Approximate Reasoning and non-Omniscient Agents

Marco Cadoli, Marco Schaerf
Dipartimento di Informatica e Sistemistica
Università di Roma "La Sapienza"
via Salaria 113, 00198 Roma, Italia
cadoli@vaxrma.roma1.infn.it   mschaerf@vaxrma.roma1.infn.it

Abstract

Modal logics have been frequently used to represent the knowledge and belief owned by an agent. However, such systems make the unrealistic assumption that agents are logically omniscent, hence capable of performing extremely complex inferences. Our goal in this paper is twofold. First of all we show that an approximation pattern can be used for a stepwise procedure which determines the satisfiability of a formula in several modal systems like $\mathcal{S}5$, $\mathcal{K}$, $\mathcal{T}$ and $\mathcal{S}4$. This method is based on a generalization of the standard possible-world semantics for modal logic. Secondly, we use this semantics to define a modal language which allows the explicit representation of the knowledge owned by a resource-limited agent.

1 Introduction

It is well known that typical reasoning problems in logic are computationally intractable. This is true in particular for modal logics for knowledge and belief [14,16]. As a consequence, when we represent agents through modal theories, we make the unrealistic assumption that they are capable of performing extremely complex inferences. This drawback is known as the *logical omniscence problem* (see [10]).

The most common technique for achieving tractability in a logical problem is *language restriction*: The expressiveness of the representational language is limited so that problems are computationally affordable (see [1,5,11] for examples of this technique). An alternative idea, which has been recently investigated, is the idea of *approximating* the answer to a logical problem, see for example [2,8,12,17]. Actually approximation techniques are widely used in many areas of Computer Science for dealing with polynomially intractable problems. For example, although finding an optimal solution to an Integer Programming problem is NP-complete, it is possible to provide sub-optimal solutions in polynomial time by means of the well-known technique of constraints relaxation. Moreover, it is possible to precisely bind the distance between the optimal solution and the sub-optimal ones.

The major difficulty about introducing approximation in reasoning problems is in that it is very hard to find a measure of the approximation which is not dependent on the particular problem at hand. It is important to point out that we deal with formalisms where the knowledge is precisely defined and certain. We are not analyzing situations involving any form of approximate or fuzzy knowledge, based on probabilistic or numerical representations, such as, for example, Zadeh's fuzzy logics [22]. We are trying instead to formalize the approximation of a logical consequence relation defined by means of an extensional semantics but too difficult to compute, through different, but simpler, consequence relations, also defined by means of an extensional semantics.

In a recent paper [2], we have proposed an approach for the approximation of the answer to a query in propositional logic, which is a co-NP-complete problem. We have started from ideas of Levesque [16,17,18] and we have defined a notion of truth assignment that generalizes both 2-valued and 3-valued propositional logics. Two different classes of entailment relations have been defined, the relations of the first class being sound and those of the second one being complete. As a result, we have obtained a stepwise procedure for determining the answer to a propositional query, having the following features:

- any intermediate step (level of approximation) provides clear information which is semantically related to the final answer;
- the intermediate steps can be efficiently computed;
- subsequent steps are computed using information obtained in previous ones.

In our method the answer to a query is reached –although in exponential time– through the computation of several simple steps.

Our goal in this paper is twofold. First of all we show that this pattern can be used for a stepwise procedure which determines the satisfiability of a formula in several modal systems like $\mathcal{S}5$, $\mathcal{K}$, $\mathcal{T}$ and $\mathcal{S}4$. In particular we define a generalization of the standard possible-world semantics for modal formulae, based on the generalization of truth assignment defined in [2]. Moreover, we use such a semantics to define a modal language which allows the explicit representation of the knowledge owned by a resource-limited agent. We also compare our approach with the related work by Levesque [16], Lakemeyer [15] and Fagin and Halpern [6].

The paper is organized as follows. In Section 2 we summarize the results reported in [2] about the approximation of propositional calculus. In Section 3 we show how our methods can be extended for approximating satisfiability in the most widely used modal logics for knowledge and belief. Finally, in the last section we present a very general modal system, where "approximate" knowledge can be explicitly represented and used.

2 Preliminaries

In this section we briefly summarize the definitions and the results presented in [2].

Throughout the paper, we denote with L a set of propositional letters. A literal is a letter l of L or its negation $\neg l$. We denote with L^* the set of all literals associated with the letters of L plus the two special atoms t and f.

Formulae are built on the set L^* by means of the usual connectives $\vee$ and $\wedge$, plus parentheses. We call formulae of this kind *negation normal form* (NNF) formulae, since in this way negation is only applied to propositional letters. This is not a restriction, since any formula in which negation is applied to subformulae can be transformed in linear time in NNF by means of well-known rewriting rules. A truth assignment is a function mapping the set of literals L^* into the set $\{0, 1\}$; the special atoms t and f are always mapped to 1 and 0, respectively. We now introduce three different forms of truth assignment, in which the mapping is differently restricted.

Definition 1 (1-, 2-, 3-interpretation)

- **(Levesque [18])** *A 3-interpretation of L^* is a truth assignment which does not map both a letter l of L and its negation $\neg l$ into 0;*
- *A 2-interpretation of L^* is a truth assignment which maps every letter l of L and its negation $\neg l$ into opposite values;*
- *An 1-interpretation of L^* is a truth assignment which maps every letter l of L and its negation $\neg l$ into 0.*

2-interpretation is the standard notion of interpretation, corresponding to having the two truth values *true* and *false*; 3-interpretations also admit the truth value *contradiction*, while 1-interpretations only admit the value *undefined.* Notice that every 2-interpretation is also a 3-interpretation, and that an 1-interpretation is neither a 2- nor a 3-interpretation.

The value assigned by a truth assignment to a NNF formula is computed by evaluating first every single literal, and then complex formulae with the usual rules for $\vee$ and $\wedge$. The value assigned by a truth assignment to any other formula requires a previous transformation into its NNF.

An 1-, 2- or 3-interpretation *satisfies* a formula iff it maps it into 1. A formula is *1-satisfiable* iff there exists an 1-interpretation satisfying it; it is *1-valid* iff every 1-interpretation satisfies it. The same definitions apply to 2- and 3-interpretations. A formula different from f is always 3-satisfiable, while it is never 1-satisfiable, except for t. t is the only formula which is 1-valid, while a formula is 3-valid iff it is 2-valid.

The definition of 1- and 3-interpretation can be generalized by restricting the possibility of mapping a letter to the truth value *undefined* or *contradiction* to a subset of L. In the following we denote with S a subset — possibly not proper — of the alphabet L.

Definition 2 (S-1-, S-3-interpretation)

- *An S-3-interpretation of L is a truth assignment which maps every letter l of S and its negation $\neg l$ into opposite values. Moreover, it does not map both a letter l of $L \setminus S$ and its negation $\neg l$ into 0.*
- *An S-1-interpretation of L is a truth assignment which maps every letter l of S and its negation $\neg l$ into opposite values. Moreover, it maps every letter l of $L \setminus S$ and its negation $\neg l$ into 0.*

Notice that, for any S, a 2-interpretation is always an S-3-interpretation, while the latter is always a 3-interpretation.

Both definitions of S-1- and S-3-interpretation are equivalent to that of 2-interpretation when $S = L$. On the other hand, S-1-interpretations are 1-interpretations when $S = \emptyset$. The notion of satisfaction and validity of formulae previously introduced applies to S-1- and S-3-interpretation as well. Notice that while a formula different from f is always 3-satisfiable, there are some formulae which are not S-3-satisfiable for a given S. On the other hand, while a formula different from t is never 1-satisfiable, some formulae are S-1-satisfiable.

We now show some interesting properties of the notions of S-3- and S-1-satisfiability concerning their adequacy to *approximate* 2-satisfiability.

Theorem 1 (monotonicity) *For any S, S' such that $S \subseteq S' \subseteq L$, if α is S-1-satisfiable, then α is S'-1-satisfiable (hence 2-satisfiable). Moreover if α is S-3-unsatisfiable, then α is S'-3-unsatisfiable (hence 2-unsatisfiable).*

Theorem 2 (convergence) *If α is 2-satisfiable, then there exists an $S \subseteq L$ such that α is S-1-satisfiable. If α is 2-unsatisfiable, then there exists an $S \subseteq L$ such that α is S-3-unsatisfiable.*

Theorem 3 (complexity) *There exists an algorithm for deciding if α is S-1-satisfiable and deciding if α is S-3-satisfiable which runs in $O(|\alpha| \cdot 2^{|S|})$ time.*

The above theorems show that the NP-complete problem of deciding whether a formula α is 2-satisfiable can be computed in a stepwise fashion by deciding the S-1- and S-3-satisfiability of α for increasing sets S, starting with $S = \emptyset$ and stopping for the least $S \subseteq L$ such that α is S-1-satisfiable or S-3-unsatisfiable. The complexity of the method is in the worst case $O(|T| \cdot 2^{|L|})$, hence competitive with the best known algorithms for deciding 2-satisfiability [4,19]. The method can be anyway stopped for any $S \subset L$, and in this case the time spent in the computation has provided interesting semantical information. This is a typical *approximation* process, since 1) a complex task is performed through several simpler steps, 2) every intermediate step provides a partial solution whose relation to the final solution (the *error*) is clearly identified, and 3) partial solutions can be computed using information obtained in the previous steps.

In our previous work [2] we were concerned mainly on the approximation of *entailment* queries in standard propositional logic. The notion of entailment is defined in the usual way for all the forms of interpretation we have defined. A formula α S-1-*entails* a formula γ ($\alpha \models^1_S \gamma$) iff every S-1-interpretation satisfying α also satisfies γ. An analogous definition applies to S-3-interpretations and 2-interpretations ($\models$), hence we have two families of entailment relations ($\models^1_S, \models^3_S$). Since it is possible to relate S-1- and S-3-entailment to S-1- and S-3-unsatisfiability, respectively, the stepwise procedure for deciding the 2-satisfiability of a formula can be adapted for deciding the validity of $T \models \gamma$, which is a co-NP-complete problem.

This method resembles Kautz and Selman's *knowledge compilation* ([12]) although significant differences exist. As an example, the compiled version of a formula may differ from the original one, hence convergence property does not hold.

3 Approximation in propositional modal logics

In this section we present the first use of our approximation methods in the fields of modal logics. In particular we show how it is possible to approximate satisfiability in the most widely used modal logics for knowledge and belief, namely $\mathcal{S}5$, $\mathcal{K}$, $\mathcal{T}$ and $\mathcal{S}4$. It is well known that modal logics can be characterized either by its axiom schemata or the equivalent restrictions on the accessibility relation. In the following we will use both characterization assuming that the reader is familiar with them. A comprehensive descriptions of the different systems can be found in [3].

A detailed analysis of the computational complexity of satisfiability problems in several propositional modal systems has been done by Ladner [14]. He showed that the problem of checking satisfiability of a formula in the systems $\mathcal{K}$, $\mathcal{T}$ and $\mathcal{S}4$ is PSPACE-complete, while the same problem is NP-complete in the system $\mathcal{S}5$. Therefore most of the modal logics frequently used for modelling knowledge and belief (see [9]) lead to computationally intractable reasoning problems.

In this section we focus on the problem of applying approximation techniques to such propositional modal systems. The main idea is to extend the method defined for propositional logic by defining two classes of interpretations which are approximations of the standard Kripke semantics.

In the following we refer to modal formulae which are built on the set L^* by means of the usual connectives $\vee$ and $\wedge$, the modal operator $\mathbf{K}$, the negation $\neg\mathbf{K}$ of the modal operator, plus parentheses. We call formulae of this kind *modal negation normal form* (MNNF) formulae. Again, we do not loose generality by restricting to such normal form, since any formula can be transformed in linear time into an equivalent one in MNNF. We call formulae not containing modal operators simply propositional.

An S-1-*Kripke interpretation* is a Kripke model (see [13]), i.e. a triple $\mathcal{M} = \langle W, R, V\rangle$ where W is a set of worlds, R an accessibility relation among worlds and V a mapping $W \rightarrow \tau$, where τ is the set of all the truth assignments of L^*. We remind that in a standard Kripke model $V(w)$ is a 2-interpretation for every world $w \in W$. We refer to standard Kripke models as 2-*Kripke interpretations*. On the other hand in an S-1-Kripke interpretation we define $V(w)$ to be an S-1-interpretation for every world $w \in W$.

The evaluation of a propositional formula γ in any world $w \in W$ of an S-1-Kripke interpretation $\mathcal{M} = \langle W, R, V\rangle$ is defined as in Section 2; in particular we write $\mathcal{M}, w \models^1_S \gamma$ iff $V(w)(\gamma) = 1$, that is $V(w)$ maps γ into 1. The value assigned by $\mathcal{M}$ to a MNNF formula α in a world $w \in W$ is defined by using the rule for propositional formulae and recursively the following rules:

- $\mathcal{M}, w \models^1_S \mathbf{K}\beta$ iff $\forall t \in W\ wRt \rightarrow \mathcal{M}, t \models^1_S \beta$;
- $\mathcal{M}, w \models^1_S \neg\mathbf{K}\beta$ iff $\exists t \in W\ wRt \wedge \mathcal{M}, t \models^1_S \neg\beta$

Notice that β may not be in MNNF, in this case we need to transform it in its MNNF equivalent. A modal formula α is S-1-*Kripke satisfiable* iff there exists an S-1-Kripke interpretation $\mathcal{M} = \langle W, R, V\rangle$ and a $w \in W$ s.t. $\mathcal{M}, w \models_S^1 \alpha$.

The notions of S-3-Kripke interpretation and S-3-Kripke satisfiability are straightforward. We refer to standard satisfiability of a modal formula as 2-*Kripke satisfiability.* We now demonstrate our definitions by means of two examples.

Example 1 We show an alphabet L, a modal formula σ on L and a subset S of L such that σ is S-3-Kripke unsatisfiable.

Let L be $\{a, b, c\}$, S be $\{a, b\}$ and σ be $(\mathbf{K}a \wedge \mathbf{K}(\neg a \vee b) \wedge \mathbf{K}\neg b \wedge \neg\mathbf{K}c)$. Let us assume that σ is S-3-Kripke satisfiable. By the above definition this implies that there exists an S-3-Kripke-interpretation $\mathcal{M} = \langle W, R, V\rangle$ and a $w \in W$ such that all the following conditions hold:

- $\mathcal{M}, w \models_S^3 \mathbf{K}a$;
- $\mathcal{M}, w \models_S^3 \mathbf{K}(\neg a \vee b)$;
- $\mathcal{M}, w \models_S^3 \mathbf{K}\neg b$;
- $\mathcal{M}, w \models_S^3 \neg\mathbf{K}c$.

The above conditions are equivalent to the following ones:

- $\forall t \in W \; wRt \rightarrow \mathcal{M}, t \models_S^3 a$;
- $\forall t \in W \; wRt \rightarrow \mathcal{M}, t \models_S^3 (\neg a \vee b)$;
- $\forall t \in W \; wRt \rightarrow \mathcal{M}, t \models_S^3 \neg b$;
- $\exists t \in W \; wRt \wedge \mathcal{M}, t \models_S^3 \neg c$.

Let t_0 be the world whose existence is implied by the last condition, that is let $t_0 \in W$, wRt_0 and $V(t_0)(\neg c) = 1$. According to the other conditions, the truth assignment $V(t_0)$ must satisfy the propositional formula $(a \wedge (\neg a \vee b) \wedge \neg b)$. Taking into account that $V(t_0)$ is an S-3-interpretation of L^* and that $S = \{a, b\}$, $V(t_0)$ satisfies $(a \wedge \neg b)$ if and only if it maps a into 1, $\neg a$ into 0, b into 0 and $\neg b$ into 1. Therefore $V(t_0)$ maps $(\neg a \vee b)$ into 0, hence it does not satisfy $(a \wedge (\neg a \vee b) \wedge \neg b)$. This contradiction proves that σ is S-3-Kripke unsatisfiable.

Notice that σ is S'-3-Kripke satisfiable, where $S' = \{a\}$. □

Example 2 We show an alphabet L, a modal formula τ on L and a subset S of L such that τ is S-1-Kripke satisfiable.

Let L be $\{a, b\}$, S be $\{b\}$ and τ be $(\neg b \wedge \mathbf{K}a \wedge \mathbf{K}(\neg a \vee b))$. By the above definition, τ is S-1-Kripke satisfiable if and only if there exists an S-1-Kripke-interpretation $\mathcal{M} = \langle W, R, V\rangle$ and a $w \in W$ such that all the following conditions hold:

- $\mathcal{M}, w \models_S^1 \neg b$;

- $\forall t \in W \; wRt \rightarrow \mathcal{M}, t \models^1_S a$;
- $\forall t \in W \; wRt \rightarrow \mathcal{M}, t \models^1_S (\neg a \vee b)$.

Let W be the singleton $\{w\}$, R be the empty set and the S-1-interpretation $V(w)$ of L^* be such that $V(w)(a) = V(w)(\neg a) = V(w)(b) = 0$ and $V(w)(\neg b) = 1$. It is easy to see that $\mathcal{M}, w \models^1_S \tau$ holds, where $\mathcal{M} = \langle W, R, V \rangle$. Therefore τ is S-1-Kripke satisfiable. □

The following are straightforward consequences of the definitions of S-1 and S-3-Kripke satisfiability (α is a MNNF formula).

Theorem 4 (monotonicity) *For any S, S' such that $S \subseteq S' \subseteq L$, if α is S-1-Kripke satisfiable, then α is S'-1-Kripke satisfiable (hence 2-Kripke satisfiable). Moreover if α is S-3-Kripke unsatisfiable, then α is S'-3-Kripke unsatisfiable (hence 2-Kripke unsatisfiable).*

Theorem 5 (convergence) *If α is 2-Kripke satisfiable, then there exists an $S \subseteq L$ such that α is S-1-Kripke satisfiable. If α is 2-Kripke unsatisfiable, then there exists an $S \subseteq L$ such that α is S-3-Kripke unsatisfiable.*

The above theorems account for a stepwise procedure for deciding 2-Kripke satisfiability of a modal formula, which is analogous to that defined in Section 2 for checking 2-satisfiability of a propositional formula. As a consequence of theorem 4 is that the formula σ of example 1 is 2-Kripke-unsatisfiable and the formula τ of example 2 is 2-Kripke-satisfiable. The following theorem shows that there exist modal systems, e.g. $S5$, in which such a stepwise procedure is interesting from a computational point of view.

Theorem 6 (complexity for $S5$) *If we restrict our attention to accessibility relations which are reflexive, transitive and euclidean, then there exists one algorithm to decide if α is S-1-Kripke satisfiable and one to decide if α is S-3-Kripke satisfiable both running in $O(m \cdot |\alpha| \cdot 2^{|S|})$ time, where m is the number of occurrences of the modal operator in α.*

PROOF (sketch): The algorithms for checking S-1- and S-3-Kripke satisfiability are based on a mapping π from any modal formula α on the alphabet L into a propositional formula $\pi(\alpha)$ on the alphabet $\pi(L)$ such that $|\pi(L)| = (m+1) \cdot |L|$ and $|\pi(\alpha)| \leq (m+1) \cdot |\alpha|$.

The alphabet $\pi(L)$ is defined as $\bigcup_{i=1}^{m+1} \bigcup_{p \in L} p^i$, that is it contains $m+1$ copies of each letter of L. If S is a subset of L, then $\pi(S)$ is defined as $\bigcup_{i=1}^{m+1} \bigcup_{p \in S} p^i$. The mapping $\pi(\alpha)$ is defined by the following rewriting rules, where $\alpha, \alpha_1, \alpha_2$ are modal formulae, and p is in L:

$$\begin{array}{rcl}
\alpha & \longrightarrow & (\alpha, 1) \\
(\alpha_1 \wedge \alpha_2, i) & \longrightarrow & (\alpha_1, i) \wedge (\alpha_2, i) \\
(\alpha_1 \vee \alpha_2, i) & \longrightarrow & (\alpha_1, i) \vee (\alpha_2, i) \\
(\neg(\alpha_1 \wedge \alpha_2), i) & \longrightarrow & (\neg\alpha_1, i) \vee (\neg\alpha_2, i) \\
(\neg(\alpha_1 \vee \alpha_2), i) & \longrightarrow & (\neg\alpha_1, i) \wedge (\neg\alpha_2, i)
\end{array}$$

$$\begin{array}{rcl} (\mathbf{K}\alpha, i) & \longrightarrow & (\alpha, 1) \wedge \cdots \wedge (\alpha, m+1) \\ (\neg\mathbf{K}\alpha, i) & \longrightarrow & (\neg\alpha, 1) \vee \cdots \vee (\neg\alpha, m+1) \\ (p, i) & \longrightarrow & p^i \\ (\neg p, i) & \longrightarrow & \neg p^i \end{array}$$

It is possible to prove that the S-1-Kripke satisfiability of the modal formula α is equivalent to the S'-1-satisfiability of the propositional formula $\pi(\alpha)$, where $S' = \pi(S)$. In fact the mapping π is based on a generalization of a property of the system $\mathcal{S}5$, stating that if β is a 2-Kripke satisfiable formula with m occurrences of the modal operator, then there exists a 2-Kripke interpretation $\mathcal{M} = \langle W, R, V\rangle$ and a $w \in W$ such that $\mathcal{M}, w \models \beta$ and where the size of W is less than $m + 1$ (see [14, Lemma 6.1]).

The S'-1-satisfiability of $\pi(\alpha)$ can be determined in $O(m \cdot |\alpha| \cdot 2^{|S|})$ time with the algorithms presented in [2]. Analogous properties hold for the S-3-Kripke satisfiability of α. □

The above theorem shows that all the considerations made in Section 2 on the approximation of the 2-satisfiability of a propositional formula also hold for the approximation of the 2-Kripke-satisfiability of any formula of the modal system $\mathcal{S}5$.

The same idea can be applied, with only minor variations, to other systems whose satisfiability check is known to be an NP-complete problem, such as $\mathcal{K}45$ and $\mathcal{KD}45$. This holds since in these systems it is possible to satisfy a formula with a 2-Kripke interpretation whose set of worlds has size bounded by a polynomial function of the size of the formula itself.

On the other hand, as proved by the following result, there exist interesting modal systems, such as $\mathcal{K}$, in which the stepwise procedure suggested by Theorems 4, 5 is not useful from a computational point of view.

Theorem 7 (complexity for $\mathcal{K}$) *If the accessibility relation is unrestricted, then deciding if α is S-1-Kripke satisfiable and deciding if α is S-3-Kripke satisfiable are PSPACE-complete problems even if $|S| = 1$.*

This result easily follows from those of [20,21], and prevents us from the development of a result analogous to Theorem 6 for unrestricted accessibility relation (unless P=PSPACE). This is not surprising, since Ladner has shown [14] that there exist formulae in the systems $\mathcal{K}$, $\mathcal{T}$ and $\mathcal{S}4$, which are satisfied only by 2-Kripke interpretations having a set of worlds whose size is exponential in the nesting of the modal operators.

A possible way to overcome this problem is to focus only on limited parts of the interpretations. We now present a semantics for approximation which further extends the possible-worlds semantics. The idea is that a Kripke interpretation $\mathcal{M}$ should satisfy a formula α in a world w iff α is satisfied in the subset $W' \subseteq W$ of the possible worlds containing only those worlds whose *distance* from w is less than or equal to i, where i is a particular integer. In this way we can limit our attention to Kripke interpretations having at most 2^i worlds. The worlds which are outside the "range" i are treated differently in the S-1- and in the S-3- case. In particular S-1-Kripke interpretations are "pessimistic", since they do not validate anything in those worlds, while S-3-Kripke interpretations are "optimistic", since they validate everything.

Let α be a formula in which each occurrence of a modal operator lies in the scope of at most n modal operators, S a subset of L, $i \leq n+1$, and $\mathcal{M} = \langle W, R, V\rangle$ an S-3-Kripke interpretation. We define a new relation $\models^3_{S,i}$ as follows (γ is a propositional formula):

- $\mathcal{M}, w \models^3_{S,i} \gamma$ if and only if $(V(w)(\gamma) = 1$ or $i < 0)$;
- $\mathcal{M}, w \models^3_{S,i} \mathbf{K}\alpha$ if and only if $\forall t \in W\ wRt \rightarrow \mathcal{M}, t \models^3_{S,i-1} \alpha$;
- $\mathcal{M}, w \models^3_{S,i} \neg\mathbf{K}\alpha$ if and only if $\exists t \in W\ wRt \wedge \mathcal{M}, t \models^3_{S,i-1} \neg\alpha$.

Notice that according to the above definition, if $i < 0$ then any formula is true in any world. A modal formula α is $\langle S, i\rangle$-3-Kripke satisfiable iff there exists an S-3-interpretation $\mathcal{M} = \langle W, R, V\rangle$ and a $w \in W$ s.t. $\mathcal{M}, w \models^3_{S,i} \alpha$.

A similar definition can be given for the relation $\models^1_{S,i}$. The only difference is that now $\mathcal{M}$ is an S-1-Kripke interpretation and the definition of the base case is:

- $\mathcal{M}, w \models^1_{S,i} \gamma$ if and only if $(V(w)(\gamma) = 1$ and $i \geq 0)$;

Notice that if $i < 0$ then a formula cannot be true in a world. It is easy to show that the analogous of Theorem 4 holds for the new definitions, when we compare pairs $\langle S, i\rangle$ and $\langle S', j\rangle$ such that $S \subseteq S' \subseteq L$ and $i \leq j \leq n+1$. In other words if a modal formula α is $\langle S, i\rangle$-3-Kripke unsatisfiable, then it is $\langle S', j\rangle$-3-Kripke unsatisfiable, and if it is $\langle S, i\rangle$-Kripke satisfiable, then it is $\langle S', j\rangle$-1-Kripke satisfiable. Moreover, there exists a subset S of L and an integer $i \leq n+1$ such that α is either $\langle S, i\rangle$-1-satisfiable or $\langle S, i\rangle$-3-unsatisfiabile.

We can also prove that there exists an algorithm for deciding if a modal formula α is $\langle S, i\rangle$-3-Kripke satisfiable which runs in $O(|\alpha| \cdot 2^{|S|\cdot i})$ time, provided that the constraints of either $\mathcal{K}$ or $\mathcal{T}$ or $\mathcal{S}4$ hold on the accessibility relation. The algorithm for determining $\langle S, i\rangle$-3-Kripke satisfiability of a modal formula α is based on a mapping of α into another modal formula $\psi^3(\alpha)$, in which the nesting of the modal operators is limited and any occurrence of a letter not in S is substituted by the literal t. More precisely, the modal formula $\psi^3(\alpha)$ is obtained by:

1. substituiting each occurrence of a letter in $L \setminus S$ with the literal t, thus obtaining the formula α';
2. substituting every subformula $\neg\mathbf{K}\beta$ of α' which is in the scope of at least i modal operators $\mathbf{K}$ with the literal t.

The $\langle S, i\rangle$-3-Kripke satisfiability of α is equivalent to the 2-Kripke satisfiability of $\psi^3(\alpha)$. The 2-Kripke satisfiability of $\psi^3(\alpha)$ can be determined with standard algorithms. The algorithm for checking $\langle S, i\rangle$-1-Kripke satisfiability is based on a similar mapping ψ^1 in which the literal f is used instead of using t.

This allows us to extend all the considerations on the approximation of the 2-satisfiability of a propositional formula to the approximation of the 2-Kripke-satisfiability of any formula of the modal systems $\mathcal{K}$, $\mathcal{T}$ and $\mathcal{S}4$.

4 Modeling resource-bounded agents

In this section we present a very general system, where "approximate" knowledge can be explicitly represented and used; this system is also compared with some of the formalisms presented in the literature.

It has been frequently pointed out that a major drawback of the possible-worlds semantics as a semantics for logics of knowledge and belief is the implicit assumption of logical omniscience. We now present a system which extends the possible-worlds framework, in order to represent non-omniscent agents.

The main idea underlying the system consists in providing language constructs for representing the kind of approximation implicit in the entailment relations $\models^3_S$ and $\models^1_S$ defined in Section 2. The system consists of two families of modal operators related to the notion of S-interpretation, where the elements of the first family are denoted as $\Box^3_S$ and the elements of the second family as $\Box^1_S$. Formulae are built using the usual connectives and the two sets of modal operators $\Box^3_S$ and $\Box^1_S$ for any $S \subseteq L$. A model is a triple $M = (Sit, R, V)$, where Sit is a set of situations, R is an accessibility relation that is reflexive, transitive and euclidean, and V a (4-valued) valuation, which maps any situation into an unrestricted truth assignment $(V : Sit \rightarrow (L^* \rightarrow \{0, 1\}))$. We denote with $\mathcal{W}(Sit)$ the set of situations $s \in Sit$ such that s is also a possible world, i.e. $V(s)$ is a 2-interpretation. Similarly, we denote with S-$3(Sit)$ and S-$1(Sit)$ the sets of situations that can be interpreted as S-3- and S-1-interpretations, respectively. The semantics is defined as follows (γ is a propositional formula):

- $\mathcal{M}, s \models \gamma$ iff $V(s)(\gamma) = 1$;
- $\mathcal{M}, s \models \Box^3_S \alpha$ iff $\forall t \in Sit\ sRt \wedge t \in S\text{-}3(Sit) \rightarrow \mathcal{M}, t \models \alpha$;
- $\mathcal{M}, s \models \neg\Box^3_S \alpha$ iff $\exists t \in Sit\ sRt \wedge t \in S\text{-}3(Sit) \wedge \mathcal{M}, t \not\models \alpha$;
- $\mathcal{M}, s \models \Box^1_S \alpha$ iff $\forall t \in Sit\ sRt \wedge t \in S\text{-}1(Sit) \rightarrow \mathcal{M}, t \models \alpha$;
- $\mathcal{M}, s \models \neg\Box^1_S \alpha$ iff $\exists t \in Sit\ sRt \wedge t \in S\text{-}1(Sit) \wedge \mathcal{M}, t \not\models \alpha$;

A formula α is valid, written $\models \alpha$, if α is true at every possible world $w \in \mathcal{W}(Sit)$ of every model $\mathcal{M} = (Sit, R, V)$. A formula α is satisfiable if there is a model $\mathcal{M} = (Sit, R, V)$ and a possible world $w \in \mathcal{W}(Sit)$ s.t. $\mathcal{M}, w \models \alpha$. The choice of an accessibility relation that is reflexive, transitive and euclidean corresponds to the modal system $S5$, which is considered as an appropriate formalisation of the notion of knowledge.

A minimal requirement for the system is its ability to represent the entailment relations $\models^1_S$ and $\models^3_S$ via the modal operators $\Box^1_S$ and $\Box^3_S$. This is in fact possible, as proven by the following result (α and γ are propositional formulae):

- $\models \Box^3_S \alpha \supset \Box^3_S \gamma$ iff $\Box^3_S \alpha \wedge \neg\Box^3_S \gamma$ is unsatisfiable iff $\alpha \models^3_S \gamma$.
- $\models \Box^1_S \alpha \supset \Box^1_S \gamma$ iff $\Box^1_S \alpha \wedge \neg\Box^1_S \gamma$ is unsatisfiable iff $\alpha \models^1_S \gamma$.

Our language is at least capable of representing the approximate entailment relations. It is anyway interesting to check whether the schemata defining the system $S5$ are valid for these new operators, in order to show their adequacy to represent resource-bounded agents.

The system $S5$ is charactherized by the usual rules and axiom schemata of the propositional calculus plus the inference rule (necessitation):

$Nec)$ $\vdash p \Rightarrow \vdash \mathbf{K}p$

and the axiom schemata:

$K)$ $\mathbf{K}(p \supset q) \supset (\mathbf{K}p \supset \mathbf{K}q)$

$T)$ $\mathbf{K}p \supset p$

$4)$ $\mathbf{K}p \supset \mathbf{KK}p$

$5)$ $\neg\mathbf{K}p \supset \mathbf{K}\neg\mathbf{K}p$

We now analyze which of these schemata are valid in our semantics when we replace $\mathbf{K}$ with $\Box^1_S$. The result is the following:

$Nec)$ $\models p \not\Rightarrow \models \Box^1_S p$

$K)$ $\models \Box^1_S(p \supset q) \supset (\Box^1_S p \supset \Box^1_S q)$

$T)$ $\not\models \Box^1_S p \supset p$

$4)$ $\models \Box^1_S p \supset \Box^1_S \Box^1_S p$

$5)$ $\models \neg\Box^1_S p \supset \Box^1_S \neg\Box^1_S p$

The validity of the schemata 4 and 5 is a straightforward consequence of the properties of the accessibility relation, while the schema K follows from the semantic definition of $\Box^1_S$. We now show counterexamples for the properties which do not hold.

$Nec)$ Let $p = q \vee \neg q$, $S = \emptyset$, $\mathcal{M} = (Sit, R, V)$, $Sit = \{s_1, s_2\}$, $R = \{(s_1, s_1), (s_1, s_2), (s_2, s_1), (s_2, s_2)\}$, $V(s_1)(q) = 1$ and $V(s_1)(\neg q) = V(s_2)(q) = V(s_2)(\neg q) = 0$. We have that $\models p$ holds but $\models \Box^1_S p$ does not hold, in fact $\mathcal{M}, s_1 \not\models \Box^1_S p$.

$T)$ Let $S = \emptyset$, $\mathcal{M} = (Sit, R, V)$, $Sit = \{s_1\}$, $R = \{(s_1, s_1)\}$, $V(s_1)(\neg p) = 1$ and $V(s_1)(p) = 0$. $\Box^1_S p \supset p$ is valid iff $\forall N\ \forall w \in \mathcal{W}(Sit)\ N, w \models \neg\Box^1_S p \vee N, w \models p$. Since by instanciating N to $\mathcal{M}$ and w to s_1 we obtain that $\mathcal{M}, s_1 \not\models \neg\Box^1_S p$ and $\mathcal{M}, s_1 \not\models p$ then it is not the case that $\Box^1_S p \supset p$ is valid.

We have shown that both the rule of necessitation and the axiom schema T do not hold in general. As a consequence we can use $\Box^1_S$ to model an agent capable of performing *at least* every sound inference, because its knowledge is closed under modus ponens (the K schema), nevertheless, the agent can do some inference which is not sound, in fact the T schema does not hold. Since both schemata 4 and 5 are valid, it follows that agents modeled in our system are fully introspective. Even if the necessitation rule and the T schema do not hold in general they are valid whenever $letters(p) \subseteq S$, so they still hold for a subset of the language.

We now analyze which schemata are valid in our semantics when we replace $\mathbf{K}$ with $\Box^3_S$.

$Nec)$ $\models p \Rightarrow \models \Box^3_S p$

$K)$ $\not\models \Box^3_S(p \supset q) \supset (\Box^3_S p \supset \Box^3_S q)$

$T)$ $\models \Box^3_S p \supset p$

4) $\models \Box^3_S p \supset \Box^3_S \Box^3_S p$

5) $\models \neg\Box^3_S p \supset \Box^3_S \neg\Box^3_S p$

Again, the schemata 4 and 5 are a straightforward consequence of the properties of the accessibility relation, while the schema T and the rule Nec follow from the semantic definition of $\Box^3_S$. We now show a counterexample for the property K. Let $S = \emptyset$, $\mathcal{M} = (Sit, R, V)$, $Sit = \{s_1, s_2\}$, $R = \{(s_1,s_1),(s_1,s_2),(s_2,s_1),(s_2,s_2)\}$, $V(s_1)(p) = V(s_1)(q) = V(s_2)(p) = V(s_2)(\neg p) = V(s_2)(\neg q) = 1$ and $V(s_1)(\neg p) = V(s_1)(\neg q) = V(s_2)(q) = 0$. We have that $\mathcal{M}, s_1 \models \Box^3_S(p \supset q)$ and $\mathcal{M}, s_1 \models \Box^3_S p$ but $\mathcal{M}, s_1 \models \Box^3_S q$ does not hold.

Even this set of modal operators does not satisfy all the rules and axiom schemata of $S5$. In this case the only property which is not satisfied is property K, that is the closure of the knowledge under modus ponens. This implies that an agent modeled with these operators is not commited to full logical omniscience. Its deductive capabilities are limited, but its inferences are always sound, as witnessed by the validity of the T schema. Again properties 4 and 5 continue to hold thus providing the agent with perfect introspection. Since $\Box^3_S$ is an approximation of $\mathbf{K}$, it is clear that whenever the set S is equal to L the two operators coincide and therefore $\Box^3_L$ will satisfy exactly the same properties of $\mathbf{K}$, including the K schema. But this schema is also satisfiedo under weaker conditions, if p is a propositional formula we have that $\Box^3_S(p \supset q) \supset (\Box^3_S p \supset \Box^3_S q)$ is valid if and only if $p \wedge \neg p$ is S-3-unsatisfiable.

There are other interesting properties which involve different sets S, for example we have that, assuming $S \subseteq S' \subseteq L$, $\models \Box^1_S p \Rightarrow \models \Box^1_{S'} p$ and $\models \Box^3_{S'} p \supset \Box^3_S p$.

We claim that the two sets of operators $\Box^3_S$ and $\Box^1_S$ account for a non-ideal reasoner with a limited amount of resources. The first set of operators accounts for a skeptical reasoner, while the second set accounts for a credulous one. We can now model an agent's reasoning capabilities avoiding the logical omniscience assumption and providing a set of semantically motivated restrictions to its deductive capabilities. In particular, we claim that our system can be used for modeling the interaction of several non-ideal agents, each having its own attitude —either credulous or skeptical— and competence, charachterized by the set S.

An example of such a situation is a distributed inference system having a common knowledge base, but in which any component has a specific competence. The supervisor of the system uses the different modal operators to represent the answers coming from the various components, and process them to obtain a more accurate answer.

Several other systems capable of representing non-omniscient agents have been presented in the literature. We now briefly compare our system with some of the best known formalisms. In Levesque's system [16] the semantics of the modal operator of implicit belief **L** is defined in terms of possible worlds. Conversely, the modal operator of explicit belief **B** is defined in terms of situations. Both operators can be represented in our semantics and with our operators, in fact, **L** is equivalent to $\Box^3_S$ or $\Box^1_S$ when $S = L$ and **B** is equivalent to $\Box^3_S$ when $S = \emptyset$, hence, $\Box^3_S\alpha$ has the intuitive reading of "α can be explicited by reasoning only on letters in S".

Lakemeyer [15] has extended Levesque's system, allowing a restricted form of nesting of operators. However, his semantics, which relies on two distinct accessibility relations R and $\overline{R}$ for interpreting negated formulae, is very different from ours, thus making difficult any comparison.

A very interesting proposal has been presented by Fagin and Halpern in [6]. In this paper they introduce a new propositional attitude, in addition to knowledge and belief: the attitude of awareness. This new modality should act as a kind of filter on the consequences that can be drawn. In their system, truth in a world is defined in terms of the relation $\models^{\Psi}$, where $\Psi \subseteq L$ is a set of propositional letters and the agent is aware only of them. The intended meaning of $\models^{\Psi}$ is to restrict the attention only to letters in Ψ, while letters in $L \setminus \Psi$ are ignored. Our notion of S-1-interpretation is exactly in the same spirit. However, in Fagin and Halpern's system there is nothing close in spirit to the notion of S-3-interpretation.

In a more recent work [7], Fagin, Halpern and Vardi present a different system which does not commit to the logical omniscience assumption. The presented system clarifies the reason why most of the non-classical semantics are not committed to logical omniscience. The main reasons are the impossibility of distinguishing either between coherent and incoherent worlds or between complete and incomplete ones. It is exactly the possibility of discerning between the various degrees of incoherence and incompleteness which has led us to the definition of our system. For example, the effect of the operator $\Box^3_S$ is exactly to select only complete situations which can be only partially incoherent, i.e. can be incoherent only in the interpretation of the propositions in $L \setminus S$. Analogous is the effect of $\Box^1_S$ which selects only coherent situations being partially incomplete.

Acknowledgements

We want to thank Marta Cialdea for her useful suggestions which helped improving both the presentation and the contents of the paper. Luigia Carlucci Aiello and Maurizio Lenzerini carefully read and commented a previous version of this paper. This work has been supported by the ESPRIT Basic Research Action N.3012-COMPULOG and by the Progetto Finalizzato Sistemi Informatici e Calcolo Parallelo of the CNR (Italian research Council).

References

[1] Cadoli M. and Lenzerini M. 1990. The Complexity of Closed World Reasoning and Circumscription, *Proceedings of the 8th Conference of the American Association for Artificial Intelligence*, pp. 550-555.

[2] Cadoli M. and Schaerf M. 1991. Approximate Entailment, *Trends in Artficial Intelligence: Proceedings of the 2nd Conference of the Italian Association for Artificial Intelligence*, Springer Verlag LNAI 549, pp. 68-77.

[3] Chellas B. 1980. *Modal Logic: an introduction*, Cambridge University Press.

[4] Davis M. and Putnam H. 1960. A Computing Procedure for Quantification Theory. *Journal of ACM*, vol. 7, pp. 201-215.

[5] Donini F.M., Lenzerini M., Nardi D. and Nutt W. 1991. Tractable Concept Languages, *Proceedings of the 12th International Joint Conference on Artificial Intelligence*, pp. 458-463.

[6] Fagin R. and Halpern J.Y. 1988. Belief, Awareness and Limited Reasoning, *Artificial Intelligence*, 34, pp. 39-76.

[7] Fagin R., Halpern J.Y. and Vardi M.Y. 1990. A Nonstandard Approach to the Logical Omniscience Problem. In R. Parikh (ed.). *Theoretical Aspects of Reasoning about Knowledge: Proceedings of the Third Conference*, San Mateo, Calif.:Morgan Kaufmann, pp. 41-55.

[8] Ginsberg M. 1991. Computational Considerations in Reasoning about Actions, *Proc. of the Second Conference on the Principles of Knowledge Representation and Reasoning*, pp. 250-261.

[9] Hintikka J. 1962. *Knowledge and belief*, Cornell University Press, Ithaca, New York.

[10] Hintikka J. 1975. Impossible Possible Worlds Vindicated, *Journal of Philosophical Logic*, vol. 4, pp. 475-484.

[11] Kautz H.A. and Selman B. 1991. Hard Problems for Simple Default Logics, *Artificial Intelligence*, Vol. 49, pp. 243-279.

[12] Kautz H.A. and Selman B. 1991. Knowledge Compilation Using Horn Approximations, *Proceedings of the 9th Conference of the American Association for Artificial Intelligence*, pp. 904-909.

[13] Kripke S.A. 1963. Semantical Considerations on Modal Logic. *Acta Philosophica Fennica*, vol. 16, pp. 83-94.

[14] Ladner R. 1977. The computational complexity of provability in systems of modal propositional logic. *SIAM Journal of Computing*, 6(3), pp. 467-480.

[15] Lakemeyer G. 1987. Tractable Meta-Reasoning in Propositional Logics of Belief. *Proceedings of the 10th International Joint Conference on Artificial Intelligence*, pp. 402-408.

[16] Levesque H.J. 1984. A Logic of Implicit and Explicit Belief. *Proceedings of the 4th Conference of the American Association for Artificial Intelligence*, pp. 198-202.

[17] Levesque H.J. 1988. Logic and the Complexity of Reasoning. *Journal of Philosophical Logic*, vol. 17, pp. 355-389.

[18] Levesque H.J. 1989. A knowledge-level account of abduction. *Proceedings of the 11th International Joint Conference on Artificial Intelligence*, pp. 1061-1067.

[19] Robinson J.A. 1965. A Machine Oriented Logic Based on the resolution Principle. *Journal of the ACM*, vol. 12, pp. 397-415.

[20] Schild K. 1991. A Correspondence Theory for Terminological Logics: Preliminary Report. *Proceedings of the 12th International Joint Conference on Artificial Intelligence*, pp. 466-471.

[21] Schmidt-Schauß M. and Smolka G. 1991. Attributive concept descriptions with complements. *Artificial Intelligence*, vol. 48, pp. 1-26.

[22] Zadeh L.A. 1979. A Theory of Approximate Reasoning. In (J. E. Hayes, D. Michie and L. I. Mikulich eds.) *Machine Intelligence*, vol. 9, Elsevier, New York, pp. 149-194.