A Model of Decidable Introspective Reasoning with Quantifying-In

Gerhard Lakemeyer*
Institut fur Informatik III
Universitat Bonn
Romerstr. 164
W-5300 Bonn 1, Germany
e-mail: gerhard@uran.informatik.uni-bonn,de

Abstract

Since knowledge is usually incomplete, agents need to introspect on what they know and do not know. The best known models of introspective reasoning suffer from intractability or even undecidability if the underlying language is first-order. To better suit the fact that agents have limited resources, we recently proposed a model of decidable introspective reasoning in first-order knowledge bases (KBs). However, this model is deficient in that it does not allow for quantifying-in, which is needed to distinguish between *knowing that* and *knowing who*. In this paper, we extend our earlier work by adding quantifying-in and equaLity to a model of limited belief that integrates ideas from possible-world semantics and relevance logic.

1 Introduction

Since agents rarely have complete information about the world, it is important for them to introspect on what they know and, more importantly, do not know. For example,

> if somebody tells you that Sue's father is a teacher and you have no other information about Sue's father, then introspection (in addition to deduction) allows you to conclude that there is a teacher and that you do *not know who* that teacher is, that is, as far as you know, Sue's father could be any of a number of individuals.

There have been various attempts to formalize introspective reasoning, most notably in the guise of the so-called *autoepisiemic logics* (e.g. [18, 17]). While providing a very elegant framework for introspection, these logics have a major drawback in that they assume an ideal reasoner with infinite resources. In particular, in the first-order case, reasoning is undecidable. It is therefore of particular interest to devise models of introspective reasoning which are better suited for agents with limited resources.

For that purpose, a model of a tractable introspective reasoner was proposed for a propositional language in [12]. Since its obvious first-order extension leads to an undecidable reasoner, we proposed a modification which retains decidability in [13]. However, this proposal is still too limited since it lacks the expressiveness to deal with incomplete knowledge as exhibited in our initial example. In particular, it does not allow us to make distinctions between *knowing that* and *knowing who* because the underlying language does not provide for *quantifytng-in* [6], that is, the ability to use variables *within* a belief[1] that are bound *outside* the belief. With quantifying-in, the above example can easily be expressed as (we use the modal operator B for belief)

$$\text{Teacher}(z) \text{ A } \text{-BTeacher}(x),$$

a sentence that should follow from an introspective KB that contains only the sentence Teacher(father(sue)).

In this paper, we extend the results of [13] by considering a language with quantifying-in and equality. It is not at all obvious whether adding quantifying-in allows us to retain a decidable reasoner. As Konolige observed [8], while introspective reasoning in classical monadic predicate calculus is decidable, it becomes undecidable if we add quantifying-in. As a result, Konolige makes the following comment:

> Thus the piesence of quantifying-in seems to pose an inherently difficult computational problem for introspective systems.

In this paper we show that, given an arbitrary first-order KB, it is decidable for a large class of sentences with quantifying-in whether or not these sentences follow from the KB.

One way to formalize reasoning is to view the problem as one of modeling belief. In a nutshell, a model of belief tells us what the possible sets of beliefs or e*pistemic states* of an agent are. One then needs to specify for any given KB which epistemic state it represents. Under this view, reasoning reduces to testing for membership in the appropriate epistemic state.

As in [12, 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

*This work was conducted at the University of Toronto.

[1] We use the terms *knowledge* and *belief* interchangeably in this paper, even though belief is the more appropriate term, since we allow an agent to have false beliefs.

will say for short , *only-believed.* We formalize this idea using a modal logic with two modal operators B and O for belief and only-believing, respectively. This allows us to say that a KB believes a sentence *a* just in case OKB D Ba is a valid sentence[2] of the logic, thus characterizing the epistemic state of the KB. The complexity of reasoning then reduces to the complexity of solving this validity problem.

In related work, Konolige [8] 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. Since we consider a limited introspective reasoner who is able to perform full introspection and is only limited in his deductive component, work on limited deduction alone is also relevant [7, 2, 19, 4]. In particular, as discussed in [13], [19] is a special case of ours. Finally, in preliminary work [11], we proposed a model of limited belief with quantifying-in yet without nested beliefs. As a result, the corresponding reasoner was purely deductive and not able to make use of quantifying-in himself.

In the next section, we introduce the logic *OBLIQUE,*[3] which defines the model of belief and only-believing. In Section 3, we take a closer look at the epistemic states of KB's as defined by *OBLIQUE.* Section 4 shows the computational pay-off of using this particular limited form of belief. In Section 5, we use the logic to define a KR service that allows a user to query a KB and to add new information to it. Finally, we end the paper with a brief summary and an outlook on future work.

2 The Logic *OBLIQUE*

We begin with a discussion of belief and only-believing.

Belief

As in in [13], belief is modeled by integrating ideas from possible-world semantics [5, 9] and relevance logic [1, 3]. Roughly, an agent believes a sentence just in case that sentence holds in all states of affairs or *situations* the agent imagines. In order to obtain agents with perfect introspection we require that, similar to a semantics of the modal logic *weak* 55, that every model has one globally accessible set of situations. Situations are a four-valued extension of classical worlds. Instead of facts being either true or false, situations assign them independent *true* and *false-support*, which corresponds to the use of four truth values {}, {true}, {false}, and {true,false}, an idea originally proposed to provide a semantics for a fragment of relevance logic called *tautological entailment* [1, 3].[4]

In order to be able to distinguish between *knowing that* and *knowing who,* we follow [17] and use a language with both *rigid* and *non-rigid* designators (see [10]). The non-rigid designators are the usual terms of a first-order language such as *father(sue),* which may vary in their interpretation. The rigid designators are special unique identifiers called *standard names.* For simplicity, the standard names are taken to be the universe of discourse in our semantics.

Employing four-valued situations instead of worlds has the effect that beliefs are no longer closed under *modus ponens,* e.g. B(p V *q)* and *B(-q* V r) may be true and B(p V r) may be false at the same time. As discussed in [13], a further restriction is needed in order to use this model of belief as a basis for a decidable reasoner. In particular, the link between disjunction and existential quantification is weakened in the sense that an agent may believe P(a) VP(6), yet fail to believe 3xP(x). In the case of beliefs without quantifying-in, this can be achieved semantically by requiring 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. More concretely, in order to believe 3xP(x) there must be a closed term *i* (e.g. *father(sue))* such that *?(t)* is true in all accessible situations (see [13]).

In the case of beliefs with quantifying-in, this idea of simply substituting terms for existentially quantified variables does not suffice. E.g., given the belief 3xTeacher(x) A -BTeacher(x), if we replace *x* by any term, say *father(sue),* then the resulting belief is inconsistent because for an introspective agent to believe that Teacher(father(sue)) A ->BTeacher(/ai/*er(sue)) means that he both believes and does not believe that Teacher(father(sue)). What is wrong is that we should not have substituted *father(sue)* for the second occurrence of *x* (within the context of B). Instead, what we really want at its place is the *denotation of father (sue)* so that, while Teacher(father(sue)) holds at every situation the agent imagines, the agent does not know *of* the denotation of *father(sue)* at any given situation that he is a teacher, that is, the agent does not know *who* the father of Sue is. To make this distinction between a term and its denotation we introduce a so-called *level marker* ,0 which is attached to a term whenever the term is substituted *within* the context of a modal operator. In our example, the substitution results in Teacher(fatherer(sue)) A -iBTeacher(/a/Aer(sue).0). Later we will return to this example and demonstrate formally how the use of level markers has the desired effect.[5]

Only-Believing

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

As demonstrated in [12, 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 se-

[3] Whenever KB occurs within a logical sentence, we mean the conjunction of all the sentences in the KB.

[3]Thanks to Hector Levesque, who suggested that name to me. It may be read as "Only Belief Logic with Quantifiers and Equality."

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

[5]In the logic, we allow an infinite number of distinct level markers. While not apparent in this paper, this choice was made for technical convenience. The reader may simply ignore all level markers other than .0.

mantics, 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 is as large as possible, i.e., if we were to add any other situation to M, the agent would no longer believe α.[6]

With the special treatment of existential quantification as outlined above, we need to pay special attention to the case of only-believing sentences that contain existential quantifiers (see also [13]). Consider the example of only-believing $\alpha = \exists x \mathrm{P}(x)$. Since α is believed, we need a term t such that $\mathrm{P}(t)$ is believed as well. However, t should not carry any information about the world (where should the information come from?). Thus t has to be a generic term much like a Skolem function.

For that reason, we introduce a special set of function symbols called *sk-functions*, which must be used when substituting existentials in the context of only-believing. To obtain the desired effect, we allow KB's to contain sk-functions while excluding them from epistemic states. This way, given an sk-function t_{sk} and a KB $= \mathrm{P}(t_{sk})$, the beliefs that follow from KB are the same as if KB $= \exists x \mathrm{P}(x)$.

2.1 The Languages $\mathcal{L}$ and $\mathcal{BL}$

We introduce a language $\mathcal{L}$, which allows us to talk *about* the beliefs of a KB and a language $\mathcal{BL}$, which is a sublanguage of $\mathcal{L}$ and which contains all the sentences that qualify as possible beliefs of a KB. For example, the operator **O** may be used to talk about the beliefs of a KB but it may not appear within a belief itself.[7]

The language $\mathcal{L}$ is a modal first-order dialect with equality and function symbols, which form a countably infinite set $\mathcal{F}$, which itself is partitioned into two countably infinite sets $\mathcal{F}_{\mathrm{REG}}$ and $\mathcal{F}_{\mathrm{SK}}$ of every arity. The latter contains the sk-functions mentioned earlier. The language also contains a countably infinite set $N = \{\#1, \#2, \ldots\}$ of standard names, which are syntactically treated like constants. Finally the **level markers** form a countably infinite and totally ordered set with a least element. They are written as $.i$, where i is a natural number. The ordering is $.0 < .1 < .2 < .3 < \ldots$

Given the usual definitions of **terms**, a **primitive term** is a term with only standard names as arguments. Given a term t, an **extended term** $\hat{t}$ is obtained from t by appending zero or more subterms of t with level markers such that, if $u.i$ occurs in $\hat{t}$, then u is not a variable and does not contain level markers.[8] Atomic formulas (or atoms) are predicate symbols whose arguments are extended terms. **Primitive formulas** are atoms with standard names as arguments.

The **formulas** of $\mathcal{L}$ are constructed in the usual way from the atomic formulas, the connectives $\neg$ and $\vee$, the quantifier $\exists$,[9] and the modal operators **B** and **O**. To simplify the technical presentation below, we also require that no variable is bound more than once within the scope of a modal operator of 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.

The language $\mathcal{BL}$, over which the epistemic states of KB's will be defined, is a sublanguage of $\mathcal{L}$ and consists of all those basic formulas of $\mathcal{L}$ that contain neither level markers nor sk-functions. We often use the terms **$\mathcal{BL}$-formulas** and **$\mathcal{BL}$-sentences** to refer to the formulas and sentences of $\mathcal{BL}$.

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 $x_1, \ldots, x_k$, $\alpha[x_1/t_1, \ldots, x_k/t_k]$ (abbreviated as $\alpha[\vec{x}/\vec{t}]$) denotes α with every occurrence of x_i replaced by t_i. If the context is clear, we omit the variables and write $\alpha[t_1, \ldots, t_k]$ or $\alpha[\vec{t}]$ instead. The truth values **true** and **false** are used as shorthand for $(\#1 = \#1)$ and $(\#1 \neq \#1)$, respectively.

The following definitions are needed for the semantics of **B** and **O**. In particular, Definitions 2 and 3 describe what terms can be substituted for the existentially quantified variables when interpreting **B** and **O**, respectively.

Definition 1 *A variable occurs (is bound) at the* **objective level** *of α if it occurs (is bound) outside the scope of any modal operator in α.*
Let α be a formula in $\mathcal{L}$. Let x be a variable that is bound at the objective level of some formula β such that either $\beta = \alpha$ or $\mathbf{B}\beta$ is a subformula of α. x is said to be **existentially (universally) quantified** *in α iff x is bound within in the scope of an* **even (odd)** *number of $\neg$-operators in β.*

For example, in $\exists x \neg \mathbf{B} \exists y \mathrm{P}(x, y)$, both x and y are considered existentially quantified.

Definition 2 *Admissible Terms*
Let α be a formula 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 occurs within the scope of y.*

If the context is clear, we often say t is admissible for *x or t* is admissible.

Definition 3 *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).*

Definition 4 *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. $\alpha^{\not\exists}$ denotes α with all $\exists x_i$ removed.*

[6] M need not be unique for the same reasons as there are multiple extensions in autoepistemic logic (see [12, 17]).

[7] This restriction of $\mathcal{BL}$ was chosen to simplify the technical treatment. Besides, there seems to be little practical use for beliefs about whether something is *all* that is believed.

[8] E.g., given a term $f(a)$ (a constant), $f(a.0)$ and $f(a).14$ are extended terms, but $f(a).1.3$ and $f(a.1).14$ are not.

[9] 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 $\exists$.

Example 2.1 Let $\alpha = \exists w \forall x \exists y \mathsf{P}(w, x, y) \wedge \mathbf{B}\exists z \mathsf{Q}(z)$. Then $\alpha^{\not\exists} = \forall x \mathsf{P}(w, x, y) \wedge \mathbf{B}\exists z \mathsf{Q}(z)$. Note that existential quantifiers within modalities are left untouched.

Definition 5 *Let α be a formula with free variables $\vec{x} = \langle x_1, \ldots, x_k \rangle$. ($\alpha$ may contain other free variables as well.) Let $\vec{t} = \langle t_1, \ldots, t_k \rangle$ be a sequence of terms.*
$\alpha\mathbf{[}\vec{x}/\vec{t}\mathbf{]}$ is α with every occurrence of x_i at the objective level replaced by t_i and every occurrence of x_i inside the scope of a modal operator replaced by $t_i.0$.

Example 2.2
Let $\alpha = \mathsf{P}(x_1) \wedge \mathbf{B}(\neg \mathsf{Q}(x_1) \vee \mathsf{R}(x_2))$. Then $\alpha\mathbf{[}x_1/a, x_2/b\mathbf{]} = \mathsf{P}(a) \wedge \mathbf{B}(\neg \mathsf{Q}(a.0) \vee \mathsf{R}(b.0))$. Note the difference to $\alpha[x_1/a, x_2/b] = \mathsf{P}(a) \wedge \mathbf{B}(\neg \mathsf{Q}(a) \vee \mathsf{R}(b))$, that is, $[\ldots]$ indicates regular substitutions, while $\mathbf{[}\ldots\mathbf{]}$ indicates that substitutions within modalities are appended with level marker $.0$.

2.2 A Formal Semantics

We first define situations, which are a four-valued extension of classical (two-valued) Kripke worlds [9]. Situations are defined over a fixed universe of discourse, which is the set of standard names of the language. This allows us to describe the true- and false-support of predicates in terms of the primitive formulas. Closed terms are interpreted by mapping them into the standard names.

Definition 6 *Denotation Function*
A **denotation function** *d is a mapping from closed terms into the standard names such that $d(n) = n$ for all $n \in N$ and $d(f(t_1, \ldots, t_k)) = d(f(d(t_1), \ldots, d(t_k)))$ for all closed non-rigid terms $f(t_1, \ldots, t_k)$. We extend d to apply to terms with level markers. If t is a closed term and $.i$ a level marker, then $d(t.i) = d(t)$.*

Definition 7 *First-Order Situations*
A situation s is a triple $s = < T, F, d >$, where T and F are subsets of the set of primitive formulas and d is a denotation function.

T and F can be arbitrary sets of primitive formulas except for equality, which has a fixed interpretation, that is, $(n = m) \in T$ [$(n = m) \in F$] iff n and m are identical [distinct] standard names.

Notice that equality has a standard two-valued interpretation. The main reason why we have chosen such a strong form of equality is to obtain Theorem 2.

Definition 8 *Worlds*
A situation w is called a **world**, *iff*
$\mathsf{P}(\vec{n}) \in T_w \Leftrightarrow \mathsf{P}(\vec{n}) \notin F_w$ for all primitive formulas $\mathsf{P}(\vec{n})$.

Definition 9 *Let α be a sentence and $s = \langle T_s, F_s, d_s \rangle$ a situation.*

- *$\alpha^{\blacktriangledown}$ is α with every level marker $.i$ replaced by $.i\text{-}1$ for all $.i > .0$.*
- *α^{s} is obtained from α by replacing every occurrence of $t.0$ by the standard name $d_s(t)$, if t is closed, and by t otherwise.*

Example 2.3 If $\alpha = \mathsf{P}(a.0) \wedge \forall x \mathbf{B}(\mathsf{Q}(f(x).0) \vee \mathsf{R}(g(x).1))$ and $s = < T_s, F_s, d_s >$ with $d_s(a) = \#1$, then $\alpha^{s\blacktriangledown} = \mathsf{P}(\#1) \wedge \forall x \mathbf{B}(\mathsf{Q}(f(x)) \vee \mathsf{R}(g(x).0))$.

Note that $\alpha^{s\blacktriangledown}$ is always well defined, since α^{s} does not contain any occurrences of the level marker $.0$.

The true- and false-support for the sentences of $\mathcal{L}$ can now be defined. Let $s = \langle T_s, F_s, d_s \rangle$ be a situation and M a set of situations. Let $\mathsf{P}(\vec{t})$ be an atomic sentence and let α and β be sentences except in rule 4 where α may contain the free variable x.

1. $M, s \models_{\mathrm{T}} \mathsf{P}(\vec{t}) \iff \mathsf{P}(d_s(\vec{t})) \in T_s$
   $M, s \models_{\mathrm{F}} \mathsf{P}(\vec{t}) \iff \mathsf{P}(d_s(\vec{t})) \in F_s$
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 $n \in N$, $M, s \models_{\mathrm{F}} \alpha[x/n]$

   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. $M, s \models_{\mathrm{T}} \mathbf{B}\alpha \iff$ there are admissible $\vec{t}$ such that
   for all s', if $s' \in M$ then $M, s' \models_{\mathrm{T}} \alpha^{s\blacktriangledown\not\exists}\mathbf{[}\vec{x}/\vec{t}\mathbf{]}$
   $M, s \models_{\mathrm{F}} \mathbf{B}\alpha \iff M, s \not\models_{\mathrm{T}} \mathbf{B}\alpha$
6. $M, s \models_{\mathrm{T}} \mathbf{O}\alpha \iff$ there is a sequence of sk-terms $\vec{t}_{\mathrm{SK}}$ for $\vec{x}$ such that
   for all s', $s' \in M$ iff $M, s' \models_{\mathrm{T}} \alpha^{s\blacktriangledown\not\exists}\mathbf{[}\vec{x}/\vec{t}_{\mathrm{SK}}\mathbf{]}$
   $M, s \models_{\mathrm{F}} \mathbf{O}\alpha \iff M, s \not\models_{\mathrm{T}} \mathbf{O}\alpha$

Example 2.4 Let $M = \{s \mid s \models_{\mathrm{T}} \mathsf{P}(a)\}$ for some constant a. Then $M \models_{\mathrm{T}} \mathbf{B}(\exists x \mathsf{P}(x) \wedge \neg \mathbf{B}\mathsf{P}(x))$.
Proof: Let $\alpha = \exists x \mathsf{P}(x) \wedge \neg \mathbf{B}\mathsf{P}(x)$. By definition, $M \models_{\mathrm{T}} \mathbf{B}\alpha$ iff $\forall s \in M$, $M, s \models_{\mathrm{T}} \alpha^{s\blacktriangledown\not\exists}\mathbf{[}x/t\mathbf{]}$ for some admissible t. Note that, in this case, $\alpha^{s\blacktriangledown} = \alpha$ because α does not contain any level markers.
Then $\alpha^{s\blacktriangledown\not\exists}\mathbf{[}x/a\mathbf{]} = \mathsf{P}(a) \wedge \neg \mathbf{B}\mathsf{P}(a.0)$. It suffices to show that for all $s \in M$, $M, s \models_{\mathrm{T}} \mathsf{P}(a) \wedge \neg \mathbf{B}\mathsf{P}(a.0)$. Let $s = \langle T_s, F_s, d_s \rangle$ be any situation in M. $M, s \models_{\mathrm{T}} \mathsf{P}(a)$ follows immediately from the definition of M.
To show that $M, s \models_{\mathrm{T}} \neg \mathbf{B}\mathsf{P}(a.0)$, assume that $d_s(a) = n$ for some standard name n. By the definition of M, there must be a situation $s' \in M$ with a different denotation function such that $s' \not\models_{\mathrm{T}} \mathsf{P}(n)$. Therefore, $M, s \models_{\mathrm{T}} \neg \mathbf{B}\mathsf{P}(n)$ which implies $M, s \models_{\mathrm{T}} \neg \mathbf{B}\mathsf{P}(a.0)$.

Truth, *logical consequence*, *validity*, and *satisfiability* are defined with respect to worlds and sets of situations. A sentence α is *true* at a 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 sentence α is *valid* ($\models\alpha$) iff α is true at every world w and every set of situations M. α is *satisfiable* iff $\neg\alpha$ is not valid.[10]

2.3 Quantifying-in

For sentences without quantifying-in and without level-markers, *OBLIQUE* reduces essentially to the logic *OBL* in [13].[11] For example, we obtain the same limitations of belief such as no *modus ponens* ($\not\models \mathbf{B}p \wedge$

[10] In this semantics, the basic beliefs of an epistemic state (represented by a set of situations) do not completely determine what is only-believed at that state. As shown in [14], this problem can be overcome. Since this issue is independent from the main concern of this paper, we have chosen to ignore it here,

[11] A minor distinction is that we allow the empty set of situations in the definition of truth and validity, while we excluded it in [13].

$\mathbf{B}(p \supset q) \supset \mathbf{B}q$) and no existential generalization from disjunctions ($\not\models \mathbf{B}(\mathtt{P}(a) \vee \mathtt{P}(b)) \supset \mathbf{B}\exists x \mathtt{P}(x)$).

Here we focus on the additional feature of ***quantifying-in***, which allows us to properly distinguish between ***knowing that*** and ***knowing who***. For example,
$\models \mathbf{B}\mathtt{Teacher}((father(sue))) \supset \mathbf{B}\exists x \mathtt{Teacher}(x)$ while
$\not\models \mathbf{B}\mathtt{Teacher}((father(sue))) \supset \exists x \mathbf{B}\mathtt{Teacher}(x)$.
In other words, knowing the non-rigid designator of a teacher ($father(sue)$) allows an agent to conclude only the existence of a teacher, but the agent does not necessarily know who it is. On the other hand, if he knows the teacher's standard name, say #27, then he knows who the teacher is:
$\models \mathbf{B}(\mathtt{Teacher}(f(sue)) \wedge f(sue) = \#27) \supset \exists x \mathbf{B}\mathtt{Teacher}(x)$.
In general, we obtain $\models \exists x \mathbf{B}\alpha \supset \mathbf{B}\exists x \alpha$ for all α.

The *Barcan formula* ($\forall x \mathbf{B}\alpha \supset \mathbf{B}\forall x \alpha$) is not valid in general. In a sense, the agent is not able to perform arbitrary universal generalization. However, if we restrict α to formulas where no existentially quantified variable is bound at the objective level of α, the Barcan formula is indeed valid. Finally, the converse of the Barcan formula, $\models \mathbf{B}\forall x \alpha \supset \forall x \mathbf{B}\alpha$, is valid for all α.

3 Epistemic States of First-Order KB's

Besides offering us a model of limited belief with quantifying-in, *OBLIQUE* specifies for every objective KB a unique corresponding epistemic state, if we take $\mathcal{BL}$ to be the agent's language of belief.[12]

Theorem 1 *Let* KB *be an objective sentence. For any basic* α *in* $\mathcal{BL}$, *exactly one of* $\models \mathrm{OKB} \supset \mathbf{B}\alpha$ *and* $\models \mathrm{OKB} \supset \neg\mathbf{B}\alpha$ *holds.*[13]

One important property of such epistemic states is that the question whether or not an arbitrary belief follows from an objective KB reduces to the question whether an *objective* belief, that is a belief that itself does not mention any **B**'s, follows from the KB. For example, let KB = $\mathtt{Teacher}(\#27) \wedge (\mathtt{Teacher}(father(sue)) \wedge (father(sue) \neq \#27)$, where #27 is a standard name. Then determining the validity of $\mathrm{OKB} \supset \mathbf{B}(\exists x \mathtt{Teacher}(x) \wedge \neg \mathbf{B}\mathtt{Teacher}(x))$ reduces to determining the validity of $\mathrm{OKB} \supset \mathbf{B}(\exists x \mathtt{Teacher}(x) \wedge \neg(x = \#27))$, that is, the formula $\mathbf{B}\mathtt{Teacher}(x)$ is replaced by an equality expression which describes the set of all standard names who are known to be teachers. In general, it turns out that, even if the set of instances that are known to have a certain property is infinite, there is always a finite equality expression that describes it (Definitions 10 and 11).[14]

Definition 10 *Let* KB *be an objective sentence and let* α *be a* $\mathcal{BL}$*-formula, possibly containing free variables. Let* $n_1, \ldots, n_m$ *be all the standard names occurring in* KB *or* α, *and let* n^* *be a standard name not occurring in* KB *or* α.

1. If α is closed then
$$\mathbf{RES_B}(\mathrm{KB}, \alpha) = \begin{cases} \mathtt{true} & \models \mathrm{OKB} \supset \mathbf{B}\alpha \\ \mathtt{false} & \models \mathrm{OKB} \supset \neg\mathbf{B}\alpha \end{cases}$$
2. Otherwise let α contain the free variable x. Then
$$\mathbf{RES_B}(\mathrm{KB}, \alpha) = \begin{cases} \bigvee((x = n_i) \wedge \mathbf{RES_B}(\mathrm{KB}, \alpha^x_{n_i})) \vee \\ (\bigwedge(x \neq n_i) \wedge \mathbf{RES_B}(\mathrm{KB}, \alpha^x_{n^*})^{n^*}_x) \end{cases}$$

The next definition shows how to apply $\mathbf{RES_B}$ recursively to all occurrences of **B** within a belief and thus obtain an objective belief.

Definition 11 *Let* KB *be an objective sentence and let* α *be a* $\mathcal{BL}$*-formula.*

$$\begin{aligned} \|\alpha\|_{\mathrm{KB}} &= \alpha, \quad \text{for objective } \alpha \\ \|\neg\alpha\|_{\mathrm{KB}} &= \neg\|\alpha\|_{\mathrm{KB}} \\ \|\alpha \vee \beta\|_{\mathrm{KB}} &= \|\alpha\|_{\mathrm{KB}} \vee \|\beta\|_{\mathrm{KB}} \\ \|\exists x \alpha\|_{\mathrm{KB}} &= \exists x \|\alpha\|_{\mathrm{KB}} \\ \|\mathbf{B}\alpha\|_{\mathrm{KB}} &= \mathbf{RES_B}(\mathrm{KB}, \|\alpha\|_{\mathrm{KB}}) \end{aligned}$$

Lemma 3.1 $\|\alpha\|_{\mathrm{KB}}$ *is objective.*

Theorem 2 *Let* KB *be an objective sentence and* α *a* $\mathcal{BL}$*-sentence. Then* $\models \mathrm{OKB} \supset \mathbf{B}\alpha$ iff $\models \mathrm{OKB} \supset \mathbf{B}\|\alpha\|_{\mathrm{KB}}$.

This result is crucial in obtaining decidability results for the reasoner specified by *OBLIQUE*.

4 A Computational Pay-Off

While *OBLIQUE* specifies an admittedly weak reasoner, we gain a clear computational pay-off in return. In particular, we obtain a decidable reasoner for a large class of sentences including ones with quantifying-in. Let us call an equality expression $t = t'$ *ground* if both t and t' are closed terms.

Theorem 3 *Let* KB = KB$' \wedge E$ *be an objective sentence, where* KB$'$ *contains no equality and* E *is a conjunction of ground equalities and inequalities. Let* α *be a* $\mathcal{BL}$*-sentence such that every free variable* x *in a subformula* $\mathbf{B}\gamma$ *in* α *is universally quantified. (In other words, only universally quantified variables may participate in quantifying-in.) In addition we require that every term in an equality expression in* α *is either a universally quantified variable or a closed term. Then the validity of* $\mathrm{OKB} \supset \mathbf{B}\alpha$ *is decidable.*

Proof : (Sketch) While the proof is rather involved, it is instructive to sketch the main ideas.

First, it can be shown that we can assume, without loss of generality, that the existentially quantified variables in KB$'$ are skolemized and KB$'$ is in prenex conjunctive normal form.

Next, given Theorem 2, a simple induction on the nesting of B's shows that $\models \mathrm{OKB} \supset \mathbf{B}\alpha$ is decidable iff the same holds for *objective* α. (Note that the $\|.\|_{\mathrm{KB}}$-transformation introduces only equalities that satisfy the restrictions of the theorem.) Thus let us assume from now on that α is objective. Also, let $\vec{x}$ and $\vec{y}$ be the existentially and universally quantified variables of α, respectively, and let α' be the matrix of α (all the quantifiers removed) in conjunctive normal form.

For an objective α it is easy to see that $\models \mathrm{OKB} \supset \mathbf{B}\alpha$ iff $\models \mathbf{B}\mathrm{KB} \supset \mathbf{B}\alpha$. Next, $\models \mathbf{B}\mathrm{KB} \supset \mathbf{B}\alpha$ iff there are admissible terms for $\vec{x}$ in α such that for all the substitutions of standard names $\vec{n} \in I^l$ $\models \mathbf{B}\mathrm{KB} \supset \forall \vec{y} \mathbf{B}(\alpha'[\vec{x}/\vec{t}])[\vec{y}/\vec{n}]$, where $I = \{n \mid n \text{ occurs in KB or } \alpha\} \cup \{n^*_1, \ldots, n^*_l\}$. ■

[12] If the KB is not objective, the logic may not give us a unique epistemic state for the same reasons as there are multiple extensions in autoepistemic logic [18].

[13] Proofs are generally omitted for lack of space (see [14]).

[14] The same process works for Levesque [17] in the case of an ideal reasoner with unlimited resources.

Allowing arbitrary equalities within a KB is problematic because equality has a 2-valued semantics, which would result in hard reasoning such as *modus ponens.*

On the other hand, while we have not yet proven that reasoning remains decidable if we allow arbitrary forms of quantifying-in, results in [14] strongly suggest that this is indeed the case. In any event, Theorem 3 can be used to prove the decidability of special cases such as $\models \mathbf{O}\mathrm{KB} \supset \mathbf{B}(\exists x\alpha \wedge \neg\mathbf{B}\beta)$, where α and β are objective formulas containing the free variable x.

5 ASK and TELL

In this section, we apply the results of this paper to the specification of a KR service in the sense of [15], which allows a user to query a KB and to add new information to it using $\mathcal{BL}$ as the interaction language.

Definition 12 ASK *and* TELL
Let KB *be an objective sentence and* α *a* $\mathcal{BL}$*-sentence.*

$$\mathbf{ASK}(\mathrm{KB},\alpha) = \begin{cases} \mathrm{YES} & \text{if } \models \mathbf{O}\mathrm{KB} \supset \mathbf{B}\alpha \wedge \neg\mathbf{B}\neg\alpha \\ \mathrm{NO} & \text{if } \models \mathbf{O}\mathrm{KB} \supset \mathbf{B}\neg\alpha \wedge \neg\mathbf{B}\alpha \\ \mathrm{UNK} & \text{if } \models \mathbf{O}\mathrm{KB} \supset \neg\mathbf{B}\alpha \wedge \neg\mathbf{B}\neg\alpha \\ \mathrm{INC} & \text{if } \models \mathbf{O}\mathrm{KB} \supset \mathbf{B}\alpha \wedge \mathbf{B}\neg\alpha \end{cases}$$

$\mathbf{TELL}(\mathrm{KB},\alpha) = \mathrm{KB} \wedge \|\alpha\|_{\mathrm{KB}}$.

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 and replaced by an (objective) equality expression. Thus, if we start out with an initially empty KB_0, successive TELL-operations are guaranteed to always produce an objective KB. However, TELL is not prevented from returning a KB that lies outside the class of KB's for which reasoning is decidable. A simple check would be to require that the KB is in the form $\mathrm{KB}' \wedge E$ as in Theorem 3 or can be easily transformed into that form by rearranging conjuncts.

$$\mathrm{KB} = \left\{ \begin{array}{l} \mathbf{Teacher}(jack), \mathbf{Teacher}(jill) \\ \mathbf{Emp}(jack), \mathbf{Emp}(jill), \mathbf{Emp}(sue) \\ \mathbf{Teach}(jack, csc378), \mathbf{Teach}(jill, csc484) \end{array} \right\}$$

$\mathbf{ASK}(\mathrm{KB}, \forall x(\mathbf{Teacher}(x) \supset \mathbf{Emp}(x))) = \mathrm{UNK}$.
The answer is UNK because there may be teachers who are not known to the KB and who are not employees.[15]
$\mathbf{ASK}(\mathrm{KB}, \forall x(\mathbf{BTeacher}(x) \supset \mathbf{Emp}(x)) = \mathrm{YES}$.
Note the difference to the previous query. Here the KB is only asked about the *known* teachers.
$\mathbf{ASK}(\mathrm{KB}, \forall x(\mathbf{BEmp}(x) \supset \mathbf{BTeacher}(x))) = \mathrm{NO}$.
The answer is NO because Sue, who is a known employee, is not known to be a teacher.
$\mathbf{ASK}(\mathrm{KB}, \forall x(\mathbf{BTeacher}(x) \supset \exists y \mathbf{Teach}(x,y))) = \mathrm{UNK}$.
The answer is UNK because there is no admissible term for y which would work for all known teachers. To obtain the answer YES, we need to rephrase the question as $\forall x[\mathbf{BTeacher}(x) \supset \exists y_1 \exists y_2(\mathbf{Teach}(x,y_1) \vee \mathbf{Teach}(x,y_2))]$. Now it is possible to substitute different admissible terms for y_1 and y_2.

6 Summary and Future Work

In this paper, we extended earlier work on limited belief by adding quantifying-in and equality. We established that reasoning is decidable if the KB is first-order and where queries can range over a large class of modal sentences with quantifying-in. In the future we hope to prove the conjecture that decidability holds if we allow arbitrary forms of quantify ing-in. It is also important to identify classes of sentences where reasoning is not just decidable but provably tractable as well. Finally, one should investigate to what extent modalities can be allowed in the KB itself without sacrificing decidability.

[15] All names in KB are assumed to be standard names.

Acknowledgements

I am grateful to Hector Levesque for many inspiring discussions on limited belief.

References

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

[2] Davis, M., Obvious Logical Inferences, in *Proc.IJCAI-81,* 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, 1986.

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

[6] Kaplan, D., Quantifying In, in L. Linsky (ed.), *Reference and Modality,* Oxford University Press, Oxford, 1971.

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

[8] Konolige, K., A Computational Theory of Belief Introspection. In *Proc. IJCA1-85,* Los Angeles, 1985, PP 502-508.

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

[10] Kripke, S. A., *Naming and Necessity,* Harvard University Press, Cambridge, MA, 1980.

[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. 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.

[13] Lakemeyer, G., Decidable Reasoning in First-Order Knowledge Bases with Perfect Introspection, in *Proc.AAAI-90,* Boston, MA, August 1990, pp. 531-537.

[14] Lakemeyer, G., *Models of Belief for Decidable Reasoning in Incomplete Knowledge Bases,* Ph.D. thesis, University of Toronto, 1990.

[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, in *Proc. AAAI-SI* Austin, TX, 1984, pp. 198-202.

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

[18] Moore, R. C, Semantical Considerations on Nonmonotonic Logic, in *Proc. IJCAI-83,* Karlsruhe, FRG, 3983, pp. 272-279.

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