Decidable Reasoning in First-Order Knowledge Bases with Perfect Introspection

Gerhard Lakemeyer
Department of Computer Science
University of Toronto
Toronto, Ontario, Canada, M5S 1A4
e-mail: gerhard@ai.toronto.edu

Abstract

Since knowledge bases (KBs) are usually incomplete, they should be able to provide information regarding their own incompleteness, which requires them to introspect on what they know and do not know. An important area of research is to devise models of introspective reasoning that take into account resource limitations. Under the view that a KB is completely characterized by the set of beliefs it represents (its *epistemic state*), it seems natural to model KBs in terms of *belief*. Reasoning can then be understood as the problem of computing membership in the epistemic state of a KB. The best understood models of belief are based on possible-world semantics. However, their computational properties are unacceptable. In particular, they render reasoning in first-order KBs undecidable. In this paper, we propose a novel model of belief, which preserves many of the advantages of possible-world semantics yet, at the same time, guarantees reasoning to be decidable, where a KB may contain sentences in full first-order logic. Moreover, such KBs have perfect knowledge about their own beliefs even though their beliefs about the world are limited.

Introduction

Since the information contained in a knowledge base (KB) is usually incomplete, a KB should not only be able to answer queries about the domain in question but also about its own state of incompleteness. In other words, a KB should possess self-knowledge, which it gains by introspection. For the purposes of this paper, we assume that a KB is capable of *perfect* introspection, that is, it knows exactly what it knows[1] and does not know. Perfect introspection is not uncontroversial. For example, one may want to restrict knowledge of what is not known to concepts that one is at least aware of (see [14] for a formalization of this idea). However, we have adopted perfect introspection here simply because it comes at no additional computational cost.

An important area of research is to devise models of introspective reasoning that take into account resource limitations [9]. Under the view that a KB is completely characterized by the set of beliefs it represents (its *epistemic state*), it seems natural to model KBs in terms of *belief*. Reasoning can then be understood as the problem of computing membership in the epistemic state of a KB.

The best understood models of belief are based on possible-world semantics [10, 8]. Most autoepistemic logics, a favorite framework to study introspection (e.g. [20, 17, 18]), specify epistemic states that have possible-world models, as shown in [21, 5, 17]. A big advantage of possible world models is their simplicity. Unfortunately, they also make the assumption that a KB's (or agent's) beliefs are closed under logical consequence, a property often referred to as *logical omniscience*, which renders reasoning undecidable in first-order KBs. An important problem then is to find models of belief with better computational properties.

While there has been some progress in finding computationally attractive models of belief in the propositional case [14], adding quantifiers to the framework in the obvious way leads to undecidability. In this paper, we remedy this situation and propose a new model of belief that preserves much of the simplicity of possible-world semantics yet, at the same time, guarantees that reasoning in first-order KBs is computable.[2]

As in [14, 17], we use an approach that allows us to model the beliefs of a KB directly within the logic. Intuitively, a KB's epistemic state can be characterized as the set of all sentences that are believed given that the sentences in the KB are *all* that is believed or, as we will say for short, *only-believed*. This idea is formalized in a modal logic with two modal operators **B** and **O** for belief and only-believing, respectively. The epistemic state of a KB is characterized by the set of sentences α for which $\mathbf{O}KB \supset \mathbf{B}\alpha$ is valid.[3] The

[1] Although this paper is concerned with belief rather than knowledge, we nevertheless use both terms interchangeably.

[2] In section "A Decidable KR Service," we also discuss restrictions under which reasoning is in fact tractable.

[3] Whenever KB occurs within a logical sentence, we

complexity of reasoning then reduces to the complexity of determining whether $\mathbf{B}\alpha$ follows logically from $\mathbf{O}$KB for a given KB and α.

The main contribution of this paper lies in the novel model-theoretic account of belief and only-believing. In particular, KBs under this model of belief may contain arbitrary first-order sentences and they have perfect knowledge about what they do and do not believe even though their beliefs about the world are limited. Most importantly, the beliefs of such KBs are computable.

In related work, Konolige [9] also addresses the issue of modeling introspection under resource limitations. However rather than proposing an actual instance of a computationally attractive reasoner, he presents a general framework in which one can be formalized. Work regarding decidable forms of first-order entailment (without introspection) is also relevant [7, 2, 22, 4], since it is a useful starting point when considering limited forms of belief. In particular, Patel-Schneider's [22] and Frisch's [4] notions of entailment, which have a model-theoretic semantics, are good candidates as a basis for belief. Indeed the logic developed in this paper[4] is shown to subsume Patel-Schneider's t-entailment (Theorem 1). By using a more expressive modal language, this paper goes beyond t-entailment by formalizing first-order beliefs *about* beliefs and agents introspecting on their own beliefs. At the same time, we are able to preserve the computational benefits of t-entailment. Finally, belief is formalized in fairly intuitive model-theoretic terms.

The rest of the paper is organized as follows. We begin by defining the syntax and semantics of *OBL*, the logic of belief and only-believing. This is followed by a brief discussion of the properties of belief. The next section formally establishes the decidability result regarding which beliefs follow logically from only-believing a KB. After that, we apply these ideas to the specification of KR service routines `ASK` and `TELL` in the sense of [15]. Finally, we summarize the results and outline extensions of the current framework and future directions.

The Logic *OBL*

The key aspects of this logic are the semantics and properties of belief and only-believing, which ultimately gives us the specification of decidable, introspective reasoning in first-order KBs. Before turning to the technical definitions, we first introduce the two notions of belief informally. (The other logical connectives and quantifiers have the usual meaning.)

mean the conjunction of all the sentences in the KB.

[4] A preliminary model which addressed neither self-knowledge nor only-believing appeared in [11]. However, ultimately it proved to be too complex and was abandoned in favor of the current, much simpler model.

Belief Belief is defined in a possible-world fashion. Roughly, an agent is assumed to imagine a set of states of affairs or *situations* M. The agent is then said to believe a sentence α just in case α (or, as we will see below, a slightly modified α) holds in all situations in M. Except for the definition of situations, which are described in more detail below, this framework is very similar to one that defines the logic *weak S5* [6]. In fact, as in *weak S5*, our approach results in agents capable of perfect introspection with respect to what they know and don't know.

In order to avoid the logical omniscience problem, we limit an agent's ability to reason by cases in the following two ways, which follow from the way situations are defined and used. For one, we allow beliefs not to be closed under *modus ponens*. For example, if p, q, and r are atomic facts, we allow an agent to believe $(p \vee q) \wedge (\neg q \vee r)$, yet fail to believe $(p \vee r)$. Another way reasoning by cases is limited is by weakening the link between disjunction and existential quantification in the sense that an agent may believe $\mathtt{P}(a) \vee \mathtt{P}(b)$ for a predicate $\mathtt{P}$ and distinct terms a and b, yet fail to believe $\exists x \mathtt{P}(x)$.

In *OBL*, closure under modus ponens is avoided by decoupling the notions of truth and falsity within situations. Instead of assigning either true or false to atomic facts (predicates applied to individuals), situations are allowed to assign independent *true-support* and *false-support* to atoms. This corresponds to using four truth values $\{\}, \{\mathtt{true}\}, \{\mathtt{false}\}$, and $\{\mathtt{true}, \mathtt{false}\}$, an idea originally proposed to provide a semantics for a fragment of relevance logic called *tautological entailment* [1, 3].[5] Note that the classical worlds of possible-world semantics are a special kind of situations, namely those where each atomic fact has either true- or false-support but not both. In *OBL*, classical worlds are used to provide the standard notions of *truth* and *validity*. Non-classical situations are only allowed to be part of what agents imagine (defining their beliefs). The fact that an agent's imagination can be incomplete and inconsistent provides some intuition for situations that assign neither true- nor false-support to an atom or both true- and false-support.

In order to weaken the link between disjunction and existential quantification, *OBL* restricts the interpretation of existential quantifiers within belief. Roughly, we require that an agent who believes the existence of an individual with a certain property must be able to name or give a description of that individual, although we do not require the agent to know who the individual is. More concretely, for $\exists x \mathtt{P}(x)$ to be believed there must be a closed term t (e.g. $father(john)$) such that $\mathtt{P}(t)$ is true in all accessible situations. In general, if the existential appears in the scope of universal quantifiers, the corresponding universals may (but need not)

[5] Levesque [16] was the first to introduce the notion of four-valued situations to model a limited form of belief in a propositional framework.

occur in the description chosen for the existential.

Only-Believing An agent who only-believes a sentence α believes α and, intuitively, believes as little else as possible. In other words, the agent is maximally ignorant while still believing α.

As demonstrated in [14, 17], if belief is modeled by a set of situations, independent of whether they are four-valued or two-valued as in classical possible-world semantics, only-believing has a particularly simple characterization: an agent only-believes a sentence α if he or she believes α and the set of situations M the agent imagines is as large as possible, i.e., if we were to add any other world to M, the agent would no longer believe α.[6]

With the special treatment of existential quantification as outlined above, there is, however, one complication that needs to addressed. Consider the sentence $\alpha = \exists x \mathtt{P}(x)$. What should it mean for an agent, whose beliefs are modeled by a set of situations M, to only-believe α? Since a necessary requirement is that the agent believes α, there must be some closed term a such that $\mathtt{P}(a)$ is believed. It may be tempting to let M be the set of *all* situations where $\mathtt{P}(a)$ holds for some fixed closed term a. But that seems too strong. For example, to say that all the detective believes is that someone is the murderer conveys a lot less information than all the detective believes is that the driver of the red car is the murderer.

One way around this problem is to require the terms that are used when only-believing an existentially quantified sentence to convey no information about the world. In other words, the terms should behave like skolem functions or internal identifiers. For that reason, we introduce a special set of function symbols which must be used when substituting existentials in the context of only-believing. Making sure that these function symbols carry no information about the world is treated as a pragmatic issue when applying the logic to knowledge bases. As we will see in section "A Decidable KR Service," a simple way of dealing with the issue is to enforce that a user interacting with a KB is prevented from asking queries or telling the system sentences containing these special function symbols.

The Language $\mathcal{L}$

The language is a modal first-order dialect with function symbols, which are partitioned into two countably infinite sets $\mathcal{F}_{\mathrm{REG}}$ and $\mathcal{F}_{\mathrm{SK}}$ of every arity. The latter contains the function symbols that will be used to interpret existential quantifiers in the context of only-believing. The members of $\mathcal{F}_{\mathrm{SK}}$ are also referred to as **sk-functions**. The language also contains a countably infinite set N of standard names, which are syntactically treated like constants. Their meaning is explained below.

[6] M need not be unique for the same reasons as there are multiple extensions in autoepistemic logic (see [14, 17]).

Given the usual definitions of **terms** and **atomic formulas**, a **primitive term (formula)** is a term (atomic formula) with only standard names as arguments. We also distinguish a special primitive formula **true** (with the obvious intended meaning).

The **formulas** of $\mathcal{L}$ are constructed in the usual way from the atomic formulas, the connectives $\neg$ and $\vee$, the quantifier $\exists$,[7] and the modal operators **B** and **O** with the restriction that formulas of the form $\mathbf{B}\alpha$ or $\mathbf{O}\alpha$ may not contain free variables. In other words, we ignore the issue of quantifying into belief (*quantifying-in*) as in $\exists x \mathtt{P}(x) \wedge \neg \mathbf{B}\mathtt{P}(x)$. This restriction will be lifted in a forthcoming paper [13]. To simplify the technical presentation below, we also require that no variable is bound more than once in a formula. Formulas without any occurrences of **B** or **O** are called **objective**, formulas without occurrences of **O** are called **basic**, and formulas whose predicate symbols all occur within the scope of a modal operator are called **subjective**. **Sentences** are, as usual, formulas without free variables.

Notation: Sequences of terms or variables are sometimes written in vector notation. E.g., a sequence of variables $\langle x_1, \ldots, x_k \rangle$ is abbreviated as $\vec{x}$. Also, $\exists \vec{x}$ stands for $\exists x_1 \ldots \exists x_k$. If a formula α contains the free variables $x_1, \ldots, x_k$, $\alpha[x_1/t_1, \ldots, x_k/t_k]$ (sometimes abbreviated as $\alpha[\vec{x}/\vec{t}]$) denotes α with every occurrence of x_i replaced by t_i. In the case of one variable, we sometimes write α^x_t instead of $\alpha[x/t]$.

Definition 1 *A quantifier within a formula α occurs at the* **objective level** *of α if it does not occur within the scope of a modal operator.*

Definition 2 *A formula α is* **existential-free** *iff α contains no existential quantifiers at the objective level within the scope of an even number of $\neg$-operators.*

A Formal Semantics

The semantics of *OBL* relies on the notion of situations, which are essentially four-valued extensions of classical (two-valued) Kripke worlds [10]. All situations are defined over a fixed universe of discourse, which we take to be the standard names of the language. None of the results in this paper depends on this choice,[8] but it greatly simplifies the technical presentation. For example, the true- and false-support of predicates can be described by the true- and false-support of primitive formulas. Also, at each situation, the closed terms of the language are interpreted by mapping them into the standard names.

Definition 3 *Denotation Functions*
A **denotation function** *d is a mapping from closed*

[7] Other logical connectives like $\wedge$, $\supset$, and $\equiv$ and the quantifier $\forall$ are used freely and are defined in the usual way in terms of $\neg$, $\vee$, and $\exists$.

[8] The use of standard names as the universe of discourse becomes much more significant in the context of quantifying-in [15, 13].

terms into the standard names such that $d(n) = n$ *for all* $n \in N$ *and* $d(f(t_1, \ldots, t_k)) = d(f(d(t_1), \ldots, d(t_k)))$. *(*$f$ *may be* 0*-ary.)*
d is canonically extended to apply to sequences as well, i.e., $d(\langle t_1, \ldots, t_k \rangle) = \langle d(t_1), \ldots d(t_k) \rangle$.

It is easy to see that denotation functions are uniquely determined by the values they assign to primitive terms.

Definition 4 *Situations*
A situation s is a triple $s = \langle T, F, d \rangle$, *where T and F are subsets of the set of primitive sentences such that* `true` $\in T$ *and* `true` $\notin F$. *d is a denotation function.*

Definition 5 *Worlds*
A situation $\langle T, F, d \rangle$ *is called a* **world**, *iff*

$\mathtt{P}(\vec{n}) \in T \iff \mathtt{P}(\vec{n}) \notin F$ *for all primitive formulas* $\mathtt{P}(\vec{n})$

The following definitions are needed for the semantics of **B** and **O**. In particular, they describe how to substitute terms for existentially quantified variables when interpreting belief (definition 9) and what kinds of terms are eligible in the context of **B** (definition 7) and **O** (definition 8).

Definition 6 *Existentially Quantified Variables*
Let α *be a formula in* $\mathcal{L}$. *A variable x is said to be* **existentially (universally) quantified** *in* α *iff x is bound in the scope of an* **even (odd)** *number of* $\neg$*-operators.*

Definition 7 *Admissible Terms*
Let α *be objective and x existentially quantified in* α. *A term t is said to be an* **admissible** *substitution for x with respect to* α *iff every variable y in t is universally quantified in* α *and x is bound within the scope of y.*
If the context is clear, we often say t is admissible for x *or t* is admissible.

Definition 8 *Sk-terms*
Let α *be a sentence and x an existentially quantified variable bound at the objective level of* α. *Let* $U(x)$ *be a sequence of the universally quantified variables in whose scope x is bound. Let* $f \in \mathcal{F}_{\mathrm{SK}}$ *be a function symbol of arity* $|U(x)|$ *occurring nowhere else in* α. *Then* $f(U(x))$ *is called an* **sk-term** *(for x).*

Note that sk-terms are also admissible.

Definition 9 *Let* α *be a sentence and let* $\vec{x} = \langle x_1, \ldots, x_k \rangle$ *be a sequence of the existentially quantified variables bound at the objective level of* α. *Let* $\vec{t} = \langle t_1, \ldots, t_k \rangle$ *be a sequence of terms s.t.* t_i *is admissible for* x_i *for all i.* $\alpha^{\sharp}[\vec{x}/\vec{t}]$ *denotes* α *with all* $\exists x_i$ *removed and with all occurrences of* x_i *replaced by* t_i.

To illustrate the previous definition, let $\alpha = \exists x(\neg\exists y(\mathtt{P}(x, y) \vee \neg\exists z\mathtt{Q}(z)))$, $t_1 = a$, and $t_2 = f(y)$. Then both t_1 and t_2 are admissible and $\alpha^{\sharp}[x/t_1, z/t_2] = (\neg\exists y(\mathtt{P}(a, y) \vee \neg\mathtt{Q}(f(y))))$.

We are now in a position to define the semantic rules for the sentences of $\mathcal{L}$. The rules except for **B** and **O** are no different from those in classical logic except that they are somewhat more longwinded because the true- and false-support of sentences have to be considered separately.

Let s be a situation and M a set of situations. The true- ($\models_{\mathrm{T}}$) and false-support ($\models_{\mathrm{F}}$) relations for sentences in $\mathcal{L}$ are defined as: (Let $\mathtt{P}(\vec{t})$ be an atomic sentence. α and β are sentences except in rule 4., where α may contain the free variable x.)

1. $M, s \models_{\mathrm{T}} \mathtt{P}(\vec{t}) \iff \mathtt{P}(d(\vec{t})) \in T$, where $s = \langle T, F, d \rangle$
   $M, s \models_{\mathrm{F}} \mathtt{P}(\vec{t}) \iff \mathtt{P}(d(\vec{t})) \in F$
2. $M, s \models_{\mathrm{T}} \neg\alpha \iff M, s \models_{\mathrm{F}} \alpha$
   $M, s \models_{\mathrm{F}} \neg\alpha \iff M, s \models_{\mathrm{T}} \alpha$
3. $M, s \models_{\mathrm{T}} \alpha \vee \beta \iff M, s \models_{\mathrm{T}} \alpha$ or $M, s \models_{\mathrm{T}} \beta$
   $M, s \models_{\mathrm{F}} \alpha \vee \beta \iff M, s \models_{\mathrm{F}} \alpha$ and $M, s \models_{\mathrm{F}} \beta$
4. $M, s \models_{\mathrm{T}} \exists x\alpha \iff$ for some $n \in N$ $M, s \models_{\mathrm{T}} \alpha^x_n$
   $M, s \models_{\mathrm{F}} \exists x\alpha \iff$ for all $n \in N$ $M, s \models_{\mathrm{F}} \alpha^x_n$
   For the following rules, let $\vec{x} = \langle x_1, \ldots, x_k \rangle$ be a sequence of the existentially quantified variables bound at the objective level of α.
5. $M, s \models_{\mathrm{T}} \mathbf{B}\alpha \iff$
   for all s', if $s' \in M$ then $M, s' \models_{\mathrm{T}} \alpha^{\sharp}[\vec{x}/\vec{t}]$,
   where $\vec{t}$ is a sequence of admissible terms.
   $M, s \models_{\mathrm{F}} \mathbf{B}\alpha \iff M, s \not\models_{\mathrm{T}} \mathbf{B}\alpha$
6. $M, s \models_{\mathrm{T}} \mathbf{O}\alpha \iff$
   for all s', $s' \in M$ iff $M, s' \models_{\mathrm{T}} \alpha^{\sharp}[\vec{x}/\vec{t}_{\mathrm{SK}}]$,
   where $\vec{t}_{\mathrm{SK}}$ is a sequence of distinct sk-terms.
   $M, s \models_{\mathrm{F}} \mathbf{O}\alpha \iff M, s \not\models_{\mathrm{T}} \mathbf{O}\alpha$

Note that, in the definition of **B** and **O**, the same $\vec{t}$ or $\vec{t}_{\mathrm{SK}}$ must be chosen for all $s' \in M$. Also note how the definition of only-believing differs only in two places from that of belief. For one, the terms that can be substituted for existentials are restricted to mention exactly one sk-function (the "internal identifiers"). The only other change involves replacing the "if" in the definition of belief by an "iff". This ensures that the set of situations M is as large as possible.

The notions of truth, logical consequence, validity, and satisfiability are defined with respect to worlds and non-empty sets of situations.

A formula α is *true* at a non-empty set of situations M and a world w if $M, w \models_{\mathrm{T}} \alpha$. α is *false* if $M, w \not\models_{\mathrm{T}} \alpha$. A formula α is *valid* ($\models\alpha$) iff α is true at every world w and every non-empty set of situations M. α is *satisfiable* iff $\neg\alpha$ is not valid.[9]

[9] In [17, 14] it is shown that using arbitrary sets of situations has the unintuitive effect that what is only-believed at an epistemic state (represented by a set of situations) is not completely determined by the basic beliefs at that state. This flaw can be overcome by using so-called *maximal* sets of situations [17, 14]. Since this issue is independent from

Finally, if α is objective, we often write $s \models_T \alpha$ instead of $M, s \models_T \alpha$, since nothing in the interpretation of α depends on M (similar for $s \models_F \alpha$).

Properties of Belief

Apart from belief and only-believing, the logic behaves much like a classical first-order logic. For example, all the substitution instances of sentences that are valid in classical FOL are also valid in *OBL*. In the rest of this section, we present important properties of belief.

The following four examples of *invalid* sentences illustrate in what ways belief is not closed under classical logical implication. Let $\mathtt{P}(a)$ and $\mathtt{Q}(b)$ be distinct atomic formulas.

$\not\models \mathbf{B}(\mathtt{P}(a) \vee \neg\mathtt{P}(a))$ Valid sentences need not be believed

$\not\models \mathbf{B}\mathtt{P}(a) \wedge \mathbf{B}(\mathtt{P}(a) \supset \mathtt{Q}(b)) \supset \mathbf{B}\mathtt{Q}(b)$ No Modus Ponens

$\not\models \mathbf{B}\mathtt{P}(a) \wedge \mathbf{B}\neg\mathtt{P}(a) \supset \mathbf{B}\mathtt{Q}(b)$ Inconsistent beliefs do not imply believing everything

$\not\models \mathbf{B}(\mathtt{P}(a) \vee \mathtt{P}(b)) \supset \mathbf{B}\exists x\mathtt{P}(x)$ No existential generalization from disjunctions

While the first three examples are a direct consequence of the four-valued situations, the fourth is a result of the special treatment of existential quantifiers within belief.

Next we list some of the *valid* sentences concerning belief, which give an indication of what can be concluded from a given belief. Let α and β be arbitrary sentences.

$\models \mathbf{B}\mathtt{true} \wedge \neg\mathbf{B}\neg\mathtt{true}$

$\models \mathbf{B}(\alpha \wedge \beta) \equiv \mathbf{B}\alpha \wedge \mathbf{B}\beta.$

$\models \mathbf{B}\alpha \supset \mathbf{B}(\alpha \vee \beta) \wedge \mathbf{B}(\beta \vee \alpha)$

$\models \mathbf{B}\forall x\alpha \supset \mathbf{B}\alpha_t^x$ for any closed term t.

$\models \mathbf{B}\alpha_t^x \supset \mathbf{B}\exists x\alpha$, where x is free in α and t is any closed term.

$\models \mathbf{O}\alpha \supset \mathbf{B}\alpha$

Believing objective sentences is strongly related to Patel-Schneider's t-entailment:

Theorem 1 *OBL subsumes t-entailment*
Let α and β be objective sentences containing neither standard names nor occurrences of `true`.[10] *Then*

$$\models \mathbf{B}\alpha \supset \mathbf{B}\beta \quad \text{iff} \quad \alpha \longrightarrow_t \beta.$$

This result is significant in itself, since it can be viewed as providing a new semantics for t-entailment. The original semantics of t-entailment has the peculiar property of interpreting disjunction in a non-standard way. In *OBL*, all the classical connectives including disjunction are interpreted in the usual way and all the non-standard aspects are pushed into the semantics of **B** and **O**.

We conclude this section with a list of properties concerning self-knowledge. Let α and β be arbitrary sentences and let ρ and σ be subjective sentences.

Perfect Introspection:
$\models \mathbf{B}\alpha \supset \mathbf{B}\mathbf{B}\alpha$ and $\models \neg\mathbf{B}\alpha \supset \mathbf{B}\neg\mathbf{B}\alpha$

Self-Knowledge is Accurate: $\models \mathbf{B}\sigma \supset \sigma$

Self-Knowledge is Complete: $\models \sigma \supset \mathbf{B}\sigma$

Self-Knowledge is consistent: $\models \mathbf{B}\sigma \supset \neg\mathbf{B}\neg\sigma$

Self-Knowledge is Closed Under MP:
$\models \mathbf{B}(\rho \wedge (\neg\rho \vee \sigma)) \supset \mathbf{B}\sigma$

The above results show that an agent with this model of belief has *perfect* knowledge about her own beliefs even if her beliefs about the world are limited.

Computing What an Objective KB Knows

The intuition behind only-believing the sentences in a knowledge base has been to capture what a KB knows or what epistemic state the KB represents. Ideally, one would like OKB to pick out a unique epistemic state from the range of states defined by the logic (in the form of sets of situations). Unfortunately, this is not the case for arbitrary KBs. For one, if the KB is not objective, OKB may be satisfied in multiple epistemic states for the same reason as there are *multiple extensions* in other autoepistemic logics such as [20]. Unlike other autoepistemic logics, OKB does not represent a unique epistemic state even if KB is objective. For example, $\mathbf{O}\exists x\mathtt{P}(x)$ is satisfied by $\{s \mid s \models_T \mathtt{P}(a)\}$ for any constant $a \in \mathcal{F}_{SK}$. On the other hand, all those states are isomorphic up to renaming of sk-functions. Moreover, they agree on all beliefs not mentioning sk-functions. In general, we obtain

Theorem 2 *Let* KB *be an objective sentence. Then for any sentence α not containing function symbols from $\mathcal{F}_{SK}$, exactly one of $\models \mathbf{O}\mathrm{KB} \supset \mathbf{B}\alpha$ or $\models \mathbf{O}\mathrm{KB} \supset \neg\mathbf{B}\alpha$ holds.*

From a KB user's point of view, this result can be explained as follows: the user is not sure which internal identifiers (sk-terms) the KB has chosen for its existentially quantified variables, thus allowing for multiple possible epistemic states. However, the beliefs of the KB that matter to a user are those that are free of sk-terms, and those, according to the theorem, are uniquely determined by the KB. In the next section, this view will be made explicit by defining routines that allow a user to interact with a KB.

In the rest of this section, we prove that it is in fact decidable whether a belief without sk-functions follows from only-believing a KB. Although the decidability result holds for beliefs containing **O**s, we restrict our

---

the main concern of this paper, we have chosen to ignore it here.

[10] These restrictions are necessary in order to match the language of t-entailment.

attention to *basic* beliefs in order to simplify the presentation.

The idea behind the procedure for deciding whether a KB believes α, i.e., whether $\mathbf{O}\text{KB} \supset \mathbf{B}\alpha$ is valid, is as follows. First we replace all occurrences of subsentences of the form $\mathbf{B}\gamma$ in α by **true** or $\neg$**true** depending on whether γ is believed or not. This evaluation proceeds from the innermost occurrence of **B** to the outermost so that at each step we are asking whether an *objective* sentence is believed, which can be computed using Patel-Schneider's decidable t-entailment [22].

To perform the reduction, we need the following definitions.

Definition 10 *Let* KB *and* α *be objective,* α *without sk-function.*

$$\mathtt{RES}[\![\text{KB}, \alpha]\!] = \begin{cases} \mathtt{true} & \text{if } \models \mathbf{O}\text{KB} \supset \mathbf{B}\alpha \\ \neg\mathtt{true} & \text{if } \models \mathbf{O}\text{KB} \supset \neg\mathbf{B}\alpha \end{cases}$$

Definition 11 *Let* KB *be objective and* α *basic.*

$$\begin{array}{lcl} ||\alpha||_{\text{KB}} & = & \alpha, \quad \text{for objective } \alpha \\ ||\neg\alpha||_{\text{KB}} & = & \neg\, ||\alpha||_{\text{KB}} \\ ||\alpha \vee \beta||_{\text{KB}} & = & ||\alpha||_{\text{KB}} \vee ||\beta||_{\text{KB}} \\ ||\exists x\alpha||_{\text{KB}} & = & \exists x\, ||\alpha||_{\text{KB}} \\ ||\mathbf{B}\alpha||_{\text{KB}} & = & \mathtt{RES}[\![\text{KB}, ||\alpha||_{\text{KB}}]\!] \end{array}$$

The following three results are key to establishing decidability.

Lemma 1 *If* KB *is objective and* α *basic without sk-functions, then* $\models \mathbf{O}\text{KB} \supset \mathbf{B}\alpha$ *iff* $\models \mathbf{O}\text{KB} \supset \mathbf{B}\, ||\alpha||_{\text{KB}}$.

Lemma 2 *If* KB *and* α *are objective, then* $\models \mathbf{O}\text{KB} \supset \mathbf{B}\alpha$ *iff* $\models \mathbf{B}\text{KB} \supset \mathbf{B}\alpha$.

Theorem 3 *(Patel-Schneider) t-entailment is decidable.*

With these intermediate results, the main theorem can be proven.

Theorem 4 *The validity problem for sentences of the form* $\mathbf{O}\text{KB} \supset \mathbf{B}\alpha$ *is decidable, assuming that* KB *is an objective sentence and* α *is a basic sentence not containing sk-functions.*

Proof: Lemma 1 implies that deciding whether a KB believes an arbitrary basic sentence reduces to deciding whether it believes an *objective* sentence. Thus let us assume that both KB and α are objective. Next, without loss of generality, we replace every standard name in KB and α by a new constant occurring nowhere else. In addition, we simplify both sentences in case they contain occurrences of **true** (e.g. $\gamma \vee \neg$**true** reduces to γ). Then $\models \mathbf{O}\text{KB} \supset \mathbf{B}\alpha$ iff $\models \mathbf{B}\text{KB} \supset \mathbf{B}\alpha$ (by lemma 2) iff (a) KB = **false** or (b) α = **true** or (c) KB $\longrightarrow_t \alpha$ (by theorem 1), which is decidable (theorem 3). ■

A Decidable KR Service

In this section, we apply the results of this paper to the specification of a KR service in the sense of [15]. The idea is that a KB can be defined in purely functional terms by two operations **ASK** and **TELL** that allow a user to ask the KB queries and to add new information to it. All a user has to know about is an *interaction language* in which to phrase queries and updates. By defining the interaction language to consist of the basic sentences of $\mathcal{L}$ that do not contain sk-functions, the results of the previous section can be readily applied to define **ASK** and **TELL**. Note that, from a user's point of view, the absence of sk-functions is of no concern since there are an infinite supply of other function symbols ($\mathcal{F}_{\text{REG}}$) at hand.

Definition 12 **ASK** *and* **TELL**
Let KB *be an objective sentence and* α *a basic sentence without sk-functions.*

$$\mathtt{ASK}[\![\text{KB}, \alpha]\!] = \begin{cases} \text{YES} & \text{if } \models \mathbf{O}\text{KB} \supset \mathbf{B}\alpha \wedge \neg\mathbf{B}\neg\alpha \\ \text{NO} & \text{if } \models \mathbf{O}\text{KB} \supset \mathbf{B}\neg\alpha \wedge \neg\mathbf{B}\alpha \\ \text{UNK} & \text{if } \models \mathbf{O}\text{KB} \supset \neg\mathbf{B}\alpha \wedge \neg\mathbf{B}\neg\alpha \\ \text{INC} & \text{if } \models \mathbf{O}\text{KB} \supset \mathbf{B}\alpha \wedge \mathbf{B}\neg\alpha \end{cases}$$

$\mathtt{TELL}[\![\text{KB}, \alpha]\!] = \text{KB} \wedge\, ||\alpha||_{\text{KB}}$.

Note that the way TELLing a sentence α to a KB is handled. Any occurrence of a $\mathbf{B}\gamma$ within α is first evaluated with respect to the *old* KB with the effect that an objective KB is always transformed into another objective KB. **ASK** and **TELL** are also implementable, which follows easily from the last section.

Corollary 1 **ASK** *and* **TELL** *are decidable.*

Apart from being decidable, are these routines also *efficient*? To answer this question, note that the complexity for both operations is dominated by the complexity of t-entailment, which follows easily from the way queries are evaluated using definition 10 and 11. Patel-Schneider [22] shows that, while t-entailment is intractable in general, it is indeed tractable under the following assumptions: the KB is in conjunctive normal form (CNF); queries, when converted into CNF, are of size at most $log(|\text{KB}|)$; individual clauses are of constant size; and finally, only $log(|\text{KB}|)$ clauses in the KB subsume a given clause in the query. From a KR point of view, these assumptions seem quite reasonable. The last condition, for example, can be satisfied if the KB uses many different predicates.

Conclusions

In this paper, we have developed a new model of belief and only-believing with perfect introspection for a full first-order language with function symbols. Most importantly, the model of belief has attractive computational properties in that it specifies first-order knowledge bases whose epistemic states are computable and, under certain assumptions, efficiently computable.

There are several ways this framework can be extended. In a forthcoming paper [13], we show how an equality predicate and quantifying-in can be incorporated, which allows us to make important distinctions between "knowing that" and "knowing what."

The deductive component of the current framework is rather weak. One way of increasing its power is by using a sorted logic approach as in [4]. Also, the work by McAllester et. al. [19] seems applicable in this context.

Finally, the logic developed here captures aspects of nonmonotonic reasoning similar to other autoepistemic logics. For example, the default assumption that Tweety flies unless known otherwise is captured by the valid[11] sentence

$$\mathbf{O}[\neg\mathbf{B}\neg\mathtt{Fly}(tweety) \supset \mathtt{Fly}(tweety)] \supset \mathbf{B}\mathtt{Fly}(tweety).$$

In the first-order case, it has so far been very difficult to investigate how default reasoning affects the overall complexity of reasoning because the underlying deductive component is already undecidable. Our framework allows, for the first time, to investigate this issue with a decidable deductive component in hand.

Acknowledgements

I would like to thank Hector Levesque for many stimulating discussions on the subject of modeling belief. His comments on the paper and those of the anonymous referees are greatly appreciated.

References

[1] Belnap, N. D., A Useful Four-Valued Logic, in G. Epstein and J. M. Dunn (eds.), *Modern Uses of Multiple-Valued Logic*, Reidel, 1977.

[2] Davis, M., Obvious Logical Inferences, in *Proceedings of the Seventh International Joint Conference on Artificial Intelligence*, Vancouver, B.C., 1981, pp. 530–531.

[3] Dunn, J. M., Intuitive Semantics for First-Degree Entailments and Coupled Trees, *Philosophical Studies* **29**, 1976, pp. 149–168.

[4] Frisch, A. M., *Knowledge Retrieval as Specialized Inference*, Ph.D. Thesis, University of Rochester, Department of Computer Science, 1986.

[5] Halpern, J. Y. and Moses, Y. O., Towards a Theory of Knowledge and Ignorance: Preliminary Report, in Proceedings of The Non-Monotonic Workshop, New Paltz, NY, 1984, pp.125–143.

[6] Halpern, J. Y. and Moses, Y. O., A Guide to the Modal Logics of Knowledge and Belief, in *Proc. of the Ninth International Joint Conference on Artificial Intelligence*, Los Angeles, CA, 1985, pp. 480–490.

[7] Ketonen, J. and Weyhrauch, R., A Decidable Fragment of Predicate Calculus, *Theoretical Computer Science* **32**, 1984, pp. 297–307.

[8] Hintikka, J., *Knowledge and Belief: An Introduction to the Logic of the Two Notions*, Cornell University Press, 1962.

[11] Note that, even though *OBL* itself is monotonic, the epistemic states of KBs, which are specified in terms of only-believing, are nonmonotonic.

[9] Konolige, K., A Computational Theory of Belief Introspection. In *Proceedings of the Ninth International Conference on Artificial Intelligence*, Los Angeles, 1985, pp. 502–508.

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

[11] Lakemeyer, G., Steps Towards a First-Order Logic of Explicit and Implicit Belief, in *Proc. of the Conference on Theoretical Aspects of Reasoning about Knowledge*, Asilomar, California, 1986, pp. 325–340.

[12] Lakemeyer, G., Decidable Reasoning in First-Order Knowledge Bases with Perfect Introspection, Technical Report, Department of Computer Science, University of Toronto, in preparation.

[13] Lakemeyer, G., A Model of Decidable, Introspective Reasoning with Quantifying-In, in preparation.

[14] Lakemeyer, G. and Levesque, H. J., A Tractable Knowledge Representation Service with Full Introspection, in *Proc. of the Second Conference on Theoretical Aspects of Reasoning about Knowledge*, Asilomar, California, 1988, pp. 145–159.

[15] Levesque, H. J., Foundations of a Functional Approach to Knowledge Representation, *Artificial Intelligence*, **23**, 1984, pp. 155-212.

[16] Levesque, H. J., A Logic of Implicit and Explicit Belief, Tech. Rep. No. 32, Fairchild Lab. for AI Research, Palo Alto, 1984.

[17] Levesque, H. J., All I Know: A Study in Autoepistemic Logic, *Artificial Intelligence*, North Holland, **42**, 1990, pp. 263–309.

[18] Marek, W. and Truszczyński, M., Relating Autoepistemic and Default Logics. in *Proc. of the First International Conference on Principles of Knowledge Representation and Reasoning*, Morgan Kaufmann, San Mateo, CA, 1989, pp. 276–288.

[19] McAllester, D., Givan, B., and Fatima, T., Taxonomic Syntax for First Order Inference, in *Proc. of the First Int. Conf. on Principles of Knowledge Representation and Reasoning*, Morgan Kaufmann, San Mateo, 1989, pp. 289–300.

[20] Moore, R. C., Semantical Considerations on Nonmonotonic Logic, in *Proc. of the Eighth International Joint Conference on Artificial Intelligence*, Karlsruhe, FRG, 1983, pp. 272–279.

[21] Moore, R. C., Possible World Semantics for Autoepistemic Logic, in *The Non-Monotonic Reasoning Workshop*, New Paltz, NY, 1984, pp. 344–354.

[22] Patel-Schneider, P. F., *Decidable, Logic-Based Knowledge Representation*, Ph.D thesis, University of Toronto, 1987.