A THEORY OF KNOWLEDGE AND IGNORANCE FOR MANY AGENTS*

Joseph Y. Halpern

IBM Research Division
Almaden Research Center, Dept. K53/802
650 Harry Road
San Jose, CA 95120–6099
halpern@almaden.ibm.com

Abstract

We extend the notion of "only knowing" introduced by Halpern and Moses [11] to many agents and to a number of modal logics. In this approach, "all an agent knows is α" is true in a structure M if, in M, the agent knows α and has a maximum set of "possibilities". To extend this approach, we need to make precise what counts as a "possibility". In the single-agent case, we can identify a possibility with a truth assignment. In the multi-agent case, things are more complicated. We consider three notions of possibility (all related). We argue that the first is most appropriate for non-introspective logics, such as K_n, T_n, and $S4_n$, the second is most appropriate for $K45_n$ and $KD45_n$, and the last is most appropriate for $S5_n$. With the appropriate notion of possibility, we show that are reasonable extensions in all cases.

Our results also shed light on the single-agent case. It was always assumed that one of the key aspects of Halpern-Moses approach in the single-agent case was its use of S5, rather than K45 or KD45. Our results show that the notion is better understood in the context of K45 (or KD45). In the single-agent case, the notion remains unchanged if we use K45 instead of S5. However, in the multi-agent case, there are significant differences between K45 and S5. Moreover, in some sense, the K45 variants behave better: all results proved for the single-agent case extend more naturally to the multi-agent case of K45 than to the multi-agent case of S5.

*This paper is essentially identical to one that will appear in *Journal of Logic and Computation*. The work of the author is sponsored in part by the Air Force Office of Scientific Research (AFSC), under Contract F49620-91-C-0080. The United States Government is authorized to reproduce and distribute reprints for governmental purposes. Some of the material in this paper appeared in preliminary form in a paper entitled "Reasoning about only knowing with many agents" that appeared in *AAAI-93 (Proceedings of the Twelfth National Conference on Artificial Intelligence, 1993).*

1 Introduction

Halpern and Moses [11] introduced a notion of "only knowing", in an effort to characterize the state of an agent that has been told only a finite number of facts. Suppose that α is the conjunction of what the agent has been told. What else does an agent who has been told only α know? It is not just the logical consequences of α. For example, if α is the primitive proposition p, then the agent does not know q, where q is some other primitive proposition. But if the agent is introspective, then the agent knows that he does not know q.

Halpern and Moses give a number of (provably equivalent) characterizations of what it means for an agent to "only know α", assuming that there is only one agent in the system, and that agent's knowledge is characterized by the modal logic S5. The latter means that the agent does not know false facts, and that the agent has complete introspective power regarding his own knowledge.

As pointed out by Halpern and Moses, extending the definitions of "only knowing" to the multi-agent case is quite subtle. They sketched arguments showing why none of the obvious ways of extending the definitions used in the single-agent case would give reasonable answers in the multi-agent case. In this paper, we show how the definitions can be generalized to the multi-agent case. Moreover, we extend the definitions to other logics besides S5.

To make the notion of "knowing only α" precise, we need to consider the Kripke structure where the agent knows α and has a maximum set of "possibilities". While the notion of "possibility" is straightforward in the single-agent case—it can be identified with a truth assignment—it becomes more subtle once we have many agents in the picture. Indeed, we argue that the right notion depends on the logic we are considering. We use three different notion of "possibility": one for K, T, and S4, one for K45 and KD45, and one for S5. The reasons we do this highlight the differences between logics with negative introspection and those without, and the added complexities involved with S5, which has both negative introspection and the *veridicality* property: only true things are known.

Our results also shed light on the single-agent S5 case. For example, it was always viewed as significant that the HM (Halpern-Moses) notion of only knowing was based on S5, while a different notion of only knowing considered by Levesque [19] is based on K45. In fact, our results show that the HM notion is better understood in the context of K45. Indeed, in the single-agent case, the HM notion remains unchanged if we use K45 (or KD45) instead of S5. However, in the multi-agent case, there are significant differences between K45 and S5. Moreover, as we show here, all the results proved by Halpern and Moses for the single-agent case extend more naturally to the multi-agent case for K45 and KD45 than they do for S5.

The key technical tool we use to define our notions of possibility, the ω-tree, is closely related to the tools used in a number of other papers that have attempted to define HM-like notions of only knowing for many agents: the *knowledge structures* defined by Fagin, Halpern, and Vardi [4], which were also used by Vardi [24], the *normal models* defined by Parikh [21], and the use of *amalgamation* by Jaspars [15]. As we shall see, our approach seems to allow for much more efficient decision procedures.

The rest of this paper is organized as follows. In the next section, we present a brief review of the relevant details of modal logic. In Section 3, we discuss general issues concerning the definition of "only knowing" in the multi-agent case. In Section 4, we define only knowing for (multi-agent) K, T, and S4, in Section 5, we do it for (multi-agent) K45 and KD45, and in Section 6, we do it for S5. In Section 7 we consider some complexity issues. We conclude in Section 8 with some further discussion of the notion of only knowing.

2 A brief review of modal logic

We briefly review some standard notions of modal logic here. Further details can be found in, for example, [2, 12, 13]. We recommend that even the reader familiar with modal logic scan this section, since it also introduces some notation that is used in the remainder of the paper.

In this paper we focus on six logics, three that do not have negative introspection, K_n, T_n, and $\mathrm{S4}_n$, and three that do, $\mathrm{K45}_n$, $\mathrm{KD45}_n$, and $\mathrm{S5}_n$.[1] In the remainder of this paper, when we speak of a modal logic $\mathcal{S}$, we are referring to one of these six logics; we refer to $\mathrm{K45}_n$, $\mathrm{KD45}_n$ and $\mathrm{S5}_n$ as *introspective logics*, and K_n, T_n, and $\mathrm{S4}_n$ as *non-introspective logics* (despite the fact that positive introspection holds in $\mathrm{S4}_n$).

The language we use for all these logics is propositional logic augmented by the modal operators $K_1, \ldots, K_n$, where $K_i\varphi$ can be read "agent i knows (or believes) φ". We denote this language $\mathcal{L}_n$.

Consider the following collection of axioms, which hold for each agent i:

P. All instances of axioms of propositional logic

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

T. $K_i\varphi \Rightarrow \varphi$

4. $K_i\varphi \Rightarrow K_iK_i\varphi$

5. $\neg K_i\varphi \Rightarrow K_i\neg K_i\varphi$

D. $\neg K_i \mathit{false}$

and rules of inference:

R1. From φ and $\varphi \Rightarrow \psi$ infer ψ

R2. From φ infer $K_i\varphi$.

The axioms **4** and **5** are called the *positive introspection axiom* and *negative introspection axiom*, respectively. They are appropriate for agents that are sufficiently introspective so that they know what they know and do not know.

We get various systems by combining some subset of **K**, **T**, **4**, **5**, and **D** with **P**, **R1**, and **R2**. In particular, we get K_n by combining **K** with **P**, **R1**, and **R2**, T_n by adding **T** to these axioms, $\mathrm{S4}_n$ by adding **4**, $\mathrm{S5}_n$ by adding **5**, $\mathrm{K45}_n$ by deleting **T** from $\mathrm{S5}_n$, and $\mathrm{KD45}_n$ by adding **D** to $\mathrm{K45}_n$. Other modal logics can be constructed by considering different combinations of axioms.

[1] The subscript n in all these logics is meant to emphasize the fact that we are considering the n-agent version of the logic. We omit it when considering the single-agent case.

We give semantics to all these logics by using Kripke structures. A Kripke structure is a tuple $(W, \pi, \mathcal{K}_1, \ldots, \mathcal{K}_n)$, where W is a set of *worlds*, π associates with each world a truth assignment to the primitive propositions, so that $\pi(w)(p) \in \{\mathbf{true}, \mathbf{false}\}$ for each world w and primitive proposition p, and $\mathcal{K}_1, \ldots, \mathcal{K}_n$ are binary *accessibility* relations. We use the notation $\mathcal{K}_i^M$ when we want to refer to the $\mathcal{K}_i$ relation in the structure M; similarly, we use W^M and π^M. We omit the superscript M if it is clear from context. We define $\mathcal{K}_i(w)$ to be $\{w' : (w, w') \in \mathcal{K}_i\}$. Thus, $\mathcal{K}_i(w)$ is the set of worlds agent i considers possible in world w.

Recall that a binary relation $\mathcal{K}$ on W is *reflexive* if $(w, w) \in \mathcal{K}$ for all $w \in W$, *transitive* if $(u, v) \in \mathcal{K}$ and $(v, w) \in \mathcal{K}$ implies $(u, w) \in \mathcal{K}$, *Euclidean* if $(u, v) \in \mathcal{K}$ and $(u, w) \in \mathcal{K}$ implies $(v, w) \in \mathcal{K}$, and *serial* if for all $w \in W$, there is some w' such that $(w, w') \in \mathcal{K}$. Let $\mathcal{M}$ be the class of all Kripke structures. We restrict $\mathcal{M}$ by using superscripts r, s, t, and e, to denote reflexive, serial, transitive, and Euclidean structures, respectively. Thus, $\mathcal{M}^{rt}$ denotes the class of all structures where the $\mathcal{K}_i$ relations are reflexive and transitive knowledge, $\mathcal{M}^{est}$ denotes the class of all structures where the $\mathcal{K}_i$ relations are Euclidean, serial, and transitive, and so on.

A *situation* is a pair (M, w) consisting of a Kripke structure and a world w in M. We give semantics to formulas with respect to situations. If p is a primitive proposition, then $(M, w) \models p$ if $\pi^M(w)(p) = \mathbf{true}$. Conjunctions and negations are dealt with in the standard way. Finally,

$$(M, w) \models K_i\alpha \text{ iff } (M, w') \models \alpha \text{ for all } w' \in \mathcal{K}_i^M(w).$$

Thus, agent i knows α exactly if α is true in all situations that the agent considers possible.

It is well known that there is a close connection between conditions placed on $\mathcal{K}$ and the axioms. In particular, **T** corresponds to the $\mathcal{K}_i$s being reflexive, **4** to the $\mathcal{K}_i$s being transitive, **5** to the $\mathcal{K}_i$s being Euclidean, and **D** to the $\mathcal{K}_i$s being serial. Thus, we get the following result (see [2, 12] for proofs):

Theorem 1 K_n (resp. T_n, $\mathrm{S4}_n$, $\mathrm{KD45}_n$, $\mathrm{K45}_n$, $\mathrm{S5}_n$) is a sound and complete axiomatization for the language $\mathcal{L}_n$ with respect to $\mathcal{M}$ (resp., $\mathcal{M}^r$, $\mathcal{M}^{rt}$, $\mathcal{M}^{est}$, $\mathcal{M}^{et}$, $\mathcal{M}^{ret}$).[2]

An $\mathcal{S}$ *situation* (for $\mathcal{S} \in \{\mathrm{K}_n, \mathrm{T}_n, \mathrm{S4}_n, \mathrm{K45}_n, \mathrm{KD45}_n, \mathrm{S5}_n\}$) is a situation (M, w) where M satisfies the appropriate restriction; thus, for example, (M, w) is a $\mathrm{S4}_n$ situation if $M \in \mathcal{M}^{rt}$. We write $\models_{\mathcal{S}} \varphi$ if the formula φ is true in all $\mathcal{S}$ situations. By Theorem 1, for a formula $\varphi \in \mathcal{L}_n$, we have $\models_{\mathcal{S}} \varphi$ iff φ is provable in $\mathcal{S}$.

It is well known (again, see [2], [13], or [12]) that in the single-agent case of $\mathrm{KD45}_n$, $\mathrm{K45}_n$, and $\mathrm{S5}_n$, we can consider a simpler class of structures. We define a *K45 situation* to be a pair (W, w), where W is a set of truth assignments that, intuitively, characterize the worlds the agent considers possible, and w is a truth assignment that, intuitively, characterizes the "real world". A *KD45 situation* is a

[2]The more common characterization is that $\mathrm{S5}_n$ is sound and complete with respect to the class of structures where the $\mathcal{K}_i$s are equivalence relations. However, as observed in [12], $\mathcal{K}$ is an equivalence relation iff $\mathcal{K}$ is reflexive, Euclidean, and transitive. Thus, $\mathcal{M}^{ret}$ in fact consists of precisely those structures where the $\mathcal{K}_i$s are equivalence relations. Moreover, it is easy to see that reflexive and Euclidean relations must be transitive, so that $\mathcal{M}^{re}$ is identical to $\mathcal{M}^{ret}$.

K45 situation (W, w) such that $W \neq \emptyset$. An *S5 situation* (W, w) is a K45 situation such that $w \in W$.

We again give semantics to formulas with respect to situations. If p is a primitive proposition, then $(W, w) \models p$ if p is true under truth assignment w. Conjunctions and negations are dealt with in the standard way. Finally,

$$(W, w) \models K\alpha \text{ iff } (W, w') \models \alpha \text{ for all } w' \in W.^3$$

It is well known [13] that a formula is provable in K45 (resp. KD45, S5) if and only if it is true in all K45 (resp. KD45, S5) situations.

3 "Only knowing" in the multi-agent case

The intuition behind the HM notion of "all I know" in the single-agent case is straightforward: In each world of a (Kripke) structure, an agent considers a number of other worlds possible. In the case of a single agent whose knowledge satisfies S5 (or K45 or KD45), as we observed in Section 2, we can identify a world with a truth assignment, and a structure with a set of truth assignments. The more worlds an agent considers possible, the less he knows. We take (W, w) to be a situation where α is all that is known if (1) $(W, w) \models K\alpha$ (so that the agent knows α) and (2) if $(W', w') \models K\alpha$, then $W' \subseteq W$. If there is no situation (W, w) satisfying (1) and (2), then α is said to be *dishonest*; intuitively, it cannot then be the case that "all the agent knows" is α. A typical dishonest formula is $Kp \vee Kq$. To see that this formula is dishonest, let W_p consist of all truth assignments satisfying p, let W_q consist of all truth assignments satisfying q, and let w satisfy $p \wedge q$. Then $(W_p, w) \models Kp \vee Kq$, and $(W_q, w) \models Kp \vee Kq$. Thus, if $Kp \vee Kq$ were honest, there would have to be a situation (W, w') such that $(W, w') \models Kp \vee Kq$ and $W \supseteq W_p \cup W_q$. It is easy to see that no such situation exists.

Notice that in the case of one agent, it does not matter if we consider S5, KD45, or K45 when speaking of honesty. If a formula that is not equivalent to *false* is honest with respect to one of these logics, then it is honest with respect to the other two.[4]

We want to extend this intuition to the multi-agent case and—in order to put these ideas into better perspective—to other modal logics. There are philosophical problems involved in dealing with a notion of "all I know" for the non-introspective logics. What does it mean for an agent to say "all I know is α" if he cannot do negative introspection, and so does not know what he doesn't know? Fortunately, there is another interpretation of this approach that makes sense for arbitrary modal logics. Suppose that a says to b, "All c knows is α" (where c is different from a and b). If b knows in addition that c's reasoning satisfies the axioms of modal logic $\mathcal{S}$, then it seems reasonable for b to say that c's knowledge is described by the "minimal" model satisfying the axioms of $\mathcal{S}$ consistent with $K_c\alpha$, and for b to view a as dishonest if there is no such minimal model.

This suggests that making sense of "all I know" (or "all agent i knows") reduces to defining what it means for a model to be "minimal". Once we consider multi-agent logics, or even nonintrospective single-agent logics, we can no longer identify a

[3] When dealing with logics like K45, where only one agent is involved, we typically do not subscript the K operator, writing, for example, $K\alpha$ rather than $K_i\alpha$.

[4] Note that *false* is honest with respect to K45 but not with respect to KD45 or S5.

possible world with a truth assignment. It is not just the truth assignment at a world that matters; we also need to consider what other worlds are accessible from that world. This makes it more difficult to define a reasonable notion of minimality. To deal with this problem, we define a canonical collection of "possibilities", i.e., objects that an agent can consider possible. These will act like the possible truth assignments in the single-agent case.

What properties should we expect the sets of possible objects to have? It seems reasonable to expect that they satisfy the following two conditions:

1. The *determination* condition: The agent's set of possibilities should determine his knowledge. That is, if agent i has the same set of possibilities in (M, w) and in (M', w'), then $(M, w) \models K_i\varphi$ iff $(M', w') \models K_i\varphi$ for all formulas φ.

2. The *union* condition: If in some situation (M_1, w_1) agent i's set of possibilities is P_1, and in some other situation (M_2, w_2) agent i's set of possibilities is P_2, then there should be some situation (M_3, w_3) where the agent's set of possibilities contains $P_1 \cup P_2$.

These conditions are certainly quite weak. For example, we might hope for a converse to the first condition, but this is too much to expect given the lack of expressive power of the modal logics we are considering. For example, consider the single-agent case. If we take the notion of "possibility" here to be a truth assignment, then clearly these two conditions hold. But now suppose we have infinitely many primitive propositions, say $p_1, p_2, \ldots$. Let W_1 consist of all truth assignments to these propositions, and let W_2 consist of all truth assignments except the one that makes all of the propositions true. It is easy to check (by induction on the structure of φ) that for all formulas φ and any truth assignment $w \in W_1$, we have $(W_1, w) \models K_i\varphi$ iff $(W_2, w) \models K_i\varphi$. Thus, what the agent knows does not determine the set of possibilities in this case. We might also hope to have the second condition hold with equality, but, as we shall see, this too turns out to be too much to expect in general (although it does hold for K_n and the introspective logics).

Despite their apparent weakness, the determination and union conditions do serve as useful guidelines for our constructions. Among other things, they are sufficient to show that we cannot use truth assignments as our notion of possibility in the multi-agent case, or even in the single-agent case for K, T, and S4, since this would violate the determination condition. For example, the set of truth assignments agent 1 considers possible clearly does not determine whether K_1K_2p holds.

Notice that these two conditions may hold for a notion of possibility with respect to one logic and not another. Indeed, we use three different notions of possibility in this paper, one for K_n, T_n, and $\mathrm{S4}_n$, another for $\mathrm{K45}_n$ and $\mathrm{KD45}_n$, and yet another for $\mathrm{S5}_n$. While the notions of possibility we use satisfy the union and determination conditions for each of these logics, it is not clear that these are the only choices that could have been made. There may be other notions of possibility that satisfy these conditions that are not isomorphic to the ones we use (for an appropriate notion of isomorphism). While we believe that we have made the "right" choices, there is no theory yet to support this. We defer further discussion of this issue to Section 8. In the next three sections, we define these three notions of possibility, and show how they can be used to define "only knowing".

4 "Only knowing" for K_n, T_n, and $\mathrm{S4}_n$

In this section, we focus on the logics K_n, T_n, and $\mathrm{S4}_n$. We start by defining a notion of possibility for these logics.

Fix a finite set Φ of primitive propositions and agents $1, \ldots, n$.[5] We define a *(rooted, directed, and labeled) k-tree (over Φ)* by induction on k: A 0-tree consists of a single node, labeled by a truth assignment to the primitive propositions in Φ. A $(k+1)$-tree consists of a root node r labeled by a truth assignment, and for each agent i, a (possibly empty) set of directed edges labeled by i leading from r to roots of distinct k-trees.[6] We say a node w' is an *i-successor* of a node w in a tree if there is an edge labeled i leading from w to w'. The *depth* of a node in a tree is the distance of the node from the root.

An ω-tree T_ω is a sequence $\langle T_0, T_1, \ldots \rangle$, where T_k is a k-tree, for $k = 0, 1, 2, \ldots$. Notice that consecutive elements T_k and T_{k+1} in the sequence may be completely unrelated. Of course, in the ω-trees that we shall be interested in, T_k will be in a precise sense a projection of T_{k+1}. However, it turns out to be unnecessary to make this requirement in the general definition, and it would complicate the definition unnecessarily.

We remark that ω-trees are closely related to the *knowledge structures* of [4, 5]. In a precise sense, an ω-tree can be viewed as a way of representing a knowledge structure. Since the details of the comparison are beyond the scope of this paper, we do not pursue this connection here. They are also much in the spirit of Parikh's *normal models* [21]. More generally, the use of tree-like structures is quite standard in modal logic. They have played a role in many contexts in modal logic, including complexity-related arguments [12, 16] and completeness proofs (for example, the *subordination frames* and *tree frames* used in completeness proofs in [14] are treelike). Hughes and Cresswell also introduce a technique of *amalgamation* of structures into what can be viewed as one treelike structure. Amalgamation is used by Jaspars [15] in his analysis of only knowing for S4.

We now show that with each situation we can associate a unique ω-tree. We start by going in the other direction. We can associate with each k-tree T ($k \neq \omega$) a Kripke structure $M(T)$ in a straightforward way: the nodes of T are the possible worlds in $M(T)$, the accessibility relation $\mathcal{K}_i^{M(T)}$ consists of all pairs (w, w') such that w' is an i-successor of w in T, and $\pi^{M(T)}(w)$ is determined by the truth assignment labeling w.

We define the *depth* of a formula by induction on structure. Intuitively, the depth measures the depth of nesting of the K_i operators. Thus, we have $depth(p) = 0$ for a primitive proposition p; $depth(\neg\varphi) = depth(\varphi)$; $depth(\varphi \wedge \psi) = \max(depth(\varphi), depth(\psi))$; $depth(K_i\varphi) = 1 + depth(\varphi)$. If M and M' are (arbitrary) structures, w is a world in M, and w' a world in M', then we say that (M, w) and (M', w') are *equivalent up to depth k*, and write $(M, w) \equiv_k (M', w')$, if, whenever φ is a formula with $depth(\varphi) \leq k$, we have $(M, w) \models \varphi$ iff $(M', w') \models \varphi$. We say that (M, w) and (M', w') are *equivalent*, and write $(M, w) \equiv (M', w')$, if $(M, w) \equiv_k (M', w')$ for all k. Finally, we say that (M, w) and (M', w') are *i-equivalent*, and write $(M, w) \equiv^i (M', w')$, if $(M, w) \models K_i\varphi$ iff $(M', w') \models K_i\varphi$ for all formulas φ. Thus, equivalent situations

[5] As we shall see below, the assumption that Φ is finite makes some of our results a little simpler to state, but all our results hold (occasionally with minor modifications) even if Φ is infinite.

[6] Since we are allowing a node to have no successors, any k-tree is also a $(k+1)$-tree.

agree on all formulas, while i-equivalent situations agree on all formulas of the form $K_i\varphi$. For convenience, if w_0 is the root of T, we take $M(T) \models \varphi$ to be an abbreviation for $(M(T), w_0) \models \varphi$, and write $(M, w) \equiv_k M(T)$ rather than $(M, w) \equiv_k (M(T), w_0)$.

Proposition 2 For each situation (M, w) and all k, there is a unique k-tree $T_{M,w,k}$ such that $(M, w) \equiv_k M(T_{M,w,k})$.[7]

Proof We construct $T_{M,w,k}$ for each world $w \in W^M$ by induction on k. For each world w, we take $T_{M,w,0}$ to consist of a single node, labeled by the truth assignment at w. Clearly $(M, w) \equiv_0 M(T_{M,w,0})$ for all worlds $w \in W^M$.

Suppose inductively that for each world $w \in W^M$, we have constructed a tree $T_{M,w,k}$ such that $(M, w) \equiv_k M(T_{M,w,k})$ and shown that it is the unique tree with this property. We construct $T_{M,w,k+1}$ as follows. We take the root of $T_{M,w,k+1}$ to be a node labeled by the truth assignment at w. For each world w' such that $(w, w') \in \mathcal{K}_i$, we construct an edge labeled i to the root of $T_{M,w',k}$. This may not give us a $(k+1)$-tree, since there may be worlds w' and w'' such that both (w, w') and (w, w'') are in $\mathcal{K}_i$, and $T_{M,w',k}$ and $T_{M,w'',k}$ are identical. We obtain the $(k + 1)$-tree $T_{M,w,k+1}$ by deleting duplicate k-xrees.

We defer the proof that $(M, w) \equiv_{k+1} M(T_{M,w,k+1})$ and that $T_{M,w,k+1}$ is the unique tree with this property to the appendix, where the proof of all other technical results can also be found.

Let $T_{M,w}$ be the ω-tree $\langle T_{M,w,0}, T_{M,w,1}, T_{M,w,2}, \ldots\rangle$. As an immediate corollary to Proposition 2, we get that two situations that are associated with the same ω-tree are equivalent.

Corollary 3 $T_{M,w} = T_{M',w'}$ iff $(M, w) \equiv (M', w')$.

Thus, $T_{M,w}$ can be viewed as providing a canonical way of representing the situation (M, w) in terms of trees.

We use ω-trees as a tool for defining what agent i considers possible in (M, w). Thus, we define i's possibilities at (M, w) for $\mathcal{S} \in \{\mathrm{K}_n, \mathrm{T}_n, \mathrm{S4}_n\}$, denoted $Poss_i^{\mathcal{S}}(M, w)$, to be $\{T_{M,w'} : (w, w') \in \mathcal{K}_i\}$. The following two propositions say that this notion of possibility does satisfy the two requirements we made. The first says that the determination condition holds, while the second says that the union condition holds.

Proposition 4 If $\mathcal{S} \in \{\mathrm{K}_n, \mathrm{T}_n, \mathrm{S4}_n\}$ and $Poss_i^{\mathcal{S}}(M, w) = Poss_i^{\mathcal{S}}(M', w')$, then $(M, w) \equiv^i (M', w')$.

Proposition 5 For $\mathcal{S} \in \{\mathrm{K}_n, \mathrm{T}_n, \mathrm{S4}_n\}$, all agents i, and all $\mathcal{S}$ situations (M_1, w_1) and (M_2, w_2), there is an $\mathcal{S}$ situation (M_3, w_3) such that $Poss_i^{\mathcal{S}}(M_3, w_3) \supseteq Poss_i^{\mathcal{S}}(M_1, w_1) \cup Poss_i^{\mathcal{S}}(M_2, w_2)$.

The proof of Proposition 5 in the appendix shows that we can have the union condition hold with equality in the case of K_n. It is not hard to see that it cannot hold with

[7] We remark that the uniqueness here depends on the fact that the set Φ of primitive propositions which forms the basis of the language is finite. If Φ is infinite, the construction in the proof gives us a way of constructing a canonical tree $T_{M,w,k}$ such that $(M, w) \equiv M(T_{M,w,k})$. All our later results go through even if Φ is infinite, using this canonical tree. However, as we shall see later in this section, we have another way of dealing with the case that Φ is infinite.

equality for T_n and $\mathrm{S4}_n$. For example, suppose M_1 consists of only one world, w_1, such that p is true at w_1 and $(w_1, w_1) \in \mathcal{K}_1^{M_1}$. Similarly, suppose that M_2 consists of only one world, w_2, such that $\neg p$ is true at w_2 and $(w_2, w_2) \in \mathcal{K}_1^{M_2}$. Now suppose that (M, w_3) is a T_n situation such that $Poss_1^{\mathcal{S}}(M_3, w_3) \supseteq Poss_1^{\mathcal{S}}(M_1, w_1) \cup Poss_1^{\mathcal{S}}(M_2, w_2)$. This means that agent 1 must consider at least two worlds possible in w_3, one in which p holds and one in which $\neg p$ holds. It is thus easy to see that T_{M_3,w_3} can be in neither $Poss_1^{\mathcal{S}}(M_1, w_1)$ nor $Poss_1^{\mathcal{S}}(M_2, w_2)$. Since (M_3, w_3) is a T_n situation, it follows that $T_{M_3,w_3} \in Poss_1^{\mathcal{S}}(M_3, w_3)$. It follows that $Poss_1^{\mathcal{S}}(M_3, w_3) \neq Poss_1^{\mathcal{S}}(M_1, w_1) \cup Poss_1^{\mathcal{S}}(M_2, w_2)$. We remark that in the next section, we show that this notion of possibility does not satisfy even this weak union condition for the introspective logics.

Intuitively, for α to be i-honest, there should be a situation (M, w) for which i has the maximum number of possibilities. Formally, we say that α is ***$\mathcal{S}$-i-honest*** if there is an $\mathcal{S}$ situation (M, w), called an *$\mathcal{S}$-i-maximum situation for α*, such that $(M, w) \models K_i\alpha$, and for all $\mathcal{S}$ situations (M', w'), if $(M', w') \models K_i\alpha$, then $Poss_i^{\mathcal{S}}(M', w') \subseteq Poss_i^{\mathcal{S}}(M, w)$. If α is $\mathcal{S}$-i-honest, we say that agent i knows β if all he knows is α, and write $\alpha \mid\!\sim_{\mathcal{S}}^{i} \beta$, if $(M, w) \models K_i\beta$ for some $\mathcal{S}$-i-maximum situation (M, w) for α.[8] So far we have defined $\mathcal{S}$-i-honesty and $\mid\!\sim_{\mathcal{S}}^{i}$ only for $\mathcal{S} \in \{\mathrm{K}_n, \mathrm{T}_n, \mathrm{S4}_n\}$; as we shall see, the definitions carry over without change to other modal logics $\mathcal{S}$, once we define $Poss_i^{\mathcal{S}}$ for these logics.

How reasonable are our notions of honesty and $\mid\!\sim_{\mathcal{S}}^{i}$? The following results give us some justification for these definitions. The first gives us a natural characterization of honesty.

Theorem 6 If $\mathcal{S} \in \{\mathrm{K}_n, \mathrm{T}_n, \mathrm{S4}_n\}$, then the formula α is $\mathcal{S}$-i-honest iff (a) $K_i\alpha$ is $\mathcal{S}$-consistent and (b) for all formulas $\varphi_1, \ldots, \varphi_k$, if $\models_{\mathcal{S}} K_i\alpha \Rightarrow (K_i\varphi_1 \vee \ldots \vee K_i\varphi_k)$, then $\models_{\mathcal{S}} K_i\alpha \Rightarrow K_i\varphi_j$ for some $j \in \{1, \ldots, k\}$.

This characterization of honesty is similar in spirit to what was called by Lemmon and Scott [18] the *rule of disjunction*. A modal logic $\mathcal{S}$ satisfies this rule if $\models_{\mathcal{S}} K_i\varphi_1 \vee \ldots K_i\varphi_k$ implies $\models_{\mathcal{S}} \varphi_j$ for some $j \in \{1, \ldots, n\}$. In [14] it is shown that K, T, and S4 satisfy this rule. The technique used, *amalgamation*, is the basis for our proof of Theorem 6. We remark that it is quite easy to show that K45, KD45, and S5 do not satisfy this rule (for example, $Kp \vee K\neg Kp$ is valid in each of these logics, even though p is not). Nevertheless, as we shall see, a somewhat analogous property holds for these logics as well.

It follows from Theorem 6 that a typical dishonest formula in the case of T_n or $\mathrm{S4}_n$ is $K_ip \vee K_iq$, where p and q are primitive propositions. If α is $K_ip \vee K_iq$, then $K_i\alpha \Rightarrow (K_ip \vee K_iq)$ is valid in T_n and $\mathrm{S4}_n$, although neither $K_i\alpha \Rightarrow K_ip$ nor $K_i\alpha \Rightarrow K_iq$ is valid. However, the validity of $K_i\alpha \Rightarrow (K_ip \vee K_iq)$ depends on the fact that $K_i\alpha \Rightarrow \alpha$. This is not an axiom of K_n. In fact, it can be shown that $K_ip \vee K_iq$ is K_n-i-honest. Thus, what is almost the archetypical "dishonest" formula is honest in the context of K_n. As the following result shows, this is not an accident.

Theorem 7 All formulas are K_n-i-honest.

[8] There may be more than one $\mathcal{S}$-i-maximum situation for α; two $\mathcal{S}$-i-maximum situations for α may differ in what $j \neq i$ considers possible. However, if (M, w) and (M', w') are two $\mathcal{S}$-i-maximum situations for α, then $(M, w) \models K_i\beta$ iff $(M', w') \models K_i\beta$. Thus, our notion of $\mid\!\sim_{\mathcal{S}}^{i}$ is well defined.

A set S of formulas is an $\mathcal{S}$-i-*stable set* if there is some $\mathcal{S}$ situation (M, w) such that $S = \{\varphi : (M, w) \models K_i\varphi\}$. We say the situation (M, w) *corresponds* to the stable set S. This definition is a generalization of the one given by Moore [20] (which in turn is based on Stalnaker's definition [22]); Moore's notion of stable set corresponds to a K45-stable set in the single-agent case. (See [7] for some discussion as to why this notion of stable set is appropriate.) Since a stable set describes what can be known in a given situation, we would expect a formula to be honest if it is in a minimum stable set. This is indeed true.

Theorem 8 If $\mathcal{S} \in \{\mathrm{K}_n, \mathrm{T}_n, \mathrm{S4}_n\}$, then α is $\mathcal{S}$-i-honest iff there is an $\mathcal{S}$-i-stable set S^α containing α which is a subset of every $\mathcal{S}$-i-stable set containing α. Moreover, if α is honest, then $\alpha \mid\!\sim^i_{\mathcal{S}} \beta$ iff $\beta \in S^\alpha$.

This characterization of honesty is closely related to one given in [11]; we discuss this in more detail in Section 5. We remark that Jaspars [15] essentially uses the characterization provided by this theorem as his definition of honesty for S4.

Our next result gives another characterization of what agent i knows if "all agent i knows is α", for an honest formula α. Basically, it shows that all agent i knows are the logical consequences of his knowledge of α. Thus, "all agent i knows" is a monotonic notion for the non-introspective logics.

Theorem 9 If $\mathcal{S} \in \{\mathrm{K}_n, \mathrm{T}_n, \mathrm{S4}_n\}$ and α is $\mathcal{S}$-i-honest, then $\alpha \mid\!\sim^i_{\mathcal{S}} \beta$ iff $\models_{\mathcal{S}} K_i\alpha \Rightarrow K_i\beta$.

Up to now we have assumed that Φ, the set of primitive propositions, is a finite set, since ω-trees were defined only under this assumption. This turns out not to be a serious restriction. As we hinted before, we could actually deal directly with the case that Φ is infinite, but it suffices to do the following. Given a formula α, let Ψ be any finite set of primitive propositions that contains all the primitive propositions that appear in α. We then define α to be $\mathcal{S}$-i-Ψ-honest if α is $\mathcal{S}$-i-honest assuming that Ψ is the set of primitive propositions are involved. If Φ is infinite, we say that α is $\mathcal{S}$-i-honest if α is $\mathcal{S}$-i-Ψ-honest for some Ψ containing all the primitive propositions that appear in α. By Theorem 6, if α is $\mathcal{S}$-i-Ψ-honest for some choice of Ψ, it is $\mathcal{S}$-i-Ψ-honest for all choices of Ψ (that contain all the primitive propositions appearing in α). Similarly, we say that $\alpha \mid\!\sim^i_{\mathcal{S}} \beta$ if this relation holds for some Ψ containing all the primitive propositions that appear in α and β. By Theorem 9, the choice of Ψ does not matter. Thus, all our definitions and results can easily be extended where Φ is infinite. We continue to assume that Φ is finite, for simplicity, in the next two sections, but using arguments similar to those above, we can easily extend to the case that Φ is infinite. (We remark that while the assumption that Φ is finite does not affect our definitions, it does have an impact on complexity; see Section 7 for details.)

This completes our discussion of the non-introspective logics. It is interesting to compare our results here to those proved by Vardi [24]. He defines a notion of "all agent i knows" for $\mathrm{S4}_n$, using the knowledge-structures approach of [4], and proves Theorem 9 for $\mathrm{S4}_n$ in the context of his definition. Given the close connection between ω-trees and knowledge structures, it is not hard to show that our definition of honesty coincides with his for $\mathrm{S4}_n$. Moreover, ω-trees seem to be a better representation for knowledge and possibility as far as proving complexity results. For example, all that Vardi was able to show was that honesty was (nonelementary-time) decidable. In Section 7, we show that in fact deciding whether a formula in $\mathrm{S4}_n$-i-honest is PSPACE-complete.

5 "Only knowing" for $\mathrm{K45}_n$ and $\mathrm{KD45}_n$

We must take a slightly different approach in dealing with the introspective logics. The notion of possibility that we used for the non-introspective logics does not satisfy the union condition in the introspective case. To see this, and the problems it causes, consider the single-agent case. Suppose Φ consists of two primitive propositions, say p and q, and suppose that all the agent knows is p. Surely p should be honest. Indeed, according to the framework of Halpern and Moses [11], there is a maximum situation where p is true where the structure consists of two truth assignments: one where both p and q are true, and the other where p is true and q is false. Call this structure M_1, and let w be the truth assignment that makes both p and q true. Let M_2 be the structure where the only truth assignment is w. Clearly, the agent knows p in M_2 as well. It is not hard to see that $T_{M_1,w}$ and $T_{M_2,w}$ are different. This follows from Corollary 3 since, for example, $(M_1, w) \models \neg K_1 q$ and $(M_2, w) \models K_1 q$. Now suppose we use $Poss^{\mathrm{K}}$ (that is, the notion of possibility $Poss^{\mathcal{S}}$ defined in the last section, with $\mathcal{S}$ being K) as our notion of possibility. It is not hard to show that there is no $\mathcal{S}$ situation (M_3, w') for $\mathcal{S} \in \{K45, KD45, S5\}$ such that $Poss_1^{\mathrm{K}}(M_3, w') \supseteq Poss_1^{\mathrm{K}}(M_1, w) \cup Poss_1^{\mathrm{K}}(M_2, w)$. For suppose there were. What truth assignments does the agent consider possible at w'? If it is only the one where both p and q are true, then it is easy to see that $Poss_1^{\mathrm{K}}(M_1, w)$ is not contained in $Poss_1^{\mathrm{K}}(M_3, w')$. If it is anything else, then it is easy to see that $Poss_1^{\mathrm{K}}(M_1, w)$ is not contained in $Poss_1^{\mathrm{K}}(M_3, w')$. The problem is introspection: the agent knows what truth assignments he considers possible, and this information is contained in the set of possibilities. We need to factor out this introspection somehow. In the single-agent case considered by Halpern and Moses [11], this was done by considering only truth assignments, not trees. We need an analogue for the multi-agent case.

We define an *i-objective k-tree* to be a k-tree whose root has no i-successors. We define an *i-objective ω-tree* to be an ω-tree all of whose components are i-objective. Given a k-tree T, let T^i be the result of removing all the i-successors of the root of T (and all the nodes reachable from these i-successors). Given an ω-tree $T = \langle T_0, T_1, \ldots\rangle$, let $T^i = \langle T_0^i, T_1^i, \ldots\rangle$. The way we factor out introspection is by considering i-objective trees. Intuitively, the i-objective tree corresponding to a situation (M, w) eliminates all the worlds that i considers possible in that situation. Notice that in the case of one agent, the i-objective trees are precisely the possible worlds.

We say a formula is *i-objective* if it is a Boolean combination of primitive propositions and formulas of the form $K_j\varphi$, $j \neq i$, where φ is arbitrary. Thus, $q \wedge K_2K_1p$ is 1-objective, but K_1p and $q \wedge K_1p$ are not. Notice that if there is only one agent, say agent 1, then the 1-objective formulas are just the propositional formulas. We say that the situations (M, w) and (M', w') are *i-objectively equivalent up to depth k*, and write $(M, w) \equiv_k^{-i} (M', w')$ if, for all i-objective formulas φ with $depth(\varphi) \leq k$, we have $(M, w) \models \varphi$ iff $(M', w') \models \varphi$. We say (M, w) and (M', w') are *i-objectively equivalent*, and write $(M, w) \equiv^{-i} (M', w')$ if $(M, w) \equiv_k^{-i} (M', w')$ for all k. Notice that two situations are equivalent iff they are both i-equivalent and i-objectively equivalent. That is, it is almost immediate from the definitions that

Lemma 10 $(M,w) \equiv_k (M',w')$ iff both $(M,w) \equiv_k^i (M',w')$ and $(M,w) \equiv_k^{-i} (M',w')$.

We now have the following analogue of Proposition 2. Since its proof is almost identical to that of Proposition 2, we omit it here.

Proposition 11 For each situation (M,w) and all k, there is a unique i-objective k-tree T such that $(M,w) \equiv_k^{-i} M(T)$; moreover, T is in fact $T^i_{M,w,k}$.

The following corollary follows immediately from Proposition 11, just as Corollary 3 followed from Proposition 2.

Corollary 12 $T^i_{M,w} = T^i_{M',w'}$ iff $(M,w) \equiv^{-i} (M',w')$.

The notion of possibility we use for $\mathrm{K45}_n$ and $\mathrm{KD45}_n$ uses i-objective trees rather than ω-trees. For $\mathcal{S} \in \{\mathrm{K45}_n, \mathrm{KD45}_n\}$, we define

$$Poss_i^{\mathcal{S}}(M,w) = \{T^i_{M,w'} : (w,w') \in \mathcal{K}_i\}.$$

The following two propositions say that this notion of possibility satisfies the union and determination condition in the case of $\mathrm{K45}_n$ and $\mathrm{KD45}_n$. As we shall see in the next section, it does not satisfy the union requirement for $\mathrm{S5}_n$ (which is why we shall use a slightly different notion of possibility for $\mathrm{S5}_n$).

Proposition 13 If (M,w) and (M',w') are $\mathcal{S}$ situations, $\mathcal{S} \in \{\mathrm{K45}_n, \mathrm{KD45}_n\}$, and $Poss_i^{\mathcal{S}}(M,w) = Poss_i^{\mathcal{S}}(M',w')$, then $(M,w) \equiv^i (M',w')$.

Proposition 14 If $\mathcal{S} \in \{\mathrm{K45}_n, \mathrm{KD45}_n\}$, then for all agents i and $\mathcal{S}$ situations (M_1,w_1) and (M_2,w_2), there is an $\mathcal{S}$ situation (M_3,w_3) such that $Poss_i^{\mathcal{S}}(M_3,w_3) = Poss_i^{\mathcal{S}}(M_1,w_1) \cup Poss_i^{\mathcal{S}}(M_2,w_2)$.

Notice that for $\mathcal{S} \in \{\mathrm{K45}_n, \mathrm{KD45}_n\}$, $Poss_i^{\mathcal{S}}(M_3,w_3)$ is actually equal to $Poss_i^{\mathcal{S}}(M_1,w_1) \cup Poss_i^{\mathcal{S}}(M_2,w_2)$, not just a superset.

Our notion of $\mathcal{S}$-i-honesty and $\mid\!\sim^i_{\mathcal{S}}$ makes perfect sense for $\mathcal{S} \in \{\mathrm{K45}_n, \mathrm{KD45}_n\}$. Of course, we now use i-objective trees as our notion of possibility. Since i-objective trees are truth assignments in the single-agent case, it is easy to see that these definitions generalize those for the single-agent case given in [11].

We now want to characterize honesty and "all agent i knows" for $\mathrm{K45}_n$ and $\mathrm{KD45}_n$. There are some significant differences from the non-introspective case. For example, if $\mathcal{S} \in \{\mathrm{K45}_n, \mathrm{KD45}_n\}$, then, as expected, the primitive proposition p is $\mathcal{S}$-1-honest. However, due to negative introspection, $\neg K_1 q \Rightarrow K_1 \neg K_1 q$ is $\mathcal{S}$-valid, so we have $\models_{\mathcal{S}} K_1 p \Rightarrow (K_1 q \vee K_1 \neg K_1 q)$. Moreover, we have neither $\models_{\mathcal{S}} K_1 p \Rightarrow K_1 q$ nor $\models_{\mathcal{S}} K_1 p \Rightarrow K_1 \neg K_1 q$. Thus, the analogue to Theorem 6 does not hold.

As the following result shows, the analogue of Theorem 6 holds for $\mathrm{K45}_n$ and $\mathrm{KD45}_n$ provided we stick to i-objective formulas.

Theorem 15 For $\mathcal{S} \in \{\mathrm{K45}_n, \mathrm{KD45}_n\}$, the formula α is $\mathcal{S}$-i-honest iff (a) $K_i\alpha$ is $\mathcal{S}$-consistent and (b) for all i-objective formulas $\varphi_1, \ldots, \varphi_k$, if $\models_{\mathcal{S}} K_i\alpha \Rightarrow (K_i\varphi_1 \vee \ldots \vee K_i\varphi_k)$ then $\models_{\mathcal{S}} K_i\alpha \Rightarrow K_i\varphi_j$ for some $j \in \{1, \ldots, k\}$.

We remark that part (a) is vacuous in the case of $\mathrm{K45}_n$, since any formula of the form $K_i\alpha$ must be $\mathrm{K45}_n$-consistent.

This result does not hold for $\mathrm{S5}_n$, at least not if we want *true* to be $\mathrm{S5}_n$-1-honest. For notice that $\models_{\mathrm{S5}_n} K_1 \mathit{true} \Rightarrow (K_1 q \vee K_1 K_2 \neg K_2 K_1 q)$. (This follows from the fact that $\models_{\mathrm{S5}_n} \neg K_1 q \Rightarrow K_1 K_2 \neg K_2 K_1 q$.) However, it is easy to see that $\not\models_{\mathrm{S5}_n} K_1 \mathit{true} \Rightarrow K_1 q$ and $\not\models_{\mathrm{S5}_n} K_1 \mathit{true} \Rightarrow K_1 K_2 \neg K_2 K_1 q$.

Theorem 15 is a direct extension of a result in [11] for the single-agent case. Two other characterizations of honesty and "all I know" are given by Halpern and Moses, that can be viewed as analogues to Theorems 8 and 9. As we now show, they also extend to $\mathrm{K45}_n$ and $\mathrm{KD45}_n$, but not $\mathrm{S5}_n$.

One of these characterizations is in terms of stable sets. The direct analogue of Theorem 8 does not hold for the introspective logics. In fact, as was already shown by Halpern and Moses [11] for the single-agent case, any two consistent stable sets are incomparable with respect to set inclusion. Again, the problem is due to introspection. For suppose we have two consistent $\mathcal{S}$-i-stable sets S and S' such that $S \subset S'$, and $\varphi \in S' - S$. By definition, there must be situations (M, w) and (M', w'), corresponding to S and S' respectively, for which we have $(M, w) \not\models K_i\varphi$ and $(M', w') \models K_i\varphi$. By introspection, we have $(M, w) \models K_i \neg K_i \varphi$ and $(M', w') \models K_i K_i \varphi$. This means that $\neg K_i\varphi \in S$ and $K_i\varphi \in S'$. Since $S \subset S'$, we must also have $\neg K_i \varphi \in S'$, which contradicts the assumption that S' is consistent.

We can get an analogue of Theorem 8 if we consider i-objective formulas. Define the *i-kernel* of an $\mathcal{S}$-i-stable set S, denoted $\mathrm{ker}_i(S)$, to consist of all the i-objective formulas in S.

Theorem 16 For $\mathcal{S} \in \{\mathrm{K45}_n, \mathrm{KD45}_n\}$, a formula α is $\mathcal{S}$-i-honest iff there is an $\mathcal{S}$-i-stable set S^α containing α such that for all $\mathcal{S}$-i-stable sets S containing α, we have $\mathrm{ker}_i(S^\alpha) \subseteq \mathrm{ker}_i(S)$. Moreover, if α is $\mathcal{S}$-i-honest, then $\alpha {\mid\!\sim}^i_{\mathcal{S}} \beta$ iff $\beta \in S^\alpha$.

Theorem 16 does not hold for $\mathrm{S5}_n$ if we want *true* to be $\mathrm{S5}_n$-1-honest. Suppose that S is an $\mathrm{S5}_n$-1-stable set that does not include q. Thus, it must include $\neg K_1 q$. Since $\models_{\mathrm{S5}_n} K_1 \neg K_1 q \Rightarrow K_1 K_2 \neg K_2 K_1 q$, the set S must also contain the objective formula $K_2 \neg K_2 K_1 q$. But there is another $\mathrm{S5}_n$-1-stable set that contains the formula $K_1 q$ and does not contain $K_2 \neg K_2 K_1 q$. This shows that no $\mathrm{S5}_n$-1-stable set that does not contain q can have a minimum 1-kernel among all stable sets. But surely an $\mathrm{S5}_n$-1-stable set containing q cannot have a minimum $\mathrm{S5}_n$-1-kernel among $\mathrm{S5}_n$-1-stable sets, since there is an $\mathrm{S5}_n$-1-stable set not containing q. It follows that there is no $\mathrm{S5}_n$-1-stable set with a minimum 1-kernel, so *true* does not satisfy the characterization above for honesty in the case of $\mathrm{S5}_n$. In fact, we can extend this example to show for *no* formula α is there an $\mathrm{S5}_n$-i-stable set containing α whose i-kernel is a minimum; we omit details here.

Finally, let us consider the analogue to Theorem 9. For $\mathcal{S} \in \{\mathrm{K45}_n, \mathrm{KD45}_n\}$, it is not hard to show that if α and β are propositional formulas, then we have $\alpha {\mid\!\sim}^i_{\mathcal{S}} \beta$ iff $\models_{\mathcal{S}} K_i\alpha \Rightarrow K_i\beta$. This is no longer true if α or β involve modal operators. For example, we have $p {\mid\!\sim}^1_{\mathcal{S}} \neg K_1 q$ even though $\not\models_{\mathcal{S}} K_1 p \Rightarrow K_1 \neg K_1 q$. This seems reasonable: If all agent 1 knows is p, then agent 1 does not know q and (by introspection) knows that he does not know this.[9] On the other hand, if agent i learns q, then he will know q, and (by introspection) knows he knows it; that is, $p \wedge q {\mid\!\sim}^1_{\mathcal{S}} K_1 q$. This shows

[9]Note that there is an implicit assumption here that the agent is *aware* of all the primitive

that, in contrast to the non-introspective case, inference from "all agent i knows" is nonmonotonic for $\mathcal{S} \in \{\mathrm{K45}_n, \mathrm{KD45}_n\}$.

As shown by Halpern and Moses [11], there is an elegant algorithmic characterization of "all agent i knows" in the single-agent case. Roughly speaking, the idea is as follows. To see if $\alpha \mid\!\sim^1 \beta$, if β is a propositional formula, we check if $\models_{\mathcal{S}} K_1\alpha \Rightarrow K_1\beta$. If β is not a propositional formula, then there must be some subformula of β of the form $K_1\gamma$, where γ is a propositional formula. Let β' be the subformula that results if we replace $K_1\gamma$ by *true* if $\models_{\mathcal{S}} K_1\alpha \Rightarrow K_1\gamma$, and by *false* otherwise. We now apply the same procedure to β'. Ultimately, we end up with a propositional formula, say β''. We then have $\alpha \mid\!\sim \beta$ iff $\models_{\mathcal{S}} K_1\alpha \Rightarrow K_1\beta''$.

We extend this idea to the multi-agent case here. The analogue of formulas like $K_1\gamma$ where γ is a propositional formula now becomes what we call a *top-level i-subformula*. A top-level i-subformula of a formula β is a subformula of β of the form $K_i\varphi$ which is not in the scope of a modal operator K_j, $j \neq i$, such that φ is i-objective. Thus, if β is $K_iK_ip \vee K_iK_jK_iq$, its top-level i-subformulas are K_ip and $K_iK_jK_iq$. We can now generalize the construction as follows:

Definition 17 Given formulas α and β, a modal logic $\mathcal{S}$, and an agent i, we define a finite sequence $\langle \beta_0, \beta_1, \ldots, \beta_m \rangle$ of formulas and a finite sequence $\langle B_1, \ldots, B_m \rangle$ of sets of formulas as follows. We take β_0 to be $K_i\beta$. Suppose we have defined $\beta_0, \ldots, \beta_k$ and $B_1, \ldots, B_k$ so that B_h consists of all the top-level i-subformulas of β_{h-1}. If $\beta_k \notin \{\mathit{true}, \mathit{false}\}$, then we define B_{k+1} to consist of all the top-level i-subformulas of β_k and define β_{k+1} to be the result of replacing each subformula $K_i\varphi$ of β_k that is in B_{k+1} by either *true* or *false*, depending on whether or not $\models_{\mathcal{S}} K_i\alpha \Rightarrow K_i\varphi$. The construction ends if $\beta_m \in \{\mathit{true}, \mathit{false}\}$. Since each formula in the sequence $\beta_0, \beta_1, \ldots$ other than *true* or *false* is of the form $K_i\beta'$, and is shorter than the previous formulas (if we view *true* and *false* as having length 1), it is clear that this construction terminates after at most $|\beta|$ steps. Define $A^i_{\mathcal{S}}(\alpha, \beta)$ to be the sequence $\langle \beta_0, \ldots, \beta_m \rangle$ (where $\beta_m \in \{\mathit{true}, \mathit{false}\}$) and $B^i_{\mathcal{S}}(\alpha, \beta) = \cup_{j=1}^m B_j$. It is easy to see that $|B^i_{\mathcal{S}}(\alpha, \beta)| \leq |\beta|$. Finally, we define $D^i_{\mathcal{S}}(\alpha)$ to consist of all those formulas β such that the last formula in the sequence $A^i_{\mathcal{S}}(\alpha, \beta)$ is *true*.[10] $\square$

Example 18 Suppose β is the formula $K_iK_ip \vee K_iK_jK_iq$. Then, as we observed, K_ip and $K_iK_jK_iq$ are the top-level i-subformulas of $K_i\beta$; it is easy to see that neither is implied by $K_i\beta$. Thus, $A^i_{\mathrm{K45}_n}(\beta, \beta)$ is the sequence $\langle K_i(K_iK_ip \vee K_iK_jK_iq), K_i(K_i(\mathit{false}) \vee \mathit{false}), K_i(\mathit{false} \vee \mathit{false}), \mathit{false} \rangle$, while $B^i_{\mathrm{K45}_n}(\beta, \beta)$ is $\{K_ip, K_iK_jK_iq, K_i(\mathit{false}), K_i(\mathit{false} \vee \mathit{false})\}$. $\square$

As the following result shows, $D^i_{\mathcal{S}}(\alpha)$ can be viewed as the set of formulas that agent i knows, given that agent i knows only α (and reasons using modal logic $\mathcal{S}$).

Theorem 19 For $\mathcal{S} \in \{\mathrm{K45}_n, \mathrm{KD45}_n\}$, the formula α is $\mathcal{S}$-i-honest iff $K_i\alpha$ is $\mathcal{S}$-consistent and $\alpha \in D^i_{\mathcal{S}}(\alpha)$. If α is $\mathcal{S}$-i-honest, then $\alpha \mid\!\sim^i_{\mathcal{S}} \beta$ iff $\beta \in D^i_{\mathcal{S}}(\alpha)$.

propositions, even if he does not know them. If the agent is not even aware of the existence of q, then he will not even know that he does not know q. This intuition is formalized in the logic of general awareness of [3].

[10] We remark that this construction of $D^i_{\mathcal{S}}(\alpha)$ is not identical to that given in [11] if we consider the single-agent case. We could have used a direct extension of the algorithm given in [11], but the variant we use here turns out to be easier to work with in the multi-agent case.

6 "Only knowing" for $\mathrm{S5}_n$

As we have seen, the approach to dealing with honesty taken for $\mathrm{K45}_n$ and $\mathrm{KD45}_n$ does not work for $\mathrm{S5}_n$. As we show in this section, a small change in the notion of possibility solves the problem. First, let us examine why the notion of "possibility" defined for $\mathrm{KD45}_n$ and $\mathrm{K45}_n$ is inappropriate for $\mathrm{S5}_n$.

Suppose (M_1, w_1) is an $\mathrm{S5}_2$ situation such that $(M_1, w_1) \models K_2K_1p$, and (M_2, w_2) is an $\mathrm{S5}_2$ situation such that $(M_2, w_2) \models \neg p$. Note that since $(M_1, w_1) \models K_2K_1p$, it follows that $(M_1, w_1) \models p$. Now suppose we try to use $Poss^{\mathrm{K45}_2}$ as our notion of possibility. If the union condition held for $\mathrm{S5}_n$, there would be an $\mathrm{S5}_n$ situation (M_3, w_3) such that $Poss_1^{\mathrm{K45}_2}(M_3, w_3) \supseteq \{T^1_{M_1,w_1}, T^1_{M_2,w_2}\}$. Suppose there were such a situation. Then agent 1 would have to consider both p and $\neg p$ possible in this situation; that is, we would have $(M_3, w_3) \models \neg K_1 p \wedge \neg K_1 \neg p$. Since $\models_{\mathrm{S5}_n} \neg K_1 p \Rightarrow K_1 \neg K_2 K_1 p$, it follows that $(M_3, w_3) \models K_1\neg K_2K_1p$. Thus, if $(w_3, w) \in \mathcal{K}_1^M$, we have $(M_3, w) \models \neg K_2K_1p$. But since we must have $T^1_{M_1,w_1} \in Poss_1^{\mathrm{K45}_2}(M_3, w_3)$, there must be some $w' \in \mathcal{K}_1^{M_3}(w_3)$ such that $T^1_{M_1,w_1} = T^1_{M_3,w'}$. But it then follows from Proposition 11 that $(M_3, w') \models K_2K_1p$. This is a contradiction.

The problem here is that in the case of $\mathrm{S5}_n$, what is true in the actual situation must be considered possible (that is, $\varphi \Rightarrow \neg K_i \neg \varphi$ is valid). This combined with negative introspection causes our difficulties. In (M_1, w_1), 2 knows that 1 knows p. Thus, 1 considers it possible at the actual situation (M_1, w_1) that 2 knows that 1 knows p. It follows that in any situation (M_3, w_3) such that $T^1_{M_1,w_1} \in Poss_1^{\mathrm{K45}_2}(M_3, w_3)$, it must be the case that 1 considers it possible that 2 knows that 1 knows p; i.e., 1 must consider it possible that K_2K_1p holds. But, in $\mathrm{S5}_n$, K_2K_1p implies K_1p, so agent 1 must consider it possible that K_1p holds. From the negative introspection property, it then follows that 1 must know p. This means that in no situation where 1 considers $T^1_{M_1,w_1}$ possible can 1 have $T^1_{M_2,w_2}$ among his set of possibilities, at least, under this notion of "possibility". This is clearly incompatible with the union property.

To solve this problem, we need to somehow factor out (what the agent considers possible in) the actual situation when we are constructing an agent's possibilities. We now present one way of doing so. Roughly speaking, we construct a tree with nodes labeled $*$ that represent the actual world (i.e., the root of the tree). Formally, we define k-$*$-trees by induction on k: a 0-$*$-tree is a 0-tree, and a $(k+1)$-$*$-tree consists of a root r labeled by a truth assignment, and for each agent, a (possibly empty) set of directed edges labeled by i leading from r to roots of distinct k-$*$-trees or to a special node labeled $*$. Intuitively, a node labeled $*$ represents the actual situation and what is considered possible at the actual situation. We can think of an edge to a node labeled $*$ as really being a backedge to the root (i.e., the "actual situation"). We define ω-$*$-trees and i-objective $*$-trees in the obvious way; we omit the formal details here. Given a $*$-tree T, we define the corresponding Kripke structure, which we continue to denote $M(T)$, just as we did for ordinary trees, except that, as suggested by the intuition above, there are now no nodes in $M(T)$ corresponding to the nodes in T labeled by $*$, and edges in T labeled by i to a node labeled $*$ are replaced in $M(T)$ by edges to the root of T. (Note that if there are no nodes labeled $*$ in T, so that T is actually an ordinary tree, then the old definition of $M(T)$ agrees with this one, justifying our abuse of notation.)

As we are about to show, there is no difficulty proving an analogue to Proposition 2 that allows us to associate with each situation a unique $*$-tree. In what sense does the $*$-tree help us factor out what the agent considers possible at the actual world? To make sense of this, it is helpful to extend the language with a family of modal operators Q_i^ξ, $i = 1, \ldots, n$, where ξ is an *i-subjective formula*, that is, a Boolean combination of formulas of the form $K_i\psi$.[11] Intuitively, $Q_i^\xi\varphi$ holds in the situation (M, w) if φ holds no matter how we modify i's accessibility relation at w, provided we do so in a way that ξ holds. Thus, the Q_i^ξ operators gives us a way—in the language—of factoring out what the agent considers possible in the actual situation. To make this precise, we say that (M, w) is *i-embedded in* (M', w'), where $M = (W, \pi, \mathcal{K}_1, \ldots, \mathcal{K}_n)$ and $M' = (W', \pi', \mathcal{K}'_1, \ldots, \mathcal{K}'_n)$, if (a) $W \subseteq W'$, (b) $\pi'|_W = \pi$, (c) $\mathcal{K}'_j|_{W\times W} = \mathcal{K}_j$ for $j = 1, \ldots, n$, (d) there is no pair $(v', v'') \in \mathcal{K}'_j$ with $v' \in W$ and $v'' \in W' - W$ for $j = 1, \ldots n$, (e) $\pi'(w') = \pi'(w)$, and (f) $\mathcal{K}_j^{M'}(w') - \{w'\} = \mathcal{K}_j^M(w) - \{w\}$ for $j \neq i$, and (g) $w' \in \mathcal{K}_j^{M'}(w')$ iff $w \in \mathcal{K}_j^M(w)$. Roughly speaking, this definition forces w' to look just like w, except that its i-successors may be different. Put another way, (M', w') is the result of changing what i considers possible when the actual situation is (M, w), without affecting anything else. Note that (M, w) is i-embedded in itself.

We now define Q_i^ξ as follows:

$(M, w) \models Q_i^\xi\varphi$ if $(M', w') \models \xi \Rightarrow \varphi$ for all situations (M', w') in which (M, w) is i-embedded.

Let $\mathcal{L}_n^Q$ be the result of extending the language $\mathcal{L}_n$ so that it includes the Q_i operators, for $i = 1, \ldots, n$.

Define a *Q_i-formula* to be one of the form $Q_i^\xi\varphi$, where $\varphi \in \mathcal{L}_n$. For convenience, we define $depth(Q_i^\xi\varphi) = depth(\varphi)$. Q_i-formulas are the analogue for $S5_n$ of i-objective formulas in the case of $K45_n$ and $KD45_n$. For example, it is not hard to show that $Q_i^\xi\varphi \equiv \varphi$ is valid in all $K45_n$ or $KD45_n$ structures if φ is i-objective and ξ is satisfiable. This is not true for $S5_n$. For example, $Q_1^{K_1 true}K_2K_1p$ is not equivalent to K_2K_1p: We can easily construct a situation in which K_2K_1p is true that can be embedded in a situation where K_1p, and hence also K_2K_1p, is false. We write $(M, w) \equiv_k^{Q_i} (M', w')$ if for all Q_i-formulas φ such that $depth(\varphi) \leq k$, we have $(M, w) \models \varphi$ iff $(M', w') \models \varphi$. Thus, if $(M, w) \equiv_k^{Q_i} (M', w')$ then (M, w) and (M', w') agree on all Q_i-formulas involving $\mathcal{L}_n$ formulas of depth at most k.

The following proposition is the analogue to Proposition 2.

Proposition 20 For each $S5_n$ situation (M, w) and for all k, there is a unique i-objective k-$*$-tree $T_{M,w,k}^{i,*}$ such that $(M, w) \equiv_k^{Q_i} M(T_{M,w,k}^{i,*})$.

Let $T_{M,w}^{i,*}$ be the ω-$*$-tree $\langle T_{M,w,0}^{i,*}, T_{M,w,1}^{i,*}, \ldots\rangle$. Another key property of embeddings is stated in the following lemma, whose proof is almost immediate from the definitions.

Lemma 21 If (M, w) is i-embedded in (M', w), then $T_{M,w}^{i,*} = T_{M',w}^{i,*}$.

We can now define our notion of possibility for $S5_n$. Let

$$Poss_i^{S5_n}(M, w) = \{T_{M,w'}^{i,*} : (w, w') \in \mathcal{K}_i\}.$$

This notion of possibility satisfies our two requirements in the case of $S5_n$.

[11]The letter Q was chosen only because most other letters already had relatively well-known modal operators associated with them.

Proposition 22 If (M, w) and (M', w') are $\mathrm{S5}_n$ situations such that $Poss_i^{\mathrm{S5}_n}(M, w) = Poss_i^{\mathrm{S5}_n}(M', w')$, then $(M, w) \equiv^i (M', w')$.

Proposition 23 For all agents i and $\mathrm{S5}_n$ situations (M_1, w_1) and (M_2, w_2), there is an $\mathrm{S5}_n$ situation (M_3, w_3) such that $Poss_i^{\mathrm{S5}_n}(M_3, w_3) = Poss_i^{\mathrm{S5}_n}(M_1, w_1) \cup Poss_i^{\mathrm{S5}_n}(M_2, w_2)$.

Again, our definitions of $\mathcal{S}$-i-honesty and $\mid\!\sim^i_{\mathcal{S}}$ now carry over for $\mathcal{S} = \mathrm{S5}_n$. We remark that we restrict the $\mid\!\sim^i_{\mathrm{S5}_n}$ relation to formulas in $\mathcal{L}_n$ here to simplify the comparison to previous sections. There is no difficulty extending it to $\mathcal{L}_n^Q$ though. Again, it is easy to see that these definitions generalize those for the single-agent case given in [11].

We now want to characterize honesty and "all agent i knows" for $\mathrm{S5}_n$. The theorems have very much the same flavor as the corresponding results for $\mathrm{K45}_n$ and $\mathrm{KD45}_n$, so we just state them here without much comment.

Theorem 24 The formula α is $\mathrm{S5}_n$-i-honest iff (a) $K_i\alpha$ is $\mathrm{S5}_n$-consistent and (b) for all Q_i-formulas $\varphi_1, \ldots, \varphi_k$, if $\models_{\mathrm{S5}_n} K_i\alpha \Rightarrow (K_i\varphi_1 \vee \ldots \vee K_i\varphi_k)$ then $\models_{\mathrm{S5}_n} K_i\alpha \Rightarrow K_i\varphi_j$ for some $j \in \{1, \ldots, k\}$.

Notice that although we have $\models_{\mathrm{S5}_n} K_1 true \Rightarrow (K_1 q \vee K_1 K_2 \neg K_2 K_1 q)$, we do not have $\models_{\mathrm{S5}_n} K_1 true \Rightarrow (K_1 Q_1^{K_1 true} q \vee K_1 Q_1^{K_1 true} K_2 \neg K_2 K_1 q)$, so the counterexample to Theorem 15 does not apply here.

A set S of formulas is a $\mathrm{S5}_n$-*i-Q-stable set* if there is some $\mathrm{S5}_n$ situation (M, w) such that $S = \{\varphi \in \mathcal{L}_n^Q : (M, w) \models K_i\varphi\}$. Define the *$Q$-kernel* of an $\mathrm{S5}_n$-i-Q-stable set S, denoted $\ker_i^Q(S)$, to consist of all the Q_i-formulas in S.

Theorem 25 A formula α is $\mathrm{S5}_n$-i-honest iff there is an $\mathrm{S5}_n$-i-Q-stable set S^α containing α such that $\ker_i^Q(S^\alpha) \subseteq \ker_i^Q(S)$. Moreover, if α is $\mathrm{S5}_n$-i-honest, then, for all $\beta \in \mathcal{L}_n$, $\alpha \mid\!\sim^i_{\mathrm{S5}_n} \beta$ iff $\beta \in S^\alpha$.

Finally, we recursively define a set $D_Q^i(\alpha)$ of formulas in $\mathcal{L}_n$ as follows:

$$\varphi \in \mathcal{D}_Q^i(\alpha) \text{ iff } \models_{\mathrm{S5}_n} K_i\alpha \Rightarrow K_i Q_i^{\xi_\varphi} \varphi,$$

where ξ_φ is the conjunction of $K_i\psi$ for all subformulas $K_i\psi$ of φ such that $\psi \in D_Q^i(\alpha)$ and $\neg K_i\psi$ for all subformulas $K_i\psi$ of φ for which $\psi \notin D_Q^i(\alpha)$. (We take ξ_φ to be $K_i true$ if there are no i-subjective subformulas of ψ.) Then we have:

Theorem 26 The formula α is $\mathrm{S5}_n$-i-honest iff $K_i\alpha$ is $\mathrm{S5}_n$-consistent and $\alpha \in D_Q^i(\alpha)$. If α is $\mathrm{S5}_n$-i-honest, then $\alpha \mid\!\sim^i_{\mathrm{S5}_n} \beta$ iff $\beta \in D_Q^i(\alpha)$.

We close this section by briefly comparing our approach to defining "all I know" for $\mathrm{S5}_n$ to two others that have appeared in the literature. Fagin, Halpern, and Vardi [4] define a notion of *i-no-information extension* that can also be viewed as characterizing a notion of "all agent i knows" in the context of $\mathrm{S5}_n$. However, it is defined only for a limited set of formulas.[12] It can be shown that these formulas are always $\mathrm{S5}_n$-i-honest

[12] Roughly speaking, these are the formulas that, in the terminology of [4], characterize finite knowledge structures.

in our sense, and, if α is one of these formulas, we have $\alpha{\mid\!\sim}^{i}_{\mathrm{S5}_n}\beta$ iff β is true in the i-no-information extension of α. The fact that these two independently motivated definitions coincide (at least, in the cases where the i-no-information extension is defined) provides further evidence for the reasonableness of our definitions.

Parikh [21] defines a notion of "all that is known" for $\mathrm{S5}_n$ much in the spirit of the definitions given here. Among other things, he also starts with k-trees (he calls them *normal models*), although he does not use i-objective trees. However, rather than focusing on all that some fixed agent i knows as we have done, Parikh treats all agents on an equal footing. This leads to some technical differences between the approaches. His approach also does not lend itself well to proving complexity results. He was also able to obtain only nonelementary-time algorithms for deciding whether a formula was honest in his sense. As we shall see in the next section, we can do much better.

7 Complexity issues

We now characterize the complexity of computing honesty and "all i knows".

Theorem 27 For $\mathcal{S} \in \{\mathrm{T}_n, \mathrm{S4}_n : n \geq 1\} \cup \{\mathrm{K45}_n, \mathrm{KD45}_n, \mathrm{S5}_n : n \geq 2\}$, the problem of computing whether α is $\mathcal{S}$-i-honest is PSPACE-complete.

Of course, the problem of computing whether α is K_n-i-honest is trivial: the answer is always "Yes".

Theorem 28 For $\mathcal{S} \in \{\mathrm{K}_n, \mathrm{T}_n, \mathrm{S4}_n : n \geq 1\} \cup \{\mathrm{K45}_n, \mathrm{KD45}_n, \mathrm{S5}_n : n \geq 2\}$, if α is $\mathcal{S}$-i-honest, then the problem of deciding if $\alpha{\mid\!\sim}^{i}_{\mathcal{S}}\beta$ is PSPACE-complete.

The requirement that $n \geq 2$ for the introspective logics is necessary. While PSPACE is, of course, still an upper bound if $n = 1$, we can do better. How much better we can do depends on whether Φ (the set of primitive propositions) is finite or infinite. This issue was not relevant in the context of Theorems 27 and 28; as our proof shows, the PSPACE result holds as long as Φ has even a single primitive proposition. It does make a difference, however, in the case of K45, KD45, and S5. In these cases, if Φ is finite, the problems we are interested in are decidable in polynomial time. On the other hand, if Φ is infinite, then the relevant complexity class turns out to be $\Delta_2^{\mathrm{p},\log(n)}$.[13] The complexity class Δ_2^{p} consists of all languages L such that membership in L can be decided by a Turing machine that runs in polynomial time, but is allowed to make queries to an NP oracle. The complexity class $\Delta_2^{\mathrm{p},\log(n)}$ consists of those languages in Δ_2^{p} such that on an input of size n, the NP oracle is queried at most $\log(n)$ times. It is easy to see that $\Delta_2^{\mathrm{p},\log(n)} \subseteq PSPACE$. It is conjectured that the containment is strict, but this has not yet been proved. Using recent results of Gottlob [6], we can show

[13] The situation here is analogous to the satisfiability problem for these logics. As shown in [9], in the case of K_n, T_n, and $\mathrm{S4}_n$, the satisfiability problem is PSPACE complete even if $n = 1$ and Φ contains only one primitive proposition. Similarly, we get PSPACE completeness for the satisfiability problem for $\mathrm{K45}_n$, $\mathrm{KD45}_n$, and $\mathrm{S5}_n$ as long as $n \geq 2$, even if Φ contains only one primitive propositions. However, for K45, KD45, and S5 (where there is only one agent), the satisfiability problem is NP-complete if Φ contains infinitely many primitive propositions, and in polynomial time if Φ contains a finite number of primitive propositions.

Theorem 29 Suppose $\mathcal{S} \in \{KD45, K45, S5\}$. If Φ (the set of primitive propositions) is finite, then the problem of deciding whether α is $\mathcal{S}$-1-honest and the problem of deciding whether $\alpha \mid\!\sim^1_{\mathcal{S}} \beta$ for an $\mathcal{S}$-1-honest α are both decidable in polynomial time. If Φ is infinite, these problems are both $\Delta_2^{\mathrm{p},\log(n)}$-complete.

8 Discussion

We have extended the HM notion of only knowing to (the multi-agent case of) a number of modal logics. The key tool was an appropriate canonical representation of the possibilities of the agents. Such a canonical representation should also be useful in other applications where we need to characterize the set of possibilities of an agent, such as in extending Levesque's notion of only knowing to multiple agents. (See [8] for a discussion of how this can be done, and [17] for an alternative approach; a synthesis can be found in [10].)

Despite its attractive properties, there is still an element of *ad hockery* to our approach.[14] For example, for the non-introspective logics, we used ω-trees to define the notion of what agent i considers possible, for $K45_n$ and $KD45_n$, we used i-objective ω-trees, and for $S5_n$, we used i-objective ω-$*$-trees. In general, it is clear that the notion of "all I know" is a function of the notion of possibility. We could, for example, define a notion of "all I know" for $K45_n$ and $KD45_n$ using ω-trees. The reason we did not, as we argued above, is that this definition would result in a rather strange notion of "all I know", with quite counterintuitive properties. This observation, of course, raises a number of questions.

- Can we get some kind of a correspondence between properties of "all I know" and properties of the notion of possibility? We conjecture that conditions like the union condition and the determination condition will be necessary to get a reasonable notion of "all I know".
- We required that our notion of possibility satisfy the union condition and the determination condition. Are these conditions sufficient to determine the notion of "possibility" uniquely for a given logic? If not, can we find additional reasonable conditions that do determine it uniquely?
- Do the results we have proved for the logics we have considered hold for any notion of possibility that satisfies the union condition and determination condition?

Although the notion of possibility used for $S5_n$ seems rather complicated, it may well be that it (or something like it) is forced by some natural requirements. It would be comforting to have a framework in which this can be made precise.

The notion of "possibility" arises in a number of contexts. For example, it can also be used to extend Levesque's notion of "only knowing" to many agents [8, 10]. For another example, consider the approaches to modal logics of normality or plausibility such as that of Boutilier [1], which have thus far only been defined in the single-agent case. Boutilier's semantics involves placing an ordering on (all) worlds. If we try to

[14] We thank Grisha Schwarz for forcing this issue to our attention.

extend his intuitions to the multi-agent case, we may well need to place an ordering on "possibilities". Thus, it would be useful to have a general theory of what counts as a "possibility". As shown in [10], the situation is in fact even more complicated. What matters is not only what counts as an individual possibility, but what sets of possibilities an agent can consider possible. We leave further consideration of these issues to future work.

A Proofs for Section 4

Proposition 2: *For each (M,w) and all k, there is a unique k-tree $T_{M,w,k}$ such that $(M,w) \equiv_k M(T_{M,w,k})$.*

Proof Recall that it remains to show that $(M,w) \equiv_{k+1} M(T_{M,w,k+1})$ and that $T_{M,w,k+1}$ is the unique tree with this property. Assume inductively that we have shown that for all situations (M',w'), we have $(M',w') \equiv_k M(T_{M',w',k})$, and that $T_{M',w',k}$ is the unique tree with this property. To show that $(M,w) \equiv_{k+1} M(T_{M,w,k+1})$, given a formula φ such that $depth(\varphi) \leq k+1$, we must show that $(M,w) \models \varphi$ iff $M(T_{M,w,k+1}) \models \varphi$. We proceed by induction on the structure of φ. If φ is a primitive proposition, the result follows since the truth assignment at the root of $T_{M,w,k+1}$ is the same as that at w. If φ is a negation or a conjunction of two other formulas, then the result follows immediately from the induction hypothesis. Finally, if φ is of the form $K_i\psi$, then we must have $depth(\psi) \leq k$. Suppose $(M,w) \models K_i\psi$. Let w_0 be the root of $T_{M,w,k+1}$, and suppose that w_1 is an i-successor of w_0. Then our construction assures us that w_1 is the root of a subtree $T_{M,w',k}$ of $T_{M,w,k+1}$ and $(w,w') \in \mathcal{K}_i$. Since $(M,w') \models \psi$, by the induction hypothesis, it follows that $M(T_{M,w',k}) \models \psi$. Since $M(T_{M,w',k})$ is a subtree of $M(T_{M,w,k+1})$, it follows that $(M(T_{M,w,k+1}),w_1) \models \psi$. Thus, $(M(T_{M,w,k+1}),w_0) \models K_i\psi$.

For the converse, suppose that $(M,w) \models \neg K_i\psi$. Thus, for some w' with $(w,w') \in \mathcal{K}_i$, we have $(M,w') \models \neg\psi$. By construction, there must be an edge labeled i in $T_{M,w,k+1}$ from w to a node which is the root of $T_{M,w',k}$. Using the induction hypothesis just as in the previous paragraph, we get that $M(T_{M,w',k}) \models \neg\psi$, and so $M(T_{M,w,k+1}) \models \neg K_i\psi$. This completes the inductive proof.

For uniqueness, it suffices to show that if all k', if T and T' are distinct k'-trees, then $M(T) \not\equiv_{k'} M(T')$. We prove this result by induction on k'. If $k'=0$, this is immediate from the fact that if T and T' are distinct 0-trees, then the truth assignments labeling the roots of T and T' must be different. If $k'>0$ and T and T' are distinct m-trees, with roots w and w' respectively. Either the truth assignments labelling w and w' must be different, or for some agent i, the set of subtrees of T rooted at the i-successors of w must be different from the set of subtrees of T' rooted at the i-successors of w'. In the former case, the result is immediate. In the latter case, we can assume without loss of generality that there is some i-successor v of w such that the subtree of T rooted at v is different from all the subtrees rooted at i-successors of w'. Let $T_1,\ldots,T_m$ be the subtrees of T' rooted at i-successors of w', and let T_v be the subtree of T rooted at v. Since all these subtrees are $k'-1$-trees, it follows from the induction hypothesis that there are formulas $\varphi_1,\ldots,\varphi_m$ such that $depth(\varphi_i) \leq k'-1$, $M(T_v) \models \varphi_i$, and $M(T_i) \models \neg\varphi_i$, for $i=1,\ldots,m$. It now follows that $M(T) \models \neg K_i \neg(\varphi_1 \wedge \ldots \varphi_m)$ and $M(T') \models K_i \neg(\varphi_1 \wedge \ldots \wedge \varphi_m)$. Thus $M(T) \not\equiv_{k'} M(T')$.[15]

Proposition 4: *If $\mathcal{S} \in \{\mathrm{K}_n, \mathrm{T}_n, \mathrm{S4}_n\}$ and $Poss_i^{\mathcal{S}}(M,w) = Poss_i^{\mathcal{S}}(M',w')$, then $(M,w) \equiv^i (M',w')$.*

Proof Suppose that $(M,w) \not\equiv^i (M',w')$. Then, without loss of generality, there exists a formula φ such that $(M,w) \models K_i\varphi$ and $(M',w') \models \neg K_i \varphi$. But then there must exist a world w'' such that $(w',w'') \in \mathcal{K}_i^{M'}$ and $(M',w') \models \neg\varphi$. Since $(M,v) \models \varphi$ for all v such that $(w,v) \in \mathcal{K}_i^M$, it follows from Corollary 3 that $T_{M',w''} \notin Poss_i^{\mathcal{S}}(M,w)$, although by definition $T_{M',w''} \in Poss_i^{\mathcal{S}}(M',w')$. Thus, $Poss_i^{\mathcal{S}}(M,w) \neq Poss_i^{\mathcal{S}}(M',w')$.

Proposition 5: *For $\mathcal{S} \in \{\mathrm{K}_n, \mathrm{T}_n, \mathrm{S4}_n\}$, all agents i, and all $\mathcal{S}$ situations (M_1,w_1) and (M_2,w_2), there is an $\mathcal{S}$ situation (M_3,w_3) such that $Poss_i^{\mathcal{S}}(M_3,w_3) \supseteq Poss_i^{\mathcal{S}}(M_1,w_1) \cup Poss_i^{\mathcal{S}}(M_2,w_2)$.*

Proof Given $\mathcal{S}$ situations (M_1,w_1) and (M_2,w_2) we want to show that there exists an $\mathcal{S}$ situation (M_3,w_3) such that $Poss_i^{\mathcal{S}}(M_3,w_3) \supseteq Poss_i^{\mathcal{S}}(M_1,w_1) \cup Poss_i^{\mathcal{S}}(M_2,w_2)$. Observe that, without loss of generality, we can assume that W^{M_1} and W^{M_2} are disjoint. We take W^{M_3} to be $W^{M_1} \cup W^{M_2} \cup \{w_3\}$, where w_3 is a new world not in $W^{M_1} \cup W^{M_2}$. We define $\pi^{M_3}(w) = \pi^{M_j}(w)$ if $w \in W^{M_j}$, for $j = 1,2$ (here we are using the fact that W^{M_1} and W^{M_2} are disjoint). The definition of $\pi^{M_3}(w_3)$ is irrelevant. We define $\mathcal{K}_j^{M_3} = \mathcal{K}_j^{M_1} \cup \mathcal{K}_j^{M_2} \cup \{(w_3,w_3)\}$ for $j \neq i$, and define $\mathcal{K}_i^{M_3} = \mathcal{K}_i^{M_1} \cup \mathcal{K}_i^{M_2} \cup \{(w_3,w_3)\} \cup (\{w_3\} \times (\mathcal{K}_i^{M_1}(w_1) \cup \mathcal{K}_i^{M_2}(w_2)))$. Thus, $\mathcal{K}_i^{M_3}$ is the least Euclidean, transitive relation that includes $\mathcal{K}_i^{M_1} \cup \mathcal{K}_i^{M_2}$ and (w_3,w') for $w' \in \mathcal{K}_i^{M_1}(w_1) \cup \mathcal{K}_i^{M_2}(w_2)$. In the language of [14], (M_3,w_3) is an *amalgamation* of (M_1,w_1) and (M_2,w_2). We leave it to the reader to check that with these definitions, (M_3,w_3) is an $\mathcal{S}$ situation (our requirement that $(w_3,w_3) \in \mathcal{K}_j$ was precisely to take care of the possibility that $\mathcal{S}$ is T or S4) and that $Poss_i^{\mathcal{S}}(M_3,w_3) \supseteq Poss_i^{\mathcal{S}}(M_1,w_1) \cup Poss_i^{\mathcal{S}}(M_2,w_2)$.

Notice that if $\mathcal{S}$ is T_n or $\mathrm{S4}_n$, then the assumption that $(w_3,w_3) \in \mathcal{K}_i$ is necessary (otherwise M_3 would not be an $\mathcal{S}$-structure). This means that $T_{(M_3,w_3)} \in Poss_i^{\mathcal{S}}(M_3,w_3)$, which in turn means that $Poss_i^{\mathcal{S}}(M_3,w_3)$ may be a strict superset of $Poss_i^{\mathcal{S}}(M_1,w_1) \cup Poss_i^{\mathcal{S}}(M_2,w_2)$. As we showed in the main text, this is unavoidable. On the other hand, if $\mathcal{S} = \mathrm{K}_n$, we do not need the assumption that $(w_3,w_3) \in \mathcal{K}_i$. So, for K_n, we can assume that $Poss_i^{\mathcal{S}}(M_3,w_3) = Poss_i^{\mathcal{S}}(M_1,w_1) \cup Poss_1^{\mathcal{S}}(M_2,w_2)$.

Theorem 6: *If $\mathcal{S} \in \{\mathrm{K}_n, \mathrm{T}_n, \mathrm{S4}_n\}$, then the formula α is $\mathcal{S}$-i-honest iff (a) $K_i\alpha$ is $\mathcal{S}$-consistent and (b) for all formulas $\varphi_1, \ldots, \varphi_k$, if $\models_{\mathcal{S}} K_i\alpha \Rightarrow (K_i\varphi_1 \vee \ldots \vee K_i\varphi_k)$, then $\models_{\mathcal{S}} K_i\alpha \Rightarrow K_i\varphi_j$ for some $j \in \{1,\ldots,k\}$.*

---

[15] This part of the proof fails if Φ, the set of primitive propositions, is infinite, since we can no longer assume that there are only finitely many distinct subtrees rooted at i-successors of w'. If there are infinitely many subtrees, we would need an infinitary formula to distinguish $M(T)$ from $M(T')$, and such formulas are not in our language.

Proof Let $\mathcal{S} \in \{\mathrm{K}_n, \mathrm{T}_n, \mathrm{S4}_n\}$. For the "only if" direction, suppose that α is $\mathcal{S}$-i-honest. Thus, there is an $\mathcal{S}$ situation (M,w) such that $(M,w) \models K_i\alpha$, and for all $\mathcal{S}$ situations (M',w'), if $(M',w') \models K_i\alpha$, then $Poss_i^{\mathcal{S}}(M',w') \subseteq Poss_i^{\mathcal{S}}(M,w)$. Clearly $K_i\alpha$ must be $\mathcal{S}$-consistent. Now suppose that $\models_{\mathcal{S}} K_i\alpha \Rightarrow (K_i\varphi_1 \vee \ldots \vee K_i\varphi_k)$. To obtain a contradiction, suppose $K_i\alpha \wedge \neg K_i\varphi_j$ is $\mathcal{S}$-consistent for $j = 1,\ldots,k$. This means that there is some $\mathcal{S}$ situation (M_j, w_j) such that $(M_j, w_j) \models K_i\alpha$ and a world w_j' such that $(w_j, w_j') \in \mathcal{K}_i$ and $(M_j, w_j') \models \neg\varphi_j$, for $j = 1,\ldots,k$. By assumption, $T_{M_j,w_j'} \in Poss_i^{\mathcal{S}}(M,w)$. Thus, there is a world $w' \in W^M$ such that $(w,w') \in \mathcal{K}_i^M$ and $T_{M,w'} = T_{M_j,w_j'}$. It follows from Corollary 3 that $(M,w') \models \neg\varphi_j$. Hence, $(M,w) \models \neg K_i\varphi_j$, for $j = 1,\ldots,k$. But this contradicts the assumption that $\models_{\mathcal{S}} K_i\alpha \Rightarrow (K_i\varphi_1 \vee \ldots \vee K_i\varphi_k)$. Thus, we must have $\models_{\mathcal{S}} K_i\alpha \Rightarrow K_i\varphi_j$ for some $j \in \{1,\ldots,k\}$.

For the converse, suppose $K_i\alpha$ satisfies conditions (a) and (b) in the theorem. Let F consist of all subformulas of α of the form $K_i\psi$ such that $K_i\alpha \wedge \neg K_i\psi$ is $\mathcal{S}$-consistent. From condition (b), it follows that it cannot be the case that $\models_{\mathcal{S}} K_i\alpha \Rightarrow (\bigvee_{\psi \in F} K_i\psi)$. Thus, there is an $\mathcal{S}$ situation (M_α, w_α) such that $(M_\alpha, w_\alpha) \models K_i\alpha \wedge \bigwedge_{\psi\in F} \neg K_i\psi$.

We now construct an $\mathcal{S}$-i-maximum situation $(M_\alpha^{max}, w_\alpha)$ for $K_i\alpha$ as follows: Intuitively, we want to start with all situations satisfying $K_i\alpha$ and "glue" them together appropriately, as was done in the proof of Proposition 5. The problem is that we cannot talk about the "set" of situations satisfying $K_i\alpha$; it is not a set (it is too large). In fact, for our purposes, it would suffice to consider only situations with countably many worlds, but rather than proving this formally, we take $\mathcal{N}_\alpha$ to be a set of situations including (M_α, w_α) such that each situation $(M,w) \in \mathcal{N}_\alpha$ satisfies $K_i\alpha$ and, if $T \in Poss_i^{\mathcal{S}}(M',w')$ for some situation (M',w') satisfying $K_i\alpha$, then $T \in Poss_i^{\mathcal{S}}(M,w)$ for some situation (M,w) in $\mathcal{N}_\alpha$. Roughly speaking, we take M_α^{max} to consist of the union of all the situations (M,w) in $\mathcal{N}_\alpha$. In more detail, observe that without loss of generality, we can assume that if (M,w) and (M',w') are two situations in $\mathcal{N}_\alpha\}$, then the worlds in W^M and $W^{M'}$ are disjoint. We take the worlds in M_α^{max} to consist of the union of all the worlds in W^M for each situation (M,w) in $\mathcal{N}_\alpha$. We define $\pi^{M_\alpha^{max}}(w) = \pi^M(w')$ if w' is a world in W^M for some $(M,w) \in \mathcal{N}_\alpha$. We define $\mathcal{K}_j^{M_\alpha^{max}}$ to be the union of the $\mathcal{K}_j^M$ relations for the situations $(M,w) \in \mathcal{N}_\alpha$. In addition, if $(w,w') \in \mathcal{K}_i^M$ for some situation (M,w) in $\mathcal{N}_\alpha$, then we add the pair (w_α, w') to $\mathcal{K}_i^{M_\alpha^{max}}$. Thus, in M_α^{max}, the set of worlds considered possible by agent i in w_α is the union of all the worlds considered possible by agent i in any situation in $\mathcal{N}_\alpha$. Again, we can view $(M_\alpha^{max}, w_\alpha)$ as an amalgamation of the situations in $\mathcal{N}_\alpha$. We leave it to the reader to check that with these definitions, $(M_\alpha^{max}, w_\alpha)$ is an $\mathcal{S}$ situation.

We now want to show $(M_\alpha^{max}, w_\alpha) \models K_i\alpha$. We actually prove a stronger claim: We show that if (M,w) is a situation in $\mathcal{N}_\alpha$ and w' is a world in W^M, then $(M,w') \models \beta$ iff $(M_\alpha^{max}, w') \models \beta$ for every subformula β of $K_i\alpha$. The proof is by induction on the structure of β. Given our construction of M_α^{max} as the "union" of the situations in $\mathcal{N}_\alpha$, the proof is completely straightforward except if β is of the form $K_i\beta'$ and $w = w_\alpha$. Suppose $(M_\alpha^{max}, w_\alpha) \models K_i\beta'$; we want to show that $(M_\alpha, w_\alpha) \models K_i\beta'$. If $(w_\alpha, w') \in \mathcal{K}_i^{M_\alpha}$, then $(w_\alpha, w') \in \mathcal{K}_i^{M_\alpha^{max}}$ by construction. Since $(M_\alpha^{max}, w_\alpha) \models K_i\beta'$, it follows that $(M_\alpha^{max}, w') \models \beta'$. By the induction hypothesis, we have $(M_\alpha, w') \models \beta'$. It follows that $(M_\alpha, w_\alpha) \models K_i\beta'$. For the converse, suppose that $(M_\alpha, w_\alpha) \models K_i\beta'$. From the choice of M_α, it follows that $\models_{\mathcal{S}} K_i\alpha \Rightarrow K_i\beta'$. Suppose $(w_\alpha, w') \in \mathcal{K}_i^{M_\alpha^{max}}$.

We want to show that $(M_\alpha^{max}, w') \models \beta'$. From the construction of $\mathcal{K}_i^{M_\alpha^{max}}$, it follows that there is a situation (M, w) in $\mathcal{N}_\alpha$ such that $(w, w') \in \mathcal{K}_i^M$. Since $(M, w) \models K_i\alpha$, we must also have $(M, w) \models K_i\beta'$. Thus, $(M, w') \models \beta'$. By the induction hypothesis, we get $(M_\alpha^{max}, w') \models \beta'$ as desired. Thus, $(M_\alpha^{max}, w_\alpha) \models K_i\beta'$. This completes the inductive proof. In particular, we get that $(M_\alpha^{max}, w_\alpha) \models K_i\alpha$.

It remains to show that $(M_\alpha^{max}, w_\alpha)$ is an $\mathcal{S}$-i-maximum situation for $K_i\alpha$. Suppose $(M', w') \models K_i\alpha$. We must show that $Poss_i^{\mathcal{S}}(M', w') \subseteq Poss_i^{\mathcal{S}}(M_\alpha^{max}, w_\alpha)$. Suppose $T \in Poss_i^{\mathcal{S}}(M', w')$. By construction, there must be some situation $(M, w) \in \mathcal{N}_\alpha$ such that $T \in Poss_i^{\mathcal{S}}(M, w)$. Thus, there is some world w' with $(w, w') \in \mathcal{K}_i^M$ such that $T = T_{M,w'}$. Our proof above shows that $(M, w') \equiv (M_\alpha^{max}, w')$. Thus, by Corollary 3, $T_{M,w'} = T_{M_\alpha^{max},w'} = T$. Again, by construction of M_α^{max}, we have $(w_\alpha, w') \in \mathcal{K}_i^{M_\alpha^{max}}$. It thus follows that $T \in Poss_i^{\mathcal{S}}(M_\alpha^{max}, w_\alpha)$, as desired.

From the proof of Theorem 6, we get the following corollary, which will be useful in our later complexity results.

Corollary 30 If $\mathcal{S} \in \{\mathrm{K}_n, \mathrm{T}_n, \mathrm{S4}_n\}$, then the formula α is $\mathcal{S}$-i-honest iff (a) $K_i\alpha$ is $\mathcal{S}$-consistent and (b$'$) for all formulas $\varphi_1, \ldots, \varphi_k$ such that $K_i\varphi_j$ is a subformula of α, for $j = 1, \ldots, k$, if $\models_{\mathcal{S}} K_i\alpha \Rightarrow (K_i\varphi_1 \vee \ldots \vee K_i\varphi_k)$, then $\models_{\mathcal{S}} K_i\alpha \Rightarrow K_i\varphi_j$ for some $j \in \{1, \ldots, k\}$.

Proof The "only if" direction follows immediately from Theorem 6. For the "if" direction, observe that in the proof that α is honest, we did not use the full strength of clause (b) in the statement of Theorem 6; rather it sufficed to consider subformulas of α of the form $K_i\varphi$.

Theorem 7: *All formulas are* K_n*-i-honest.*

Proof Consider any formula α. Clearly $K_i\alpha$ is satisfiable in the K_n situation (M, w) where w is the only world in W^M and $\mathcal{K}_i^M$ is the empty relation. Thus $K_i\alpha$ is K_n-consistent. Next, suppose that $\models_{\mathrm{K}_n} K_i\alpha \Rightarrow (K_i\varphi_1 \vee \ldots K_i\varphi_k)$ and suppose, by way of contradiction, that $K_i\alpha \wedge \neg K_i\varphi_j$ is K_n-consistent for $j = 1, \ldots, k$. Thus, there must be a K_n situation (M_j, w_j) such that $(M_j, w_j) \models K_i\alpha \wedge \neg K_i\varphi_j$. We now proceed much in the spirit of the proof of Theorem 6. Again, without loss of generality, we can assume that if $1 \leq j < j' \leq k$, then W^{M_j} and $W^{M_{j'}}$ are disjoint. We take $W^M = \{w\} \cup (\bigcup_{j=1}^k W^{M_j})$, where w is a fresh world that does not appear in $\bigcup_{j=1}^k W^{M_j}$. We define $\pi^M(w') = \pi^{M_j}(w')$ if $w' \in W^{M_j}$; the definition of $\pi^M(w)$ is irrelevant. For $j' \neq i$, define $\mathcal{K}_{j'}^M$ to be $\cup_{j=1}^k \mathcal{K}_{j'}^{M_j}$. Finally, we define $\mathcal{K}_i^M$ to consist of the union of the $\mathcal{K}_i^{M_j}$ relations together with (w, w') for each world $w' \in \cup_{j=1}^k \mathcal{K}_i^{M_j}(w_j)$. It is easy to show that for a world $w' \in W^{M_j}$, we have $(M_j, w') \equiv (M, w')$. Since $(M_j, w_j) \models K_i\alpha \wedge \neg K_i\varphi_j$, it is easy to see that $(M, w') \models \neg\varphi_j$ for some w' such that $(w_j, w') \in \mathcal{K}_i^{M_j}$. By construction, $(w, w') \in \mathcal{K}_i^M$, so $(M, w) \models \neg K_i\varphi_j$. It is also easy to see that $(M, w) \models K_i\alpha$. Thus, $(M, w) \models K_i\alpha \wedge \neg K_i\varphi_1 \wedge \ldots \wedge \neg K_i\varphi_k$. This contradicts the assumption that $\models_{\mathrm{K}_n} K_i\alpha \Rightarrow (K_i\varphi_1 \vee \ldots \vee K_i\varphi_k)$. It follows that if $\models_{\mathrm{K}_n} K_i\alpha \Rightarrow (K_i\varphi_1 \vee \ldots \vee K_i\varphi_k)$, then $\models_{\mathrm{K}_n} K_i \Rightarrow K_i\varphi_j$ for some $j \in \{1, \ldots, k\}$. From Theorem 6, it follows that α is K_n-honest.

Theorem 8: *If $\mathcal{S} \in \{\mathrm{K}_n, \mathrm{T}_n, \mathrm{S4}_n\}$, then α is $\mathcal{S}$-i-honest iff there is an $\mathcal{S}$-i-stable set S^α containing α which is a subset of every $\mathcal{S}$-i-stable set containing α. Moreover, if α is honest, then $\alpha \mid\!\sim^i_{\mathcal{S}} \beta$ iff $\beta \in S^\alpha$.*

Proof Suppose α is $\mathcal{S}$-i-honest. Then, by definition, there is an $\mathcal{S}$-i-maximum situation for α, say (M_α, w_α). Let $S^\alpha = \{\varphi : (M_\alpha, w_\alpha) \models K_i\varphi\}$. Clearly S^α is an $\mathcal{S}$-i-stable set containing α. Let S be any other $\mathcal{S}$-i-stable set containing α. Suppose $\varphi \notin S$. Let (M, w) be an $\mathcal{S}$ situation (M, w) corresponding to S; we must have $(M, w) \models K_i\alpha \wedge \neg K_i\varphi$. Thus, there is a world $w' \in W^M$ such that $(w, w') \in \mathcal{K}_i^M$ and $(M, w') \models \neg\varphi$. Since (M_α, w_α) is an $\mathcal{S}$-i-maximum situation, $T_{M,w'} \in Poss_i^{\mathcal{S}}(M_\alpha, w_\alpha)$. Thus, there is some world u such that $(w_\alpha, u) \in \mathcal{K}_i^M$ and $T_{M_\alpha,u} = T_{M,w'}$. By Corollary 3, we have $(M_\alpha, u) \models \neg\varphi$. Thus, $(M_\alpha, w_\alpha) \models \neg K_i\varphi$. This means that $\varphi \notin S^\alpha$. Thus, we have shown that S^α is indeed a subset of every $\mathcal{S}$-i-stable set containing α. Moreover, it is clear from this argument that $\alpha \mid\!\sim^i_{\mathcal{S}} \beta$ iff $\beta \in S^\alpha$.

For the converse, suppose that there is an $\mathcal{S}$-i-stable set S^α containing α which is a subset of every $\mathcal{S}$-i-stable set containing α. Clearly $K_i\alpha$ is $\mathcal{S}$-consistent, since it is satisfied in every $\mathcal{S}$ situation corresponding to S^α. We want to show that α is $\mathcal{S}$-i-honest. Suppose that $\models_{\mathcal{S}} K_i\alpha \Rightarrow (K_i\varphi_1 \vee \ldots \vee K_i\varphi_k)$. It follows that every $\mathcal{S}$-i-stable set containing α must contain one of $\varphi_1, \ldots, \varphi_k$. In particular, this is true of S^α. So we can suppose without loss of generality that $\varphi_1 \in S^\alpha$. By definition of S^α, this means that φ_1 is in every $\mathcal{S}$-i-stable set containing α. We must have $\models_{\mathcal{S}} K_i\alpha \Rightarrow K_i\varphi_1$, for if $K_i\alpha \wedge \neg K_1\varphi_1$ were $\mathcal{S}$-consistent, there would be an $\mathcal{S}$ situation (M, w) such that $(M, w) \models K_i\alpha \wedge \neg K_i\varphi_1$. But then $S = \{\beta : (M, w) \models K_i\beta\}$ is an $\mathcal{S}$-i-stable set containing α and not φ_1, a contradiction. It now follows from Theorem 6 that α is honest.

Theorem 9: *If $\mathcal{S} \in \{\mathrm{K}_n, \mathrm{T}_n, \mathrm{S4}_n\}$ and α is $\mathcal{S}$-i-honest, then $\alpha \mid\!\sim^i_{\mathcal{S}} \beta$ iff $\models_{\mathcal{S}} K_i\alpha \Rightarrow K_i\beta$.*

Proof Suppose α is $\mathcal{S}$-i-honest. Clearly if $\models_{\mathcal{S}} K_i\alpha \Rightarrow K_i\beta$, then $\alpha \mid\!\sim^i_{\mathcal{S}} \beta$. For the converse, suppose that $K_\alpha \wedge \neg K_i\beta$ is $\mathcal{S}$-consistent. It follows that there is an $\mathcal{S}$-i-stable set containing α and not β. From Theorem 8, it follows that it is not the case that $\alpha \mid\!\sim^i_{\mathcal{S}} \beta$.

B Proofs for Section 5

Proposition 13: *If (M, w) and (M', w') are $\mathcal{S}$ situations, $\mathcal{S} \in \{\mathrm{K45}_n, \mathrm{KD45}_n\}$, and $Poss_i^{\mathcal{S}}(M, w) = Poss_i^{\mathcal{S}}(M', w')$, then $(M, w) \equiv^i (M', w')$.*

Proof Suppose the hypotheses of the proposition hold and the conclusion does not. Then, without loss of generality, there exists a formula φ such that $(M, w) \models K_i\varphi$ and $(M', w') \models \neg K_i\varphi$. Without loss of generality, we can assume that φ is the formula of minimum depth with this property, so that for all formulas ψ with $depth(\psi) < depth(\varphi)$, we have $(M, w) \models K_i\psi$ iff $(M', w') \models K_i\psi$. Using the following equivalences, it is not hard to show that φ must be an i-objective formula:

- $\models_{\mathcal{S}} K_i(\varphi_1 \wedge \varphi_2) \Leftrightarrow (K_i\varphi_1 \wedge K_i\varphi_2)$
- $\models_{\mathcal{S}} K_i(\varphi_1 \vee K_i\varphi_2) \Leftrightarrow (K_i\varphi_1 \vee K_i\varphi_2)$
- $\models_{\mathcal{S}} K_i(\varphi_1 \vee \neg K_i\varphi_2) \Leftrightarrow (K_i\varphi_1 \vee \neg K_i\varphi_2)$
- $\models_{\mathcal{S}} K_i\varphi \Leftrightarrow K_iK_i\varphi$
- $\models_{\mathcal{S}} \neg K_i \mathit{false} \Rightarrow (\neg K_i\varphi \Leftrightarrow K_i\neg K_i\varphi)$.

But then there must exist a world w'' such that $(w', w'') \in \mathcal{K}_i^{M'}$ and $(M', w'') \models \neg\varphi$. Since $(M, v) \models \varphi$ for all v such that $(w, v) \in \mathcal{K}_i^M$, it follows from Corollary 12 that $T^i_{M',w''} \notin \mathit{Poss}_i^{\mathcal{S}}(M, w)$, although by definition $T^i_{M',w''} \in \mathit{Poss}_i^{\mathcal{S}}(M', w')$. Thus, $\mathit{Poss}_i^{\mathcal{S}}(M, w) \neq \mathit{Poss}_i^{\mathcal{S}}(M', w')$.

Before proving Proposition 14, we need one technical result. For the purposes of this appendix, define a situation (M, w) to be *i-special* if for all $j \neq i$ and all $w' \in W^M$, we have $\mathcal{K}_i^M(w) \cap \mathcal{K}_j(w') = \emptyset$. Thus, (M, w) is i-special if the i-successors of w are not j-successors of any other world, for $j \neq i$. The following result says that for K45$_n$ and KD45$_n$, we can restrict attention to i-special situations without loss of generality.

Lemma 31 If $\mathcal{S} \in \{\mathrm{K45}_n, \mathrm{KD45}_n\}$, then for all $\mathcal{S}$ situations (M, w) and all agents i, there exists an i-special situation (M', w') such that $(M, w) \equiv (M', w')$.

Proof Given (M, w) and i, we construct (M', w') as follows. Let W_i be a disjoint copy of $\mathcal{K}_i^M(w) \cup \{w\}$. More formally, W_i is a set of worlds that do not appear in W^M of the same cardinality as $\mathcal{K}_i^M(w) \cup \{w\}$. Let e be a 1-1 function from W_i to $\mathcal{K}_i(w) \cup \{w\}$. Suppose w' is the world in W_i such that $e(w') = w$. Let W' be $W^M \cup W_i$. Extend e to all of W' by defining $e(v) = v$ for $v \in W$. (Of course, e is no longer 1-1.) Let $M' = (W', \mathcal{K}_1', \ldots, \mathcal{K}_n', \pi')$, where $\mathcal{K}_j'(v) = \mathcal{K}_j(e(v))$ if either $j \neq i$ or $v \notin W_i$, $\mathcal{K}_i(v) = W_i$ for $v \in W_i$, and $\pi'(v) = \pi(e(v))$. By construction, (M', w') is i-special. A straightforward induction on structure of formulas shows that for all formulas φ and all $v \in W'$, we have $(M', v) \models \varphi$ iff $(M, e(v)) \models \varphi$. Since $e(w') = w$, it follows that $(M, w) \equiv (M', w')$.

Proposition 14: *If $\mathcal{S} \in \{\mathrm{K45}_n, \mathrm{KD45}_n\}$, then for all agents i and $\mathcal{S}$ situations (M_1, w_1) and (M_2, w_2), there is an $\mathcal{S}$ situation (M_3, w_3) such that $\mathit{Poss}_i^{\mathcal{S}}(M_3, w_3) = \mathit{Poss}_i^{\mathcal{S}}(M_1, w_1) \cup \mathit{Poss}_i^{\mathcal{S}}(M_2, w_2)$.*

Proof Given $\mathcal{S}$ situations (M_1, w_1) and (M_2, w_2) we want to show that there exists an $\mathcal{S}$ situation (M_3, w_3) such that $\mathit{Poss}_i^{\mathcal{S}}(M_3, w_3) = \mathit{Poss}_i^{\mathcal{S}}(M_1, w_1) \cup \mathit{Poss}_i^{\mathcal{S}}(M_2, w_2)$. Recall that in Proposition 5 we defined a situation (M_3, w_3) such that $W^{M_3} = W^{M_1} \cup W^{M_2} \cup \{w_3\}$ and $\mathcal{K}_i^{M_3}(w_3) = \mathcal{K}_i^{M_1}(w_1) \cup \mathcal{K}_1^{M_2}(w_2)$, and showed that $\mathit{Poss}_i^{\mathcal{S}}(M_3, w_3) \supseteq \mathit{Poss}_i^{\mathcal{S}}(M_1, w_1) \cup \mathit{Poss}_i^{\mathcal{S}}(M_2, w_2)$, for $\mathcal{S} \in \{\mathrm{K}_n, \mathrm{T}_n, \mathrm{S4}_n\}$. We might hope that a similar construction would work here. The naive construction does not quite work, since $\mathcal{K}_i^{M_3}$ will not be Euclidean if $\mathcal{K}_i^{M_1}(w_1)$ and $\mathcal{K}_i^{M_2}(w_2)$ are both nonempty: If $v_1 \in \mathcal{K}_i^{M_1}(w_1)$

and $v_2 \in \mathcal{K}_i^{M_2}(w_2)$, then the construction has the property that both v_1 and v_2 are in $\mathcal{K}_i^{M_3}(w_3)$. Euclideanity then requires that $(v_1, v_2) \in \mathcal{K}_i^{M_3}$, which does not follow from our construction. This means that (M_3, w_3) is not an $\mathcal{S}$-situation, for $\mathcal{S} \in \{\mathrm{K45}_n, \mathrm{KD45}_n\}$.

This problem is easily fixed, by appropriately modifying the $\mathcal{K}_i^{M_3}$ relation so that it is Euclidean. However, doing this may cause problems if M_1 and M_2 are not i-special. For example, suppose that $w' \in \mathcal{K}_i^{M_1}(w_1) \cap \mathcal{K}_j^{M_1}(w')$ for some $j \neq i$. Since $\mathcal{K}_i^{M_3}$ is Euclidean, we must have $\mathcal{K}_i^{M_3}(w') = \mathcal{K}_i^{M_3}(w_1)$. Moreover, by construction, $\mathcal{K}_i^{M_3}(w_1) \neq \mathcal{K}_i^{M_1}(w_1)$. Thus, $T^i_{M_3,w'} \neq T^i_{M_1,w'}$. Both have a j-successor of the root corresponding to w', but the i-successors of this j-successor must be different in the two trees. It follows that $Poss_i^{\mathcal{S}}(M_3, w_3)$ will be incomparable to $Poss_i^{\mathcal{S}}(M_1, w_1)$, rather than being a superset of it.

We can get around this problem by assuming that (M_1, w_1) and (M_2, w_2) are both i-special. (If not, then by Lemma 31, we can find i-special situations (M_j', w_j') such that $(M_j', w_j') \equiv (M_j, w_j)$, for $j = 1, 2$. Since equivalent situations have the same set of possibilities, we can use these instead.) As in Proposition 5, we now define (M_3, w_3) so that $W^{M_3} = W^{M_1} \cup W^{M_2} \cup \{w_3\}$ and $\mathcal{K}_j^{M_3} = \mathcal{K}_j^{M_1} \cup \mathcal{K}_j^{M_2}$ for $j \neq i$. We define $\mathcal{K}_i^{M_3} = \mathcal{K}_i^{M_1} \cup \mathcal{K}_i^{M_2} \cup \{(u, v) : u \in \{w_3\} \cup \mathcal{K}_i^{M_1}(w_1) \cup \mathcal{K}_i^{M_2}(w_2), v \in \mathcal{K}_i^{M_1}(w_1) \cup \mathcal{K}_i^{M_2}(w_2)\}$. We leave it to the reader to check that with this definition, (M_3, w_3) is an $\mathcal{S}$-situation and $Poss_i^{\mathcal{S}}(M_3, w_3) = Poss_i^{\mathcal{S}}(M_1, w_1) \cup Poss_i^{\mathcal{S}}(M_2, w_2)$.

Theorem 15: *For $\mathcal{S} \in \{\mathrm{K45}_n, \mathrm{KD45}_n\}$, the formula α is $\mathcal{S}$-i-honest iff (a) $K_i\alpha$ is $\mathcal{S}$-consistent and (b) for all i-objective formulas $\varphi_1, \ldots, \varphi_k$, if $\models_{\mathcal{S}} K_i\alpha \Rightarrow (K_i\varphi_1 \vee \ldots \vee K_i\varphi_k)$ then $\models_{\mathcal{S}} K_i\alpha \Rightarrow K_i\varphi_j$ for some $j \in \{1, \ldots, k\}$.*

Proof The proof follows lines similar to those of Theorem 6. Suppose $\mathcal{S} \in \{\mathrm{K45}_n, \mathrm{KD45}_n\}$. For the "only if" direction, suppose that α is $\mathcal{S}$-i-honest. Then there is an $\mathcal{S}$ situation (M, w) such that $(M, w) \models K_i\alpha$, and for all $\mathcal{S}$ situations (M', w'), if $(M', w') \models K_i\alpha$, then $Poss_i^{\mathcal{S}}(M', w') \subseteq Poss_i^{\mathcal{S}}(M, w)$. Clearly $K_i\alpha$ is $\mathcal{S}$-consistent. Now suppose that $\models_{\mathcal{S}} K_i\alpha \Rightarrow (K_i\varphi_1 \vee \ldots \vee K_i\varphi_k)$, where $\varphi_1, \ldots, \varphi_k$ are i-objective. Suppose, by way of contradiction, that $K_i\alpha \wedge \neg K_i\varphi_j$ is $\mathcal{S}$-consistent for $j = 1, \ldots, k$. This means that there is some $\mathcal{S}$ situation (M_j, w_j) such that $(M_j, w_j) \models K_i\alpha$ and a world w_j' such that $(w_j, w_j') \in \mathcal{K}_i$ and $(M_j, w_j') \models \neg\varphi_j$, for $j = 1, \ldots, k$. By assumption, $T^i_{M_j,w_j'} \in Poss_i^{\mathcal{S}}(M, w)$. Thus, there is a world $w' \in W^M$ such that $(w, w') \in \mathcal{K}_i^M$ and $T^i_{M,w'} = T^i_{M_j,w_j'}$. Since φ_j is i-objective, it follows from Proposition 11 that $(M, w') \models \neg\varphi_j$. Hence, $(M, w) \models \neg K_i\varphi_j$, for $j = 1, \ldots, k$. But this contradicts the assumption that $\models_{\mathcal{S}} K_i\alpha \Rightarrow (K_i\varphi_1 \vee \ldots \vee K_i\varphi_k)$. Thus $\models_{\mathcal{S}} K_i\alpha \Rightarrow K_i\varphi_j$ for some $j \in \{1, \ldots, k\}$.

For the converse, suppose $K_i\alpha$ satisfies conditions (a) and (b) in the theorem. Let F consist of all the formulas of the form $K_i\varphi$ in the set $B^i_{\mathcal{S}}(\alpha, \alpha)$ (see Definition 17) such that $K_i\alpha \wedge \neg K_i\varphi$ is consistent. By (a) and (b), there is an $\mathcal{S}$ situation (M_α, w_α) satisfying $K_i\alpha \wedge \bigwedge_{\varphi \in F} \neg K_i\varphi$. Suppose $A^i_{\mathcal{S}}(\alpha, \alpha) = \langle \alpha_0, \ldots, \alpha_m \rangle$ (see Definition 17). Let $C_{i,\alpha} = \{K_i\psi : \models_{\mathcal{S}} K_i\alpha \Rightarrow K_i\psi\}$. (Note that $C_{i,\alpha}$ is independent of whether we take $\mathcal{S}$ to be K45$_n$ or KD45$_n$.) Let ψ_α be the formula

$$\left(\bigwedge_{K_i\psi \in B^i_{\mathcal{S}}(\alpha,\alpha)\cap C_{i,\alpha}} K_i\psi\right) \wedge \left(\bigwedge_{K_i\psi \in B^i_{\mathcal{S}}(\alpha,\alpha)-Ci,\alpha} \neg K_i\psi\right).$$

Since $(M_\alpha, w_\alpha) \models K_i\alpha$, it is easy to see that in fact $(M_\alpha, w_\alpha) \models \psi_\alpha$. Our construction of $A^i_{\mathcal{S}}(\alpha,\alpha)$ also guarantees that $\models_{\mathcal{S}} \psi_\alpha \Rightarrow (\alpha_j \Leftrightarrow \alpha_{j+1})$, for $j = 0, \ldots, m-1$. Since $\alpha_0 = K_i\alpha$ and $(M_\alpha, w_\alpha) \models K_i\alpha$, it follows that $(M_\alpha, w_\alpha) \models \alpha_m$, so $\alpha_m = \mathit{true}$.

We now construct an $\mathcal{S}$-i-maximum situation $(M^{max}_\alpha, w_\alpha)$ for $K_i\alpha$ much as in Theorem 6. Again, let $\mathcal{N}_\alpha$ be a set of i-special situations including (M_α, w_α) such that each situation $(M, w) \in \mathcal{N}_\alpha$ satisfies $K_i\alpha$, and if $T \in \mathit{Poss}^{\mathcal{S}}_i(M', w')$ for some situation (M', w') satisfying $K_i\alpha$, then $T \in \mathit{Poss}^{\mathcal{S}}_i(M, w)$ for some situation $(M, w) \in \mathcal{N}_\alpha$. (The fact that we can assume that all the situation in $\mathcal{N}_\alpha$ are i-special follows from Lemma 31.) Again, we assume that if (M, w) and (M', w') are two situations in $\mathcal{N}_\alpha$, then the worlds in W^M and $W^{M'}$ are disjoint. We take the worlds in M^{max}_α to consist of the union of all the worlds in W^M for each situation (M, w) in $\mathcal{N}_\alpha$. We define $\pi^{M^{max}_\alpha}(w) = \pi^M(w')$ if w' is a world in W^M for some $(M, w) \in \mathcal{N}_\alpha$. We define $\mathcal{K}^{M^{max}_\alpha}_j$ to be the union of the $\mathcal{K}^M_j$ relations for the situations $(M, w) \in \mathcal{N}_\alpha$ for $j \neq i$. Finally, $\mathcal{K}^{M^{max}_\alpha}_i$ is least Euclidean, transitive relation that includes $\mathcal{K}^M_i$ for each situation $(M, w) \in \mathcal{N}_\alpha$, together with (w_α, w'), where $w' \in \mathcal{K}^M_i(w)$ for some $(M, w) \in \mathcal{N}_\alpha$. We leave it to the reader to check that with these definitions, $(M^{max}_\alpha, w_\alpha)$ is an $\mathcal{S}$ situation.

We now want to show $(M^{max}_\alpha, w_\alpha) \models K_i\alpha$. Again, we do this by proving that if (M, w) is a situation in $\mathcal{N}_\alpha$ and w' is a world in W^M, then $(M, w') \models \beta$ iff $(M^{max}_\alpha, w') \models \beta$ for every subformula β of $K_i\alpha$. The details are much as in the proof of Theorem 6, so are omitted here. It follows that $(M^{max}_\alpha, w_\alpha)$ is an $\mathcal{S}$-i-maximum situation for $K_i\alpha$.

As a corollary to the proof of Theorem 15, we get the following analogue to Corollary 30, whose proof is essentially identical to that of Corollary 30.

Corollary 32 If $\mathcal{S} \in \{\mathrm{K45}_n, \mathrm{KD45}_n\}$, then the formula α is $\mathcal{S}$-i-honest iff (a) $K_i\alpha$ is $\mathcal{S}$-consistent and (b$'$) for all formulas $\varphi_1, \ldots, \varphi_k$ in $B^i_{\mathcal{S}}(\alpha,\alpha)$, if $\models_{\mathcal{S}} K_i\alpha \Rightarrow (K_i\varphi_1 \vee \ldots \vee K_i\varphi_k)$, then $\models_{\mathcal{S}} K_i\alpha \Rightarrow K_i\varphi_j$ for some $j \in \{1, \ldots, k\}$.

Theorem 16: *For $\mathcal{S} \in \{\mathrm{K45}_n, \mathrm{KD45}_n\}$, a formula α is $\mathcal{S}$-i-honest iff there is an $\mathcal{S}$-i-stable set S^α containing α such that for all $\mathcal{S}$-i-stable sets S containing α we have* $\ker_i(S^\alpha) \subseteq \ker_i(S)$. *Moreover, if α is $\mathcal{S}$-i-honest, then $\alpha \mid\!\sim^i_{\mathcal{S}} \beta$ iff $\beta \in S^\alpha$.*

Proof The proof follows almost the same lines as that of Theorem 8. Suppose α is $\mathcal{S}$-i-honest. Then, by definition, there is an $\mathcal{S}$-i-maximum situation for α, say (M_α, w_α). Let $S^\alpha = \{\varphi : (M_\alpha, w_\alpha) \models K_i\varphi\}$. Clearly S^α is an $\mathcal{S}$-i-stable set containing α. Let S be any other $\mathcal{S}$-i-stable set containing α. Suppose $\varphi \notin \ker_i(S)$. Let (M, w) be an $\mathcal{S}$ situation (M, w) corresponding to S; we must have $(M, w) \models K_i\alpha \wedge \neg K_i\varphi$. Thus, there is a world $w' \in W^M$ such that $(w, w') \in \mathcal{K}^M_i$ and $(M, w') \models \neg\varphi$. Since (M_α, w_α) is an $\mathcal{S}$-i-maximum situation, $T^i_{M,w'} \in \mathit{Poss}^{\mathcal{S}}_i(M_\alpha, w_\alpha)$. Thus, there is some world u

such that $(w_\alpha, u) \in \mathcal{K}_i^M$ and $T^i_{M_\alpha,u} = T^i_{M,w'}$. By Corollary 3, we have $(M_\alpha, u) \models \neg\varphi$. Thus, $(M_\alpha, w_\alpha) \models \neg K_i\varphi$. This means that $\varphi \notin S^\alpha$. Since φ was chosen arbitrarily, it follows that $\ker_i(S^\alpha) \subseteq \ker_i(S)$. Moreover, it is clear from this argument that $\alpha \mid\!\sim^i_{\mathcal{S}} \beta$ iff $\beta \in S^\alpha$.

For the converse, suppose that there is an $\mathcal{S}$-i-stable set S^α containing α such that $\ker_i(S^\alpha) \subseteq \ker_i(S)$ for every other $\mathcal{S}$-i-stable set containing α. We want to show that α is $\mathcal{S}$-i-honest. Suppose that $\models_{\mathcal{S}} K_i\alpha \Rightarrow (K_i\varphi_1 \vee \ldots \vee K_i\varphi_k)$, where $\varphi_1, \ldots, \varphi_k$ are i-objective formulas. It follows that every $\mathcal{S}$-i-stable set containing α must contain one of $\varphi_1, \ldots, \varphi_k$. In particular, this is true of S^α. So we can suppose without loss of generality that $\varphi_1 \in S^\alpha$. Since φ_1 is i-objective, we in fact have $\varphi_1 \in \ker_i(S^\alpha)$. By definition of S^α, this means that φ_1 is in $\ker_i(S)$ for every other $\mathcal{S}$-i-stable set S containing α. We must have $\models_{\mathcal{S}} K_i\alpha \Rightarrow K_i\varphi_1$, for if $K_i\alpha \wedge \neg K_1\varphi_1$ were consistent, there would be an $\mathcal{S}$ situation (M, w) such that $(M, w) \models K_i\alpha \wedge \neg K_i\varphi_1$. But then $S = \{\beta : (M, w) \models K_i\beta\}$ is an $\mathcal{S}$-i-stable set containing α and not φ_1, a contradiction. It now follows from Theorem 15 that α is $\mathrm{S5}_n$-i-honest.

Theorem 19: *For $\mathcal{S} \in \{\mathrm{K45}_n, \mathrm{KD45}_n\}$, the formula α is $\mathcal{S}$-i-honest iff $K_i\alpha$ is $\mathcal{S}$-consistent and $\alpha \in D^i_{\mathcal{S}}(\alpha)$. If α is $\mathcal{S}$-i-honest, then $\alpha \mid\!\sim^i_{\mathcal{S}} \beta$ iff $\beta \in D^i_{\mathcal{S}}(\alpha)$.*

Proof Suppose α is $\mathcal{S}$-i-honest. Then there is an $\mathcal{S}$-i-maximum situation for α, say $(M^{max}_\alpha, w_\alpha)$. It follows that $K_i\alpha$ is $\mathcal{S}$-consistent. Moreover, we claim that $\beta \in D^i_{\mathcal{S}}(\alpha)$ iff $(M^{max}_\alpha, w_\alpha) \models K_i\beta$. Using the notation in the proof of Theorem 15, let ψ_β be the formula

$$\left(\bigwedge_{K_i\psi \in B^i_{\mathcal{S}}(\alpha,\beta) \cap C_{i,\alpha}} K_i\psi \right) \wedge \left(\bigwedge_{K_i\psi \in B^i_{\mathcal{S}}(\alpha,\beta) - C_{i,\alpha}} \neg K_i\psi \right).$$

It is easy to see (and is actually shown in the proof of Theorem 15) that $(M^{max}_\alpha, w_\alpha) \models \psi_\beta$. Moreover, suppose that $A^i_{\mathcal{S}}(\alpha, \beta) = \langle \beta_0, \ldots, \beta_m \rangle$. Our construction guarantees that $\models_{\mathcal{S}} \psi_\beta \Rightarrow (\beta_j \Leftrightarrow \beta_{j+1})$, for $j = 0, \ldots, m-1$. In particular, this means that $(M^{max}_\alpha, w_\alpha) \models \beta_0 \Leftrightarrow \beta_m$. Recall that $\beta_0 = K_i\beta$ and β_m is either *true* or *false*. If $(M^{max}_\alpha, w_\alpha) \models K_i\beta$, then we must have $\beta_m = \mathit{true}$, so $\beta \in D^i_{\mathcal{S}}(\alpha)$. If $(M^{max}_\alpha, w_\alpha) \models \neg K_i\beta$, then we must have $\beta_m = \mathit{false}$, so $\beta \notin D^i_{\mathcal{S}}(\alpha)$. Taking $\beta = \alpha$, since $(M^{max}_\alpha, w_\alpha) \models K_i\alpha$, we have that $\alpha_m = \mathit{true}$, so $\alpha \in D^i_{\mathcal{S}}(\alpha)$. It also immediately follows that if α is $\mathcal{S}$-i-honest, then $\alpha \mid\!\sim^i_{\mathcal{S}} \beta$ iff $\beta \in D^i_{\mathcal{S}}(\alpha)$.

Now suppose that $\alpha \in D^i_{\mathcal{S}}(\alpha)$ and $K_i\alpha$ is consistent. Let $A^i_{\mathcal{S}}(\alpha, \alpha) = \langle \alpha_0, \ldots, \alpha_m \rangle$. Since $\alpha \in D^i_{\mathcal{S}}(\alpha)$, we must have $\alpha_m = \mathit{true}$. Construct the situation $(M^{max}_\alpha, w_\alpha)$ as in the proof of Theorem 15. (Since $K_i\alpha$ is $\mathcal{S}$-consistent, we can carry out this construction.) As shown in that proof, $(M^{max}_\alpha, w_\alpha) \models \psi_\alpha$. Moreover, as we have observed, $\models_{\mathcal{S}} \psi_\alpha \Rightarrow (\alpha_0 \Leftrightarrow \alpha_m)$. It follows that $(M^{max}_\alpha, w_\alpha) \models K_i\alpha$. Thus, α is $\mathcal{S}$-i-honest, since $(M^{max}_\alpha, w_\alpha)$ is clearly an $\mathcal{S}$-i-maximum situation for $K_i\alpha$.

C Proofs for Section 6

Proposition 20: *For each* $\mathrm{S5}_n$ *situation* (M,w) *and for all* k*, there is a unique* i*-objective* k*-*$*$*-tree* $T^{i,*}_{M,w,k}$ *such that* $(M,w) \equiv^{Q_i}_k M(T^{i,*}_{M,w,k})$.

Proof The inductive construction proceeds much like that in Proposition 2. The only difference is that we give the world w special treatment in our construction. Given a structure M and worlds $w, w' \in W^M$, we construct a $*$-tree that we call $T_{M,w',w,k}$ by induction on k. $T_{M,w',w,0}$ is just $T_{M,w',0}$, and $T_{M,w',w,k+1}$ consists of a root labeled by the truth assignment at w', and for each pair $(w', w'') \in \mathcal{K}_i$, an edge labeled i to the root of $T_{M,w'',w,k}$, unless $w'' = w$. In that case, we construct an edge labeled i to a node labeled $*$. We then again eliminate duplicate trees, as in the proof of Proposition 2.

We take $T^{i,*}_{M,w,k} = T^i_{M,w,w,k}$. We now must show that $(M,w) \equiv^{Q_i}_k M(T^{i,*}_{M,w,k})$. Using techniques similar to those of Proposition 2, it is easy to show that $(M,w) \equiv_k M(T_{M,w,w,k})$. It easily follows that if (M,w) can be i-embedded in (M',w'), then $M(T^{i,*}_{M,w,k})$ can be i-embedded in a situation (M'',w'') such that $(M',w') \equiv_k (M'',w'')$. Conversely, we can show that if $M(T^{i,*}_{M,w,k})$ can be i-embedded in a situation (M'',w''), then (M,w) can be i-embedded in a situation (M',w') such that $(M',w') \equiv_k (M'',w'')$. The desired result immediately follows.

The uniqueness of $T^{i,*}_{M,w,k}$ follows along the same lines as in Proposition 2; we omit details here.

We need the following lemma to prove a number of the remaining results.

Lemma 33 Suppose (M,w) and (M',w') are $\mathrm{S5}_n$ situations such that $T^{i,*}_{M,w} = T^{i,*}_{M',w'}$ and for every i-subjective subformula ψ of β, we have $(M,w) \models \psi$ iff $(M',w') \models \psi$. Then $(M,w) \models \beta$ iff $(M',w') \models \beta$.

Proof Suppose (M,w) and (M',w') satisfy the hypotheses of the lemma. We prove the result by induction on the structure of ψ, but we need a somewhat stronger induction hypothesis. We prove by induction on the structure of ψ that if ψ is a subformula of β such that $depth(\psi) \leq k$, then (a) $(M,w) \models \psi$ iff $(M',w') \models \psi$ and (b) for arbitrary worlds $v \in W^M$ and $v' \in W^{M'}$, if $\mathcal{T}_{M,v,w,k} = \mathcal{T}_{M',v',w',k}$, then $(M,v) \models \psi$ iff $(M',v') \models \psi$. We focus on part (a) here. The only nontrivial case is if ψ is of the form $K_j\psi'$. Suppose that $(M,w) \models K_j\psi'$. We want to show $(M',w') \models K_j\psi'$. If $j = i$ the result is immediate by assumption, so suppose $j \neq i$. Suppose that $u' \in \mathcal{K}^{M'}_j(w')$. If $u' = w'$, it is immediate from the induction hypothesis that $(M',w') \models \psi'$. So suppose that $u' \neq w'$. Since $T^{i,*}_{M,w} = T^{i,*}_{M',w'}$, there must be some $u \in \mathcal{K}^M_i(w)$ such that $T_{M,u,w,k-1} = T_{M,u',w',k-1}$. Since $(M,w) \models K_j\psi'$, we must have $(M,u) \models \psi'$. Since $depth(\psi) \leq k-1$, by part (b) of our main induction hypothesis, it follows that $(M',u') \models \psi'$. Hence, $(M',w') \models K_j\psi'$. The converse follows by a symmetric argument. The proof of part (b) is similar and left to the reader.

Proposition 22: *If* (M,w) *and* (M',w') *are* $\mathrm{S5}_n$ *situations such that* $Poss^{\mathrm{S5}_n}_i(M,w) = Poss^{\mathrm{S5}_n}_i(M',w')$*, then* $(M,w) \equiv^i (M',w')$.

Proof Suppose $Poss_i^{S5_n}(M,w) = Poss_i^{S5_n}(M',w')$. We say that $w_1 \in \mathcal{K}_i^M(w)$ and $w_2 \in \mathcal{K}_i^{M'}(w')$ *correspond* if $T^{i,*}_{M,w_1} = T^{i,*}_{M',w_2}$. Since $Poss_i^{S5_n}(M,w) = Poss_i^{S5_n}(M',w')$, for each world in $\mathcal{K}_1^M(w)$, there is a corresponding world in $\mathcal{K}_1^{M'}(w')$, and for each world in $\mathcal{K}_1^{M'}(w')$, there is a corresponding world in $\mathcal{K}_1^M(w)$. We show that if w_1 and w_2 are corresponding worlds in W^M and $W^{M'}$ respectively, then $(M,w_1) \equiv (M',w_2)$. We proceed by induction on structure. Suppose $(M,w_1) \models \varphi$. We must show that $(M',w_2) \models \varphi$. The result follows immediately from Lemma 33 and the induction hypothesis except in the case that φ is of the form $K_i\varphi'$. In this case, suppose that $u' \in \mathcal{K}_i^{M'}(w_2) = \mathcal{K}_i^{M'}(w')$. This means there is some $u \in \mathcal{K}_i^M(w)$ corresponding to u'. Since we must have $(M,u) \models \varphi$, by the induction hypothesis it follows that $(M',u') \models \varphi$. Hence $(M',w_2) \models K_i\varphi$. A symmetric argument applies to the converse implication.

Proposition 23: *For all agents i and* $S5_n$ *situations (M_1,w_1) and (M_2,w_2), there is an* $S5_n$ *situation (M_3,w_3) such that* $Poss_i^{S5_n}(M_3,w_3) = Poss_i^{S5_n}(M_1,w_1) \cup Poss_i^{S5_n}(M_2,w_2)$.

Proof Without loss of generality, we can assume W^{M_1} and W^{M_2} are disjoint. Let $W^{M_3} = W^{M_1} \cup W^{M_2}$. Define π^{M_3} so that $\pi^{M_3}|_{W^{M_1}} = \pi^{M_1}$ and $\pi_3|_{W^{M_2}} = \pi^{M_2}$. Let $\mathcal{K}_j^{M_3} = \mathcal{K}_j^{M_1} \cup \mathcal{K}_j^{M_2}$ for $j \neq i$, and define $\mathcal{K}_i^{M_3}$ to be the smallest reflexive, Euclidean, and transitive relation (i.e., equivalence relation) containing $\mathcal{K}_i^{M_1}$, $\mathcal{K}_i^{M_2}$ and the pair (w_1,w_2). It is easy to see that $Poss_i^{S5_n}(M_3,w_1) = Poss_i^{S5_n}(M_1,w_1) \cup Poss_i^{S5_n}(M_2,w_2)$.

Theorem 24: *The formula α is* $S5_n$*-i-honest iff (a) $K_i\alpha$ is* $S5_n$*-consistent and (b) for all Q_i-formulas $\varphi_1,\ldots,\varphi_k$, if $\models_{S5_n} K_i\alpha \Rightarrow (K_i\varphi_1 \vee \ldots \vee K_i\varphi_k)$ then $\models_{S5_n} K_i\alpha \Rightarrow K_i\varphi_j$ for some $j \in \{1,\ldots,k\}$.*

Proof The proof again follows very similar lines to that of Theorem 6. For the "if" direction, suppose that α is $S5_n$-i-honest. Thus, there is an $S5_n$ situation (M,w) such that $(M,w) \models K_i\alpha$, and for all $S5_n$ situations (M',w'), if $(M',w') \models K_i\alpha$, then $Poss_i^{S5_n}(M',w') \subseteq Poss_i^{S5_n}(M,w)$. Clearly $K_i\alpha$ must be $S5_n$-consistent. Now suppose that $\models_{S5_n} K_i\alpha \Rightarrow (K_i\varphi_1 \vee \ldots \vee K_i\varphi_k)$, where $\varphi_1,\ldots,\varphi_k$ are Q_i-formulas. To obtain a contradiction, suppose $K_i\alpha \wedge \neg K_i\varphi_j$ is $S5_n$-consistent for $j = 1,\ldots,k$. This means that there is some $S5_n$ situation (M_j,w_j) such that $(M_j,w_j) \models K_i\alpha$ and a world $w_j' \in \mathcal{K}_i^{M_j}(w_j)$ such that $(M_j,w_j') \models \neg\varphi_j$, for $j = 1,\ldots,k$. By assumption, $T^{i,*}_{M_j,w_j'} \in Poss_i^{S5_n}(M,w)$. Thus, there is a world $w' \in \mathcal{K}_i^M(w)$ such that $T^{i,*}_{M,w'} = T^{i,*}_{M_j,w_j'}$. It follows from Proposition 20 that $(M,w') \models \neg\varphi_j$. Hence, $(M,w) \models \neg K_i\varphi_j$, for $j = 1,\ldots,k$. But this contradicts the assumption that $\models_{S5_n} K_i\alpha \Rightarrow (K_i\varphi_1 \vee \ldots \vee K_i\varphi_k)$.

For the converse, suppose $K_i\alpha$ satisfies conditions (a) and (b) in the theorem. Recall the definition of $D_Q^i(\alpha)$ given just before Theorem 26 in the main text. Let F consist of all formulas of the form $Q_i^{\xi_\psi}\psi$, where $K_i\psi$ is a subformula of α and $\psi \notin D_Q^i(\alpha)$. This means that $K_i\alpha \wedge \neg K_i Q_i^{\xi_\psi}\psi$ is consistent for $\psi \in F$. It follows from (a) and (b) that there is an $S5_n$ situation (M_α,w_α) satisfying $K_i\alpha \wedge \bigwedge_{\varphi\in F} \neg K_i\varphi$.

We claim that $(M_\alpha, w_\alpha) \models \xi_\alpha$. To show this, we prove by induction that for each subformula φ of α, we have that $(M_\alpha, w_\alpha) \models \xi_\varphi$. This requires showing that for each subformula of φ of the form $K_i\psi$, we have $(M_\alpha, w_\alpha) \models K_i\psi$ iff $\psi \in D_i^Q(\alpha)$. Suppose that $\psi \in D_i^Q(\alpha)$. Thus, $\models_{\mathrm{S5}_n} K_i\alpha \Rightarrow K_iQ_i^{\xi_\psi}\psi$. Since $(M_\alpha, w_\alpha) \models K_i\alpha$, it follows that $(M_\alpha, w_\alpha) \models K_iQ_i^{\xi_\psi}\psi$. From the induction hypothesis, it follows that $(M_\alpha, w_\alpha) \models \xi_\psi$. Since ξ_ψ is an i-subjective formula, it follows that $(M_\alpha, w) \models \xi_\psi$ for all $w \in \mathcal{K}_i^{M_\alpha}(w_\alpha)$. Since (M_α, w) can be embedded in itself and $(M_\alpha, w) \models Q_i^{\xi_\psi}\psi$, it follows that $(M_\alpha, w) \models \psi$. Since this is true for all $w \in \mathcal{K}_i^{M_\alpha}(w_\alpha)$, we have that $(M_\alpha, w_\alpha) \models K_i\psi$. On the other hand, if $\psi \notin D_i^Q(\alpha)$, then by construction, $(M_\alpha, w_\alpha) \models \neg K_iQ_i^{\xi_\psi}\psi$. Thus, there must be some $w \in \mathcal{K}_i^{M_\alpha}(w_\alpha)$ such that $(M_\alpha, w) \models \neg Q_i^{\xi_\psi}\psi$. Thus, there must be a situation (M, w') such that (M_α, w) is i-embedded in (M, w') and $(M, w') \models \xi_\psi \wedge \neg\psi$. By Lemma 21, $T^{i,*}_{M_\alpha,w} = T^{i,*}_{M,w'}$. By the induction hypothesis, $(M_\alpha, w) \models \xi_\psi$. Thus, it follows from Lemma 33 that $(M_\alpha, w) \models \neg\psi$. Thus, $(M_\alpha, w_\alpha) \models \neg K_i\psi$ as desired.

We now construct an $\mathrm{S5}_n$-i-maximum situation $(M_\alpha^{max}, w_\alpha)$ for $K_i\alpha$ using amalgamation, much as in the proof of Proposition 6 and 15. Let $\mathcal{N}_\alpha$ be a set of situations including (M_α, w_α) such that each situation $(M, w) \in \mathcal{N}_\alpha$ satisfies $K_i\alpha$, and if $T \in Poss_i^{\mathrm{S5}_n}(M', w')$ for some situation (M', w') satisfying $K_i\alpha$, then $T \in Poss_i^{\mathrm{S5}_n}(M, w)$ for some situation $(M, w) \in \mathcal{N}_\alpha$. Again, we assume that if (M, w) and (M', w') are two situations in $\mathcal{N}_\alpha$, then the worlds in W^M and $W^{M'}$ are disjoint. We take the worlds in M_α^{max} to be an amalgamation of the situations in $\mathcal{N}_\alpha$, just as in Theorem 15. The only difference is that we now define $\mathcal{K}_i^{M_\alpha^{max}}$ to be the least equivalence relation that includes $\mathcal{K}_i^M$ for each situation $(M, w) \in \mathcal{N}_\alpha$ together with (w_α, w'), where $w' \in \mathcal{K}_i^M(w)$ for some $(M, w) \in \mathcal{N}_\alpha$.

It remains to show that $(M_\alpha^{max}, w_\alpha) \models K_i\alpha$. We first show that $(M_\alpha^{max}, w_\alpha) \models \xi_\alpha$. As in the case of (M_α, w_α), we actually show that for every subformula φ of α, we have that $(M_\alpha^{max}, w_\alpha) \models \xi_\varphi$. We do this by showing that for every subformula ψ of φ, we have $(M_\alpha^{max}, w_\alpha) \models K_i\psi$ iff $\psi \in D_i^Q(\alpha)$. Suppose that $\psi \in D_i^Q(\alpha)$. Thus, $\models_{\mathrm{S5}_n} K_i\alpha \Rightarrow K_iQ_i^{\xi_\psi}\psi$. It easily follows from Proposition 20 that $(M_\alpha^{max}, w) \models Q_i^{\xi_\psi}\psi$ for all $w \in \mathcal{K}_i^{M_\alpha^{max}}(w_\alpha)$. By the induction hypothesis, $(M_\alpha^{max}, w_\alpha) \models \xi_\psi$. Since ξ_ψ is i-subjective, $(M_\alpha^{max}, w) \models \xi_\psi$ for all $w \in \mathcal{K}_i^{M_\alpha^{max}}(w_\alpha)$. It follows that $(M_\alpha^{max}, w) \models \psi$ for all $w \in \mathcal{K}_i^{M_\alpha^{max}}(w_\alpha)$, and thus $(M_\alpha^{max}, w_\alpha) \models K_i\psi$. On the other hand, if $\psi \notin D_i^Q(\alpha)$, then by definition there is some situation (M, w) such that $(M, w) \models K_i\alpha \wedge \neg Q_i^{\xi_\psi}\psi$. By construction, there must be some world $w' \in \mathcal{K}_i^{M_\alpha^{max}}(w_\alpha)$ such that $T^{i,*}_{M_\alpha^{max},w'} = T^{i,*}_{M,w}$. By Proposition 20, we have $(M_\alpha^{max}, w') \models \neg Q_i^{\xi_\psi}\psi$. Using Lemmas 21 and 33 and the induction hypothesis just as we did for (M_α, w_α), we can now show $(M_\alpha^{max}, w') \models \neg\psi$, and so $(M_\alpha^{max}, w_\alpha) \models \neg K_i\psi$. Thus, $(M_\alpha^{max}, w_\alpha) \models \xi_\alpha$, as desired.

Since $T^{i,*}_{M_\alpha^{max},w_\alpha} = T^{i,*}_{M_\alpha,w_\alpha}$, and both $(M_\alpha^{max}, w_\alpha)$ and (M_α, w_α) satisfy ξ_α, it follows from Lemma 33 that $(M_\alpha^{max}, w_\alpha)$ and (M_α, w_α) agree on the truth value of $K_i\alpha$. Since $(M_\alpha, w_\alpha) \models K_i\alpha$, it follows that $(M_\alpha^{max}, w_\alpha) \models K_i\alpha$.

Theorem 25: *A formula α is $\mathrm{S5}_n$-i-honest iff there is an $\mathrm{S5}_n$-i-Q-stable set S^α containing α such that $\ker_i^Q(S^\alpha) \subseteq \ker_i^Q(S)$. Moreover, if α is $\mathrm{S5}_n$-i-honest, then, for* *all* $\beta \in \mathcal{L}_n$, $\alpha \mid\!\sim^i_{\mathrm{S5}_n} \beta$ *iff* $\beta \in S^\alpha$.

Proof The proof is almost identical to that of Theorem 16. Suppose α is $\mathrm{S5}_n$-i-honest. Then, by definition, there is an $\mathrm{S5}_n$-i-maximum situation for α, say (M_α, w_α). Let $S^\alpha = \{\varphi \in \mathcal{L}_n^Q : (M_\alpha, w_\alpha) \models K_i\varphi\}$. Clearly S^α is an $\mathrm{S5}_n$-i-Q-stable set containing α. Let S be any other $\mathrm{S5}_n$-i-Q-stable set containing α. Suppose $\varphi \notin \ker_i^Q(S)$. Let (M, w) be an $\mathrm{S5}_n$ situation corresponding to S; we must have $(M, w) \models K_i\alpha \wedge \neg K_i\varphi$. Thus, there is a world $w' \in W^M$ such that $(w, w') \in \mathcal{K}_i^M$ and $(M, w') \models \neg\varphi$. Since (M_α, w_α) is an $\mathrm{S5}_n$-i-maximum situation, $T^{i,*}_{M,w'} \in Poss_i^{\mathrm{S5}_n}(M_\alpha, w_\alpha)$. Thus, there is some world u such that $(w_\alpha, u) \in \mathcal{K}_i^M$ and $T^{i,*}_{M_\alpha,u} = T^{i,*}_{M,w'}$. By Proposition 20, we have $(M_\alpha, u) \models \neg\varphi$. Thus, $(M_\alpha, w_\alpha) \models \neg K_i\varphi$. This means that $\varphi \notin S^\alpha$. Since φ was chosen arbitrarily, it follows that $\ker_i^Q(S^\alpha) \subseteq \ker_i^Q(S)$. Moreover, from the definition of $\mid\!\sim^i_{\mathrm{S5}_n}$, it follows that $\alpha \mid\!\sim^i_{\mathrm{S5}_n} \beta$ iff $\beta \in S^\alpha$.

For the converse, suppose that there is an $\mathrm{S5}_n$-i-Q-stable set S^α containing α such that $\ker_i^Q(S^\alpha) \subseteq \ker_i^Q(S)$ for every other $\mathrm{S5}_n$-i-stable set containing α. We want to show that α is $\mathrm{S5}_n$-i-honest. Suppose that $\models_{\mathrm{S5}_n} K_i\alpha \Rightarrow (K_i\varphi_1 \vee \ldots \vee K_i\varphi_k)$, where $\varphi_1, \ldots, \varphi_k$ are Q_i-formulas. It follows that every $\mathrm{S5}_n$-i-Q-stable set containing α must contain one of $\varphi_1, \ldots, \varphi_k$. In particular, this is true of S^α. So we can suppose without loss of generality that $\varphi_1 \in S^\alpha$. Thus, $\varphi_1 \in \ker_i(S^\alpha)$. By definition of S^α, this means that φ_1 is in $\ker_i^Q(S)$ for every other $\mathrm{S5}_n$-i-stable set S containing α. Just as in Theorem 16, it now follows that $\models_{\mathrm{S5}_n} K_i\alpha \Rightarrow K_i\varphi_1$. By Theorem 24, α is $\mathrm{S5}_n$-i-honest.

Theorem 26: *The formula* α *is* $\mathrm{S5}_n$-i*-honest iff* $\alpha \in \mathcal{D}_1^i(\alpha)$. *If* α *is* $\mathrm{S5}_n$-i*-honest, then* $\alpha \mid\!\sim^i_{\mathrm{S5}_n} \beta$ *iff* $\beta \in D_Q^i(\alpha)$.

Proof Suppose α is $\mathrm{S5}_n$-i-honest. Then there is an $\mathrm{S5}_n$-i-maximum situation for α, say (M_α, w_α). The proof of Theorem 24 shows that $\psi \in D_Q^i(\alpha)$ iff $(M_\alpha, w_\alpha) \models K_i\psi$. Since $(M_\alpha, w_\alpha) \models K_i\alpha$, it immediately follows that $\alpha \in D_Q^i(\alpha)$. Notice that this argument also shows that if α is $\mathrm{S5}_n$-i-honest, then $\alpha \mid\!\sim_i \beta$ iff $\beta \in D_Q^i(\alpha)$.

Now suppose that $\alpha \in D_Q^i(\alpha)$ and $K_i\alpha$ is $\mathrm{S5}_n$-consistent. Let $(M_\alpha^{max}, w_\alpha)$ be the situation constructed in the proof of Theorem 24. (Since $K_i\alpha$ is $\mathrm{S5}_n$-consistent, we can carry out this construction.) The proof of that theorem shows that $(M_\alpha^{max}, w_\alpha) \models K_i\beta$ iff $\beta \in D_Q^i(\alpha)$. Since $\alpha \in D_Q^i(\alpha)$, we must have that $(M_\alpha^{max}, w) \models K_i\alpha$. Thus, $(M_\alpha^{max}, w_\alpha)$ is an $\mathrm{S5}_n$-i-maximum situation for α. It follows that α is $\mathrm{S5}_n$-i-honest, as desired.

D Proofs for Section 7

Theorem 27: *For* $\mathcal{S} \in \{\mathrm{T}_n, \mathrm{S4}_n : n \geq 1\} \cup \{\mathrm{K45}_n, \mathrm{KD45}_n, \mathrm{S5}_n : n \geq 2\}$*, the problem of computing whether* α *is* $\mathcal{S}$-i*-honest is PSPACE-complete.*

Proof For $\mathcal{S} \in \{\mathrm{T}_n, \mathrm{S4}_n : n \geq 1\}$, the upper bound is almost immediate from Corollary 30, together with the fact (proved in [12]) that the satisfiability problem for T_n and $\mathrm{S4}_n$ is in PSPACE: Let F' consist of all the subformulas of $K_i\alpha$ of the form $K_i\varphi$ such that $K_i\alpha \wedge \neg K_i\varphi$ is satisfiable. We then check if $K_i\alpha \bigwedge_{K_i\varphi\in F'} \neg K_i\varphi$ if satisfiable. If it is, then α is honest; otherwise α is not honest. These tests can all be done in PSPACE.

If $\mathcal{S} \in \{\mathrm{K45}_n, \mathrm{KD45}_n : n \geq 2\}$, then the upper bound follows using similar reasoning from Corollary 32 and the fact that the validity problem for $\mathrm{K45}_n$ and $\mathrm{KD45}_n$ is in PSPACE [12].[16]

Finally, for the case of $\mathrm{S5}_n$, we first remark that the techniques of [12] can be used to show that validity in $\mathrm{S5}_n$ situations for the language $\mathcal{L}^Q$ can be determined in PSPACE. Thus, the tests required to check whether $\alpha \in D^i_Q(\alpha)$ can all be carried out in PSPACE. The upper bound now follows from Theorem 26.

For the lower bound in the case that $\mathcal{S}$ is T_n or $\mathrm{S4}_n$, let β be an arbitrary formula and let q be a primitive proposition that does not appear in β. We claim that β is $\mathcal{S}$-valid iff $\alpha = K_1 q \vee K_1(q \Rightarrow \beta)$ is $\mathcal{S}$-1-honest. Clearly, if β is $\mathcal{S}$-valid, then α is equivalent to *true*, and so is honest. For the converse, suppose β is not $\mathcal{S}$-valid. Since $K_1\alpha \Rightarrow \alpha$ is $\mathcal{S}$-valid, by Theorem 6, it suffices to show that neither $K_1\alpha \Rightarrow K_1 q$ nor $K_1\alpha \Rightarrow K_1(q \Rightarrow \beta)$ is $\mathcal{S}$-valid. Since β is not $\mathcal{S}$-valid, there must be some $\mathcal{S}$ situation (M, w) such that $(M, w) \models \neg\beta$. Since q does not appear in β, there are $\mathcal{S}$-structures M_1 and M_2 that are identical to M except that in M_1, q is true at all the worlds, while in M_2, $\neg q$ is true at all worlds. Clearly $(M_1, w) \models K_1\alpha \wedge \neg K_1(q \Rightarrow \beta)$ while $(M_2, w) \models K_1\alpha \wedge \neg K_1 q$. This proves that α is not honest. Since deciding validity in T_n or $\mathrm{S4}_n$ is PSPACE-hard for $n \geq 1$ as long as there is at least one primitive proposition in Φ [9],[17] it follows that deciding honesty is PSPACE hard if there are at least two primitive propositions in the language (since we have assumed that q does not appear in β). We can improve this slightly. Using the techniques of [9], we can show that the PSPACE lower bound holds even if $|\Phi| = 1$. The idea is that, with only one primitive proposition and the modal operator K_1, we can write an infinite family of formulas that have all the properties that we really need of primitive propositions. Thus we can simulate the argument above using only one primitive proposition. We refer the reader to [9] for details.

For the lower bound in the case that $\mathcal{S}$ is $\mathrm{K45}_n$, $\mathrm{KD45}_n$, or $\mathrm{S5}_n$, let β be an arbitrary formula and let p be an arbitrary primitive proposition (which may appear in β). We claim that $\mathit{true}{\mid\!\sim}^1_{\mathcal{S}}\beta$ iff $\alpha = K_1 p \vee K_1\neg p \vee K_1\beta$ is $\mathcal{S}$-1-honest. If $\mathit{true}{\mid\!\sim}^1_{\mathcal{S}}\beta$, then $K_1\beta$ is true in an $\mathcal{S}$-1-maximum situation for *true*, say (M, w). Clearly we also have $(M, w) \models K_1\alpha$. Thus, there is an $\mathcal{S}$-1-maximum situation for $K_1\alpha$, namely (M, w), and hence α is $\mathcal{S}$-1-honest. Conversely, suppose that α is $\mathcal{S}$-1-honest. Then it has a maximum $\mathcal{S}$-1 situation, say (M, w). Consider any i-objective tree (or, in the case of $\mathrm{S5}_n$, i-objective $*$-tree) T. Either p or $\neg p$ must be true at the root of T. If it is p, then clearly T is in $\mathit{Poss}^{\mathcal{S}}_1(M', w')$ for some $\mathcal{S}$ situation (M', w') that satisfies $K_1 p$. Similarly, if the root satisfies $\neg p$, then T is in $\mathit{Poss}^{\mathcal{S}}_1(M', w')$ for some $\mathcal{S}$ situation (M', w') that satisfies $\neg p$. In either case, we must have $T \in \mathit{Poss}^{\mathcal{S}}_1(M, w)$. It follows that (M, w) is a maximum $\mathcal{S}$-1 situation for *true*. Moreover, we have

[16] Actually, $\mathrm{K45}_n$ is not considered in [12]. However, the proof that the validity problem for $\mathrm{K45}_n$ is PSPACE-complete is a trivial modification of that for $\mathrm{KD45}_n$.

[17] A PSPACE lower bound is also proved in [12], but that proof requires that Φ, the set of primitive propositions, be infinite. In [9], it is shown that the result holds even if there is one primitive proposition in the language.

$(M,w) \models \neg K_1 p \wedge \neg K_1 \neg p$. Since $(M,w) \models K_1\alpha$, we must have $(M,w) \models K_1\beta$, so $true \mid\!\sim^1_{\mathcal{S}} \beta$. This completes the proof of the claim.

Thus, for $\mathcal{S} \in \{\mathrm{K45}_n, \mathrm{KD45}_n, \mathrm{S5}_n\}$, we have shown that checking honesty is as hard as checking whether $true \mid\!\sim^1_{\mathcal{S}} \beta$. For the case of $\mathcal{S} \in \{\mathrm{K45}_n, \mathrm{KD45}_n : n \geq 2\}$, it is easy to see that if φ is a 1-objective formula, then $true \mid\!\sim^1_{\mathcal{S}} \varphi$ iff φ is valid. Thus, to get the PSPACE lower bound, it suffices to show that it is PSPACE hard to decide the validity of 1-objective formulas in the case of $\mathrm{K45}_n$ and $\mathrm{KD45}_n$. This follows from the lower bound proof given in [9]. There (just as in [12]) it is shown that we can effectively translate a QBF (*Quantified Boolean Formula*) A to a modal formula φ_A involving only the modal operators K_1 and K_2 so that so that A is true iff φ_A is $\mathcal{S}$-satisfiable, for $\mathcal{S} \in \{\mathrm{K45}_n, \mathrm{KD45}_n, \mathrm{S5}_n\}$. The PSPACE lower bound for satisfiability for $\mathrm{K45}_n$ and $\mathrm{KD45}_n$, $n \geq 2$, then follows from the PSPACE lower bound for QBF [23]. The formula φ_A is in fact 1-objective and can be assumed to mention only one primitive proposition. Thus, deciding satisfiability and validity for 1-objective formulas is PSPACE-hard. It follows that deciding whether a formula is $\mathcal{S}$-1-honest is PSPACE-hard, even if Φ consists of a single primitive proposition, for $\mathcal{S} \in \{\mathrm{K45}_n, \mathrm{KD45}_n\}$.

For $\mathrm{S5}_n$, in the case that $n \geq 3$, we can get a lower bound in a similar way: From Theorem 26, it follows that if φ is a formula that only involves the modal operators K_2 and K_3, then φ is $\mathrm{S5}_2$-valid iff $true \mid\!\sim^1_{\mathrm{S5}_n} \varphi$. The PSPACE lower bound now follows from the PSPACE lower bound for $\mathrm{S5}_2$. To get the PSPACE lower bound in the case that $n = 2$, we need to look even more closely at the PSPACE lower bound proof in [9]. The formulas φ_A that arise in the proof are easily seen to have the following property: either $\neg\varphi_A$ is valid (if the QBF A is false), and hence so is $K_2 K_1 \neg\varphi$, or φ_A is satisfiable and it is not the case that $true \mid\!\sim^1_{\mathrm{S5}_n} K_2 K_1 \neg\varphi_A$. Thus, A is false iff $true \mid\!\sim^1_{\mathrm{S5}_n} K_2 K_1 \neg\varphi_A$. This gives us the PSPACE lower bound for checking honesty in $\mathrm{S5}_2$.

Theorem 28: *For $\mathcal{S} \in \{\mathrm{K}_n, \mathrm{T}_n, \mathrm{S4}_n : n \geq 1\} \cup \{\mathrm{K45}_n, \mathrm{KD45}_n, \mathrm{S5}_n : n \geq 2\}$, if α is $\mathcal{S}$-i-honest, then the problem of deciding if $\alpha \mid\!\sim^i_{\mathcal{S}} \beta$ is PSPACE-complete.*

Proof For $\mathcal{S} \in \{\mathrm{K}_n, \mathrm{T}_n, \mathrm{S4}_n : n \geq 1\}$, the upper bound follows from Theorem 9 and the fact that checking validity for $\mathcal{S}$ is in PSPACE [12]. For the lower bound, observe that α is $\mathcal{S}$-valid iff $K_1\alpha$ is valid, and thus, by Theorem 9, α is valid iff $true \mid\!\sim^1_{\mathcal{S}} \alpha$. The result now follows from the PSPACE lower bound for checking validity for $\mathcal{S}$, which holds even if Φ consists of a single proposition [9, 12].

For $\mathcal{S} \in \{\mathrm{K45}_n, \mathrm{KD45}_n, \mathrm{S5}_n : n \geq 2\}$, the upper bound follows from Theorems 19 and 26. The lower bound follows from the observation made in the proof of Theorem 27 that proving whether $true \mid\!\sim^i_{\mathcal{S}} \beta$ is already PSPACE hard.

Theorem 29: *Suppose $\mathcal{S} \in \{KD45, K45, S5\}$. If Φ (the set of primitive propositions) is finite, then the problem of deciding whether α is $\mathcal{S}$-1-honest and the problem of deciding whether $\alpha \mid\!\sim^1_{\mathcal{S}} \beta$ for an $\mathcal{S}$-1-honest α are both decidable in polynomial time. If Φ is infinite, these problems are both $\Delta_2^{\mathrm{p},\log(n)}$-complete.*

Proof Since, as we have observed, the notions of honesty and $\mid\!\sim^i_{\mathcal{S}}$ coincide for K45, KD45, and $S5$ (except that *false* is K45-honest and not KD45- or S5-honest), it suffices to focus on KD45.

Clearly if Φ is finite, there are $2^{|\Phi|}$ truth assignment to the propositions in Φ. Thus, there are no more than $2^{2^{|\Phi|}}2^{|\Phi|}$ KD45 situations. To decide if α is KD45-1-honest, we must see if there is a maximum situation satisfying $K\alpha$. This can be done by exhaustively checking all situations, in time linear in $|\alpha|$, the length of α. (Of course, the constant here will be some multiple of $2^{2^{|\Phi|}}$, but this is independent of $|\alpha|$.) Similar arguments show that deciding if $\alpha\mid\!\sim^1_{KD45}\beta$ can be done in linear time.

If Φ is infinite, we must work harder. Before we go into details, we briefly review Gottlob's [6] results. Extend the language so that it includes a modal operator $\Box$. Roughly speaking, $\Box\varphi$ says that φ is valid. Formally,

- $(W,w) \models \Box\varphi$ if $(W',w') \models \varphi$ for all KD45 situations (W',w').

Thus, if φ does not contain any occurrences of $\Box$, then $(W,w) \models \Box\varphi$ iff φ is KD45-valid. (We could, of course, similarly extend K45 or S5.) If we do not allow any occurrences of K (so that the only modal operator is $\Box$), we get what Gottlob called *Carnap's logic*. Gottlob showed that the validity problem for Carnap's logic is $\Delta_2^{\mathrm{p},\log(n)}$-complete. Gottlob's argument in fact shows that if we start with any base logic whose satisfiability problem is NP-complete and extend it with a $\Box$ operator that denotes validity as above, then the validity problem of the resulting logic $\Delta_2^{\mathrm{p},\log(n)}$-complete. Since the satisfiability problem for KD45 is also NP-complete, Gottlob's argument shows that the validity problem for the full logic with both K and $\Box$ operators is $\Delta_2^{\mathrm{p},\log(n)}$-complete.

For the lower bound, given a formula φ in Carnap's logic, let φ^* be the result of replacing all $\Box$ operators in φ by K_1. Let W^* consist of all truth assignments and let $w \in W^*$. An easy induction on the structure of φ that $w \models \varphi$ iff $(W^*,w) \models \varphi$. It follows that $true\mid\!\sim^1_{KD45}\varphi^*$ iff φ is valid. This shows that the problem of deciding if $\alpha\mid\!\sim^1_{KD45}\beta$ is $\Delta_2^{\mathrm{p},\log(n)}$ hard.

Let q be a primitive proposition that does not appear in φ. It is easy to see that φ is valid in Carnap's logic iff $\alpha = K_1\varphi^* \vee K_1 q \vee K_1\neg q$ is KD45-1-honest. For if φ is valid in Carnap's logic, then there is an KD45-1-maximum situation satisfying $K_1\alpha$ (namely, the situation (W^*,w), where W^* consisting of all truth assignments and $w \in W^*$). Conversely, suppose φ is not valid in Carnap's structure. Clearly every truth assignment satisfying q is in some situation satisfying $K_1 q$ (and thus $K_1\alpha$); similarly every truth assignment satisfying $\neg q$ is in some situation satisfying $K_1\neg q$. Thus, if $K_1\alpha$ were KD45-1-honest, then the only KD45-1-maximum structure satisfying it would be W^*. But, clearly $W^* \models \neg K_1 q \wedge \neg K_1\neg q$. Since φ is not valid in Carnap's logic, it also follows from our earlier argument that there is some truth assignment $w \in W^*$ such that $(W^*,w) \models \neg\varphi^*$. Hence, $(W^*,w) \models \neg K_1\varphi^*$. It follows that $(W^*,w) \models \neg K_1\alpha$, so α is not KD45-1-honest. The $\Delta_2^{\mathrm{p},\log(n)}$ lower bound on checking whether a formula α is KD45-1-honest now follows from Gottlob's results.

For the upper bounds, we use the ideas of Definition 17 and Theorem 19. We now define an *extended objective formula* to be a Boolean combination of propositional formulas and formulas of the form $\Box\varphi$. A *top-level subformula* of a formula β is a subformula of the form $K_i\varphi$ such that φ is an extended objective formula. Given formulas α and β, we construct a finite sequence $\langle \beta'_0, \beta'_1, \ldots, \beta'_m \rangle$ of formulas and a finite sequence $\langle B'_1, \ldots, B'_m \rangle$ of sets of formulas of the form $K\varphi$ where φ is an extended objective formula, as follows: We take β'_0 to be $K_i\beta$. Suppose we have defined $\beta'_0, \ldots, \beta'_k$ and $B'_1, \ldots, B'_k$ so that B'_h consists of all the top-level subformulas of β'_{h-1}. If β'_k is not of the form $\Box\varphi$, then we define B'_{k+1} to consist of all the top-level extended subformulas of β'_k and define β'_{k+1} to be the result of replacing each subformula $K_i\varphi$ of β'_k that is in B_{k+1} by $\Box(\alpha \Rightarrow \varphi)$. The construction ends if β'_m is of the form $\Box\varphi$. Recall that in Definition 17, we defined a sequence $A^1_{KD45}(\alpha, \beta) = \langle \beta_0, \ldots, \beta_m \rangle$. It is straightforward to check that our construction guarantees that $\models_{KD45} \beta_j \equiv \beta'_j$. (The proof is by induction on j.) Suppose $A^1_{KD45}(\alpha, \alpha) = \langle \alpha_0, \ldots, \alpha_m \rangle$, and the corresponding sequence using $\Box$ is $\langle \alpha'_0, \ldots, \alpha'_m \rangle$. By Theorem 19, α is KD45-1-honest iff $\alpha_m = true$, and by the observation above, this is true iff α'_m is valid. In addition, we have $\alpha {\mid\!\sim}^1_{KD45} \beta$ iff $\beta_m = true$, which holds iff β'_m is valid. Thus, we have reduced both the question of checking for honesty and checking nonmonotonic consequence to deciding validity in the extended logic. This proves the $\Delta_2^{\mathrm{p},\log(n)}$ upper bound.

Acknowledgments

I would like to thank Ron Fagin, Gerhard Lakemeyer, Grisha Schwarz, and Moshe Vardi for their helpful comments on an earlier draft of this paper, and the two anonymous referees for their careful reading and very constructive comments.

References

[1] C. Boutilier. Conditional logics of normality: a modal approach. *Artificial Intelligence*, 68:87–154, 1994.

[2] B. F. Chellas. *Modal Logic*. Cambridge University Press, Cambridge, U.K., 1980.

[3] R. Fagin and J. Y. Halpern. Belief, awareness, and limited reasoning. *Artificial Intelligence*, 34:39–76, 1988.

[4] R. Fagin, J. Y. Halpern, and M. Y. Vardi. A model-theoretic analysis of knowledge. *Journal of the ACM*, 91(2):382–428, 1991. A preliminary version appeared in *Proc. 25th IEEE Symposium on Foundations of Computer Science*, 1984.

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

[6] G. Gottlob. NP trees and Carnap's modal logic. *Journal of the ACM*, 42(2):421–457, 1995.

[7] J. Y. Halpern. A critical reexamination of default logic, autoepistemic logic, and only knowing. In *Proceedings, 3rd Kurt Gödel Colloquium*, pages 43–60. Springer-Verlag, 1993.

[8] J. Y. Halpern. Reasoning about only knowing with many agents. In *Proc. National Conference on Artificial Intelligence (AAAI '93)*, pages 655–661, 1993.

[9] J. Y. Halpern. The effect of bounding the number of primitive propositions and the depth of nesting on the complexity of modal logic. *Artificial Intelligence*, 75(2):361–372, 1995.

[10] J. Y. Halpern and G. Lakemeyer. Multi-agent only knowing. In Y. Shoham, editor, *Theoretical Aspects of Rationality and Knowledge: Proc. Sixth Conference*. Morgan Kaufmann, San Francisco, Calif., 1996.

[11] J. Y. Halpern and Y. Moses. Towards a theory of knowledge and ignorance. In *Proc. AAAI Workshop on Non-monotonic Logic*, pages 125–143, 1984. Reprinted in K. R. Apt (Ed.), *Logics and Models of Concurrent Systems*, Springer-Verlag, Berlin/New York, pp. 459–476, 1985.

[12] J. Y. Halpern and Y. Moses. A guide to completeness and complexity for modal logics of knowledge and belief. *Artificial Intelligence*, 54:319–379, 1992.

[13] G. E. Hughes and M. J. Cresswell. *An Introduction to Modal Logic*. Methuen, London, 1968.

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

[15] J. Jaspars. A generalization of stability and its applications to circumscription of positive introspective knowledge. In *Proceedings of the Ninth Workshop on Computer Science Logic (CSL '90)*, pages 289–299. Springer-Verlag, Berlin, 1991.

[16] R. E. Ladner. The computational complexity of provability in systems of modal propositional logic. *SIAM Journal on Computing*, 6(3):467–480, 1977.

[17] G. Lakemeyer. All they know: a study in multi-agent autoepestemic reasoning. In *Proc. Thirteenth International Joint Conference on Artificial Intelligence (IJCAI '93)*, pages 376–381, 1993.

[18] E. J. Lemmon. *The "Lemmon Notes": An Introduction to Modal Logic*. Basil Blackwell, Oxford, U.K., 1977. Written in collaboration with Dana Scott; edited by Krister Segerberg. American Philosophical Quarterly Monograph Series. Monograph No. 11.

[19] H. J. Levesque. All I know: a study in autoepistemic logic. *Artificial Intelligence*, 42(3):263–309, 1990.

[20] R. C. Moore. Semantical considerations on nonmonotonic logic. *Artificial Intelligence*, 25:75–94, 1985.

[21] R. Parikh. Monotonic and nonmonotonic logics of knowledge. *Fundamenta Informaticae*, 15(3,4):255–274, 1991.

[22] R. Stalnaker. A note on nonmonotonic modal logic. *Artificial Intelligence*, 64:183–196, 1993. This is a slightly revised version of a note that was originally circulated in 1980.

[23] L. J. Stockmeyer. The polynomial-time hierarchy. *Theoretical Computer Science*, 3:1–22, 1977.

[24] M. Y. Vardi. A model-theoretic analysis of monotonic knowledge. In *Proc. Ninth International Joint Conference on Artificial Intelligence (IJCAI '85)*, pages 509–512, 1985.