On the decidability and complexity of reasoning about only knowing

Riccardo Rosati

Dipartimento di Informatica e Sistemistica
Università di Roma "La Sapienza"
Via Salaria 113, 00198 Roma, Italy
email: rosati@dis.uniroma1.it

Abstract

We study reasoning in Levesque's logic of only knowing. In particular, we first prove that extending a decidable subset of first-order logic with the ability of reasoning about only knowing preserves decidability of reasoning, as long as *quantifying-in* is not allowed in the language, and define a general method for reasoning about only knowing in such a case. Then, we show that the problem of reasoning about only knowing in the propositional case lies at the second level of the polynomial hierarchy. Thus, it is as hard as reasoning in the majority of propositional formalisms for nonmonotonic reasoning, like default logic, circumscription, and autoepistemic logic, and it is easier than reasoning in propositional formalisms based on the minimal knowledge paradigm, which is strictly related to the notion of only knowing. Finally, we identify a syntactic restriction in which reasoning about only knowing is easier than in the general propositional case, and provide a specialized deduction method for such a restricted setting.

1 Introduction

Research in the formalization of commonsense reasoning through epistemic logics [21,23,18] has pointed out the need for providing systems (agents) with the ability of introspecting on their own knowledge and ignorance. To this aim, an *epistemic closure assumption* is generally adopted, which informally can be stated as follows: "the logical theory formalizing the agent is a *complete* specification of the agent's knowledge". As a consequence, any fact that is not logically implied by such a theory is assumed to be *not known* by the agent. [1]

[1] The use of the notion of logical implication here may be misleading: to be more precise, the closure assumption acts by "maximizing ignorance" (in a way that changes according to the different proposals) in each possible *epistemic state* of the

As shown in [18], this paradigm underlies the vast majority of the logical formalizations of nonmonotonic reasoning. Roughly speaking, there exist two different ways to embed such a principle into a logic:

(i) by considering a *nonmonotonic* formalism, whose semantics *implicitly* realizes such a “closed” interpretation of the logical theory representing the agent’s knowledge;
(ii) by representing the closure assumption *explicitly* in the framework of a monotonic logic, suitably extending its syntax and semantics.

The first approach has been pursued in the definition of several *modal* formalizations of nonmonotonic reasoning, e.g. McDermott and Doyle’s nonmonotonic modal logics [20], Halpern and Moses’ logic of minimal epistemic states [9] and Lifschitz’s logic of minimal belief and negation as failure [19]. On the other hand, the second approach has been followed by Levesque [18] in the definition of the logic of *only knowing*.

The logic of only knowing is obtained by adding an “all-I-know” modal operator O to modal logic $\mathsf{K45}$. Informally, such an interpretation of the modality O is obtained through a maximization of the set of successors of each world satisfying O-formulas.

There is a strict similarity between the interpretation of the modality O and the semantics of nonmonotonic modal logics. Let φ be a modal formula specifying the knowledge of the agent: in the logic of only knowing, satisfiability of the formula $O\varphi$ in a world w requires maximization of the possible worlds connected to w and satisfying φ; an analogous kind of maximization is generally realized by the preference semantics of nonmonotonic modal logics, by choosing, among the models for φ, only the models having a “maximal” set of possible worlds, where such a notion of maximality changes according to the different proposals. In a nutshell, the logic of only knowing is a monotonic formalism, in which the modality O allows for an explicit representation of the epistemic closure assumption at the object level (i.e. in the language of the logic), whereas in nonmonotonic formalisms the closure assumption is a meta-level notion.

The studies investigating the relationship between only knowing and nonmonotonic logics [1] have stressed the analogies between the two approaches from an epistemological viewpoint. An analogous analysis from the computational viewpoint has not been pursued so far. Indeed, there exist several studies concerning the computational properties of nonmonotonic logics, in particular propositional nonmonotonic modal formalisms (e.g., [3,5,20,22]). On the other hand, the computational properties of only knowing in the propositional case have not been thoroughly investigated. The only related studies appearing

---

agent.

in the literature concern a fragment of $\mathcal{OL}$ built upon a very restricted subclass of propositional formulas, for which satisfiability is tractable [15], and a computational study of a framework in which only knowing is added to a formal model of *limited* reasoning [13]. Moreover, a lower bound for reasoning in the propositional fragment of $\mathcal{OL}$ (Σ_2^p) is known, due to the fact that autoepistemic logic [21] can be embedded in polynomial time into $\mathcal{OL}$.

The goal of the present work is to provide algorithms for computing satisfiability in the logic of only knowing. To this end, we exploit the similarity between this formalism and nonmonotonic modal logics, in order to identify a finite characterization of the models of a formula in the logic $\mathcal{OL}$.

The main results of the paper concern both decidability and complexity of reasoning about only knowing. Specifically, we first prove that extending a decidable subset of first-order logic (without equality) with the ability of reasoning about only knowing preserves decidability of reasoning, as long as *quantifying-in*, i.e. the presence of modalities inside quantifiers, is not allowed. Moreover, we define a general method for computing satisfiability in $\mathcal{OL}$ without quantifying-in. To the best of our knowledge, such an algorithm is the first terminating procedure for reasoning about only knowing in any decidable fragment of first-order logic (e.g. in the full propositional fragment of $\mathcal{OL}$).

Then, we show that the problem of reasoning about only knowing in the propositional case lies at the second level of the polynomial hierarchy. More precisely, satisfiability in the propositional fragment of $\mathcal{OL}$ is a Σ_2^p-complete problem. Thus, reasoning about only knowing is as hard as reasoning in the majority of propositional formalisms for nonmonotonic reasoning, like autoepistemic logic [22,5], default logic [5], circumscription [4], and several McDermott and Doyle's logics [20]. Moreover, reasoning about only knowing is easier (unless the polynomial hierarchy collapses) than reasoning in nonmonotonic modal formalisms based on the *minimal knowledge* paradigm [27], like Halpern and Moses' logic of minimal epistemic states [2], Lifschitz's logic *MBNF* [24], and the moderately grounded version of autoepistemic logic [3].

We also define an interesting syntactic restriction of the propositional fragment of $\mathcal{OL}$ in which deduction is easier than in the general case. Specifically, we identify a subset of formulas in $\mathcal{OL}$ for which the satisfiability (and validity) problem is $\mathrm{P}^{\mathrm{NP}}[O(\log n)]$-complete, i.e. can be reduced to a logarithmic number of propositional satisfiability problems. This case is particularly interesting, since it can be viewed as a generalization of the problem of answering epistemic queries to a propositional knowledge base [23,6].

In the following, we first briefly introduce the modal logic of only knowing $\mathcal{OL}$. Then, in Section 3 we present a finite characterization of the models of a sentence in $\mathcal{OL}$, which provides the basis for the definition of reasoning methods for $\mathcal{OL}$. In Section 4 we define a deduction method for satisfiability (validity) in decidable fragments of $\mathcal{OL}$, and analyze the computational properties of $\mathcal{OL}$ in the propositional fragment of $\mathcal{OL}$; we also define a syntactic restriction of $\mathcal{OL}$, showing that reasoning in this setting is easier than in the general case. Finally, in Section 5 we investigate the relationship between only knowing and the minimal knowledge paradigm, and conclude in Section 6.

2 The logic $\mathcal{OL}$

In this section we briefly recall the formalization of only knowing [18]. We assume that the reader is familiar with the basic notions of modal logic. We recall that $\mathsf{K45}$ denotes the modal logic interpreted on Kripke structures whose accessibility relation among worlds is transitive and euclidean, while modal logic $\mathsf{KD45}$ imposes in addition that the relation be serial; finally, modal logic $\mathsf{S5}$ also imposes reflexivity on the accessibility relation (see e.g. [10,20] for more details).

We use $\mathcal{L}$ to denote the language of first-order logic without equality, i.e. $\mathcal{L}$ is the set of first-order sentences built in the usual way upon connectives $\wedge, \neg$ (the symbols $\vee, \supset, \equiv$ are used as abbreviations), an existential quantifier, an infinite set of variables, an infinite set $\mathcal{A}$ of propositional symbols, an infinite set of predicate symbols of every arity, and an infinite set of function symbols. [2] We assume that $\mathcal{A}$ contains the symbols true, false. We call *objective* any sentence from $\mathcal{L}$.

Following [18], we interpret sentences from $\mathcal{L}$ with respect to a fixed, countably infinite interpretation domain Δ. As shown in [17], imposing a countably infinite domain does not influence satisfiability/validity of first-order sentences without equality, i.e. the set of satisfiable sentences is the same as in classical first-order logic. In the following, we call *interpretation* a usual first-order interpretation for $\mathcal{L}$ over Δ. An interpretation is also called *world*. For each interpretation w, $w(\mathsf{true}) = \mathit{TRUE}$ and $w(\mathsf{false}) = \mathit{FALSE}$. The evaluation $w(\varphi)$ of a sentence φ in an interpretation w is defined in the usual way. We say that a sentence $\varphi \in \mathcal{L}$ is *satisfiable* if there exists an interpretation w such that $w(\varphi) = \mathit{TRUE}$ (which we also denote as $w \models \varphi$).

Definition 1 *We denote as $\mathcal{L}_O$ the modal extension of $\mathcal{L}$ with the modalities K and O inductively defined as follows:*

(i) if $\varphi \in \mathcal{L}$, then $\varphi \in \mathcal{L}_O$;

[2] The assumption done in [18] that constants are *rigid designators* (i.e., in each interpretation, each constant denotes the same element of the interpretation domain) can be omitted here, since the case of quantifying-in is not dealt with in this paper.

(ii) *if* $\varphi \in \mathcal{L}_O$*, then* $K\varphi \in \mathcal{L}_O$*;*
(iii) *if* $\varphi \in \mathcal{L}_O$*, then* $O\varphi \in \mathcal{L}_O$*;*
(iv) *if* $\varphi \in \mathcal{L}_O$*, then* $\neg\varphi \in \mathcal{L}_O$*;*
(v) *if* $\varphi_1 \in \mathcal{L}_O$ *and* $\varphi_2 \in \mathcal{L}_O$*, then* $\varphi_1 \wedge \varphi_2 \in \mathcal{L}_O$*;*
(vi) *nothing else belongs to* $\mathcal{L}_O$*.*

Informally, the above definition does not allow *quantifying-in*, i.e., in all sentences from $\mathcal{L}_O$, each occurrence of the modalities K and O lies outside the scope of quantifiers. E.g., the sentence $\forall x O(p(x))$ does not belong to $\mathcal{L}_O$, while the sentence $O(\forall x\, p(x))$ belongs to $\mathcal{L}_O$.

We also use $\mathcal{L}_K$ to denote the analogous extension of $\mathcal{L}$ with the only modality K. We call *O-sentence* a sentence from $\mathcal{L}_O$ of the form $O\varphi$. Notice that, with respect to [18], we slightly change the language of the logic, using the modality K instead of B.

The semantics of a sentence $\varphi \in \mathcal{L}_O$ is defined in terms of satisfiability in a structure (w, M) where w is an interpretation (called *initial world*) and M is a set of interpretations.

Definition 2 *Let* w *be an interpretation on* $\mathcal{L}$*, and let* M *be a set of such interpretations. We say that a sentence* $\varphi \in \mathcal{L}_O$ *is* satisfied *in* (w, M)*, and write* $(w, M) \models \varphi$*, iff the following conditions hold:*

(i) *if* $\varphi \in \mathcal{L}$*, then* $(w, M) \models \varphi$ *iff* $w(\varphi) = TRUE$*;*
(ii) *if* $\varphi = \neg\varphi_1$*, then* $(w, M) \models \varphi$ *iff* $(w, M) \not\models \varphi_1$*;*
(iii) *if* $\varphi = \varphi_1 \wedge \varphi_2$*, then* $(w, M) \models \varphi$ *iff* $(w, M) \models \varphi_1$ *and* $(w, M) \models \varphi_2$*;*
(iv) *if* $\varphi = K\varphi_1$*, then* $(w, M) \models \varphi$ *iff for every* $w' \in M$*,* $(w', M) \models \varphi_1$*;*
(v) *if* $\varphi = O\varphi_1$*, then* $(w, M) \models \varphi$ *iff for every* w'*,* $w' \in M$ *iff* $(w', M) \models \varphi_1$*.*

We say that $\varphi \in \mathcal{L}_O$ is *weakly* $\mathcal{OL}$-satisfiable if there exists (w, M) such that $(w, M) \models \varphi$. Since the initial world does not influence satisfiability of a sentence of the form $K\varphi$ or $O\varphi$, we write $M \models K\varphi$ (resp. $M \models O\varphi$) iff $(w, M) \models K\varphi$ (resp. $(w, M) \models O\varphi$) for any interpretation w.

The above semantics is not actually the one originally proposed in [18]: in addition to the above rules, a pair (w, M) must satisfy a maximality condition for the set M, as described below. However, as mentioned in [7], the above, weaker notion of satisfiability is also meaningful.

In the following, $Th(M)$ denotes the set of sentences $K\varphi$ such that $\varphi \in \mathcal{L}_K$ and, for each $w \in M$, $(w, M) \models K\varphi$. Given two sets of interpretations M_1, M_2, we say that M_1 is *equivalent* to M_2 iff $Th(M_1) = Th(M_2)$.

Definition 3 *A set of interpretations* M *is* maximal *iff, for each set of interpretations* M'*, if* M' *is equivalent to* M *then* $M' \subseteq M$*.*

Definition 4 *A sentence* $\varphi \in \mathcal{L}_O$ *is* $\mathcal{OL}$-satisfiable *iff there exists a pair* (w, M) *such that* $(w, M) \models \varphi$ *and* M *is maximal.*

Roughly speaking, the maximality condition prevents from the existence of models which agree on all basic beliefs, yet disagree on whay they only know [18, Section 2.2].

We say that a sentence $\varphi \in \mathcal{L}_O$ is $\mathcal{OL}$*-valid* iff $\neg\varphi$ is not $\mathcal{OL}$-satisfiable. In the next section we will prove that the notions of $\mathcal{OL}$-satisfiability and weak $\mathcal{OL}$-satisfiability of a sentence from $\mathcal{L}_O$ coincide. Notice, however, that $\mathcal{OL}$-satisfiability and weak $\mathcal{OL}$-satisfiability for *infinite* theories are, in general, different (see [18, Section 2.4]).

As for reasoning in $\mathcal{OL}$, we give the following definition.

Definition 5 *A sentence* $\varphi \in \mathcal{L}_O$ logically implies *a sentence* $\psi \in \mathcal{L}_O$ *in* $\mathcal{OL}$ *(and write* $\varphi \models_{\mathcal{OL}} \psi$*) iff* $\varphi \supset \psi$ *is* $\mathcal{OL}$*-valid.*

Based on the above definition, we can immediately reduce reasoning to unsatisfiability in $\mathcal{OL}$.

Remark 6 *An alternative definition of logical implication is given in several studies on epistemic and nonmonotonic modal logics (see e.g. [20, Definition 7.9]). Such a notion is based on the following notion of* validity *of a modal sentence in a model: a formula* φ *is valid in a Kripke model* $\mathcal{M}$ *iff, for each world* w *in* $\mathcal{M}$*,* $(w, \mathcal{M}) \models \varphi$*. The notion of logical implication is then expressed as follows: "*ψ *is logically implied by* φ *iff* ψ *is valid in every model in which* φ *is valid". The two notions are in general different, and such a difference also holds for the logic* $\mathcal{OL}$*. However, since in* $\mathcal{OL}$ *the accessibility relation of each interpretation structure is transitive, it can immediately be shown [20, Remark 7.11] that the last notion of logical implication can be reduced to the one given in Definition 5, and hence to validity in* $\mathcal{OL}$*. In particular,* ψ *is logically implied by* φ *according to the last notion iff* $(\varphi \wedge K\varphi) \supset \psi$ *is* $\mathcal{OL}$*-valid.*

Notice that the above semantics strictly relates the logic $\mathcal{OL}$ with modal logic K45, since there is a precise correspondence between the pairs (w, M) used in the above definition and K45 models. We recall that, with respect to the satisfiability problem, a K45 model can be considered without loss of generality as a pair $(w, \mathcal{M})$, where w is a world, $\mathcal{M}$ is a set of worlds (possibly empty), w is connected to all the worlds in $\mathcal{M}$, the worlds in $\mathcal{M}$ are connected with each other (i.e. $\mathcal{M}$ is a universal S5 model) and no world in $\mathcal{M}$ is connected to w [20] (in the case of KD45 models, M is required to be non-empty). Thus, in the following we will refer to a pair (w, M) as a K45 model whose S5 component is M. Notice also that, if $\Sigma \in \mathcal{L}_K$, then Σ is $\mathcal{OL}$-satisfiable if and only if it is K45-satisfiable, which is shown by the fact that, if a K45 model (w, M)

satisfies such a Σ, then there exists a maximal set M' equivalent to M, hence (w, M') satisfies Σ.

Informally, the interpretation of the O modality is obtained through the maximization of the set of successors of each world satisfying an O-sentence. As pointed out e.g. in [14], the meaning of an O-sentence $O\varphi$ such that φ is non-modal is intuitive, whereas it is more difficult to understand the semantics of an O-sentence with nested modalities.

Example 7 *Suppose $\varphi \in \mathcal{L}$. Then, (w, M) is a model for $O\varphi$ iff $M = \{w : w \models \varphi\}$. Hence, the effect of prefixing φ with the modality O is that of maximizing the possible worlds in M, which contains all the interpretations consistent with φ.*

Example 8 *Suppose $\varphi \in \mathcal{L}$ and φ is not a tautology. Then, the sentence $OK\varphi$ is not $\mathcal{OL}$-satisfiable. In fact, suppose $OK\varphi$ is $\mathcal{OL}$-satisfiable. Then, there exists (w, M) such that $(w, M) \models OK\varphi$. Now, it is easy to see that, by Definition 2, M cannot contain any interpretation w' such that $w' \not\models \varphi$. On the other hand, since φ is not a tautology, there exists such an interpretation w'; moreover, $(w', M) \models K\varphi$, since the interpretation of $K\varphi$ in (w', M) does not depend on the initial world, hence by Definition 2 it follows that $w' \in M$. Contradiction. Hence, $OK\varphi$ is not $\mathcal{OL}$-satisfiable. On the contrary, $O(K\varphi \wedge \varphi)$ is $\mathcal{OL}$-satisfiable, under the assumption that φ is satisfiable.*

3 Characterizing $\mathcal{OL}$-satisfiability

In this section we present a finite characterization of the models of a sentence $\Sigma \in \mathcal{L}_O$ which is based on the use of partitions of modal sentences occurring in Σ. Similar techniques are used in several methods for reasoning in nonmonotonic modal logics (e.g. [5,20,3,22,2]): in such methods, partitions of subformulas of a modal theory are generally used for providing a finite characterization of the epistemic states of the agent, which correspond to infinite modal theories. In fact, such partitions can also be used in order to provide a finite characterization of an S5 model. In particular, a partition satisfying certain properties identifies a particular S5 model $\mathcal{M}$, by uniquely determining a non-modal theory (called the *objective knowledge* of $\mathcal{M}$). $\mathcal{M}$ is then defined as the set of all interpretations satisfying such objective knowledge.

Now, in order to check whether an O-sentence $O\varphi$ is satisfied in a K45 model (w, M), we exploit the possibility of expressing, by means of an objective sentence, the objective knowledge of the S5 component M of (w, M). This allows us to establish whether φ is "all that is known" in the set of interpretations M.

We first introduce some preliminary definitions. Following [6], we say that an occurrence of a sentence ψ in a sentence $\varphi \in \mathcal{L}_O$ is *strict* if it is not in the scope of a modal operator. We also call *modal atom* a sentence of the form $K\varphi$ or $O\varphi$, with $\varphi \in \mathcal{L}_O$, and call *modal atoms of* Σ (denoted by $MA(\Sigma)$) the set of modal atoms occurring in Σ.

Definition 9 *Let $\Sigma \in \mathcal{L}_O$ and let P, N be sets of modal atoms such that $P \cup N \supseteq MA(\Sigma)$ and $P \cap N = \emptyset$. We denote with $\Sigma|_{P,N}$ the objective sentence obtained from Σ by substituting each strict occurrence in Σ of a sentence in P with* true*, and each strict occurrence in Σ of a sentence in N with* false*.*

Notice that only the occurrences in Σ of modal atoms which are not within the scope of another modality are replaced; notice also that $\Sigma|_{P,N}$ is an objective sentence. Informally, the pair (P, N) identifies a "guess" on the modal atoms from Σ, and $\Sigma|_{P,N}$ represents the "objective knowledge" implied by Σ under such a guess.

Definition 10 *Let (P, N) be a partition of $MA(\Sigma)$. Then, we denote with $ob(P)$ the following objective sentence:*

$$ob(P) = \left(\bigwedge_{K\varphi \in P} \varphi|_{P,N} \right) \wedge \left(\bigwedge_{O\varphi \in P} \varphi|_{P,N} \right)$$

Roughly speaking, the objective sentence $ob(P)$ represents the objective knowledge implied by the guess (P, N) on the modal atoms belonging to P.

Example 11 *Suppose $\Sigma = a \wedge O(\neg a \vee Kb)$. Then, $MA(\Sigma) = \{O(\neg a \vee Kb), Kb\}$. One possible partition of $MA(\Sigma)$ is the following:*

$$P = \{O(\neg a \vee Kb)\}$$
$$N = \{Kb\}$$

Then, $\Sigma|_{P,N} = a \wedge \mathsf{true} = a$, and $ob(P) = (\neg a \vee Kb)|_{P,N} = \neg a \vee \mathsf{false} = \neg a$.

Definition 12 *Let S be a set of modal atoms. We say that a set of interpretations M* induces *the partition (P, N) on S if, for each modal atom $K\varphi \in S$, $K\varphi \in P$ iff $M \models K\varphi$, and for each modal atom $O\varphi \in S$, $O\varphi \in P$ iff $M \models O\varphi$.*

We now define the notion of partition of a set of modal atoms induced by an objective sentence.

Definition 13 *Let $\Sigma \in \mathcal{L}_O$, $\varphi \in \mathcal{L}$. We denote with $(P_\varphi(\Sigma), N_\varphi(\Sigma))$ the partition of $MA(\Sigma)$ induced by $M = \{w : w \models \varphi\}$.*

Notice that the above definition associates a maximal set of interpretations M with the sentence φ and the partition $(P_\varphi(\Sigma), N_\varphi(\Sigma))$.

In order to establish a characterization of $\mathcal{OL}$-satisfiability based on the use of partitions of modal atoms, we prove some preliminary properties of such partitions.

Lemma 14 *Let $\varphi \in \mathcal{L}_O$, let w be an interpretation, let M be a set of interpretations, and let (P, N) be the partition induced by M on a set of modal atoms S. Then, $(w, M) \models \varphi$ iff $(w, M) \models \varphi|_{P,N}$.*

Proof. Follows immediately from Definition 9 and from Definition 2. □

Lemma 15 *Let $\Sigma \in \mathcal{L}_O$, $\varphi \in \mathcal{L}$. Then:*

(i) each modal atom $K\psi$ of $MA(\Sigma)$ belongs to $P_\varphi(\Sigma)$ iff $\varphi \supset \psi|_{P_\varphi(\Sigma),N_\varphi(\Sigma)}$ is a valid objective sentence;
(ii) each modal atom $O\psi$ of $MA(\Sigma)$ belongs to $P_\varphi(\Sigma)$ iff $\varphi \equiv \psi|_{P_\varphi(\Sigma),N_\varphi(\Sigma)}$ is a valid objective sentence.

Proof. Follows immediately from Definition 2. □

Lemma 16 *Let $\Sigma \in \mathcal{L}_K$. If Σ is* K45*-satisfiable and* KD45*-unsatisfiable, then the partition (P, N) induced by the empty set of interpretations is such that $ob(P)$ is unsatisfiable.*

Proof. Suppose Σ is K45-satisfiable and KD45-unsatisfiable, and let (w, M) be a K45 model such that $(w, M) \models \Sigma$. Then, $M = \emptyset$. Let (P, N) be the partition of $MA(\Sigma)$ induced by M: since $M = \emptyset$, it follows that $P = MA(\Sigma)$ (each sentence of the form $K\varphi$ is trivially satisfied by M). Now let $M' = \{w : w \models ob(P)\}$. We prove that $(P, N) = (P_{ob(P)}(\Sigma), N_{ob(P)}(\Sigma))$. The proof is by induction on the modal depth of the sentences in $MA(\Sigma)$. First, let $K\varphi$ be a modal atom of $MA(\Sigma)$ such that $\varphi \in \mathcal{L}$. Then, since $K\varphi \in P$, Definition 10 implies that $ob(P) \supset \varphi$ is a valid objective sentence, hence $K\varphi \in P_{ob(P)}(\Sigma)$. Suppose now that (P, N) and $(P_{ob(P)}(\Sigma), N_{ob(P)}(\Sigma))$ agree on all modal atoms of modal depth less or equal to i. Consider a modal atom $K\varphi$ of $MA(\Sigma)$ of modal depth $i + 1$. Again, since $K\varphi \in P$, by Definition 10 it follows that $ob(P) \supset \varphi|_{P,N}$ is a valid objective sentence, hence $M' \models \varphi|_{P,N}$, and since by Definition 9 the value of the sentence $\varphi|_{P,N}$ only depends on the value of the modal atoms of modal depth less or equal to i in (P, N), by the induction hypothesis and Lemma 14 it follows that $M' \models \varphi$, hence $K\varphi \in P_{ob(P)}(\Sigma)$. Consequently, $(P, N) = (P_{ob(P)}(\Sigma), N_{ob(P)}(\Sigma))$, which in turn implies that $\Sigma|_{P,N} = \Sigma|_{P_{ob(P)}(\Sigma),N_{ob(P)}(\Sigma)}$. Now, since $(w, M) \models \Sigma$, by Lemma 14 $w \models \Sigma|_{P,N}$, hence $w \models \Sigma|_{P_{ob(P)}(\Sigma),N_{ob(P)}(\Sigma)}$ and, by the same lemma, $(w, M') \models \Sigma$. Since $M' = \{w : w \models ob(P)\}$ and Σ is KD45-unsatisfiable, it follows that M' is empty, hence $ob(P)$ is unsatisfiable. □

We say that a sentence $\varphi \in \mathcal{L}_O$ has *modal depth* i if each occurrence of an objective sentence in φ lies within the scope of at most i modalities, and there is an occurrence of an objective sentence in φ which lies within the scope of exactly i modalities.

Lemma 17 *Let $\Sigma \in \mathcal{L}_O$, $\varphi \in \mathcal{L}$. Let (P, N) be the partition of $MA(\Sigma)$ built as follows:*

(i) start from $P = N = \emptyset$;
(ii) for each modal atom $K\psi$ in $MA(\Sigma)$ such that $\psi|_{P,N} \in \mathcal{L}$, if $\varphi \supset \psi|_{P,N}$ is a valid objective sentence, then add $K\psi$ to P, otherwise add $K\psi$ to N;
(iii) for each modal atom of the form $O\psi$ in $MA(\Sigma)$ such that $\psi|_{P,N} \in \mathcal{L}$, if $\varphi \equiv \psi|_{P,N}$ is a valid objective sentence, then add $O\psi$ to P, otherwise add $O\psi$ to N;
(iv) iteratively apply the above rules until $P \cup N = MA(\Sigma)$.

Then, $(P, N) = (P_\varphi(\Sigma), N_\varphi(\Sigma))$.

Proof. The proof is by induction on the structure of the sentences in $MA(\Sigma)$. First, from the fact that $(P_\varphi(\Sigma), N_\varphi(\Sigma))$ is the partition induced by $M = \{w : w \models \varphi\}$, and from Definition 2, it follows that, if $\psi \in \mathcal{L}$, then $M \models K\psi$ if and only if $\varphi \supset \psi$ is a valid objective sentence, and $M \models O\psi$ if and only if $\varphi \equiv \psi$ is a valid objective sentence. Therefore, (P, N) agrees with $(P_\varphi(\Sigma), N_\varphi(\Sigma))$ on all modal atoms of modal depth 1. Suppose now that (P, N) and $(P_\varphi(\Sigma), N_\varphi(\Sigma))$ agree on all modal atoms of modal depth less or equal to i. Consider a modal atom $K\psi$ of $MA(\Sigma)$ of modal depth $i + 1$. From Lemma 15 it follows that $M \models K\psi$ if and only if $\varphi \supset \psi|_{P_\varphi(\Sigma),N_\varphi(\Sigma)}$ is a valid objective sentence, and since by Definition 9 the value of the sentence $\psi|_{P_\varphi(\Sigma),N_\varphi(\Sigma)}$ only depends on the guess of the modal atoms of modal depth less or equal to i in $(P_\varphi(\Sigma), N_\varphi(\Sigma))$, by the induction hypothesis it follows that $\psi|_{P_\varphi(\Sigma),N_\varphi(\Sigma)} = \psi|_{P,N}$, hence $K\psi$ belongs to P if and only if it belongs to $P_\varphi(\Sigma)$. Analogously, it can be proven that any modal atom of depth $i+1$ of the form $O\psi$ belongs to P if and only if it belongs to $P_\varphi(\Sigma)$. Therefore, (P, N) and $(P_\varphi(\Sigma), N_\varphi(\Sigma))$ agree on all modal atoms of modal depth $i + 1$. □

In the following, the sentence $\Sigma[O/\mathsf{false}]$ stands for the sentence obtained from Σ by replacing all O-sentences occurring in Σ with false. Notice that $\Sigma[O/\mathsf{false}] \in \mathcal{L}_K$, hence such a sentence is $\mathcal{OL}$-satisfiable if and only if it is K45-satisfiable.

We are now ready to provide a characterization of the notion of satisfiability in $\mathcal{OL}$, based on the existence of a partition (P, N) of $MA(\Sigma)$ which satisfies the property $(P, N) = (P_{ob(P)}(\Sigma), N_{ob(P)}(\Sigma))$.

Theorem 18 *Let $\Sigma \in \mathcal{L}_O$. Then, Σ is $\mathcal{OL}$-satisfiable iff at least one of the following two conditions holds:*

(a) $\Sigma[O/\mathsf{false}]$ *is* KD45*-satisfiable;*
(b) *there exists a partition (P, N) of $MA(\Sigma)$ such that $\Sigma|_{P,N}$ is satisfiable and $(P, N) = (P_{ob(P)}(\Sigma), N_{ob(P)}(\Sigma))$.*

Proof. *If part.* Suppose that either condition (a) or condition (b) of the theorem holds. Then, there are two possible cases:

(i) $\Sigma[O/\mathsf{false}]$ is KD45-satisfiable. Then, there exists a K45 model (w, M) such that $M \neq \emptyset$ and $(w, M) \models \Sigma[O/\mathsf{false}]$. Let p' be a propositional symbol not appearing in Σ: without loss of generality, we can assume that M contains at least one world evaluating p' to *TRUE*, and at least one world evaluating p' to *FALSE*. Let M' be the maximal set equivalent to M: since $Th(M') = Th(M)$, it follows that (w, M') satisfies $\Sigma[O/\mathsf{false}]$. Now let M'' be the set of interpretations obtained from M' by eliminating all interpretations w' such that $w'(p') = \mathit{FALSE}$. By construction, M'' is maximal and non-empty. Consider the model (w, M''): on the one hand, (w, M'') satisfies $\Sigma[O/\mathsf{false}]$, since p' does not appear in Σ; on the other hand, since all interpretations in M'' satisfy p', $M'' \models Kp'$. Now consider a modal atom of the form $O\varphi$ in $MA(\Sigma)$, and suppose $M'' \models O\varphi$. Let (P, N) be the partition induced by M'' on $MA(\Sigma)$: from Lemma 14 it follows that $M'' \models O\varphi|_{P,N}$. Then, since $M'' \models Kp'$, from Definition 2 it follows that $\varphi|_{P,N} \supset p'$ is a valid objective sentence. Since p' does not occur in φ, $\varphi|_{P,N} \supset p'$ is valid iff $\varphi|_{P,N}$ is unsatisfiable, but, if we assume $\varphi|_{P,N}$ unsatisfiable, then by Definition 2 $M'' \models K\mathsf{false}$, thus contradicting the hypothesis that M'' be non-empty. Therefore, for each modal atom $O\varphi$ in $MA(\Sigma)$, $M'' \not\models O\varphi$, and since (w, M'') satisfies $\Sigma[O/\mathsf{false}]$, Lemma 14 implies that (w, M'') satisfies Σ. Since M'' is maximal, it follows that Σ is $\mathcal{OL}$-satisfiable;

(ii) there exists a partition (P, N) of $MA(\Sigma)$ such that $\Sigma|_{P,N}$ is satisfiable and $(P, N) = (P_{ob(P)}(\Sigma), N_{ob(P)}(\Sigma))$. Now, since $\Sigma|_{P,N}$ is satisfiable, there exists an interpretation satisfying $\Sigma|_{P,N}$. Let w be such an interpretation, and let $M = \{w' : w' \models ob(P)\}$. Since $(P, N) = (P_{ob(P)}(\Sigma), N_{ob(P)}(\Sigma))$, it follows that (P, N) is the partition of $MA(\Sigma)$ induced by M. Therefore, since w satisfies $\Sigma|_{P,N}$, from Lemma 14 it follows that $(w, M) \models \Sigma$. Moreover, M is maximal by construction, hence Σ is $\mathcal{OL}$-satisfiable.

Only-If part. Suppose Σ is $\mathcal{OL}$-satisfiable. Then, there exists a K45 model (w, M) such that $(w, M) \models \Sigma$ and M is maximal. Let (P, N) be the partition induced by M on $MA(\Sigma)$. There are two possible cases.

(i) All modal atoms of $MA(\Sigma)$ of the form $O\varphi$ belong to N. Then, from

Lemma 14, it follows that $(w, M) \models \Sigma[O/\mathsf{false}]$, i.e. $\Sigma[O/\mathsf{false}]$ is $\mathsf{K45}$-satisfiable. Now, there are two possibilities:

- $\Sigma[O/\mathsf{false}]$ is $\mathsf{KD45}$-satisfiable. In this case, condition (a) of the theorem holds;
- $\Sigma[O/\mathsf{false}]$ is not $\mathsf{KD45}$-satisfiable. In this case, from Lemma 16 it follows that $ob(P)$ is unsatisfiable. Moreover, M is empty, hence $M = \{w : w \models ob(P)\}$. Therefore, (P, N) coincides with the partition induced by $ob(P)$, that is, $(P, N) = (P_{ob(P)}(\Sigma), N_{ob(P)}(\Sigma))$. Furthermore, since $(w, M) \models \Sigma$ and (P, N) is the partition induced by M on $MA(\Sigma)$, from Lemma 14 it follows that the interpretation w satisfies $\Sigma|_{P,N}$, hence condition (b) of the theorem holds.

(ii) At least one modal atom of $MA(\Sigma)$ of the form $O\varphi$ belongs to P. Then, since (w, M) satisfies Σ, it follows that $M = \{w : w \models \varphi|_{P,N}\}$; moreover, by definition of $ob(P)$ it follows that $ob(P)$ is equivalent to $\varphi|_{P,N}$, thus (P, N) coincides with the partition induced by $ob(P)$, that is, $(P, N) = (P_{ob(P)}(\Sigma), N_{ob(P)}(\Sigma))$. Furthermore, since $(w, M) \models \Sigma$ and (P, N) is the partition induced by M on $MA(\Sigma)$, from Lemma 14 it follows that the interpretation w satisfies $\Sigma|_{P,N}$, hence condition (b) of the theorem holds. □

Intuitively, the above theorem provides for a characterization of the notion of $\mathcal{OL}$-satisfiability of a formula Σ in terms of properties of partitions of the modal atoms of Σ. Specifically, the theorem states that a formula $\Sigma \in \mathcal{L}_O$ is $\mathcal{OL}$-satisfiable iff either $\Sigma[O/\mathsf{false}]$ is $\mathsf{KD45}$-satisfiable, which informally corresponds to checking whether it is consistent to assume as false every O-sentence occurring in Σ, or there exists a partition (P, N) of $MA(\Sigma)$ such that $\Sigma|_{P,N}$ is satisfiable and $(P, N) = (P_{ob(P)}(\Sigma), N_{ob(P)}(\Sigma))$, which corresponds to checking whether there exists a guess of the modal atoms of Σ which is both consistent with Σ and not self-contradictory.

From the above theorem, it is easy to prove that the two notions of $\mathcal{OL}$-satisfiability and weak $\mathcal{OL}$-satisfiability coincide in the case of formulas from $\mathcal{L}_O$.

Theorem 19 *Let $\Sigma \in \mathcal{L}_O$. Then, Σ is $\mathcal{OL}$-satisfiable iff it is weakly $\mathcal{OL}$-satisfiable.*

Proof. Follows immediately from the fact that the proof of the only-if part of Theorem 18 holds even if the assumption that M is maximal is discarded, since such an assumption is not used in the proof. This in turn implies that Σ is weakly $\mathcal{OL}$-satisfiable iff conditions (a) and (b) of Theorem 18 hold. Thus, from the same theorem, it follows that Σ is weakly $\mathcal{OL}$-satisfiable iff Σ is $\mathcal{OL}$-satisfiable. □

A property analogous to the above theorem has been proved in [8] for the propositional fragment of $\mathcal{OL}$.

4 Reasoning in $\mathcal{OL}$

In this section we study reasoning in $\mathcal{OL}$. In particular, we first show that extending a decidable fragment of first-order logic with only knowing preserves decidability of reasoning. Then, we establish an upper bound for the satisfiability problem in the propositional fragment of $\mathcal{OL}$, and finally analyze a restriction of the propositional case in which reasoning is computationally easier.

We briefly introduce the complexity classes mentioned in the following (refer e.g. to [11] for further details). All the classes we use reside in the *polynomial hierarchy*. In particular, the complexity class Σ_2^p is the class of problems that are solved in polynomial time by a nondeterministic Turing machine that uses an NP-oracle (i.e., that solves in constant time any problem in NP), and Π_2^p is the class of problems that are complement of a problem in Σ_2^p. The class P^{NP}, also known as Δ_2^p, is the class of problems that are solved in polynomial time by a *deterministic* Turing machine that uses an NP-oracle, while the class $\mathrm{P}^{\mathrm{NP}}[O(\log n)]$, also known as Θ_2^p, is the class of problems that are solved in polynomial time by a deterministic Turing machine that makes a number of calls to an NP-oracle which is logarithmic in the size of the input. Hence, the class $\mathrm{P}^{\mathrm{NP}}[O(\log n)]$ is “mildly” harder than the class NP, since a problem in $\mathrm{P}^{\mathrm{NP}}[O(\log n)]$ can be solved by solving “few” (i.e. a logarithmic number of) instances of problems in NP. It is generally assumed that the polynomial hierarchy does not collapse, and that a problem in the class $\mathrm{P}^{\mathrm{NP}}[O(\log n)]$ is computationally easier than a Σ_2^p-hard or Π_2^p-hard problem.

4.1 Reasoning method

As for effective methods for reasoning in $\mathcal{OL}$, we recall that $\mathcal{OL}$-satisfiability in unrestricted $\mathcal{L}_O$ is not a decidable problem, since establishing $\mathcal{OL}$-satisfiability of objective sentences corresponds to solving the satisfiability problem for full first-order logic. However, the characterization provided by Theorem 18 allows for the definition of an algorithm for reasoning in subsets of $\mathcal{L}_O$ built upon decidable fragments of first-order logic. In the following, we say that a language $\mathcal{L}' \subseteq \mathcal{L}$ is *closed under boolean composition* if, for each $\varphi_1, \varphi_2 \in \mathcal{L}'$, $\varphi_1 \wedge \varphi_2 \in \mathcal{L}'$ and $\neg\varphi_1 \in \mathcal{L}'$. Moreover, we denote as $\mathcal{L}'_O$ the subset of $\mathcal{L}_O$ built upon $\mathcal{L}'$, i.e., the modal extension of $\mathcal{L}'$ obtained according to Definition 1.

To the aim of identifying decidable fragments of $\mathcal{L}_O$, we prove the following lemma.

Lemma 20 *Let $\Sigma \in \mathcal{L}_K$. Then, Σ is* KD45*-satisfiable iff there exists a partition (P, N) of $MA(\Sigma)$ such that:*

(a) $\Sigma|_{P,N}$ is satisfiable;
(b) for each $K\varphi \in N$, $ob(P) \wedge \neg\varphi|_{P,N}$ is satisfiable;
(c) $ob(P)$ is satisfiable.

Proof. *If part.* Let (P, N) be a partition of $MA(\Sigma)$ satisfying conditions (a), (b), and (c) of the theorem, and let $M = \{w : w \models ob(P)\}$. Condition (c) implies $M \neq \emptyset$. Moreover, since $\Sigma|_{P,N}$ is satisfiable, there exists an interpretation w such that $w \models \Sigma|_{P,N}$. Now we prove, by induction on the modal depth of the modal atoms in $MA(\Sigma)$, that $(P, N) = (P_{ob(P)}(\Sigma), N_{ob(P)}(\Sigma))$. First, from Lemma 15 and condition (b), it follows that each modal atom $K\varphi$ in N of modal depth 1 (i.e. such that $\varphi \in \mathcal{L}$) also belongs to $N_{ob(P)}(\Sigma)$; moreover, Definition 10 and Lemma 15 imply that each modal atom $K\varphi$ in P of modal depth 1 belongs to $P_{ob(P)}(\Sigma)$. Now suppose that (P, N) and $(P_{ob(P)}(\Sigma), N_{ob(P)}(\Sigma))$ agree on all modal atoms of modal depth less or equal to i. Consider a modal atom $K\varphi$ of $MA(\Sigma)$ of modal depth $i + 1$. Since by Definition 9 the value of the sentence $\varphi|_{P,N}$ only depends on the value of the modal atoms of modal depth less or equal to i in (P, N), by the induction hypothesis it follows that $\varphi|_{P,N} = \varphi|_{P_{ob(P)}(\Sigma), N_{ob(P)}(\Sigma)}$, hence condition (b) and Lemma 15 imply that, if $K\varphi \in N$, then $K\varphi \in N_{ob(P)}(\Sigma)$, while Definition 10 and Lemma 15 imply that, if $K\varphi \in P$, then $K\varphi \in P_{ob(P)}(\Sigma)$. Therefore, $(P, N) = (P_{ob(P)}(\Sigma), N_{ob(P)}(\Sigma))$, and since $w \models \Sigma|_{P,N}$ and $(P_{ob(P)}(\Sigma), N_{ob(P)}(\Sigma))$ is the partition of $MA(\Sigma)$ induced by M, from Lemma 14 it follows that $(w, M) \models \Sigma$, which proves that Σ is KD45-satisfiable.

Only-if part. Suppose Σ is KD45-satisfiable. Then, there exists a model (w, M) such that $(w, M) \models \Sigma$ and $M \neq \emptyset$. Let (P, N) be the partition of $MA(\Sigma)$ induced by M. Then, from Lemma 14 it follows that $w \models \Sigma|_{P,N}$, hence condition (a) holds. We now prove, by induction on the modal depth of the modal atoms in $MA(\Sigma)$, that $(P, N) = (P_{ob(P)}(\Sigma), N_{ob(P)}(\Sigma))$. Let $M' = \{w : w \models ob(P)\}$. First, by Definition 10 it follows that, for each $K\varphi \in MA(\Sigma)$ such that $\varphi \in \mathcal{L}$, $M \models K\varphi$ iff $M' \models K\varphi$, hence $K\varphi \in P$ iff $K\varphi \in P_{ob(P)}(\Sigma)$. Now suppose that (P, N) and $(P_{ob(P)}(\Sigma), N_{ob(P)}(\Sigma))$ agree on all modal atoms of modal depth less or equal to i. Consider a modal atom $K\varphi$ of $MA(\Sigma)$ of modal depth $i+1$. Since $K\varphi \in P$, by Definition 10 it follows that $ob(P) \supset \varphi|_{P,N}$ is a valid objective sentence, hence $M'' \models \varphi|_{P,N}$, and since by Definition 9 the value of the sentence $\varphi|_{P,N}$ only depends on the value of the modal atoms of modal depth less or equal to i in (P, N), by the induction hypothesis and Lemma 14 it follows that $M' \models \varphi$, hence $K\varphi \in P_{ob(P)}(\Sigma)$. Consequently,

$(P, N) = (P_{ob(P)}(\Sigma), N_{ob(P)}(\Sigma))$, and by Lemma 15 it follows that condition (b) holds. Finally, since $M \neq \emptyset$, it follows that $ob(P)$ is satisfiable, hence condition (c) holds. □

We are now ready to prove decidability of $\mathcal{OL}$-satisfiability for subsets of $\mathcal{L}_O$ built upon decidable subsets of the first-order language $\mathcal{L}$.

Theorem 21 *Let $\mathcal{L}' \subset \mathcal{L}$. If $\mathcal{L}'$ is closed under boolean composition and satisfiability in $\mathcal{L}'$ is decidable, then $\mathcal{OL}$-satisfiability in $\mathcal{L}'_O$ is decidable.*

Proof. Let $\Sigma \in \mathcal{L}'_O$. Theorem 18 implies that $\mathcal{OL}$-satisfiability of Σ can be decided through the following steps.

(i) First, checking KD45-satisfiability of $\Sigma[O/\mathsf{false}]$. From Lemma 20, this can be accomplished by verifying the existence of a partition (P, N) of $MA(\Sigma[O/\mathsf{false}])$ such that:
  (a) $\Sigma[O/\mathsf{false}]|_{P,N}$ is satisfiable. Since $\mathcal{L}'$ is closed under boolean composition, it follows that $\Sigma[O/\mathsf{false}]|_{P,N} \in \mathcal{L}'$, and since satisfiability in $\mathcal{L}'$ is decidable, this check is decidable;
  (b) for each $K\varphi \in N$, $ob(P) \wedge \neg\varphi|_{P,N}$ is satisfiable. Again, since $\mathcal{L}'$ is closed under boolean composition, it follows that, for each $K\varphi \in N$, $ob(P) \wedge \neg\varphi|_{P,N} \in \mathcal{L}'$, hence this check is decidable;
  (c) $ob(P)$ is satisfiable. Again, since $ob(P) \in \mathcal{L}'$, this check is decidable.

(ii) Verifying the existence of a partition (P, N) of $MA(\Sigma)$ such that $\Sigma|_{P,N}$ is satisfiable and $(P, N) = (P_{ob(P)}(\Sigma), N_{ob(P)}(\Sigma))$. Again, since $\mathcal{L}'$ is closed under boolean composition, $\Sigma|_{P,N} \in \mathcal{L}'$, hence verifying satisfiability of $\Sigma|_{P,N}$ is decidable. Moreover, since $\mathcal{L}'$ is closed under boolean composition, and since satisfiability in $\mathcal{L}'$ is decidable, Lemma 17 provides an effective method to build the partition $(P_{ob(P)}(\Sigma), N_{ob(P)}(\Sigma))$ in a finite amount of time, hence checking whether $(P_{ob(P)}(\Sigma), N_{ob(P)}(\Sigma))$ is equal to (P, N) is decidable. □

Therefore, the above theorem states that reasoning about only knowing in the modal extension (without quantifying-in) of a decidable fragment of first-order logic closed under boolean composition is decidable.

In Figure 1 we present the algorithm $\mathcal{OL}$-Sat for computing satisfiability in any fragment $\mathcal{L}'_O$ of $\mathcal{L}_O$ satisfying the conditions of Theorem 21. The algorithm is based on Theorem 18, and relies on both Lemma 20, which provides a method for computing KD45-satisfiability in $\mathcal{L}'_K$ by using a procedure for computing satisfiability in $\mathcal{L}'$, and Lemma 17, which provides a constructive way to build the partition $(P_\varphi(\Sigma), N_\varphi(\Sigma))$ starting from the sentences Σ and φ, again using a procedure for satisfiability in $\mathcal{L}'$. Therefore, the algorithm

Algorithm $\mathcal{OL}$-Sat(Σ)
Input: sentence $\Sigma \in \mathcal{L}'_O$;
Output: true if Σ is $\mathcal{OL}$-satisfiable, false otherwise.
begin
if $\Sigma[O/\mathsf{false}]$ is $\mathsf{KD45}$-satisfiable
then return true
else if there exists partition (P, N) of $MA(\Sigma)$ **such that**
(a) $\Sigma|_{P,N}$ is satisfiable **and**
(b) $(P, N) = (P_{ob(P)}(\Sigma), N_{ob(P)}(\Sigma))$
then return true
else return false
end

Fig. 1. Algorithm $\mathcal{OL}$-Sat.

computes $\mathcal{OL}$-satisfiability in $\mathcal{L}'_O$ by reducing such a problem to a number of satisfiability problems in $\mathcal{L}'$. Correctness of the algorithm follows immediately from Theorem 18.

Informally, the algorithm first checks whether it is possible to satisfy Σ by assuming as false all O-sentences occurring in Σ, that is, by making no closure assumptions about what is known. If in this way it is not possible to satisfy Σ, that is, the sentence $\Sigma[O/\mathsf{false}]$ is not $\mathsf{KD45}$-satisfiable, then the algorithm checks whether there exists a partition (P, N) of $MA(\Sigma)$ satisfying certain conditions. Intuitively, the partition must be consistent with Σ (condition (a)) and cannot be self-contradictory (condition (b)). In particular, the condition $(P, N) = (P_{ob(P)}(\Sigma), N_{ob(P)}(\Sigma))$ establishes that the objective knowledge implied by the partition (P, N) (that is, the sentence $ob(P)$) identifies a set of interpretations which induces the same partition (P, N) on $MA(\Sigma)$. We illustrate the algorithm through the following simple example.

Example 22 *Let us consider the sentence Σ defined in Example 11. Then, $\Sigma[O/\mathsf{false}] = a \wedge \mathsf{false} = \mathsf{false}$, hence $\Sigma[O/\mathsf{false}]$ is not $\mathsf{KD45}$-satisfiable. Now, consider the partition $(P, N) = (\{O(\neg a \vee Kb)\}, \{Kb\})$ of $MA(\Sigma)$. As shown in Example 11, $\Sigma|_{P,N} = a$, hence (P, N) satisfies condition (a) of the algorithm. Now, since $ob(P) = \neg a$, and $(\neg a \vee Kb)|_{P,N} = \neg a$, it follows that $O(\neg a \vee Kb) \in P_{ob(P)}(\Sigma)$. Moreover, the objective sentence $\neg a \supset b$ is not valid, hence $Kb \in N_{ob(P)}(\Sigma)$. Therefore, $(P, N) = (P_{ob(P)}(\Sigma), N_{ob(P)}(\Sigma))$, i.e. condition (b) of the algorithm is satisfied. Consequently, $\mathcal{OL}$-Sat(Σ) returns true. In fact, the partition (P, N) identifies the set of interpretations $M = \{w : w \models \neg a\}$. Moreover, since $\Sigma|_{P,N}$ is satisfiable, it follows that there exists an interpretation w satisfying $\Sigma|_{P,N}$, which implies that $(w, M) \models \Sigma$.*

4.2 Propositional case: complexity

We now study the complexity of reasoning about only knowing in the propositional case. To this aim, we analyze the complexity of the algorithm $\mathcal{OL}$-Sat, reported in Figure 1, under the restriction that Σ is a modal propositional formula. To the best of our knowledge, such an algorithm is the first terminating method for reasoning about only knowing in the full propositional case. [3] In the following, we denote as $\mathcal{L}^p$ the propositional fragment of $\mathcal{L}$, and with $\mathcal{L}_O^p$ the propositional fragment of $\mathcal{L}_O$.

Observe that, if $\Sigma \in \mathcal{L}_O^p$, then all formulas involved in the conditions reported in the algorithm are propositional, hence all such conditions can be checked by solving propositional satisfiability/validity problems. In particular:

- satisfiability in propositional KD45 can be computed in nondeterministic polynomial time, since such a problem is NP-complete [10]. Membership in NP is also an immediate consequence of Lemma 20;
- condition (a) can be checked in time linear with respect to the size of Σ;
- given (P, N), the formula $ob(P)$ can be computed in time linear with respect to the size of P. Moreover, by Lemma 17 it follows that, since $MA(\Sigma)$ has size linear with respect to the size of Σ, construction of the partition $(P_{ob(P)}(\Sigma), N_{ob(P)}(\Sigma))$ can be performed through a linear number (with respect to the size of Σ) of calls to an NP-oracle for propositional satisfiability. Therefore, condition (b) can be checked in linear time (with respect to the size of Σ) using an NP-oracle.

Now, since the guess of the partition (P, N) of $MA(\Sigma)$ requires a nondeterministic choice, it follows that the algorithm $\mathcal{OL}$-Sat, if considered as a nondeterministic procedure, is able to establish satisfiability of a formula $\Sigma \in \mathcal{L}_O^p$ in nondeterministic polynomial time (with respect to the size of Σ), using an NP-oracle for propositional satisfiability. Thus, we obtain an upper bound of Σ_2^p for the problem.

Lemma 23 *Let $\Sigma \in \mathcal{L}_O^p$. The problem of establishing $\mathcal{OL}$-satisfiability of Σ is in Σ_2^p.*

As for the lower bound of the satisfiability problem in propositional $\mathcal{OL}$, we recall that reasoning in Moore's autoepistemic logic (AEL) [21] can be reduced to reasoning in $\mathcal{OL}$. We now briefly recall the notion of stable expansion in AEL. In order to keep notation to a minimum, we change the language of *AEL*, using the modality K instead of L: thus, in the following a formula of

[3] In fact, Levesque's axiomatization of the propositional fragment of $\mathcal{OL}$ [18] does not directly imply the existence of a terminating procedure for reasoning in the propositional fragment of $\mathcal{OL}$.

AEL is a formula belonging to the propositional fragment of $\mathcal{L}_K$ (denoted as $\mathcal{L}_K^p$).

A set of formulas T from $\mathcal{L}_K^p$ is a *stable expansion* for a formula $\Sigma \in \mathcal{L}_K^p$ if T satisfies the equation

$$T = Cn(\{\Sigma\} \cup \{K\varphi : \varphi \in T\} \cup \{\neg K\varphi : \varphi \notin T\})$$

where Cn is the logical consequence operator of propositional logic.

Proposition 24 *[18, Theorem 3.9] Let $\varphi \in \mathcal{L}_K^p$. Then, $O\varphi$ is $\mathcal{OL}$-satisfiable iff there exists a stable expansion for φ.*

Lemma 25 *Let $\Sigma \in \mathcal{L}_O^p$. The problem of establishing $\mathcal{OL}$-satisfiability of Σ is Σ_2^p-hard.*

Proof. Let $\varphi \in \mathcal{L}_K^p$. By Proposition 24, $O\varphi$ is $\mathcal{OL}$-satisfiable iff there exists a stable expansion for φ. And since the problem of establishing whether a formula $\varphi \in \mathcal{L}_K^p$ admits a stable expansion is Σ_2^p-hard [5, Theorem 4.3], this proves the thesis. $\square$

The last two lemmas imply the following property.

Theorem 26 *Let $\Sigma \in \mathcal{L}_O^p$. The problem of establishing $\mathcal{OL}$-satisfiability of Σ is Σ_2^p-complete.*

The previous theorem implies that validity in propositional $\mathcal{OL}$ is Π_2^p-complete, and that the logical implication problem $\varphi \models_{\mathcal{OL}} \psi$ is Π_2^p-complete as well (with respect to the size of $\varphi \wedge \psi$). Consequently, the algorithm $\mathcal{OL}$-Sat is "optimal" with respect to the complexity of satisfiability in propositional $\mathcal{OL}$, in the sense that it matches the lower bound of the problem.

4.3 Propositional case: restrictions

We now define an interesting subset of propositional $\mathcal{OL}$ in which the modality O can be used in a restricted way. We prove that reasoning in such a fragment of $\mathcal{OL}$ is easier than in the general propositional case.

First, from Lemma 25 it follows that if we impose no restrictions on formulas which lie within the scope of the operator O, then $\mathcal{OL}$-satisfiability is a Σ_2^p-hard problem. Hence, in order to find a fragment of $\mathcal{L}_O^p$ for which satisfiability is computationally easier, we need to impose some restrictions on the structure of O-subformulas.

The first significant restriction corresponds to the case of formulas of the form $O\varphi \wedge \neg K\psi$ in which $\varphi \in \mathcal{L}^p$, $\psi \in \mathcal{L}^p_K$. Satisfiability of this kind of formulas in $\mathcal{OL}$ is analogous to a reasoning problem which has been analyzed in several different settings (e.g. [18,23,5]), and corresponds to posing an epistemic query $\psi \in \mathcal{L}^p_K$ to the propositional knowledge base φ, interpreting queries under the following intuitive epistemic closure assumption:

- for any $\xi \in \mathcal{L}^p$, if ξ is logically implied (in propositional logic) by φ then $K\xi$ is implied by φ, otherwise $\neg K\xi$ is implied by φ;
- the interpretation of an epistemic query ψ with nested occurrences of the modal operator is obtained by iteratively checking all modal subformulas $K\xi$ such that $\xi \in \mathcal{L}^p$, then substituting all such subformulas with true or false in ψ accordingly, thus obtaining new modal subformulas in ψ without nested occurrences of the modality; when all modal subformulas in ψ have been replaced in this way, it can be checked whether ψ is implied by φ.

It can be shown (see [18, Corollary 3.13]) that $O\varphi \wedge \neg K\psi$ is satisfiable if and only if ψ is not implied by φ under the above semantics for epistemic queries. Moreover, it is known that the problem is $\mathrm{P}^{\mathrm{NP}}[O(\log n)]$-complete [6], that is, it can be solved in polynomial time through a number of calls to the NP-oracle which is logarithmic in the size of the formula $\varphi \wedge \psi$. Therefore, satisfiability in $\mathcal{OL}$ of a formula of the form $O\varphi \wedge \neg K\psi$ in which $\varphi \in \mathcal{L}^p$, $\psi \in \mathcal{L}^p_K$, is $\mathrm{P}^{\mathrm{NP}}[O(\log n)]$-complete as well, hence it is easier than the problem of $\mathcal{OL}$-satisfiability of a generic formula in $\mathcal{L}^p_O$.

We now define a large superclass of the above set of formulas, and show that satisfiability in $\mathcal{OL}$ for such kind of formulas is still easier than in the general case.

Definition 27 *Let $\mathcal{L}^S_O$ denote the set of formulas belonging to $\mathcal{L}^p_O$ in which each propositional symbol lies within the scope of a modality. Then, we denote with $\mathcal{L}^-_O$ the set of formulas belonging to $\mathcal{L}^p_O$ in which each O-subformula $O\varphi$ is such that φ is of the form $f \wedge \psi$ or $f \vee \psi$, with $f \in \mathcal{L}^p$, $\psi \in \mathcal{L}^S_O$.*

Notice that the only restriction imposed by the above definition is on the form of O-subformulas: roughly speaking, in each O-subformula φ it must be possible to isolate an "objective" (i.e. belonging to $\mathcal{L}^p$) and a "subjective" (i.e. belonging to $\mathcal{L}^S_O$) subformula. For instance, the formula $\Sigma = a \wedge O(\neg a \vee Kb)$ defined in Example 11 belongs to the set $\mathcal{L}^-_O$, since the O-subformula $O(\neg a \vee Kb)$ has an objective subformula $\neg a$ and a subjective subformula Kb. Conversely, the formula $a \wedge O(\neg a \vee (Kb \wedge c))$ does not belong to the set $\mathcal{L}^-_O$, since the subformula $Kb \wedge c$ is neither objective nor subjective.

The language $\mathcal{L}^-_O$ allows for a nice formalization of a generalization of the above mentioned setting of epistemic queries, in which one can express queries regarding the epistemic state of a number (say n) of propositional knowledge bases. A multimodal language with n operators $K_1, \ldots, K_n$ can be used for expressing the epistemic state of each of the knowledge bases. Given a set of n propositional knowledge bases $\mathcal{K} = \{KB_1, \ldots, KB_n\}$, in which each KB_i is a formula from $\mathcal{L}^p$, we define an epistemic query to $\mathcal{K}$ as a boolean combination of epistemic queries to the single knowledge bases. E.g., we can pose a query of the form

$$Q = K_1\varphi \wedge (\neg K_2\psi \vee K_3\xi)$$

such that φ is an epistemic query to KB_1 (i.e. in which the only modality K_1 is used), ψ is an epistemic query to KB_2, and ξ is an epistemic query to KB_3. Q is implied by $\mathcal{K}$ if and only if φ is implied by KB_1 and either ψ is not implied by KB_2 or ξ is implied by KB_3.

It is immediate to see that the evaluation of such forms of epistemic queries can be reduced to checking validity of formulas in $\mathcal{L}_O^-$. In the case of the above example, Q is implied by $\mathcal{K}$ iff the formula

$$(OKB_1 \supset K\varphi') \wedge ((OKB_2 \supset \neg K\psi') \vee (OKB_3 \supset K\xi'))$$

is $\mathcal{OL}$-valid, where φ' is obtained from φ by substituting each occurrence of K_1 with K, and ψ', ξ' are obtained in a similar way from ψ and ξ.

We now prove that $\mathcal{OL}$-satisfiability for a formula Σ belonging to $\mathcal{L}_O^-$ is easier than for generic formulas in $\mathcal{L}_O^p$. Informally, the key point is that the syntactic restriction satisfied by a formula in $\mathcal{L}_O^-$ allows for easily identifying a "small" (i.e., linear in the size of Σ) number of possible sets of propositional interpretations, each one represented in terms of a propositional formula. In particular, given $\Sigma \in \mathcal{L}_O^-$, there is a finite number (say n) of occurrences of O-subformulas in Σ, and each of such formulas is of the form $f_i \wedge \psi_i$ or $f_i \vee \psi_i$, with $f_i \in \mathcal{L}^p$ and $\psi_i \in \mathcal{L}_O^S$, for $i = 1, \ldots, n$: it is then possible to show that a model (w, M) for Σ *must* be such that M is one of the maximal sets of interpretations represented by one of the f_i's (plus the formulas true, false). This property simplifies the problem of finding a model for Σ, since in this case the search can be restricted to a linear number of candidate sets of interpretations, while in the general case there is an exponential number of such candidate sets (represented in the algorithm $\mathcal{OL}$-Sat by all the possible partitions of the modal atoms of Σ).

In Figure 2 we present the algorithm $\mathcal{L}_O^-$-Sat for computing $\mathcal{OL}$-satisfiability of formulas in $\mathcal{L}_O^-$. In the algorithm, we assume without loss of generality that the set $MA(\Sigma)$ contains $n \geq 0$ modal atoms prefixed by the operator O, of the form $O(f_i \wedge \psi_i)$ or $O(f_i \vee \psi_i)$, for $i = 1, \ldots, n$, with $f_i \in \mathcal{L}^p$, $\psi_i \in \mathcal{L}_O^S$.

Example 28 *Let us again consider the formula* $\Sigma = a \wedge O(\neg a \vee Kb)$ *defined in Example 11. As shown before,* $\Sigma[O/\mathsf{false}]$ *is not* $\mathsf{KD45}$*-satisfiable. Moreover,* $S_\Sigma = \{\neg a, \mathsf{true}, \mathsf{false}\}$. *Now let* $\varphi = \neg a$. *As shown before,* $P_\varphi(\Sigma) = \{O(\neg a \vee$

Algorithm $\mathcal{L}_O^-$-Sat(Σ)
Input: formula $\Sigma \in \mathcal{L}_O^-$;
Output: true if Σ is $\mathcal{OL}$-satisfiable, false otherwise.
begin
if $\Sigma[O/\mathsf{false}]$ is $\mathsf{KD45}$-satisfiable
then return true
else if there exists φ **in** $S_\Sigma = \{f_1, f_2, \ldots, f_n, \mathsf{true}, \mathsf{false}\}$ **such that**
$\Sigma|_{P_\varphi(\Sigma),N_\varphi(\Sigma)}$ is satisfiable
then return true
else return false
end

Fig. 2. Algorithm $\mathcal{L}_O^-$-Sat.

$Kb)\}$, $N_\varphi(\Sigma) = \{Kb\}$, and $\Sigma|_{P_\varphi(\Sigma),N_\varphi(\Sigma)} = a \wedge \mathsf{true}$ is satisfiable. Therefore, $\mathcal{L}_O^-$-Sat(Σ) returns true.

Correctness of the algorithm is established by the following theorem.

Theorem 29 *Let $\Sigma \in \mathcal{L}_O^-$. Then, Σ is $\mathcal{OL}$-satisfiable iff $\mathcal{L}_O^-$-Sat(Σ) returns* true.

Proof. *If part.* Suppose $\mathcal{L}_O^-$-Sat(Σ) returns true. Then, there are two possible cases:

(i) $\Sigma[O/\mathsf{false}]$ is $\mathsf{KD45}$-satisfiable. As shown in the proof of Theorem 18, this implies that Σ is $\mathcal{OL}$-satisfiable;
(ii) there exists a formula φ in the set $S_\Sigma = \{f_1, \ldots, f_n, \mathsf{true}, \mathsf{false}\}$ such that $\Sigma|_{P_\varphi(\Sigma),N_\varphi(\Sigma)}$ is satisfiable. Now let $M = \{w' : w' \models \varphi\}$; moreover, let w be an interpretation satisfying the satisfiable propositional formula $\Sigma|_{P_\varphi(\Sigma),N_\varphi(\Sigma)}$. From the definition of $(P_\varphi(\Sigma), N_\varphi(\Sigma))$ it follows that each modal atom in $MA(\Sigma)$ is satisfied by (w, M) iff it belongs to $P_\varphi(\Sigma)$. Therefore, (w, M) satisfies Σ, and since M is maximal by construction, it follows that Σ is $\mathcal{OL}$-satisfiable.

Only-If part. Suppose Σ is $\mathcal{OL}$-satisfiable. Then, there is a model (w, M) satisfying Σ and such that M is maximal. Let (P, N) be the partition induced by M on $MA(\Sigma)$. Then, there are two possible cases.

(i) There is no modal atom of the form $O\varphi$ in P. Then, $\Sigma[O/\mathsf{false}] \in \mathcal{L}_K^p$ is $\mathsf{K45}$-satisfiable, since (w, M) satisfies it. Now, there are two possibilities:
   - $\Sigma[O/\mathsf{false}]$ is $\mathsf{KD45}$-satisfiable. In this case, the algorithm $\mathcal{L}_O^-$-Sat(Σ) returns true;
   - $\Sigma[O/\mathsf{false}]$ is not $\mathsf{KD45}$-satisfiable. In this case, from Lemma 16 it follows that $ob(P)$ is unsatisfiable, hence $ob(P)$ is equivalent to false. Moreover,

M is empty, hence $M = \{w : w \models \mathsf{false}\}$. Therefore, (P, N) coincides with the partition induced by false, that is,

$$(P, N) = (P_{\mathsf{false}}(\Sigma), N_{\mathsf{false}}(\Sigma))$$

Furthermore, since $(w, M) \models \Sigma$ and (P, N) is the partition induced by M on $MA(\Sigma)$, from Lemma 14 it follows that the interpretation w satisfies $\Sigma|_{P,N}$, hence the algorithm $\mathcal{L}_O^-$-Sat(Σ) returns true.

(ii) There exists $O\varphi \in P$. Then, since (w, M) satisfies Σ and M is maximal, it follows that $M = \{w' : w' \models \varphi|_{P,N}\}$. Now, since by hypothesis $\Sigma \in \mathcal{L}_O^-$, $O\varphi$ is of the form $f_i \wedge \psi$ or $f_i \vee \psi$, with $1 \leq i \leq n$ and $\psi \in \mathcal{L}_O^S$, therefore $\psi|_{P,N}$ is equivalent either to true or to false, which implies that $\varphi|_{P,N}$ is equivalent to one of the formulas in the set $S_\Sigma = \{f_1, \ldots, f_n, \mathsf{true}, \mathsf{false}\}$. Consequently, (P, N) is induced by one of the formulas in S_Σ, and since the formula $\Sigma|_{P,N}$ is satisfiable, it follows that the algorithm $\mathcal{L}_O^-$-Sat(Σ) returns true. □

We now analyze the complexity of the algorithm $\mathcal{L}_O^-$-Sat reported in Figure 2. As noticed above, the partition $(P_\varphi(\Sigma), N_\varphi(\Sigma))$ can be computed through a linear number (with respect to the size of Σ) of calls to an NP-oracle for propositional satisfiability. Now, since the cardinality of $MA(\Sigma)$ (and hence the number of formulas in the set S_Σ) is also linearly bounded by the size of Σ, it follows that the algorithm $\mathcal{L}_O^-$-Sat is able to establish $\mathcal{OL}$-satisfiability of a formula $\Sigma \in \mathcal{L}_O^-$ in *deterministic* polynomial time (in the size of Σ), using an NP-oracle for propositional satisfiability.

More precisely, it can be shown that the problem of $\mathcal{OL}$-satisfiability of a formula Σ in $\mathcal{L}_O^-$ is $\mathrm{P}^{\mathrm{NP}}[O(\log n)]$-complete, namely it can be computed in polynomial time by a *logarithmic* (in the size of Σ) number of calls to an NP-oracle. To this aim, we recall the decision problem TREES(SAT) and the notion of *NP-tree* [6]. An NP-tree is a triple $\langle Var, G, R\rangle$ in which:

- Var is a set of propositional variables $v_1, \ldots, v_n$ (called the *linking* variables);
- $G = (V, E)$ is a directed tree, with edges directed from the leaves to the root. Each element of the set of nodes $V = \{F_1, \ldots, F_n\}$ contains a propositional formula F_i built upon a set of private propositional symbols (i.e., symbols which do not appear in any other node) and the linking variables v_j such that $(F_j, F_i) \in E$;
- F_r is a distinguished node, called terminal node.

The truth assignment σ to the propositional variables in an NP-tree is defined as follows: $\sigma(v_i) = \mathit{TRUE}$ iff the formula F_i' is satisfiable, where F_i' stands for for the formula obtained from F_i by replacing each propositional linking variable v_j occurring in F_i with true if $\sigma(v_j) = \mathit{TRUE}$, and with false if $\sigma(v_j) = \mathit{FALSE}$. The result value of an NP-tree is the value $\sigma(v_r)$.

The decision problem TREES(SAT) is the problem of establishing, given an NP-tree T, whether the result value of T is TRUE. It has been shown [6, Theorem 4.5] that TREES(SAT) is $\mathrm{P}^{\mathrm{NP}}[O(\log n)]$-complete.

Theorem 30 *Let $\Sigma \in \mathcal{L}_O^-$. Then, the problem of establishing $\mathcal{OL}$-satisfiability of Σ is $\mathrm{P}^{\mathrm{NP}}[O(\log n)]$-complete.*

Proof. Hardness follows from the aforementioned fact that satisfiability in $\mathcal{OL}$ of a formula of the form $O\varphi \wedge \neg K\psi$ such that $\varphi \in \mathcal{L}^p$, $\psi \in \mathcal{L}_K^p$, corresponds to verify whether the epistemic query ψ is not implied by φ. In turn, this last problem corresponds (see [6]) to check non-membership of the formula $K\psi$ in the stable set identified by the formula φ, which is a $\mathrm{P}^{\mathrm{NP}}[O(\log n)]$-complete problem [6, Theorem 5.3.6].

As for membership in $\mathrm{P}^{\mathrm{NP}}[O(\log n)]$, we show that the condition expressed in the innermost **if-then-else** statement in the algorithm $\mathcal{L}_O^-$-Sat, namely the existence of a formula φ among $\{f_1, f_2, \ldots, f_n, \mathsf{true}, \mathsf{false}\}$ such that $\Sigma|_{P_\varphi(\Sigma), N_\varphi(\Sigma)}$ is satisfiable, can be computed in $\mathrm{P}^{\mathrm{NP}}[O(\log n)]$. To this aim, we reduce, in time polynomial in the size of Σ, the problem of checking whether such a statement returns true to TREES(SAT). We construct the NP-tree $T(\Sigma)$ as follows. [4] First, let $m = n + 2$, $f_{n+1} = \mathsf{true}$, $f_{n+2} = \mathsf{false}$. We start from the following tree:

- $F_r = v_1 \vee v_2 \vee \ldots \vee v_m$;
- $F_i = \Sigma$, for $i = 1, \ldots, m$;
- $E = \{(F_i, F_r) : i = 1, \ldots, m\}$.

Then, we obtain $T(\Sigma)$ by expanding each node F_i ($i = 1, \ldots, m$) of the above tree as follows. Let F be the node F_i or any successor node of F_i. Now:

- For each strict occurrence of a formula $K\varphi$ in F, create a new node $F_j = \neg(f_i \supset \varphi)$, replace such an occurrence of $K\varphi$ in F with the linking variable v_j, and add the edge (F_j, F) to E;
- for each strict occurrence of a formula $O\varphi$ in F, create a new node $F_j = \neg(f_i \equiv \varphi)$, replace such an occurrence of $O\varphi$ in F with the linking variable v_j, and add the edge (F_j, F) to E.

We repeat the above expansion until there are no nodes in $T(\Sigma)$ which contain modal operators. Now, let v_j be a linking variable occurring in a node

[4] The construction trivially extends the technique employed in [6] in the case of Carnap's modal logic.

F_i or in a successor node of F_i, and let φ be the modal atom that generates v_j in the above construction. From Lemma 15 it immediately follows that $\sigma(v_j) = \mathit{TRUE}$ iff $\varphi \in P_{f_i}(\Sigma)$. Consequently, for each node F_i such that $1 \leq i \leq m$, F_i is satisfiable iff $\Sigma|_{P_\varphi(\Sigma),N_\varphi(\Sigma)}$ is satisfiable. And since $F_r = v_1 \vee \ldots \vee v_m$, it follows that the result value of $T(\Sigma)$ is TRUE iff there exists $\varphi \in \{f_1, \ldots, f_m\}$ such that $\Sigma|_{P_\varphi(\Sigma),N_\varphi(\Sigma)}$ is satisfiable. Moreover, it is immediate to verify that $T(\Sigma)$ can be constructed in time polynomial in the size of Σ. Therefore, the condition expressed by the innermost **if-then-else** statement of the algorithm $\mathcal{L}_O^-$-Sat can be computed in $\mathrm{P}^{\mathrm{NP}}[O(\log n)]$. And since satisfiability in propositional KD45 can be computed in nondeterministic polynomial time, Theorem 29 implies that the satisfiability problem for formulas in $\mathcal{L}_O^-$ is in $\mathrm{P}^{\mathrm{NP}}[O(\log n)]$. □

5 Only knowing vs. minimal knowledge

As mentioned in the introduction, only knowing is strictly related to the *minimal knowledge* principle. We now compare these two notions from the computational viewpoint.

The principle of minimal knowledge is a very general notion which can be phrased as follows: "In each possible epistemic state, the agent has *minimal objective knowledge*", that is, the agent has as much ignorance as possible about the current state of the world. As a consequence, there exists no epistemic state whose objective knowledge logically implies the objective knowledge of another epistemic state.

There are several proposals in the literature based on the minimal knowledge paradigm (see e.g. [9,19,12,26]). Among them, the first attempt is due to Halpern and Moses [9] and is the most similar to the notion of only knowing. Informally, Halpern and Moses apply minimal knowledge to modal logic S5: thus, they define a preference semantics [27] over S5, by considering as intended models of a modal theory Σ only those S5 models satisfying Σ in which the set of possible worlds is maximal with respect to set containment. Hence, in this case the notion of maximization lies at the semantic level.

Recently, it has been proven [2] that reasoning in Halpern and Moses' version of S5 (also known as ground nonmonotonic modal logic $\mathsf{S5}_G$) lies at the third level of the polynomial hierarchy. In particular, logical implication in $\mathsf{S5}_G$ is a Π_3^p-complete problem. Moreover, many other formalisms based on the minimal knowledge paradigm share the same computational properties of $\mathsf{S5}_G$ [3,2,24]. Hence, we can conclude that (unless the polynomial hierarchy collapses) minimal knowledge is computationally harder than only knowing. In particular, $\mathsf{S5}_G$ cannot be "polynomially embedded" into $\mathcal{OL}$. This is a surprising result, since the logic of only knowing is generally considered as a very expressive formalism, due to its powerful ability of explicitly expressing minimization of knowledge.

On the other hand, it can be shown that the reason why $\mathsf{S5}_G$ (and more generally all logics based on $\mathsf{S5}_G$) is computationally harder than $\mathcal{OL}$ (and all major propositional nonmonotonic formalisms) is that $\mathsf{S5}_G$ allows for expressing minimal knowledge states in a more compact form than $\mathcal{OL}$. See [25] for a detailed study of the epistemological properties of $\mathsf{S5}_G$.

6 Conclusions

In this paper we have defined a general method for reasoning about only knowing based on deduction techniques developed for nonmonotonic modal logics, which proves the strict similarity between these logics and Levesque's monotonic formalism. Based on such a reasoning method, we have investigated the computational properties of the propositional fragment of Levesque's modal logic of only knowing. Our analysis shows that the problem of reasoning about only knowing in the propositional case lies at the second level of the polynomial hierarchy, just like reasoning in most of the propositional formalisms for nonmonotonic reasoning. We have also studied syntactic restrictions in which reasoning about only knowing is easier than in the general case, and have shown the connections between such a restricted setting and the framework of epistemic queries to "classical" knowledge bases [23].

We remark that a computational analysis of reasoning about only knowing is interesting not only from a theoretical perspective, but also for the development of automated reasoning procedures in the setting of reasoning about actions, where the logic of only knowing has been recently applied [14,16].

One further development of the present work is towards the analysis of reasoning about only knowing in the presence of quantifying-in: in particular, it should be interesting to see whether it is possible to extend the techniques presented here for fragments of such a more expressive case. This analysis may also take advantage of recent results on reasoning with quantifying-in in standard modal logics [28].

Furthermore, the problem of embedding only knowing into nonmonotonic formalisms (as autoepistemic logic or the logic $\mathsf{S5}_G$) is very interesting from the theoretical viewpoint, in order to establish further relationships between reasoning about only knowing and other forms of nonmonotonic reasoning.

Acknowledgements

We gratefully thank Daniele Nardi, for his comments on an earlier version of this paper. We are also thankful to the anonymous referees, for their invaluable remarks and observations.

References

[1] J. Chen. The logic of only knowing as a unified framework for non-monotonic reasoning. *Fundamenta Informaticae*, 21:205–220, 1994.

[2] F. M. Donini, D. Nardi, and R. Rosati. Ground nonmonotonic modal logics. *J. of Logic and Computation*, 7(4):523–548, Aug. 1997.

[3] T. Eiter and G. Gottlob. Reasoning with parsimonious and moderately grounded expansions. *Fundamenta Informaticae*, 17(1,2):31–54, 1992.

[4] T. Eiter and G. Gottlob. Propositional circumscription and extended closed world reasoning are Π_2^p-complete. *Theoretical Computer Science*, 114:231–245, 1993.

[5] G. Gottlob. Complexity results for nonmonotonic logics. *J. of Logic and Computation*, 2:397–425, 1992.

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

[7] J. Y. Halpern and G. Lakemeyer. Levesque's axiomatization of only knowing is incomplete. *Artificial Intelligence*, 74(2):381–387, 1995.

[8] J. Y. Halpern and G. Lakemeyer. Multi-agent only knowing. In *Proc. of the 6th Conf. on Theoretical Aspects of Reasoning about Knowledge (TARK'96)*. Morgan Kaufmann, 1996.

[9] J. Y. Halpern and Y. Moses. Towards a theory of knowledge and ignorance: Preliminary report. In K. Apt, editor, *Logic and models of concurrent systems*. Springer-Verlag, 1985.

[10] 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.

[11] D. S. Johnson. A catalog of complexity classes. In J. van Leuven, editor, *Handbook of Theoretical Computer Science*, volume A, chapter 2. Elsevier Science Publishers (North-Holland), Amsterdam, 1990.

[12] M. Kaminski. Embedding a default system into nonmonotonic logics. *Fundamenta Informaticae*, 14:345–354, 1991.

[13] G. Lakemeyer. Limited reasoning in first-order knowledge bases with full introspection. *Artificial Intelligence*, 84:209–255, 1996.

[14] G. Lakemeyer. Only knowing in the situation calculus. In L. C. Aiello, J. Doyle, and S. C. Shapiro, editors, *Proc. of the 5th Int. Conf. on the Principles of Knowledge Representation and Reasoning (KR'96)*, pages 14–25. Morgan Kaufmann, Los Altos, 1996.

[15] G. Lakemeyer and H. J. Levesque. A tractable knowledge representation service with full introspection. In *Proc. of the 2nd Conf. on Theoretical Aspects of Reasoning about Knowledge (TARK'88)*, pages 145–159. Morgan Kaufmann, Los Altos, 1988.

[16] G. Lakemeyer and H. J. Levesque. AOL: a logic of acting, sensing, knowing, and only knowing. In *Proc. of the 6th Int. Conf. on the Principles of Knowledge Representation and Reasoning (KR'98)*, pages 316–327. Morgan Kaufmann, Los Altos, 1998.

[17] H. J. Levesque. Foundations of a functional approach to knowledge representation. *Artificial Intelligence*, 23:155–212, 1984.

[18] H. J. Levesque. All I know: a study in autoepistemic logic. *Artificial Intelligence*, 42:263–310, 1990.

[19] V. Lifschitz. Minimal belief and negation as failure. *Artificial Intelligence*, 70:53–72, 1994.

[20] W. Marek and M. Truszczyński. *Nonmonotonic Logics – Context-Dependent Reasoning*. Springer-Verlag, 1993.

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

[22] I. Niemelä. On the decidability and complexity of autoepistemic reasoning. *Fundamenta Informaticae*, 17(1,2):117–156, 1992.

[23] R. Reiter. What should a database know? *J. of Logic Programming*, 14:127–153, 1990.

[24] R. Rosati. Reasoning with minimal belief and negation as failure: Algorithms and complexity. In *Proc. of the 14th Nat. Conf. on Artificial Intelligence (AAAI'97)*, pages 430–435. AAAI Press/The MIT Press, 1997.

[25] R. Rosati. Reasoning about minimal knowledge in nonmonotonic modal logics. *J. of Logic, Language and Information*, 8(2):187–203, 1999.

[26] G. Schwarz and M. Truszczyński. Minimal knowledge problem: a new approach. *Artificial Intelligence*, 67:113–141, 1994.

[27] Y. Shoham. Nonmonotonic logics: Meaning and utility. In *Proc. of the 10th Int. Joint Conf. on Artificial Intelligence (IJCAI'87)*, pages 388–392, 1987.

[28] F. Wolter and M. Zakharyaschev. Decidable fragments of first-order modal logics. Available at `www.informatik.uni-leipzig.de/~wolter`, 1999.