_\mu(3) = \{\mathbf{e} : \mathbf{e} \sim_3 \mathbf{h}_{<\mu} \text{ and } \mathbf{e} \in U_\mu\}$, and for all m such that $1 \leq m < l$ we have that $h_{\mu+m}(2) = \{\mathbf{e} : \mathbf{e} \sim_2 \mathbf{h}_{<\mu+m} \text{ and } \mathbf{e} \in V_{\mu+m}\}$, and $h_{\mu+m}(3) = \{\mathbf{e} : \mathbf{e} \sim_3 \mathbf{h}_{<\mu+m} \text{ and } \mathbf{e} \in V_{\mu+m}\}$. A $(\mu + \omega)$-world $\mathbf{h}$ is in $V_{\mu+\omega}$ if $h_{\mu+l}(1)$ is the 1-equivalence class of $\mathbf{h}_{<\mu+l}$ for all $l > 1$, and $\mathbf{h}_{<\mu+l}$ is in $V_{\mu+l}$ for all $l > 1$.

To prove the existence of $\mathbf{f}$ and $\mathbf{g}$ we have to first prove several properties of the U_μ's and V_μ's. The proof requires a fairly technical induction hypothesis. We also need define the classes U'_μ for certain successor ordinals. Let μ be a limit ordinal. A world $\mathbf{h}$ is in $U'_{\mu+1}$ if $\mathbf{h}_{<\mu}$ is in U_μ and $h_\mu(1)$ is the 1-equivalence class of $\mathbf{h}_{<\mu}$.

Claim A. Let $\mu < \omega \sup 2$.

1. If $\mathbf{h} \in U_\mu$, then there is some $\mathbf{h}' \in U_{\mu+1}$ such that $\mathbf{h}'_{<\mu} = \mathbf{h}$.
2. If μ is a successor ordinal and $\mathbf{h} \in V_\mu$, then there is some $\mathbf{h}' \in V_{\mu+1}$ such that $\mathbf{h}'_{<\mu} = \mathbf{h}$.
3. If μ is a limit ordinal and $\mathbf{h} \in U_\mu$, then there is some $\mathbf{h}' \in V_{\mu+1}$ such that $\mathbf{h}'_{<\mu} = \mathbf{h}$.
4. If μ is a successor ordinal and $\mathbf{h} \in V_\mu$, then there is some $\mathbf{h}' \in U_{\mu+1}$ such that $\mathbf{h}'_{<\mu} = \mathbf{h}$.

5. If μ is a limit ordinal and $\mathbf{h} \in U_\mu$, then there is some $\mathbf{h}' \in U'_{\mu+1}$ such that $\mathbf{h}'_{<\mu} = \mathbf{h}$.

6. The classes U_μ and V_μ are nonempty.

The proof is by induction on μ. We first prove part (1). If μ is a successor ordinal and $\mathbf{h} \in U_\mu$, then $\mathbf{h} \sup \mu + 1$ is the desired extension. Assume now that $\mu = \omega \times k$ is a limit ordinal. Let $\mathbf{h} \in U_\mu$. We construct a $(\mu+1)$-world $\mathbf{h}'$ in $U_{\mu+1}$. Let $\mathbf{h}'_{<\mu} = \mathbf{h}$, $h'_\mu(2)$ (resp., $h'_\mu(3)$) be the 2-no-information (resp., 3-no-information) extension of $\mathbf{h}$, and $h'_\mu(1) = \{\mathbf{e} : \mathbf{e} \sim_1 \mathbf{h} \text{ and } \mathbf{e} \in U_\mu\}$. We have to show that $\mathbf{h}'$ satisfies (K3$'$) for all agents. That (K3$'$) holds for agents 2 and 3 is obvious, so we focus on agent 1. Let $\mathbf{e} \in h_\kappa(1)$. Without loss of generality we can assume that $\kappa > \omega \times (k-1)$. It is easy to see that $\mathbf{e} \sup \mu \in U_\mu$ and $\mathbf{e} \sim_1 \mathbf{h}$, so $\mathbf{e} \sup \mu \in h'_\mu(1)$ and (K3$'$) holds. This completes the proof of part (1).

We now prove part (2). We first consider the case $\mu = l < \omega$. The claim clearly holds for $l = 1$. For $l > 1$, let $\mathbf{h}'$ be an $(l+1)$-world defined as follows: $\mathbf{h}'_{<l} = \mathbf{h}$, $h'_l(1)$ is the 1-no-information extension of $\mathbf{h}$, $h'_l(2) = \{\mathbf{e} : \mathbf{e} \sim_2 \mathbf{h} \text{ and } \mathbf{e} \in V_l\}$, and $h'_l(3) = \{\mathbf{e} : \mathbf{e} \sim_3 \mathbf{h} \text{ and } \mathbf{e} \in V_l\}$. We claim that $\mathbf{h}' \in V_{l+1}$. (K1$'$) and (K2$'$) clearly holds, so only (K3$'$) remains to be verified. That (K3$'$) holds for agent 1 is obvious, so we focus on agent 2 (the argument for agent 3 is analogous). Let $\mathbf{e} \in h_{l-1}(2)$. Since $\mathbf{h} \in V_l$, we have that $\mathbf{e} \in V_{l-1}$. Thus, by the inductive hypothesis, there is $\mathbf{e}' \in V_l$ such that $\mathbf{e}'_{<l-1} = \mathbf{e}$. In particular, $e'_{l-1}(2) = \{\mathbf{e}'' : \mathbf{e}'' \sim_2 \mathbf{e} \text{ and } \mathbf{e}'' \in V_{l-1}\}$. But $\mathbf{e} \sim_2 \mathbf{h}_{<l-1}$, so $e'_{l-1}(2) = h_{l-1}(2)$, and consequently $\mathbf{e}' \sim_2 \mathbf{h}$. Thus, $\mathbf{e}' \in h'_l(2)$ and (K3$'$) holds. It follows that $\mathbf{h}' \in V_{l+1}$.

Now let μ be a limit ordinal, and let $\mathbf{h} \in V_{\mu+l}$, $l \geq 1$. We define $\mathbf{h}'$ to be a $(\mu+l+1)$-world defined as follows: $\mathbf{h}'_{<\mu+l} = \mathbf{h}$, $h'_{\mu+l}(1)$ is the 1-no-information extension of $\mathbf{h}$, $h'_{\mu+l}(2) = \{\mathbf{e} : \mathbf{e} \sim_2 \mathbf{h} \text{ and } \mathbf{e} \in V_{\mu+l}\}$, and $h'_{\mu+l}(3) = \{\mathbf{e} : \mathbf{e} \sim_3 \mathbf{h} \text{ and } \mathbf{e} \in V_{\mu+l}\}$. We claim that $\mathbf{h}' \in V_{\mu+l+1}$. (K1$'$) and (K2$'$) clearly holds, so only (K3$'$) remains to be verified. That (K3$'$) holds for agent 1 is obvious, so we focus on agent 2 (the argument for agent 3 is analogous). Let $\mathbf{e} \in h_{\mu+l-1}(2)$. If $l = 1$, then $\mathbf{e} \in U_\mu$, and if $l > 1$, then $\mathbf{e} \in V_{\mu+l-1}$. In either case there is $\mathbf{e}' \in V_{\mu+l}$ such that $\mathbf{e}'_{<\mu+l-1} = \mathbf{e}$. In particular, if $l = 1$ then $e'_{\mu+l-1}(2) = \{\mathbf{e}'' : \mathbf{e}'' \sim_2 \mathbf{e} \text{ and } \mathbf{e}'' \in U_\mu\}$, and if $l > 1$, then $e'_{\mu+l-1}(2) = \{\mathbf{e}'' : \mathbf{e}'' \sim_2 \mathbf{e} \text{ and } \mathbf{e}'' \in V_{\mu+l-1}\}$. But $\mathbf{e} \sim_2 \mathbf{h}_{<\mu+l-1}$, so $e'_{\mu+l-1}(2) = h_{\mu+l-1}(2)$, and consequently $\mathbf{e}' \sim_2 \mathbf{h}$. Thus, $\mathbf{e}' \in h'_{\mu+l}(2)$. It follows that $\mathbf{h}' \in V_{\mu+l+1}$ and (K3$'$) holds. This completes the proof of part (2).

We now prove part (3). Let $\mathbf{h} \in U_\mu$. We construct a $(\mu+1)$-world $\mathbf{h}'$ in $V_{\mu+1}$. Let $\mathbf{h}'_{<\mu} = \mathbf{h}$, and $h'_\mu(i) = \{\mathbf{e} : \mathbf{e} \sim_i \mathbf{h} \text{ and } \mathbf{e} \in U_\mu\}$ for $i \in \mathcal{P}$. We verify that $\mathbf{h}'$ satisfies (K3$'$) as above. This completes the proof of part (3).

Let μ be a successor ordinal and let $\mathbf{h} \in V_\mu$. Then $\mathbf{h} \sup \mu + 1 \in U_{\mu+1}$. Let μ be a limit ordinal and let $\mathbf{h} \in U_\mu$. Then $\mathbf{h} \sup \mu + 1 \in U'_{\mu+1}$. This completes the proof of parts (4) and (5). Finally, part (6) follows from parts (1), (2), and (3). This completes the proof of Claim A.

Let θ_k be the formula $(\neg C_{\{2,3\}} K_1) \sup kp$, where $(\neg C_{\{2,3\}} K_1) \sup 0p$, is p, and $(\neg C_{\{2,3\}} K_1) \sup k+1p$ is $(\neg C_{\{2,3\}} K_1)(\neg C_{\{2,3\}} K_1) \sup kp$.

Claim B. Let $\mu < \omega \sup 2$, $\mu = \omega \times k + l$.

1. Let $k = 0$ and $l = 2$. If $\mathbf{h} \in U_2'$, then $\mathbf{h} \models \neg K_1 p$. If $\mathbf{h} \in U_2$, then $\mathbf{h} \models K_1 p$.

2. Let $k = 0$ and $l > 2$. If $\mathbf{h} \in U_\mu$, then $\mathbf{h} \models \neg E_{\{2,3\}} \sup l - 2K_1 p$. If $\mathbf{h} \in V_\mu$, then $\mathbf{h} \models E_{\{2,3\}} \sup l - 2K_1 p$.

3. Let $k > 0$ and $l = 0$. If $\mathbf{h} \in U_\mu$, then $\mathbf{h} \models \theta_k$. If $\mathbf{h} \in V_\mu$, then $\mathbf{h} \models \neg\theta_k$.

4. Let $k > 0$ and $l = 1$. If $\mathbf{h} \in U_\mu'$, then $\mathbf{h} \models \neg K_1\theta_k$. If $\mathbf{h} \in U_\mu$, then $\mathbf{h} \models K_1\theta_k$.

5. Let $k > 0$ and $l > 1$. If $\mathbf{h} \in U_\mu$, then $\mathbf{h} \models \neg E_{\{2,3\}} \sup l - 2K_1\theta_k$. If $\mathbf{h} \in V_\mu$, then $\mathbf{h} \models E_{\{2,3\}} \sup l - 2K_1\theta_k$.

Part (1) of the claim is obvious. We now prove part (2) by induction on l. Let $\mathbf{h} \in U_3$ and assume that $h_2(2)$ is the 2-equivalence class of $\mathbf{h}_{<2}$. Let $\mathbf{e}$ be the 2-matching extension of $\mathbf{h}_{<1}$ with respect to $\mathbf{h}_{<2}$. It is easy to see that $\mathbf{e} \in U_2'$ and $\mathbf{e} \sim_2 \mathbf{h}_{<2}$. Thus, $\mathbf{e} \in h_2(2)$, so $\mathbf{h} \models \neg E_{\{2,3\}} K_1 p$. Inductively, let $\mathbf{h} \in U_l$ and assume that $h_{l-1}(2)$ is the 2-equivalence class of $\mathbf{h}_{<l-1}$. Let $\mathbf{e}$ be the 2-matching extension of $\mathbf{h}_{<l-2}$ with respect to $\mathbf{h}_{<l-1}$. Then $\mathbf{e} \in U_{l-1}$ and $\mathbf{e} \in h_{l-1}(2)$. By induction, $\mathbf{e} \models \neg E_{\{2,3\}} \sup l - 3K_1 p$. It follows that $\mathbf{h} \models \neg E_{\{2,3\}} \sup l - 2K_1 p$.

We now prove that if $\mathbf{h} \in V_l$, $l > 1$, then $\mathbf{h} \models E_{\{2,3\}} \sup l - 2K_1 p$. This clearly holds for $l = 2$. Suppose now that $\mathbf{h} \in V_l$, $l > 2$. Let $\mathbf{e} \in h_{l-1}(2)$. By definition, $\mathbf{e} \in V_{l-1}$, so $\mathbf{e} \models E_{\{2,3\}} \sup l - 3K_1 p$. Thus, $\mathbf{h} \models K_2 E_{\{2,3\}} \sup l - 3K_1 p$. Similarly, $\mathbf{h} \models K_3 E_{\{2,3\}} \sup l - 3K_1 p$. It follows that $\mathbf{h} \models E_{\{2,3\}} \sup l - 2K_1 p$. This complete the proof of part (2).

Part (3) follows from parts (2) and (5) and the definition of U_μ and V_μ for limit ordinals μ.

We now prove part (4). Let $\nu = \omega \times k$. Let $\mathbf{h} \in U_{\nu+1}$ and $\mathbf{e} \in h_\nu(1)$. By definition, $\mathbf{e} \in U_\nu$, so by induction (part (3)) $\mathbf{e} \models \theta_k$. Thus, $\mathbf{h} \models K_1\theta_k$. Let $\mathbf{h} \in U_{\nu+1}'$. Suppose first that $k = 1$. Let $\mathbf{e} \in V_\omega$. We have that $\mathbf{e} \in h_\nu(1)$, since $\mathbf{e} \sim_1 \mathbf{h}_{<\omega}$ by construction. But $\mathbf{e} \models \neg\theta_1$ (by part (3)), so $\mathbf{h} \models \neg K_1\theta_1$. Suppose now that $k > 1$. Let $\mathbf{d}$ be $\mathbf{h}_{<\omega\times(k-1)}$. We know that $\mathbf{d} \in U_{\omega\times(k-1)}$. By Claim A, $\mathbf{d}$ can be extended to a world $\mathbf{e} \in V_\nu$ such that $\mathbf{e}_{<\omega\times(k-1)} = \mathbf{d}$ and $\mathbf{e} \sim_1 \mathbf{h}$. But $\mathbf{e} \models \neg\theta_k$, so $\mathbf{h} \models \neg K_1\theta_k$.

We prove Part (5) by induction on l. Let $\nu < \omega \sup 2$ be a limit ordinal. Let $\mathbf{h} \in U_{\nu+2}$ and assume that $h_{\nu+1}(2)$ is the 2-equivalence class of $\mathbf{h}_{<\nu+1}$. Let $\mathbf{e}$ be the 2-matching extension of $\mathbf{h}_{<\nu}$ with respect to $\mathbf{h}_{<\nu+1}$. It is easy to see that $\mathbf{e} \in U_{\nu+1}'$ and $\mathbf{e} \sim_2 \mathbf{h}_{<\nu+1}$. Thus, $\mathbf{e} \in h_{\nu+1}(2)$, so $\mathbf{h} \models \neg E_{\{2,3\}} K_1\theta_k$, since, by part (4), $\mathbf{e} \models \neg K_1\theta_k$. Inductively, let $\mathbf{h} \in U_{\nu+l}$, $l > 2$, and assume that $h_{\nu+l-1}(2)$ is the 2-equivalence class of $\mathbf{h}_{<\nu+l-1}$. Let $\mathbf{e}$ be the 2-matching extension of $\mathbf{h}_{<\nu+l-2}$ with respect to $\mathbf{h}_{<\nu+l-1}$. We have that $\mathbf{e} \in U_{\nu+l-1}$ and $\mathbf{e} \in h_{\nu+l-1}(2)$. By induction, $\mathbf{e} \models \neg E_{\{2,3\}} \sup l - 3K_1\theta_k$. It follows that $\mathbf{h} \models \neg E_{\{2,3\}} \sup l - 2K_1\theta_k$.

If $\mathbf{h} \in V_{\nu+1}$, then we also have $\mathbf{h} \in U_{\nu+1}$ by construction. Thus, by part (4), we have $\mathbf{h} \models K_1\theta_k$. Suppose now that $\mathbf{h} \in V_{\nu+l}$, $l > 1$. Let $\mathbf{e} \in h_{\nu+l-1}(2)$. By definition,

$\mathbf{e} \in V_{\nu+l-1}$, so $\mathbf{e} \models E_{\{2,3\}} \sup l - 3K_1\theta_k$ by induction. Thus, $\mathbf{h} \models K_2 E_{\{2,3\}} \sup l - 3K_1\theta_k$. Similarly, $\mathbf{h} \models K_3 E_{\{2,3\}} \sup l - 3K_1\theta_k$. It follows that $\mathbf{h} \models E_{\{2,3\}} \sup l - 2K_1\theta_k$. This completes the proof of part (5) and of Claim B.

Let $\mu = \omega \times k + l$, $k > 0$. Consider first the case that μ is a limit ordinal. Let $\mathbf{h} \in U_\mu$. By Claim A, there are worlds $\mathbf{f} \in U_{\mu+1}$ and $\mathbf{g} \in U'_{\mu+1}$ such that $\mathbf{f}_{<\mu} = \mathbf{g}_{<\mu} = \mathbf{h}$. By Claim B, we have $\mathbf{f} \models \neg K_1\theta_k$ and $\mathbf{g} \not\models \neg K_1\theta_k$. Consider now the case that μ is a successor ordinal. Let $\mathbf{h} \in V_\mu$. By Claim A, there are worlds $\mathbf{f} \in U_{\mu+1}$ and $\mathbf{g} \in V_{\mu+1}$ such that $\mathbf{f}_{<\mu} = \mathbf{g}_{<\mu} = \mathbf{h}$. By Claim B, we have $\mathbf{f} \models \neg E_{\{2,3\}} \sup l - 1K_1\theta_k$ and $\mathbf{g} \not\models \neg E_{\{2,3\}} \sup l - 1K_1\theta_k$. ∎

References

[Au] R. Aumann, Agreeing to disagree, *Annals of Statistics*, 4(1976), pp. 1236–1239.

[Ba] J. Barwise, Three views of common knowledge, *Proc. 2nd Conference on Theoretical Aspects of Reasoning about Knowledge*, M. Y. Vardi, ed., 1988, pp. 365–379.

[BD1] A. Brandenburger and E. Dekel, Common knowledge with probability one, *J. Math. Economics* 16(1987), pp. 237–245.

[BD2] A. Brandenburger and E. Dekel, *Hierarchies of beliefs and common knowledge*, Research Paper No. 841, Department of Economics, Stanford University, 1985.

[BS] R. Brachman and B. Smith (eds.) *SIGART Newsletter*, No. 70, Special issue on knowledge representation, 1980.

[CM] H.H. Clark and C.R. Marshall, Definite reference and mutual knowledge, in *Elements of Discourse Understanding* (A.K. Joshi, B.L. Webber, and I.A. Sag, eds.), Cambridge University Press, Cambridge, 1981.

[DM] C. Dwork and Y. Moses, Knowledge and common knowledge in a Byzantine environment: crash failures (extended abstract), *Proc. Conf. on Theoretical Aspects of Reasoning about Knowledge* (J. Y. Halpern, ed.), Morgan Kaufmann Publishers, 1986, pp. 149–170, to appear in *Information and Computation*.

[DS] D. Dolev and H.R. Strong, Authenticated algorithms for Byzantine agreement, *SIAM J. of Computing* 12(1983), pp. 656–666.

[FH] R. Fagin and J. Y. Halpern, Reasoning about knowledge and probability: preliminary report, *Proc. 2nd Conference on Theoretical Aspects of Reasoning about Knowledge* (M. Y. Vardi, ed.), 1988, pp. 277–293.

[FHV] R. Fagin, J. Y. Halpern, and M. Y. Vardi, What can machines know? On the epistemic properties of machines, *Proc. National Conf. on Artificial Intelligence (AAAI-86)*, 1986, pp. 428–434 (a revised and expanded version appears as IBM Research Report RJ 6250, 1988).

[FV1] R. Fagin and M. Y. Vardi, An internal semantics for modal logic, *Proc. 17th ACM Symp. on Theory of Computing*, Providence, 1985, pp. 305–315.

[FV2] R. Fagin and M. Y. Vardi, Knowledge and implicit knowledge in a distributed environment: preliminary report, *Proc. Conference on Theoretical Aspects of Reasoning about Knowledge* (J. Y. Halpern, ed.), 1986, pp. 187–206.

[Ha1] J. Y. Halpern, Reasoning about knowledge: an overview, *Proc. Conference on Theoretical Aspects of Reasoning about Knowledge* (J. Y. Halpern, ed.), 1986, pp. 1–17.

[Ha2] J. Y. Halpern, Using reasoning about knowledge to analyze distributed systems, in *Annual Review of Computer Science, Vol. 2*, J. Traub et al., eds., Annual Reviews Inc., 1987, pp. 37–68.

[HM1] J.Y. Halpern and Y.O. Moses, Knowledge and common knowledge in a distributed environment, *Proc. 3rd ACM Symp. on Principles of Distributed Computing*, 1984, pp. 50–61 (a revised and expanded version appears as IBM Research Report RJ 4421, 1988).

[HM2] J.Y. Halpern and Y.O. Moses, A guide to the modal logic of knowledge, *Proc. 9th Int'l Joint Conference on Artificial Intelligence (IJCAI-85)*, 1985, pp. 480–490.

[HV] J. Y. Halpern and M. Y. Vardi, The complexity of reasoning about knowledge and time, *Proc. 18th ACM Symp. on Theory of Computing*, 1986, pp. 304–315.

[Hi] J. Hintikka, *Knowledge and belief*, Cornell University Press, 1962.

[Ho] J. E. Hopcroft, An $n \log n$ algorithm for minimizing the states in a finite-automaton, in *Theory of Machines and Computations* (Z. Kohavi, ed.), Academic Press, 1971, pp. 189–196.

[HC] G.E. Hughes and M.J. Cresswell, *Modal Logic*, Methuen, 1968.

[Kr] S. Kripke, Semantical analysis of modal logic, *Zeitschrift fur Mathematische Logik und Grundlagen der Mathematik* 9 (1963), pp. 67–96.

[La] R. E. Ladner, The computational complexity of provability in systems of modal propositional logic, *SIAM J. Comput.* 6(1977), pp. 467–480.

[Le] D. Lehmann, Knowledge, common knowledge, and related puzzles, *Proc. 3rd ACM Symp. on Principles of Distributed Computing*, 1984, pp. 62–67.

[Mc] J. McCarthy, personal communication, 1984.

[MH] J. McCarthy and P. Hayes, Some philosophical problems from the standpoint of artificial intelligence, in *Machine Intelligence 4* (D. Michie, ed.), American Elsevier, 1969, pp. 463–502.

[MSHI] J. McCarthy, M. Sato, T. Hayashi, S. Igarashi, On the model theory of knowledge, Stanford AI Lab Memo AIM–312, Stanford University, 1978.

[MZ] J. F. Mertens and S. Zamir, Formulation of Bayesian analysis for games with incomplete information, *Int'l J. Game Theory*, 14(1985), pp. 1–29.

[Mi] P. Milgrom, An axiomatic characterization of common knowledge, *Econometrica* 49, 1(1981), pp. 219–222.

[MS] D. Monderer and D. Samet, Approximating common knowledge with common beliefs, *Games and Economic Behavior* 1(1989), pp. 170–190.

[Mo1] R. C. Moore, A formal theory of knowledge and action, in *Formal Theories of the Commonsense World*, J. Hobbs and R. C. Moore, eds., Ablex Publishing Corp., 1985.

[Mo2] R.C. Moore, Semantical considerations on nonmonotonic logic, *Artificial Intelligence* 25(1985), pp. 75–94.

[MT] Y. Moses and M. Tuttle, Programming simultaneous actions using common knowledge, *Proceedings of the 27th Annual Symposium on Foundations of Computer Science*, 1986, pp. 208–221.

[Pa] D. Park, Concurrency and automata on infinite sequences, in *Proc. 5th GI Conf. on Theoretical Computer Science* (P. Deussen, ed.), Lecture Notes in Computer Science 104, Springer-Verlag, 1981.

[PSL] M. Pease, R. Shostak, and L. Lamport, Reaching agreement in the presence of faults, *J. ACM* 27(1980), pp. 228–234.

[Ro] S.J. Rosenschein, Formal theories of knowledge in AI and robotics, *New Generation Computing* **3**, 1985, pp. 345–357.

[RK] S.J. Rosenschein and L.P. Kaelbling, The synthesis of digital machines with provable epistemic properties, *Theoretical Aspects of Reasoning about Knowledge* (J. Y. Halpern, ed.), Morgan Kaufmann, 1986, pp. 83–98.

[Sa] M. Sato, *A Study of Kripke-type Models for Some Modal Logics*, Research Institute for Mathematical Science, Kyoto University, Kyoto, Japan, 1976.

[TW] T. Tan and S. Werlang, On Aumann's notion of common knowledge – an alternative approach, Discussion Paper in Economics and Econometrics 85-26, University of Chicago, 1985.

[VSG] P. van Emde Boas, J. Groenendijk, M. Stokhof, The Conway paradox: its solution in an epistemic framework. *Proc. 3rd Amsterdam Montague Symp.*, 1980.

[Va1] M. Y. Vardi, A model-theoretic analysis of monotonic knowledge, *Proc. 9th Int'l Joint Conference on Artificial Intelligence (IJCAI-85)*, 1985, pp. 509–512.

[Va2] M. Y. Vardi, On the complexity of epistemic reasoning. Proc. 4th IEEE Symp. on Logic in Computer Science, June 1989, pp. 243–252.