On the Complexity of Epistemic Reasoning

Moshe Y. Vardi

IBM Almaden Research

Abstract

A fundamental problem with epistemic logics based on Hintikka's possible-worlds semantics is the logical omniscience problem. A partial solution to the problem is to use Montague and Scott's approach to modal logic; in that approach, agent's knowledge is characterized not by a set of possible worlds but rather by a set of propositions known by the agent. By imposing various closure conditions on the set of propositions known by the agent, we can model agents with different reasoning powers.

In this paper we study the complexity of the decision problem for epistemic logics based on Montague and Scott's semantics. We are mainly interested in finding out how assumptions about the agents' reasoning power affects the complexity of reasoning about the agents' knowledge. We study a spectrum of assumptions and show that the complexity of the logic under different assumptions is always in NP or PSPACE. Furthermore, we seem to pinpoint the "mental faculty" that raises the complexity of the logic from NP to PSPACE; it is the ability to combine distinct items of knowledge, namely, to infer from the facts p and q the fact $p \wedge q$.

1 Introduction

Epistemic logic, the logic of epistemic notions such as *knowledge* and *belief*, is a major area of research in artificial intelligence (cf. [HM85,

MH69, Moo85, XW83]). The main existing formal model for epistemic notions, originated by Hintikka [Hin62], is based on the possible world approach. The basic notions of this approach are a set W of *worlds* and relations of *possibility* (also called *accessibility*) between them. Knowledge and belief of an agent who is situated in an actual world w is characterized by a subset W_w of W as the set of worlds that are epistemic alternatives to w; the agent knows or believes a fact φ if φ holds in all the worlds in W_w.

A fundamental problem with this model is the so called *logical omniscience* problem [Hin75]. Essentially, the problem is that an agent always knows all the consequences of her knowledge. Formally, the axiom $Kp \wedge K(p \rightarrow q) \rightarrow Kq$, is fundamental to Hintikka's semantics. This situation is, of course, unintuitive and unrealistic. Dealing with the logical omniscience problem is a major issue in epistemic logic.

In [Var86] we argued that the source of the logical omniscience problem is Hintikka's interpretation of knowledge in terms of possibility relations. We suggested an alternative formal model, based on a model proposed, in a somewhat different context, by Montague [Mon68, Mon70] and Scott [Sco70]. In this model we have *epistemic structures*[1] that have multiple worlds but there is no notion of possibility among worlds. Agent's knowledge is characterized not by a set of possible

---

[1] Epistemic structures are called *belief* structures in [Var86]

worlds but rather by a set of propositions known by the agent. By imposing various closure conditions on the set of propositions known by the agent, we can model agents with different reasoning powers. In fact, by imposing enough conditions we can model even logically omniscient agents. We should note, however, that this approach does not completely solve the logical omniscience problem. While the axiom $Kp \wedge K(p \rightarrow q) \rightarrow Kq$, is fundamental to Hintikka's semantics, the (weaker) inference rule $p \equiv q \vdash Kq \equiv Kq$ is fundamental to Montague's and Scott's semantics.

In this paper we study the complexity of the decision problem for epistemic structures. We are mainly interested in finding out how assumptions about the agents' reasoning power affects the complexity of reasoning about the agents' knowledge. For example, we consider agents that are able to draw inferences from their knowledge (characterized by the axiom $K(p \wedge q) \rightarrow Kp$) and agents that are introspective of their knowledge (characterized by the axiom $Kp \rightarrow KKp$). While it is known that for certain assumptions the complexity of epistemic logic is PSPACE-complete [Lad77, HM85], the best known complexity under other assumptions is nondeterministic doubly exponential time (this follows from results in [Seg71]). We study a spectrum of assumptions and show that the complexity of the logic under different assumptions is always in NP or PSPACE. Furthermore, we seem to pinpoint the "mental faculty" that raises the complexity of the logic from NP to PSPACE; it is the ability to combine distinct items of knowledge, namely, to infer from the facts p and q the fact $p \wedge q$ (this is characterized by the axiom $Kp \wedge Kq \rightarrow K(p \wedge q)$).

Our main technical tool is the ubiquitous *tableau*, developed for classical logic by Beth [Bet59] and Smullyan [Smu68] and extended to modal logic by Kripke [Kri63]. Unfortunately, the modal tableau technique is quite messy (as Kripke himself acknowledged), because it combines trees that deal with the nondeterminism of propositional disjunction (i.e., if the formula $p \vee q$ is true, then either p is true or q is true, but we do not know which one is true) with trees that deal with the multiplicity of worlds in a modal setting. By combining the tool of *valuations* [Hin57, KP81, Lop78] and the idea of the tableau as an *update process* [Fit83], we manage

to eliminate trees altogether. We believe that our approach to tableaux is of interest for its own sake.

2 Epistemic Structures

The most straightforward way to model epistemic notions is by means of *epistemic sets*. Assume that we have an assertion language L. Then an epistemic set for an agent is simply a set of sentences of L; these are the sentences that the agent knows or believes. This approach is purely syntactical. To obtain a semantical analogue, we assume a set W of *worlds* and a relation $\models$ of *satisfaction* between worlds in W and sentences in L (i.e., $w \models \varphi$ if w satisfies φ). The *intension* of a sentence φ is the set of worlds in which φ is satisfied, i.e, $\{w \mid w \models \varphi\}$. Semantically, a sentence expresses the *proposition* that is its intension, where a proposition is simply a set of worlds. An epistemic set for an agent is now a set of propositions; these are the propositions that the agent knows or believes.

We formalize this approach. We assume a set $\mathcal{P}$ of atomic propositions and a set $\mathcal{A}$ of agents. An *epistemic frame* is a pair $F = (W, N)$, where W is a nonempty set, which we take to be the set of *worlds*, and $N : \mathcal{A} \times W \rightarrow 2^{2^W}$ is an *epistemic assignment*, which assigns to every agent in a world a set of propositions, i.e., an epistemic set. An *epistemic structure* is a triple $M = (W, N, I)$, where (W, I) is an epistemic frame and $I : \mathcal{P} \rightarrow 2^W$ is an *intension assignment*, which gives the intensions of atomic propositions.[2]

The language L is the smallest set that contains $\mathcal{P}$, is closed under propositional connectives, and contains $E_a\varphi$ (we use E as a generic epistemic modality, without having to commit to knowledge or belief) if φ is in L and a is in $\mathcal{A}$. The *epistemic depth* of a formula ψ, denoted $depth(\psi)$, is the level of nesting of epistemic modalities in ψ. The *epistemic size* of a formula ψ, denoted $size(\psi)$ is the number of epistemic modalities in ψ. Note that the epistemic size of ψ is greater or equal to the depth of ψ.

---

[2]If there is only one agent then N can be taken as a function $N : W \rightarrow 2^{2^W}$. Such structures are called in the literature *neighbourhood* structures (the sets in $N(u)$, for $u \in W$, are called *neighbourhoods*) [Seg71] or *minimal* structures [Che80].

We can now define what it means for worlds to satisfy formulas of L.

- $M, w \models p$, where $p \in \mathcal{P}$, if $w \in I(p)$.
- $M, w \models \neg\varphi$ if $M, w \not\models \varphi$.
- $M, w \models \varphi \lor \psi$ if $M, w \models \varphi$ or $M, w \models \psi$.
- $M, w \models E_a\varphi$ if $\{u \mid M, u \models \varphi\} \in N(a, w)$.

We can extend the intension assignment I to arbitrary formulas by defining $I(\varphi)$ to be the set $\{u \mid M, u \models \varphi\}$. The crucial clause in the definition of satisfaction is the clause for $E_a\varphi$. It can be read as saying that a believes φ in w if $I(\varphi)$ is in the epistemic set of a in w.[3]

It is important to understand that epistemic structure can be viewed as a generalization of Kripke structures. A Kripke structure M is of the form (W, R, I), where W is a set of worlds, I is an intension assignment for atomic propositions, and $R : \mathcal{A} \to 2^{W^2}$ assigns a possibility relation among worlds to each agent. The Kripke semantics is $M, w \models E_a\varphi$ if $M, u \models \varphi$ for every u such that $(w, u) \in R(a)$. Define $R(a, w)$ to be the set $\{u \mid (w, u) \in R(a)\}$, i.e., the set of "$a$-neighbours" of w. We can assign epistemic sets to agents and worlds by $N(a, w) = \{U \mid R(a, w) \subseteq U\}$. It is easy to see that w satisfies φ according to our epistemic semantics precisely when w satisfies φ according to Kripke semantics.

Does our model solve the logical omniscience problem? It does and it does not. It does, because it is no longer the case that if a believes φ and φ implies ψ then a believes ψ. In particular, an agent does not have to believe in all valid sentences and she can believe in contradictory propositions. Nevertheless, if a believes φ, and φ is equivalent to ψ, then a believes ψ. This, however, is unavoidable as long as one wishes to see epistemic modalities as operators on intensions. We refer the reader to [Cre85, Tho80] where different approaches to propositional attitudes (which include epistemic attitudes) are described.

The agents in our model have very little reasoning power. This is clearly demonstrated by the

---

[3] We remark that in [Var86] we criticized the above approach because it takes the notion of a world as a primitive notion. We described there an alternative constructive approach. This has no bearing on our investigation here, since the two approaches are equivalent with respect to the decision problem.

fact that the formulas $E_a\varphi \land E_a\psi$ and $E_a(\varphi \land \psi)$ are incomparable, i.e., neither of them implies the other. In fact, it is instructive to characterize all the *valid* sentences (φ is valid if it satisfied by any world w in any epistemic structure M).

Proposition 2.1: *The following formal system is sound and complete for validity in epistemic structures:*

E0 *All propositional tautologies.*
R0 *From φ and $\varphi \supset \psi$ infer ψ.*
R2 *From $\varphi \equiv \psi$ infer $E_a\varphi \equiv E_a\psi$.*

Proposition 2.1 was proven for the single agent case by Segerberg [Seg71] (Segerberg called that logic the *weakest classical modal logic*, denoted E). The proof used the *canonical model* technique (cf. [Che80]).

By Proposition 2.1, validity is fully characterized by propositional reasoning plus substitutivity of equivalents. We often would like, however, to consider agents with somewhat more developed "mental faculties". We now list several modes of reasoning that we may wish to endow our agents with:

E1 $\neg E_a$false
E2 $E_a$true
E3 $E_a(p \land q) \supset E_a q$
E4 $E_a p \land E_a q \supset E_a(p \land q)$
E5 $E_a p \supset E_a E_a p$
E6 $\neg E_a p \supset E_a \neg E_a p$.
E7 $E_a p \supset p$

E1-E4 express aspects of logical omniscience. E1 says that agents do not believe contradictory propositions, while E2 says that agents believe self-evident truth. E3 says that agents can make inferences, while E4 says that agents can combine facts. E5-E6 express *introspection*. E5 says that agents are aware of their beliefs and E6 says that agents are aware of their ignorance. E7 does not really reflect reasoning; rather, it is the axiom of knowledge - it says that if a knows φ then φ is true (this distinguishes knowledge from belief). Note that E1-E7 are not necessarily independent. For example, it is easy to see that E7 entails E1.

It is easy to express the above reasoning modes by closure conditions on epistemic sets $N(a, w)$ in an epistemic frame $F = (W, N)$:

C1 $\emptyset \notin N(a, w)$

C2 $W \in N(a, w)$

C3 If $U \in N(a, w)$ and $U \subseteq V$, then $V \in N(a, w)$.

C4 If $U \in N(a, w)$ and $V \in N(a, w)$, then $U \cap V \in N(a, w)$.

C5 If $U \in N(a, w)$ then $\{u \mid U \in N(a, u)\} \in N(a, w)$.

C6 If $U \notin N(a, w)$ then $\{u \mid U \notin N(a, u)\} \in N(a, w)$.

C7 If $U \in N(a, w)$, then $w \in U$.

The conditions C1-C7 captures E1-E7 is a precise sense. An epistemic frame $F = (W, I)$ is said to *validate* a formula φ if φ is satisfied in all worlds of all structures (W, N, I) (i.e, all structures obtained by adding an intension assignment to the frame F).

Proposition 2.2: *For $1 \leq i \leq 7$, an epistemic structure F validates Ei if and only if F satisfies Ci.*

Let $\mathcal{E}$ be the class of all epistemic structures. If S is a subset of $\{1, \ldots, 7\}$, then $\mathcal{E}_S$ is the class of all epistemic structures that satisfy Cj for all $j \in S$. For example $\mathcal{E}_{\{1,2\}}$ is the class of all epistemic structures that satisfy C1 and C2. This gives rise to 128 (not necessarily distinct) classes of structures. It is easy to see that epistemic structures that correspond to Kripke structures are in $\mathcal{E}_{\{2,3,4\}}$. It is not hard to show that every epistemic structure in $\mathcal{E}_{\{2,3,4\}}$ correspond to a Kripke structure. Thus, the logic of $\mathcal{E}_{\{2,3,4\}}$ is the multi-agent extension of the modal logic K (cf. [Che80]). Similarly, it can be shown that the logic in $\mathcal{E}_{\{1,2,3\}}$ is the logic of local reasoning in [FH88] (without, however, the modalities for *implicit belief*; see [Var86]),[4] and the logic of $\mathcal{E}_{\{3\}}$ is the multi-agent extension of the modal logic U (cf. [Fit83]).

A formula φ is $\mathcal{E}_S$-satisfiable if it is satisfied by a world in a structure in $\mathcal{E}_S$. We are interested in studying the complexity of the satisfiability problem for these classes $\mathcal{E}_S$ of epistemic structures.

---

[4] The logic of local reasoning is interpreted over epistemic structures. An agent a is said to *explicitly* believe φ if there is a set $U \in N(a, w)$ such that $M, u \models \varphi$ for all $u \in U$. An agent a is said to *implicitly* believe φ if $M, u \models \varphi$ for all $u \in \bigcap_{U \in N(a,w)} U$.

In particular, we are interested in the relationship between S and the complexity of satisfiability problem for $\mathcal{E}_S$.

Theorem 2.3: *If S is a subset of $\{1, \ldots, 7\}$, then the satisfiability problem for $\mathcal{E}_S$ is in PSPACE. If S is a subset of $\{1, \ldots, 7\}$ and $4 \notin S$, then the satisfiability problem for $\mathcal{E}_S$ is in NP.*

The proof of this theorem is by analyzing the 128 (not necessarily distinct) cases. Of course, we analyze only a few cases here. The focus is on upper bounds. We discuss lower bounds in our concluding remarks.

3 The Satisfiability Problem

3.1 $\mathcal{E}$, $\mathcal{E}_{\{1\}}$, $\mathcal{E}_{\{2\}}$, $\mathcal{E}_{\{3\}}$, and $\mathcal{E}_{\{5\}}$

Segerberg [Seg71] proved the *finite-model property* for the logic E (i.e., the class $\mathcal{E}$ with one agent). More precisely, he showed that if a formula φ is satisfiable, then it is satisfiable in a structure where the number of worlds is at most exponential in the length of the formula. Since the size of an epistemic set can be exponential in the number of worlds, Segerberg's finite-model property yields a nondeterministic doubly exponential-time decision procedure. We obtain here a nondeterministic polynomial-time decision procedure.

We start by stating an obvious lemma.

Lemma 3.1: *Let $M = (W, N, I) \in \mathcal{E}$, let φ be a formula, let $w \in W$, and let $a \in \mathcal{A}$. Then $M, w \not\models E_a \varphi$ if and only iff $U \neq I(\varphi)$ for all $U \in N(a, w)$.*

If φ is a formula, then $sub(\varphi)$ is the set of all subformulas of φ and their negations (we identify the formula $\neg\neg\psi$ with ψ). A *valuation* for φ is a function $\nu : sub(\varphi) \rightarrow \{0, 1\}$ that satisfies the following conditions:[5]

1. $\nu(\psi) = 1$ iff $\nu(\neg\psi) = 0$,

2. $\nu(\psi_1 \vee \psi_2) = 1$ iff $\nu(\psi_1) = 1$ or $\nu(\psi_2) = 1$, and

3. $\nu(\varphi) = 1$.

---

[5] Our valuations are called *model sets* in [Hin57], *atoms* in [KP81] and *semi-valuations* in [Lop78].

Proposition 3.2: *A formula φ is $\mathcal{E}$-satisfiable if and only if there exists a valuation ν for φ such that if $E_a\psi_1$ and $E_a\psi_2$ are in $sub(\varphi)$, $\nu(E_a\psi_1) = 1$, and $\nu(E_a\psi_2) = 0$, then $(\psi_1 \wedge \neg\psi_2) \vee (\neg\psi_1 \wedge \psi_2)$ is $\mathcal{E}$-satisfiable.*

Proof: Suppose first that φ is satisfied by a world w of a structure $M = (W, N, I)$, i.e., $M, w \models \varphi$. Define a valuation ν for φ by: $\nu(\psi) = 1$ if $M, w \models \psi$ and $\nu(\psi) = 0$ if $M, w \not\models \psi$. It is easy to check that ν is indeed a valuation for φ. Suppose now that $E_a\psi_1$ and $E_a\psi_2$ are in $sub(\varphi)$, $\nu(E_a\psi_1) = 1$, and $\nu(E_a\psi_2) = 0$. Then $M, w \models E_a\psi_1$ and $M, w \not\models E_a\psi_2$. Thus, by definition, $I(\psi_1) \in N(a, w)$ and $I(\psi_2) \notin N(a, w)$. By Lemma 3.1, it follows that $I(\psi_1) \neq I(\psi_2)$. Thus, there is a world u such that either $u \in I(\psi_1) - I(\psi_2)$ or $u \in I(\psi_2) - I(\psi_1)$. In the first case we have $M, u \models \psi_1 \wedge \neg\psi_2$ and in the second case we have $M, u \models \neg\psi_1 \wedge \psi_2$. Thus, $M, u \models (\psi_1 \wedge \neg\psi_2) \vee (\neg\psi_1 \wedge \psi_2)$.

Suppose now that the conditions of the proposition holds. That is, there exists a valuation ν for φ such that

> if $E_a\psi_1$ and $E_a\psi_2$ are in $sub(\varphi)$, $\nu(E_a\psi_1) = 1$, and $\nu(E_a\psi_2) \doteq 0$, then there exists a structure $M_{\psi_1,\psi_2} = (W_{\psi_1,\psi_2}, N_{\psi_1,\psi_2}, I_{\psi_1,\psi_2})$ and a world $w_{\psi_1,\psi_2} \in W_{\psi_1,\psi_2}$ such that $M_{\psi_1,\psi_2}, w_{\psi_1,\psi_2} \models (\psi_1 \wedge \neg\psi_2) \vee (\neg\psi_1 \wedge \psi_2)$.

Let $M_1, \ldots, M_n$ be an enumeration of the structures M_{ψ_1,ψ_2} from clause (2) above, i.e., we take one structure for each pair ψ_1, ψ_2, where $M_i = (W_i, N_i, I_i)$, and let $w_1, \ldots, w_n$ be an enumeration of the worlds w_{ψ_1,ψ_2} from clause (2) above, where $w_i \in W_i$. We can assume without loss of generality that $W_i \cap W_j = \emptyset$ for $i \neq j$.

We define a structure $M = (W, N, I)$ as follows. Intuitively, we construct M by taking a union of the M_i's with the addition of a new world w that will satisfy φ. The nontrivial part in the construction is defining the epistemic assignment. We do it in such a way that the truth values of formulas in $sub(\varphi)$ are preserved (though it may change for other formulas). The set W of worlds is the set $\bigcup_{j=1}^{n} W_j \cup \{w\}$, where w is a new world. Also define $I_0 : sub(\varphi) \to 2^{\{w\}}$ by

$$I_0(\psi) = \begin{cases} \{w\} & \text{if } \nu(\psi) = 1 \\ \emptyset & \text{otherwise} \end{cases}$$

Before defining I and N, we define an intension assignment $J : sub(\varphi) \to 2^W$ by $J(\psi) = \bigcup_{j=0}^{n} I_j(\psi)$. Note that by construction we have that $J(\neg\psi) = W - J(\psi)$ and $J(\psi_1 \vee \psi_2) = J(\psi_1) \cup J(\psi_2)$. We now define I as the projection of J on $\mathcal{P}$ (we can assume without loss of generality that $\mathcal{P} \subseteq sub(\varphi)$), that is, $I(p) = J(p)$ for $p \in \mathcal{P}$. It remains to define N.

For $u \in W_j$ we put $U \subseteq W$ in $N(a, u)$ precisely when $M_j, u \models E_a\psi_U$ and $U = J(\psi_U)$ for some $E_a\psi_U \in sub(\varphi)$. We claim that if $V \in N(a, u)$ and $V = J(\psi)$ for some $E_a\psi \in sub(\varphi)$, then $M_j, u \models E_a\psi$. Indeed, since $V = J(\psi) \in N(a, u)$, we must have that $M_j, u \models E_a\psi_V$ and $V = J(\psi_V)$ for some $E_a\psi_V \in sub(\varphi)$. But since $J(\psi) = J(\psi_V)$, we also have $I_j(\psi) = I_j(\psi_V)$, so $M_j, u \models E_a\psi$ iff $M_j, u \models E_a\psi_U$. It follows that $M_j, u \models E_a\psi$.

Also, we put $U \subseteq W$ in $N(a, w)$ precisely when $\nu(E_a\psi_U) = 1$ and $U = J(\psi_U)$ for some $E_a\psi_U \in sub(\varphi)$. We claim that if $V \in N(a, w)$ and $V = J(\psi)$ for some $E_a\psi \in sub(\varphi)$, then $\nu(E_a\psi) = 1$. Indeed, since $V = J(\psi) \in N(a, u)$, we must have that $\nu(E_a\psi_V) = 1$ and $V = J(\psi_V)$ for some $E_a\psi_V \in sub(\varphi)$. Suppose now that $\nu(E_a\psi) = 0$. Then, by assumption, there exists a structure $M_j = (W_j, N_j, I_j)$ and a world $w_j \in W_j$ such that $M_j, w_j \models (\psi_V \wedge \neg\psi) \vee (\neg\psi_V \wedge \psi)$. It follows that $I_j(\psi_V) \neq I_j(\psi)$. Consequently $J(\psi_V) \neq J(\psi)$, which is a contradiction.

We now show by induction on the structure of formulas that I and J agree on $sub(\varphi)$. This holds by construction for atomic propositions. It is easy to deal with propositional connectives, since we know that $J(\neg\psi) = W - J(\psi)$ and $J(\psi_1 \vee \psi_2) = J(\psi_1) \cup J(\psi_2)$ and similarly for I. Assume inductively that $I(\psi) = J(\psi)$. Suppose first that $u \in J(E_a\psi)$. Then either $u = w$ and $\nu(E_a\psi) = 1$ or $u \in W_j$ and $M_j, u \models E_a\psi$. In either case we have that $J(\psi) \in N(a, u)$. Since $I(\psi) = J(\psi)$, it follows that $M, u \models E_a\psi$, i.e., $u \in I(E_a\psi)$. Suppose now that $u \in I(E_a\psi)$, i.e, $M, u \models E_a\psi$, or, equivalently, $I(\psi) \in N(a, u)$. Since $I(\psi) = J(\psi)$ it follows that either $u = w$ and $\nu(E_a\psi) = 1$ or $u \in W_j$ and $M_j, u \models E_a\psi$. In either case we have that $u \in J(E_a\psi)$.

Since $\nu(\varphi) = 1$, we have that $w \in J(\varphi)$, and consequently $w \in I(\varphi)$. That is, $M, w \models \varphi$. ∎

It is easy to see from Proposition 3.2 that satis-

fiability for $\mathcal{E}$ can be decided in polynomial space. Consider the following algorithm:

1. Nondeterministically guess a valuation ν for φ.

2. For all $E_a\psi_1$ and $E_a\psi_2$ in $sub(\varphi)$, such that $\nu(E_a\psi_1) = 1$ and $\nu(E_a\psi_2) = 0$, nondeterministically and recursively check that either $(\psi_1 \wedge \neg\psi_2)$ or $\neg\psi_1 \wedge \psi_2$ is satisfiable.

It is easy to see that this is an alternating polynomial-time algorithm, since the depth of formulas the algorithm deals with decreases at each level of the recursion. Thus, we can simulate this algorithm by a deterministic polynomial space algorithm [CKS81]; intuitively, the deterministic algorithm conducts a depth-first search for an accepting computation tree of the alternating algorithm.

To appreciate the simplicity of our approach, the reader should compare it with the standard modal tableau approach, as in [HM85, Lad77, Kri63]. In the standard approach, given a formula φ one constructs a tree to find all valuations for φ. Here we just quantify existentially over all valuations for φ. Also, instead of using trees to deal with multiplicity of worlds, Proposition 3.2 just gives a recursive condition in the spirit of Fitting's update rules [Fit83]. The result is a "tree-less" tableau.

It turns out, however, that the use here of an alternating polynomial-time algorithm is an overkill.

Theorem 3.3: *The satisfiability problem for $\mathcal{E}$ is in NP.*

Proof: To determine satisfiability of φ, we use Proposition 3.2 and the fact there are only quadratically many formulas of the form $\psi_1 \wedge \neg\psi_2$, for formulas ψ_1 and ψ_2 in $sub(\varphi)$. Thus, we can use dynamic programming techniques to determine satisfiability of such formulas:

- At stage i determine satisfiability of formulas of the form $\psi_1 \wedge \neg\psi_2$, $depth(\psi_1 \wedge \neg\psi_2) = i$, for formulas ψ_1 and ψ_2 in $sub(\varphi)$.

By Proposition 3.2, every stage can be performed in nondeterministic polynomial time, and only $depth(\varphi)$ stages have to be performed. ∎

We now show how to deal with $\mathcal{E}_{\{1\}}$ and $\mathcal{E}_{\{2\}}$.

Proposition 3.4: *A formula φ is $\mathcal{E}_{\{1\}}$-satisfiable if and only if there exists a valuation ν for φ such that*

1. *if $E_a\psi$ is in $sub(\varphi)$ and $\nu(E_a\psi) = 1$, then ψ is $\mathcal{E}_{\{1\}}$-satisfiable, and*

2. *if $E_a\psi_1$ and $E_a\psi_2$ are in $sub(\varphi)$, $\nu(E_a\psi_1) = 1$, and $\nu(E_a\psi_2) = 0$, then $(\psi_1 \wedge \neg\psi_2) \vee (\neg\psi_1 \wedge \psi_2)$ is $\mathcal{E}_{\{1\}}$-satisfiable.*

Proposition 3.5: *A formula φ is $\mathcal{E}_{\{2\}}$-satisfiable if and only if there exists a valuation ν for φ such that*

1. *if $E_a\psi$ is in $sub(\varphi)$ and $\nu(E_a\psi) = 0$, then $\neg\psi$ is $\mathcal{E}_{\{2\}}$-satisfiable, and*

2. *if $E_a\psi_1$ and $E_a\psi_2$ are in $sub(\varphi)$, $\nu(E_a\psi_1) = 1$, and $\nu(E_a\psi_2) = 0$, then $(\psi_1 \wedge \neg\psi_2) \vee (\neg\psi_1 \wedge \psi_2)$ is $\mathcal{E}_{\{2\}}$-satisfiable.*

Corollary 3.6: *The satisfiability problems for $\mathcal{E}_{\{1\}}$ and $\mathcal{E}_{\{2\}}$ are in NP.*

To deal with $\mathcal{E}_{\{3\}}$ we first adapt Lemma 3.1.

Lemma 3.7: *Let $M = (W, N, I) \in \mathcal{E}_{\{3\}}$, let φ be a formula, let $w \in W$, and let $a \in \mathcal{A}$. Then $M, w \not\models E_a\varphi$ if and only iff $U \not\subseteq I(\varphi)$ for all $U \in N(a, w)$.*

Proof: By Lemma 3.1, $M, w \not\models E_a\varphi$ if and only iff $U \neq I(\varphi)$ for all $U \in N(a, w)$. If $U \not\subseteq I(\varphi)$ for all $U \in N(a, w)$, then in particular $U \neq I(\varphi)$ for all $U \in N(a, w)$. Suppose that $U \neq I(\varphi)$ for all $U \in N(a, w)$, but $V \subseteq I(\varphi)$ for some $V \in N(a, w)$. By C3, $V \cup I(\varphi) \in N(a, w)$. But $V \cup I(\varphi) = I(\varphi)$ - contradiction. ∎

We can now state the condition for $\mathcal{E}_{\{3\}}$-satisfiability.

Proposition 3.8: *A formula φ is $\mathcal{E}_{\{3\}}$-satisfiable if and only if there exists a valuation ν for φ such that if $E_a\psi_1$ and $E_a\psi_2$ are in $sub(\varphi)$, $\nu(E_a\psi_1) = 1$, and $\nu(E_a\psi_2) = 0$, then $(\psi_1 \wedge \neg\psi_2)$ is $\mathcal{E}_{\{3\}}$-satisfiable.*

Corollary 3.9: *The satisfiability problem for $\mathcal{E}_{\{3\}}$ is in NP.*

It is easy to combine Propositions 3.4, 3.5 and 3.8 to characterize $\mathcal{E}_S$-satisfiability and obtain an NP decision procedure for all $S \subseteq \{1, 2, 3\}$. Since the logic of $\mathcal{E}_{\{1,2,3\}}$ is the logic of local reasoning in [FH88] (without, however, modalities for implicit belief), this establishes that the satisfiability problem for that logic is in NP. (Note, however, that with modalities for implicit belief the satisfiability problem for that logic is PSPACE-complete [FH88].)

We conclude this section by demonstrating how to deal with introspection.

Proposition 3.10 : *A formula φ is $\mathcal{E}_{\{5\}}$-satisfiable if and only if there exists a valuation ν for φ such that if $E_a\psi_1$ and $E_a\psi_2$ are in $sub(\varphi)$, $\nu(E_a\psi_1) = 1$, and $\nu(E_a\psi_2) = 0$, then both $(\psi_1 \wedge \neg\psi_2) \vee (\neg\psi_1 \wedge \psi_2)$ and $(E_a\psi_1 \wedge \neg\psi_2) \vee (\neg E_a\psi_1 \wedge \psi_2)$ are $\mathcal{E}_{\{5\}}$-satisfiable.*

Corollary 3.11: *The satisfiability problem for $\mathcal{E}_{\{5\}}$ is in NP.*

Proof: To determine satisfiability of φ, we use Proposition 3.10 and the fact that there are only quadratically many formulas of the form $\psi_1 \wedge \neg\psi_2$ for formulas ψ_1 and ψ_2 in $sub(\varphi)$.

As before, we can use dynamic programming techniques to determine satisfiability of such formulas:

- At stage i determine $\mathcal{E}_{\{5\}}$-satisfiability of formulas of the form $\psi_1 \wedge \neg\psi_2$, where $size(\psi_1 \wedge \neg\psi_2) = i$, for formulas ψ_1 and ψ_2 in $sub(\varphi)$.

Clearly, every stage can be performed in nondeterministic polynomial time, and at most $2size(\varphi)$ stages have to be performed. ∎

3.2 $\mathcal{E}_{\{4\}}$, $\mathcal{E}_{\{2,3,4\}}$, and $\mathcal{E}_{\{2,3,4,5\}}$

The main feature of the logics in the previous subsection is that the recursive condition in the characterization of satisfiability is expressed in terms of combinations of one or two subformulas. As a result, there are only polynomially many distinct nodes in the computation tree of the natural alternating polynomial-time algorithm for satisfiability, which is the basis for our nondeterministic polynomial-time upper bounds. This is not the case any more when agents can epistemically combine facts.

We start by adapting Lemma 3.1 to $\mathcal{E}_{\{4\}}$.

Lemma 3.12: *Let $M = (W, N, I) \in \mathcal{E}_{\{4\}}$, let φ be a formula, let $w \in W$, and let $a \in \mathcal{A}$. Then $M, w \not\models E_a\varphi$ if and only iff $\bigcap_{j=1}^k U_j \neq I(\varphi)$ for all nonempty subsets $\{U_1, \ldots, U_k\} \subseteq N(a, w)$.*

Proof: By Lemma 3.1, $M, w \not\models E_a\varphi$ if and only iff $U \neq I(\varphi)$ for all $U \in N(a, w)$. If $\bigcap_{j=1}^k U_j \neq I(\varphi)$ for all subsets $\{U_1, \ldots, U_k\} \subseteq N(a, w)$, in particular $U \neq I(\varphi)$ for all $U \in N(a, w)$. Suppose that $U \neq I(\varphi)$ for all $U \in N(a, w)$, but $\bigcap_{j=1}^k U_j = I(\varphi)$ for some subset $\{U_1, \ldots, U_k\} \subseteq N(a, w)$. By C4, $\bigcap_{j=1}^k U_j \in N(a, w)$. But $\bigcap_{j=1}^k U_j = I(\varphi)$ - contradiction. ∎

We can now characterize $\mathcal{E}_{\{4\}}$-satisfiability.

Proposition 3.13 : *A formula φ is $\mathcal{E}_{\{4\}}$-satisfiable if and only if there exists a valuation ν for φ such that if $E_a\psi_1, \ldots, E_a\psi_k$ are in $sub(\varphi)$, $\nu(E_a\psi_j) = 1$, for $1 \leq j < k$, and $\nu(E_a\psi_k) = 0$, then either $\bigwedge_{j=1}^{k-1} \psi_j \wedge \neg\psi_k$ is $\mathcal{E}_{\{4\}}$-satisfiable or $\neg\psi_j \wedge \psi_k$ is $\mathcal{E}_{\{4\}}$-satisfiable for some j, $1 \leq j < k$.*

Theorem 3.14: *The satisfiability problem for $\mathcal{E}_{\{4\}}$ is in PSPACE.*

Proof: Consider the following algorithm:

1. Nondeterministically guess a valuation ν for φ.

2. For all $E_a\psi_1, \ldots, E_a\psi_k$ in $sub(\varphi)$, such that $\nu(E_a\psi_j) = 1$, for $1 \leq j < k$, and $\nu(E_a\psi_k) = 0$, nondeterministically and recursively check that either $\bigwedge_{j=1}^{k-1} \psi_j \wedge \neg\psi_k$ is $\mathcal{E}_{\{4\}}$-satisfiable or $\neg\psi_j \wedge \psi_k$ is $\mathcal{E}_{\{4\}}$-satisfiable for some j, $1 \leq j < k$.

Again, it is easy to see that this is an alternating polynomial-time algorithm, and therefore can be simulated by a deterministic polynomial space algorithm. Thus, the satisfiability problem for $\mathcal{E}_{\{4\}}$ is in PSPACE. ∎

We now consider $\mathcal{E}_{\{2,3,4\}}$. Recall that the logic for this class is the multi-agent extension of the modal logic K. The satisfiability problem for this logic is known to be PSPACE-complete [HM85, Lad77].

Lemma 3.15: *Let $M = (W, N, I) \in \mathcal{E}_{\{2,3,4\}}$, let φ be a formula, let $w \in W$, and let $a \in \mathcal{A}$. Then $M, w \not\models E_a\varphi$ if and only iff $\bigcap_{U \in N(a,w)} U \not\subseteq I(\varphi)$.*

Proof: First note that by C2, the set $\{U \in N(a, w)\}$ is nonempty.

By Lemma 3.7, $M, w \not\models E_a\varphi$ if and only iff $U \not\subseteq I(\varphi)$ for all $U \in N(a, w)$. If $\bigcap_{U \in N(a,w)} U \not\subseteq I(\varphi)$, then $U \not\subseteq I(\varphi)$ for all $U \in N(a, w)$. Suppose that $U \not\subseteq I(\varphi)$ for all $U \in N(a, w)$, but $\bigcap_{U \in N(a,w)} U \subseteq I(\varphi)$. By C4, $\bigcap_{U \in N(a,w)} U \in N(a, w)$ - contradiction. ∎

Proposition 3.16 : *A formula φ is $\mathcal{E}_{\{2,3,4\}}$-satisfiable if and only if there exists a valuation ν for φ such that if $E_a\psi \in sub(\varphi)$ and $\nu(E_a\psi_j) = 0$, then $\bigwedge_{\nu(E_a\psi')=1} \psi' \wedge \neg\psi$ is $\mathcal{E}_{\{2,3,4\}}$-satisfiable.*[6]

Corollary 3.17: *The satisfiability problem for $\mathcal{E}_{\{2,3,4\}}$ is in PSPACE.*

We note that valuations were used in [Lop78] to obtain a decision procedure for the modal logic K. The method there is much more complicated than ours and its running time is exponential.

We conclude this section by demonstrating how to deal with introspection. We consider the logic $\mathcal{E}_{\{2,3,4,5\}}$, which is the multi-agent extension of the modal logic K4 (cf. [Che80]).

Proposition 3.18 : *A formula φ is $\mathcal{E}_{\{2,3,4,5\}}$-satisfiable if and only if there exists a valuation ν for φ such that if $E_a\psi \in sub(\varphi)$ and $\nu(E_a\psi_j) = 0$, then both $\bigwedge_{\nu(E_a\psi')=1} \psi' \wedge \neg\psi$ and $\bigwedge_{\nu(E_a\psi')=1} E_a\psi' \wedge \neg\psi$ are $\mathcal{E}_{\{2,3,4,5\}}$-satisfiable.*

Corollary 3.19: *The satisfiability problem for $\mathcal{E}_{\{2,3,4,5\}}$ is in PSPACE.*

Proof: At first attempt one may try to come up with the obvious alternating algorithm that correspond to Proposition 3.18, as in the proof of Theorem 3.14. Unfortunately, this does not work, since when the algorithm recursively check that $\bigwedge_{\nu(E_a\psi')=1} E_a\psi' \wedge \neg\psi$ is $\mathcal{E}_{\{2,3,4,5\}}$-satisfiable,

---

[6] If the index set for the conjunction $\bigwedge_{\nu(E_a\psi')=1} \psi'$ is empty then the formula $\bigwedge_{\nu(E_a\psi')=1} \psi' \wedge \neg\psi$ degenerates to $\neg\psi$.

neither the epistemic depth nor nor the epistemic size are decreased. As a result, the algorithm may not terminate.

To achieve termination we describe a *dual algorithm* for $\mathcal{E}_{\{2,3,4,5\}}$-*unsatisfiabity*:

> For every valuation ν for φ guess a formula $E_a\psi$ in $sub(\varphi)$ such that $\nu(E_a\psi) = 0$, and recursively and nondeterministically check that either $\bigwedge_{\nu(E_a\psi')=1} \psi' \wedge \neg\psi$ or $\bigwedge_{\nu(E_a\psi')=1} E_a\psi' \wedge \neg\psi$ is $\mathcal{E}_{\{2,3,4,5\}}$-unsatisfiable.

There is a subtlety here in showing that this alternating algorithm terminates in polynomial time. We first have to define the notion of a *computation forest* for a formula φ, whose unsatisfaiability is being tested. Nodes in a computation forest are labeled by formulas and valuations for the formulas. The roots are labeled by the formula φ and all valuations for φ. Suppose now that x is a node in the computation forest labeled by the formula ψ_x and the valuation ν. Then there is a formula θ of the form $\bigwedge_{\nu(E_a\psi')=1} \psi' \wedge \neg\xi$, in which case we say that θ is of the *first kind*, or $\bigwedge_{\nu(E_a\psi')=1} E_a\psi' \wedge \neg\xi$, in which case we say that θ is of the *second kind*, where ξ is a formula such that $E_a\xi \in sub(\varphi)$ and $\nu(E_a\xi) = 0$, such that the children of x are labeled by θ and all valuations for θ. In this case we say that x is an *a-parent*. In particular, if θ has no valuations, then x is a leaf. It is easy to see that φ is $\mathcal{E}_{\{2,3,4,5\}}$-unsatisfiable iff there is a finite computation forest for φ. It remains to prove that if there is a finite computation forest, then there is a forest whose depth is polynomial in the length of φ.

Some observations are now in order. First, if θ is of the first kind then its epistemic depth is smaller than the epistemic depth of ψ. Second, if x is an *a-parent* of y, and y is an *b-parent* of z, where $a \neq b$, and x, y, and z are labeled by the formulas ψ_x, ψ_y, ψ_z, correspondingly, then then all the conjuncts of ψ_z are *proper* subformulas of ψ_x. It follows that the only way the forest can have superpolynomially long branches is if there is some a such that the forest has a superpolynomially long sequence of nodes that are all *a-parents* labeled by formulas of the second type. Finally, if x is an *a-parent* of y, x and y are labeled by the formulas ψ_x and ψ_y, correspondingly, where ψ_y is of the second kind, and

by valuations ν_x and ν_y, and if $E_a \alpha \in sub(\psi_x)$, then $\nu_x(\mathcal{E}_a \alpha) = 1$ entails $\nu_y(\mathcal{E}_a \alpha) = 1$. We call this property the *monotonicity of valuations*.

Since the algorithm recursively try to prove unsatisfiability of formulas, we can assume without loss of generality that the computation forest is nonredundant, that is, if x is an ancestor of y, then x and y are labeled by distinct formulas. Let m be the number of formulas of the form $E_a \xi$ in $sub(\varphi)$. Let $E_a\xi_1, \ldots, E_a\xi_m$ be an enumeration of these formulas. Suppose that $x_1, \ldots, x_{m(m+1)+1}$ is a sequence of nodes in the computation forest, where x_i is an a-parent of x_{i+1} for $1 \le i \le m(m+1)$, and x_i is labeled by the formula ψ_i and the valuation ν_i, for $1 \le i \le m(m+1)+1$. Because of the monotonicity of valuations, there is some j, $1 \le j \le m^2 - 1$, such that $\nu_j, \nu_{j+1}, \ldots, \nu_{j+m+1}$ agree on $E_a\xi_i$, for $1 \le i \le m$. Without loss of generality, we can assume that all these valuations assign 1 to $E_a\xi_1, \ldots, E_a\xi_k$ and assign 0 to $E_a\xi_{k+1}, \ldots, E_a\xi_m$, where $1 \le k < m$. It follows that $\psi_{j+1}, \ldots, \psi_{j+m+1}$ are all of the form $\bigwedge_{i=1}^{k} E_a\xi_i \wedge \neg\xi_l$, where $k < l \le m$. Consequently, for some j', j'', $j < j' < j'' \le j+m+1$, we have that $\psi_{j'}$ and $\psi_{j''}$ are identical - contradiction.

It follows from our analysis, that if φ has a nonredundant computation forest, then it has a forest whose depth is at most cubic in the length of φ. This proves that the algorithm terminates in polynomial time. ∎

4 Concluding remarks

We have described a methodology for establishing upper bounds on complexity of epistemic reasoning. Since all our logics extend propositional logic, the satisfiability problem is clearly NP-hard. We conjecture that all our bounds are tight. That is, for $S \subseteq \{1, \ldots, 7\}$, if $4 \in S$, then the satisfiability problem for $\mathcal{E}_S$ is PSPACE-complete, otherwise, the satisfiability problem is NP-complete. Intuitively, the conjecture is that it is the ability to epistemically combine facts that causes epistemic reasoning to be harder than propositional reasoning.

Finally, we note that it is possible to obtain characterizations for satisfiability in the spirit of Proposition 3.16 for logics such as *propositional dynamic logic* [FL79]. These characterizations, however, do not yields decision procedures in a straightforward manner as here, since the obvious algorithms for either satisfiability or unsatisfiability do not terminate (cf. [Pra80, VW86]).

References

[Bet59] E. W. Beth, *The Foundation of Mathematics*, North-Holland, 1959.

[Che80] B. F. Chellas, *Modal Logic*, Cambridge University Press, 1980.

[CKS81] A. K. Chandra, D. Kozen, and L. J. Stockmeyer, Alternation, *Journal of the ACM* **28**, 1981, pp. 114–133.

[Cre85] M. J. Cresswell, *Structured Meanings*, MIT Press, 1985.

[FH88] R. Fagin and J. Y. Halpern, Belief, awareness, and limited reasoning, *Artificial Intelligence* **34**, 1988, pp. 39–76.

[Fit83] M. Fitting, *Proof Methods for Modal and Intuitionistic Logics*, Synthese Library 169, Reidel, 1983.

[FL79] M. J. Fischer and R. E. Ladner, Propositional dynamic logic of regular programs, *Journal of Computer and System Sciences* **18**:2, 1979, pp. 194–211.

[Hin57] J. Hintikka, Quantifiers in deontic logic, *Societas Scientiarumi Fennica, Commentationes Humanarum Literarum* **23**:4, 1957, pp. 1–23.

[Hin62] J. Hintikka, *Knowledge and Belief*, Cornell University Press, 1962.

[Hin75] J. Hintikka, Impossible worlds vindicated, *Journal of Philosophy* **4**, 1975, pp. 475–484.

[HM85] J. Y. Halpern and Y. Moses, A guide to the modal logics of knowledge and belief, *Ninth International Joint Conference on Artificial Intelligence (IJCAI-85)*, 1985, pp. 480–490.

[KP81] D. Kozen and R. Parikh, An elementary proof of the completeness of PDL, *Theoretical Computer Science* **14**:1, 1981, pp. 113–118.

[Kri63] S. Kripke, A semantical analysis of modal logic I: normal modal propositional calculi, *Zeitschrift für Mathema-*