Proceedings of the Twenty-Third International Joint Conference on Artificial Intelligence

Decidable Reasoning in a Logic of Limited Belief with Introspection and Unknown Individuals

Gerhard Lakemeyer
Dept. of Computer Science
RWTH Aachen
52056 Aachen, Germany
gerhard@cs.rwth-aachen.de

Hector J. Levesque
Dept. of Computer Science
University of Toronto, Toronto, Ontario
Canada M5S 3A6
hector@cs.toronto.edu

Abstract

There are not very many existing logics of belief which have both a perspicuous semantics and are computationally attractive. An exception is the logic $\mathcal{SL}$, proposed by Liu, Lakemeyer, and Levesque, which allows for a decidable and often even tractable form of reasoning. While the language is first-order and hence quite expressive, it still has a number of shortcomings. For one, beliefs about beliefs are not addressed at all. For another, the names of individuals are rigid, that is, their identity is assumed to be known. In this paper, we show how both shortcomings can be overcome by suitably extending the language and its semantics. Among other things, we show that determining the beliefs of a certain kind of fully introspective knowledge bases is decidable and that unknown individuals in the knowledge base can be accommodated in a decidable manner as well.

1 Introduction

One of the long-standing aims in the area of Knowledge Representation and Reasoning has been to devise computationally attractive reasoning mechanisms for very expressive knowledge bases (*KB*s). The need for an expressive representation language arises when one needs to deal with incomplete information. For example, a *KB* may know[1] that either Sue or Sam is a teacher or that Sid is not a teacher without knowing for any particular person that he or she is a teacher. While first-order logic (FOL) is able to represent such incompleteness, it is well known that reasoning in FOL based on classical logical entailment is undecidable. A principled way to arrive at weaker forms of entailment is to come up with appropriate models of limited belief [Levesque, 1984b; Konolige, 1986; Vardi, 1986; Fagin and Halpern, 1988; Fagin *et al.*, 1990; Lakemeyer, 1996; Cadoli and Schaerf, 1992; Delgrande, 1995; Liu et al., 2004], where reasoning can be studied as the question of which beliefs follow logically from believing the sentences in the *KB*. In order to be able to study the properties of such a model of belief, it seems particularly desirable to have a perspicuous semantics.

Despite many years of research in this area, there have not been that many proposals of this kind. One notable exception is the work by Liu, Lakemeyer, and Levesque [2004] (henceforth abbreviated as LLL), which is also the starting point of this paper. Based on earlier work by Lakemeyer and Levesque [Lakemeyer and Levesque, 2002], LLL propose the logic $\mathcal{SL}$, which uses as semantic primitive a *setup*, which is a possibly infinite set of ground clauses closed under unit propagation. Roughly, the clauses in a setup can be viewed as those which the agent believes explicitly. They then consider a sequence of modalities $\boldsymbol{B}_k$, for $k \geq 0$, where $\boldsymbol{B}_k\phi$ should be read as "ϕ is believed at level k." For example, given a clause c, $\boldsymbol{B}_0 c$ is satisfied by a setup s just in case there is a clause in s contained in c. In other words, at level 0, belief essentially reduces to retrieval wrt. s. At level 1, the agent believes everything that is believed at level 0 and, in addition, the agent is able to reason by cases allowing to split a single clause in s. At belief level 2, the number of possible case splits increases to 2, and so on.

For example, consider a setup s consisting of a single clause $(\mathit{Teacher}(\mathit{sue}) \vee \mathit{Teacher}(\mathit{sam}))$. Then s satisfies $\boldsymbol{B}_0(\mathit{Teacher}(\mathit{sue}) \vee \mathit{Teacher}(\mathit{sam}) \vee \mathit{Teacher}(\mathit{zoe}))$ by subsumption, but s does not satisfy $\boldsymbol{B}_0 \exists x.\mathit{Teacher}(x)$ because there is no name n for which $\boldsymbol{B}_0\mathit{Teacher}(n)$ holds at s. On the other hand, s does satisfy $\boldsymbol{B}_1 \exists x.\mathit{Teacher}(x)$ because we can split the clause in s, and in one case we can choose $n = \mathit{sue}$ and in the other $n = \mathit{sam}$. LLL present other, more complicated examples which demonstrate that this kind of existential reasoning may easily require many case splits in classical FOL. In fact, this problem alone is known to be undecidable in general [Patel-Schneider, 1985].

In this setting, a reasoning service for a *KB* can be defined as the problem of computing, for a given k, whether $\boldsymbol{B}_k\phi$ is logically implied by $\boldsymbol{B}_0 KB$. LLL show that for a certain kind of so-called proper$^+$ *KB*s, this reasoning service is decidable and indeed tractable in the propositional case.

While this work is an interesting contribution towards a computationally attractive and a semantically well-founded reasoning service, it still has a number of limitations. For one, the beliefs considered by LLL may not themselves refer to beliefs, which is needed to account for introspective agents. For example, continuing the above example, we may want to say

[1] Throughout the paper, we will use the terms knowledge and belief interchangeably.

that s satisfies $\boldsymbol{B}_1(\exists x.\mathit{Teacher}(x) \wedge \neg \boldsymbol{B}_0(\exists x.\mathit{Teacher}(x)))$ and $\boldsymbol{B}_1(\exists x.\mathit{Teacher}(x) \wedge \neg \boldsymbol{B}_1\mathit{Teacher}(x))$. The latter is particularly interesting as it expresses that the agent believes at level 1 that someone is a teacher but it does not know who that teacher is. In other words, introspection allows the agent to distinguish between *de dicto* and *de re* beliefs. Another shortcoming of $\mathcal{SL}$ directly related to this distinction is the assumption that all constants of the language are rigid, that is, they are standard names whose identity is known by definition. For example, $\boldsymbol{B}_0\mathit{Teacher}(\mathit{bestFriendOfSue})$ logically entails $\exists x.\boldsymbol{B}_0\mathit{Teacher}(x)$, which seems too strong.

In this paper we will address both of these issues by extending the logic $\mathcal{SL}$ to the new logic $\mathcal{LB}$. The idea of a setup and how beliefs at different levels come about will be lifted from $\mathcal{SL}$. Full introspection will be built into the new logic allowing an arbitrary nesting of beliefs at all levels and in any order. One of the main results of this paper will be that introspective reasoning remains decidable and that reasoning about nested beliefs is reducible to reasoning about non-nested beliefs in $\mathcal{SL}$. Our proposal to deal with unknown individuals is inspired by recent work by De Giacomo et al [2011]. Again we will provide a decidable reduction to reasoning in $\mathcal{SL}$. To keep the technical treatment of unknown individuals simple, we will consider them only for non-nested beliefs.

Considering introspection is perhaps most interesting when an agent is able to draw conclusions about what it does not believe. To specify a reasoning service with this feature, the assumption of $\boldsymbol{B}_0\mathit{KB}$ turns out to be too weak, as it only says that *at least* the sentences of the *KB* are believed explicitly so that nothing is entailed about what is not believed. What is needed instead is a notion that the sentences in the *KB* are *all* that is believed or that the *KB* is *only-known.* For this purpose we will introduce another operator $\boldsymbol{O}$ so that the reasoning service is then characterized by the beliefs which are entailed by $\boldsymbol{O}\mathit{KB}$. To simplify matters for the purposes of this paper, we only consider $\boldsymbol{O}$ at level 0 and hence leave out the subscript altogether.

1.1 Related Work

Modelling belief has had a long tradition starting with Hintikka's possible-world approach [Hintikka, 1962] (see also [Halpern and Moses, 1992]). While intuitively appealing, the possible-world model suffers from the logical omniscience problem [Hintikka, 1975] so that reasoning services based on this model are intractable in the propositional case and undecidable when the language is first order. The approaches addressing the logical omniscience problem can roughly be characterized as syntactic or semantic. The syntactic approaches include [Konolige, 1986; Vardi, 1986; Fagin and Halpern, 1988] and either include sets of sentences as part of the interpretation of belief or use notions such as awareness to rule out certain beliefs. Examples of the semantic approach are [Levesque, 1984b; Cadoli and Schaerf, 1992; Lakemeyer, 1996], which are based on *tautological entailment* [Anderson and Belnap, 1975; Dunn, 1976; Patel-Schneider, 1985] using four truth values instead of the usual two. While the syntactic approach can be criticized for being perhaps too fine-grained and not providing enough guidance as to which beliefs to include and which to leave out, the semantic approaches are much more in line with the possible-world tradition, but they also have a significant downside: the beliefs that follow from believing the sentences in a knowledge base are extremely weak, as *Modus ponens* is completely ruled out. For example, from p and $p \supset q$ it is not possible to infer q. Moreover, these approaches also need to deal with the fact that reasoning about existentials is undecidable even without Modus ponens [Patel-Schneider, 1985]. Hence reasoning becomes even weaker once decidability is restored. The model of belief underlying the logic $\mathcal{SL}$ can be seen as a compromise between the two camps. On the one hand, there is a syntactic flavour in that setups consist of clauses. On the other hand, there are well-motivated semantic rules which determine the beliefs at any level. Most importantly, reasoning becomes more and more powerful with increasing belief levels while remaining decidable. For these reasons, $\mathcal{SL}$ is an attractive starting point for further investigations into limited, semantically coherent reasoning.

The rest of the paper is organized as follows. In the next chapter, we introduce the logic $\mathcal{LB}$ and demonstrate its connection to $\mathcal{SL}$ and further properties. In Section 4 we show how reasoning about nested beliefs in the context of proper$^+$ *KB*s can be reduced to reasoning in $\mathcal{SL}$. In Section 5, we consider unknown individuals and then we conclude.

2 The Logic $\mathcal{LB}$

The language is a first-order modal dialect with an infinite supply of predicate symbols of every arity, including $=$, and an infinite supply of standard names $\#1, \#2, \#3, \ldots$, which are syntactically treated like constants but which are intended to be isomorphic to the (fixed) domain of discourse.[2] For now, no other constants or function symbols are allowed. Besides the usual connectives $\neg, \vee$ and the quantifier $\exists$ we have modal operators $\boldsymbol{B}_k$ for $k = 0, 1, 2, \ldots$ and the modal operator $\boldsymbol{O}$. The terms of the language are variables and standard names. An atomic formula is either a predicate symbol P with terms as arguments or of the form $(t = t')$ where t and t' are terms. Formulas are the least set which contain the atomic formulas, and if α and β are formulas and x a variable, then so are $\neg\alpha, \alpha \vee \beta, \exists x.\alpha, \boldsymbol{B}_k\alpha$, and $\boldsymbol{O}\alpha$, with the restriction that $\boldsymbol{O}$ may not appear within the scope of any modal operator and that for any $\boldsymbol{O}\alpha$, α contains no modal operators. We also allow the special symbol $\Box$, intended to represent the empty clause, to appear as a subformula within the scope of a modal operator. $\boldsymbol{B}_k\alpha$ should be read as "α is believed at level k" and $\boldsymbol{O}\phi$ as "ϕ is all that is known (or only-known)."

As is customary, we will freely use the logical connectives $\wedge, \supset, \equiv$ and the quantifier $\forall$ as the usual syntactic abbreviations. Given a formula α, we write α^x_n to mean the result of replacing every free occurrence of x within α by n.

Predicate symbols applied to standard names are called *primitive formulas.* Sentences are formulas without free variables. Formulas without modal operators are called *objective,* formulas where all predicate symbols other than $=$ appear

[2] In other words, standard names can be thought of as constants that satisfy the unique name assumption and an infinitary version of domain closure. See [Levesque, 1984a] for a discussion of why the assumption of a countably infinite domain is not really a limitation.

within the scope of a modal operator are called *subjective*, and formulas which do not mention $\boldsymbol{O}$ are called *basic*.

As clauses play an essential role in the semantics, we introduce them here together with some other conventions. A *clause* is a disjunction of literals, where a literal is either an atomic formula or its negation. The complement of a literal l is often denoted as $\bar{l}$. We will identify a clause with the set of literals it contains. A *primitive clause* is a clause which consists only of primitive formulas and their negations. In other words, primitive clauses mention neither variables nor $=$. As already mentioned, the empty clause is denoted as $\Box$. A unit clause is a clause with a single literal.

Given a set s of primitive clauses, the closure of s under unit propagation, which we denote as $\mathsf{UP}(s)$, is defined as the least set s' which contains s, and if unit clause $l \in s'$ and $\{\bar{l}\} \cup c \in s'$, then $c \in s'$. $\mathsf{VP}(s)$ is defined as the set $\{c \mid c$ is a primitive clause and there exists a $c' \in \mathsf{UP}(s)$ such that $c' \subseteq c\}$.

The semantics of $\mathcal{LB}$ is based on *worlds,* which determine what is true apart from the agents beliefs, and *setups,* which determine what the agent believes. A world is a mapping from the set of all primitive formulas to $\{0, 1\}$. A setup is a possibly infinite set of primitive clauses. Roughly, these represent what the agent believes explicitly (at level 0). As we will see, beliefs at higher levels will be obtained by case splitting from the given setup.

Given a world w and a setup s, the truth of a sentence α, written as $s, w \models \alpha$ is defined as follows:

1. $s, w \models P(\vec{n})$ iff $w[P(\vec{n})] = 1$;
2. $s, w \models (n = m)$ iff n, m are identical standard names;
3. $s, w \models \neg\alpha$ iff $s, w \not\models \alpha$;
4. $s, w \models (\alpha \vee \beta)$ iff $s, w \models \alpha$ or $s, w \models \beta$;
5. $s, w \models \exists x.\alpha$ iff $s, w \models \alpha^x_n$ for some standard name n;
6. $s, w \models \boldsymbol{B}_k\alpha$ iff $s, s, k \models \alpha$ (defined below);
7. $s, w \models \boldsymbol{O}\alpha$ iff $s, w \models \boldsymbol{B}_0\alpha$ and for all s', if $\mathsf{VP}(s') \subsetneq s$ then $s', w \not\models \boldsymbol{B}_0\alpha$.

Compared to classical logic, only the last three rules are non-standard. Rule 5 is somewhat special as it gives a substitutional account of quantification. This is possible because the set of standard names is essentially the universe of discourse and part of the language. While there are philosophical arguments criticizing substitutional interpretations [Kripke, 1976], it allows us to make *de dicto* vs. *de re* distinctions in a simple manner and it greatly simplifies the overall technical apparatus. Let us now turn to the semantics of belief.

Let the length of a basic formula α, denoted as $|\alpha|$, be defined in the usual way except that we let $|\boldsymbol{B}_k\alpha| = (k + 1) + |\alpha|$. It is easy to see that $|\boldsymbol{B}_k\alpha| > |\alpha|$ and $|\boldsymbol{B}_{k+1}\alpha| > |\boldsymbol{B}_k\alpha|$ for all k. To make sure that the following inductive definition of $s, s', k \models \alpha$ is well-founded, we need to define a measure that not only takes into account the length of α but also k. A simple way to achieve this is as follows: for any pair (k, α), let $\| (k, \alpha) \| = (k + 1) + |\alpha|$.

For any setups s, s', any $k \geq 0$ and any basic sentence α, we let $s, s', k \models \alpha$ be the least relation that satisfies the following:

8. $s, s', k \models \alpha$ if $\Box \in \mathsf{UP}(s')$;
9. $s, s', k \models c$ if $k = 0$, c is a clause and $c \in \mathsf{VP}(s')$;
10. $s, s', k \models \alpha$ if $k > 0$ and there is a $c \in s'$ s.t. for all $l \in c$, $s, s' \cup \{l\}, k - 1 \models \alpha$.
11. $s, s', k \models (n = m)$ if n, m are identical std. names; $s, s', k \models \neg(n = m)$ if n, m are distinct std. names;
12. $s, s', k \models \neg\neg\alpha$ if $s, s', k \models \alpha$;
13. $s, s', k \models (\alpha \vee \beta)$ if $s, s', k \models \alpha$ or $s, s', k \models \beta$; $s, s', k \models \neg(\alpha \vee \beta)$ if $s, s', k \models \neg\alpha$ and $s, s', k \models \neg\beta$;
14. $s, s', k \models \exists x.\alpha$ if $s, s', k \models \alpha^x_n$ for some n; $s, s', k \models \neg\exists x.\alpha$ if $s, s', k \models \neg\alpha^x_n$ for all n;
15. $s, s', k \models \boldsymbol{B}_{k'}\alpha$ if $s, s, k' \models \alpha$; $s, s', k \models \neg\boldsymbol{B}_{k'}\alpha$ if $s, s', k \not\models \boldsymbol{B}_{k'}\alpha$.

The above rules state the various ways sentences can be believed at level k. Rule 8 simply says that as soon as the empty clause is derivable by unit propagation, then everything is believed. Rule 9 deals with the base case of belief at level 0, that is, a clause is explicitly believed if it is a member of $\mathsf{VP}(s')$ or, equivalently, a superset of a clause in $\mathsf{UP}(s')$. Rule 10 deals with the the case where $k > 0$ and it is possible to establish that α is believed at level $k - 1$ after splitting a clause. Rule 11 means that we have perfect reasoning about equalities and this is independent of any setup. Rules 12–14 establish certain "obvious" beliefs, for example from either believing α or believing β one can conclude believing $\alpha \vee \beta$ (Rule 13). Finally, Rule 15 deals with nested beliefs. As we will see, the effect is a fully introspective agent. Notice that s' is replaced by s on the RHS. In fact, this rule is the reason why we require two setups. While s' may change due to Rule 10, s remains fixed and determines the interpretation of all nested beliefs.

With the semantics of $\boldsymbol{B}_k\alpha$ in hand, let us now take a brief look at the meaning of $\boldsymbol{O}$. Intuitively, the definition says that to only-know α the agent believes α explicitly and no other setup with truly fewer explicit beliefs believes α explicitly. We will come back to $\boldsymbol{O}$ in more detail in Section 4.

To conclude the semantic definitions, we say that a sentence α is valid (written $\models \alpha$) iff $s, w \models \alpha$ for all setups s and all worlds w. When α is subjective, the world w plays no role and we will often write $s \models \alpha$ instead of $s, w \models \alpha$.

3 The connection with the logic $\mathcal{SL}$

We begin our investigations of the properties of $\mathcal{LB}$ by introducing a slight variant of $\mathcal{SL}$ and establishing that the two logics agree on all sentences of $\mathcal{SL}$. Its formulas consist of all basic subjective sentences of $\mathcal{LB}$ yet without nesting of beliefs.

We begin by introducing $(\boldsymbol{B}_k\phi)\downarrow$, which, roughly, denotes the $\mathcal{SL}$ formula resulting from pushing the belief operator into ϕ. The purpose is to allow obvious conclusions from $(\boldsymbol{B}_k\phi)\downarrow$ to $\boldsymbol{B}_k\phi$. For any $\phi \in \mathcal{L}$, the $\mathcal{SL}$ formula $(\boldsymbol{B}_k\phi)\downarrow$ is defined as follows:

1. $(\boldsymbol{B}_k c)\downarrow = \boldsymbol{B}_k c$, where c is a clause;
2. $(\boldsymbol{B}_k(t = t'))\downarrow = (t = t')$;
3. $(\boldsymbol{B}_k\neg(t = t'))\downarrow = \neg(t = t')$;

4. $(\boldsymbol{B}_k\neg\neg\phi)\downarrow = \boldsymbol{B}_k\phi$;
5. $(\boldsymbol{B}_k(\phi\vee\psi))\downarrow = (\boldsymbol{B}_k\phi\vee\boldsymbol{B}_k\psi)$, where ϕ or ψ is not a clause;
6. $(\boldsymbol{B}_k\neg(\phi\vee\psi))\downarrow = (\boldsymbol{B}_k\neg\phi\wedge\boldsymbol{B}_k\neg\psi)$;
7. $(\boldsymbol{B}_k\exists x.\phi)\downarrow = \exists x.\boldsymbol{B}_k\phi$;
8. $(\boldsymbol{B}_k\neg\exists x.\phi)\downarrow = \forall x.\boldsymbol{B}_k\neg\phi$.

As discussed above, the semantics of $\mathcal{SL}$ is based on setups. Worlds are not needed as only subjective formulas are considered. Let s be a setup. Then for any sentence $\alpha\in\mathcal{SL}$, $s\models_{\text{SL}}\alpha$ (read "s satisfies α in $\mathcal{SL}$") is defined inductively[3] as follows:

1. $s\models_{\text{SL}}(d=d')$ iff d and d' are the same constant;
2. $s\models_{\text{SL}}\neg\alpha$ iff $s\not\models_{\text{SL}}\alpha$;
3. $s\models_{\text{SL}}\alpha\vee\beta$ iff $s\models_{\text{SL}}\alpha$ or $s\models_{\text{SL}}\beta$;
4. $s\models_{\text{SL}}\exists x.\alpha$ iff for some constant d, $s\models_{\text{SL}}\alpha^x_d$;
5. $s\models_{\text{SL}}\boldsymbol{B}_k\phi$ iff one of the following holds:
   (a) *inconsistent*: $\Box\in\mathsf{UP}(s)$;
   (b) *subsume*: $k=0$, ϕ is a clause c, and $c\in\mathsf{VP}(s)$;
   (c) *reduce*: ϕ is not a clause and $s\models_{\text{SL}}(\boldsymbol{B}_k\phi)\downarrow$;
   (d) *split*: $k>0$ and there is some $c\in s$ such that for all $\rho\in c$, $s\cup\{\rho\}\models_{\text{SL}}\boldsymbol{B}_{k-1}\phi$.

A sentence α is *valid* in $\mathcal{SL}$ ($\models_{\text{SL}}\alpha$) if for every setup s, we have that $s\models_{\text{SL}}\alpha$.

We deviate slightly from the original proposal of $\mathcal{SL}$ by adding the rule (5a), which says that everything is believed as soon as $\Box$ is part of the setup after unit propagation. While this changes some properties of beliefs about $=$,[4] it is easy to see that all other results, including those about decidability, remain the same.

In order to establish that $\mathcal{LB}$ and $\mathcal{SL}$ coincide on all sentences of $\mathcal{SL}$, the following lemma is needed.

Lemma 1 *For any objective ϕ, setups s and s' and $k\geq 0$, $s\models_{SL}\boldsymbol{B}_k\phi$ iff $s',s,k\models\phi$*

Proof: The proof is by induction on k and $|\phi|$. Let $k=0$. If ϕ is a clause c, then $s\models_{\text{SL}}\boldsymbol{B}_0c$ iff $c\in\mathsf{VP}(s)$ iff $s',s,0\models c$. The other cases for ϕ follow easily by induction. Here we consider only $\exists$: $s\models_{\text{SL}}\boldsymbol{B}_0\exists x.\phi$ iff $s\models_{\text{SL}}\exists x\boldsymbol{B}_0\phi$ iff $s\models\boldsymbol{B}_0\phi^x_n$ for some n iff (by ind.) $s',s,0\models\phi^x_n$ for some n iff $s',s,0\models\exists x\phi$.

Suppose the lemma holds for $k-1$. Here we only consider the case for splitting a clause. (The other cases follow again by a simple induction on $|\phi|$.) Let $s\models_{\text{SL}}\boldsymbol{B}_k\phi$ and suppose there is a $c\in s$ such that for all $\rho\in c$, $s\cup\{\rho\}\models_{\text{SL}}\boldsymbol{B}_{k-1}\phi$. Then, by induction, $s',s\cup\{\rho\},k-1\models\phi$ for all $\rho\in c$ and hence $s',s,k\models\phi$. The reverse direction in case of clause splitting is analogous. ∎

Using this lemma it is easy to show that

Lemma 2 *For every $\alpha\in\mathcal{SL}$, setup s, and world w, $s\models_{SL}\alpha$ iff $s,w\models\alpha$.*

As an immediate consequence, we obtain

Theorem 1 *$\models_{SL}\alpha$ iff $\models\alpha$.*

In other words, for basic non-nested sentences, $\mathcal{LB}$ inherits all the properties of $\mathcal{SL}$. For example, we have that for any i,j there is a k s.t. $\models\boldsymbol{B}_i\phi\wedge\boldsymbol{B}_j(\phi\supset\psi)\supset\boldsymbol{B}_k\psi$.[5]

We now consider proper$^+$ *KB*s, which were originally introduced in [Lakemeyer and Levesque, 2002] and are an extension of the proper *KB*s proposed in [Levesque, 1998]. Let θ denote a substitution of all variables by standard names. We write $\alpha\theta$ as the result of applying the substitution to α. We use $\forall\alpha$ to mean the universal closure of α. We let e range over ewffs, which are quantifier-free formulas containing no predicate symbols other than $=$.

Let e be an ewff and c a clause. Then a formula of the form $\forall(e\supset c)$ is called a $\forall$-clause. A *KB* is called proper$^+$ if it is a finite non-empty set of $\forall$-clauses. Given a proper$^+$ *KB*, we define gnd(*KB*) as the possibly infinite setup $\{c\theta\,|\,\forall(e\supset c)\in KB \text{ and } \models e\theta\}$. In the following we will use *KB* both as a set of sentences and as a conjunction of the sentences it contains.

An example proper$^+$ *KB* is $\{\forall x,y.(x=\#1\wedge y=\#2)\supset(\mathit{Teacher}(x)\vee\mathit{Teacher}(y),\forall x.x\neq\#1\supset\mathit{Female}(x)\}$.

LLL established that determining the beliefs at any level k of a proper$^+$ *KB* is decidable.

Theorem 2 (LLL) *For any proper$^+$ KB and objective ϕ, $\models_{SL}\boldsymbol{B}_0KB\supset\boldsymbol{B}_k\phi$ is decidable.*

Given Thm. 1, we obtain the same decidability result for $\mathcal{LB}$:

Corollary 1 *For any proper$^+$ KB and objective ϕ, the validity of $\boldsymbol{B}_0KB\supset\boldsymbol{B}_k\phi$ is decidable in $\mathcal{LB}$.*

LLL proved that the reasoning service as defined by the above implications, is always sound with respect to classical reasoning and they showed tractability in the propositional case. Given the undecidability of classical reasoning, reasoning is also necessarily incomplete. For example, we have that $\not\models\boldsymbol{B}_0p\supset\boldsymbol{B}_k(q\vee\neg q)$ for all k.

Let us now go beyond $\mathcal{SL}$ and consider nested beliefs.

4 Nested beliefs

We begin by generalizing a result by LLL, which says that anything that is believed at level k is also believed at levels higher than k.

Proposition 1 *For any basic α, $\models\boldsymbol{B}_k\alpha\supset\boldsymbol{B}_{k+1}\alpha$.*

Proof: Suppose $s,w\models\boldsymbol{B}_k\alpha$. We need to consider two cases for s. If s is empty, then a simple induction on $|\alpha|$ shows that for all k,j, $s,s,k\models\alpha$ iff $s,s,j\models\alpha$, from which the lemma follows. Now let s be non-empty. Again, a simple induction on $|\alpha|$ shows that for all s,s' and literal ρ, if $s,s',k\models\alpha$ then $s,s'\cup\{\rho\},k\models\alpha$. By assumption, we have $s,s,k\models\alpha$. We can then pick an arbitrary clause $c\in s$ such that for all $\rho\in c$, $s,s\cup\{\rho\},k\models\alpha$, from which $s,w\models\boldsymbol{B}_{k+1}\alpha$ follows using Rule 15 of the semantics. ∎

[3] LLL demonstrate that the induction is indeed well defined.

[4] For example, while $\boldsymbol{B}_ke\equiv e$ is valid in the original $\mathcal{SL}$ for all e mentioning only $=$, we only have that $\neg\boldsymbol{B}_k\Box\supset(\boldsymbol{B}_ke\equiv e)$ is valid.

[5] We remark that, given our slight change of the semantics of $\mathcal{SL}$, this property holds for all objective ψ and not just equality-free ψ as in the original $\mathcal{SL}$.

It is easy to see that our model of belief is fully introspective. For example, if p is believed at some level k then at any level j it is believed that p is believed at k. We also obtain the Barcan formula since standard names are the fixed universe of discourse.

Proposition 2

1. $\models \boldsymbol{B}_k\alpha \supset \boldsymbol{B}_j\boldsymbol{B}_k\alpha$
2. $\models \neg\boldsymbol{B}_k\alpha \supset \boldsymbol{B}_j\neg\boldsymbol{B}_k\alpha$
3. $\models \forall x\boldsymbol{B}_k\alpha \supset \boldsymbol{B}_k\forall x\alpha$.

Proof: Here we only prove (1.). Let $s, w \models \boldsymbol{B}_k\alpha$. Then $s, s, k \models \alpha$. By Rule 15, we then also have $s, s, j \models \boldsymbol{B}_k\alpha$ from which $s, w \models \boldsymbol{B}_j\boldsymbol{B}_k\alpha$ follows. ∎

4.1 A reasoning service for introspective *KB*s

We now turn to the issue of specifying a reasoning service for introspective proper$^+$ knowledge bases in $\mathcal{LB}$. As we argued already in the beginning, being able to draw conclusions about its own ignorance, an agent needs to assume that its *KB* is *all* that it knows, and it is for this purpose that we included the only-knowing operator $\boldsymbol{O}$ in our logic. The specification of the reasoning service will then be in terms of the valid sentences of the form

$$\boldsymbol{O}KB \supset \boldsymbol{B}_k\alpha \text{ for proper}^+ \textit{ KB}\text{s and basic } \alpha.$$

Let us begin by considering which setups s only-know a proper$^+$ *KB*. It turns out that there is essentially a unique s with this property:

Theorem 3 $s, w \models \boldsymbol{O}KB$ *iff* $\mathsf{VP}(s) = \mathsf{VP}(\mathit{gnd}(KB))$.

Proof: For the only-if direction, let $s \models \boldsymbol{O}KB$. Then $s \models \boldsymbol{B}_0KB$. Hence $s \models \boldsymbol{B}_0c$ for all $c \in \text{gnd}(KB)$ and, therefore, $c \in \mathsf{VP}(s)$. As shown by LLL, this implies $\mathsf{VP}(\text{gnd}(KB)) \subseteq \mathsf{VP}(s)$. To show that $\mathsf{VP}(s) \subseteq \mathsf{VP}(\text{gnd}(KB))$, suppose otherwise. Then $\mathsf{VP}(\text{gnd}(KB)) \subsetneq \mathsf{VP}(s)$. However, $\mathsf{VP}(\text{gnd}(KB)) \models \boldsymbol{B}_0KB$, contradicting the assumption that $s \models \boldsymbol{O}KB$.

Conversely, let $\mathsf{VP}(s) = \mathsf{VP}(\text{gnd}(KB))$. Then clearly $s \models \boldsymbol{B}_0KB$. Now consider a setup s' with $\mathsf{VP}(s') \subsetneq \mathsf{VP}(s)$. Then there is a $c \in \text{gnd}(KB)$ such that $c \notin \mathsf{VP}(s')$. But then $s' \not\models \boldsymbol{B}_0c$ and hence $s' \not\models \boldsymbol{B}_0KB$. ∎

Hence it suffices to consider only the possibly infinite setup gnd(*KB*) when determining whether implications of the form $\boldsymbol{O}KB \supset \boldsymbol{B}_k\alpha$ are valid. As a consequence, $\boldsymbol{O}$ can be replaced by $\boldsymbol{B}_0$ in the case of objective beliefs.

Theorem 4 *For any proper$^+$ KB and objective ϕ,* $\models \boldsymbol{O}KB \supset \boldsymbol{B}_k\phi$ *iff* $\models \boldsymbol{B}_0KB \supset \boldsymbol{B}_k\phi$.

Proof: The if direction is immediate since $\models \boldsymbol{O}KB \supset \boldsymbol{B}_0KB$. Conversely, let $\models \boldsymbol{O}KB \supset \boldsymbol{B}_k\phi$ and $s \models \boldsymbol{B}_0KB$. Then $\mathsf{VP}(\text{gnd}(KB)) \subseteq \mathsf{VP}(s)$. By Theorem 3, $\mathsf{VP}(\text{gnd}(KB)) \models \boldsymbol{O}KB$. Thus, by assumption, $\mathsf{VP}(\text{gnd}(KB)) \models \boldsymbol{B}_k\phi$. By Prop. 4 of LLL, $s \models \boldsymbol{B}_k\phi$. ∎

We are now ready to consider nested beliefs of a *KB*. For example, suppose *sam* and *sue* are standard names and let $KB = (\mathit{Teacher}(\mathit{sue}) \vee \mathit{Teacher}(\mathit{sam}))$, which can easily be massaged into proper$^+$ form. As expected, we have

$$\models \boldsymbol{O}KB \supset \boldsymbol{B}_1(\exists x.\mathit{Teacher}(x) \wedge \neg\boldsymbol{B}_1(\mathit{Teacher}(x))),$$

that is, the *KB* believes at level 1 that someone is a teacher but does not know who it is.

Proof: By Theorem 3, it suffices to consider $s \models \boldsymbol{O}KB$ with $s = \{\mathit{Teacher}(\mathit{sue}) \vee \mathit{Teacher}(\mathit{sam})\}$. Then we have $s, s, 1 \models \exists x.\mathit{Teacher}(x)$ because we can split the clause in s and obtain $s, s\cup\{\mathit{Teacher}(\mathit{sam})\}, 0 \models \exists x.\mathit{Teacher}(x)$ (choose $x = \mathit{sam}$) in one case and $s, s \cup \{\mathit{Teacher}(\mathit{sue})\}, 0 \models \exists x.\mathit{Teacher}(x)$ (choose $x = \mathit{sue}$) in another.

However, $s, s \cup \{\mathit{Teacher}(\mathit{sam})\}, 0 \models \neg\boldsymbol{B}_1\mathit{Teacher}(\mathit{sam})$ because $s, s, 1 \not\models \boldsymbol{B}_1\mathit{Teacher}(\mathit{sam})$, and similar for $x = \mathit{sue}$.

Putting things together, we therefore have that $s \models \boldsymbol{B}_1(\exists x.\mathit{Teacher}(x) \wedge \neg\boldsymbol{B}_1(\mathit{Teacher}(x)))$. ∎

Note also that $\models \boldsymbol{O}KB \supset \neg\boldsymbol{B}_0(\exists x.\mathit{Teacher}(x))$ since at level 0 we are not allowed to reason by cases.

The main question now is how to automate introspective reasoning for proper$^+$ *KB*s. As we will see, we can leverage an idea originally proposed by Levesque [1984a] for a classical modal reasoner, where reasoning about nested beliefs is reduced to reasoning about objective beliefs.

The key idea is to replace occurrences of nested beliefs such as $\boldsymbol{B}_1(\mathit{Teacher}(x))$ by a description of the *known* teachers. In our example, there are no known teachers and hence the formula is replaced by FALSE.[6] As we will see, nested beliefs such as $\boldsymbol{B}_1\neg\boldsymbol{B}_0(\mathit{Teacher}(x))$ can be handled recursively.

The approach then is to define, given an objective formula ϕ with free variables $\vec{x}$, a function RES$[k, \phi, KB]$, which produces an ewff e which describes for which standard names $\vec{n}$ the sentence $\phi^{\vec{x}}_{\vec{n}}$ is believed at level k.

Let ϕ be an objective formula and *KB* be proper$^+$. Suppose that $n_1, \ldots, n_k$, are all the names in ϕ or in *KB*, and that n' is some name that does not appear in ϕ or in *KB*. Then RES$[k, \phi, KB]$ is defined by:

1. If ϕ has no free variables, then RES$[k, \phi, KB]$ is TRUE, if $\models \boldsymbol{B}_0KB \supset \boldsymbol{B}_k\phi$, and FALSE, otherwise.
2. If x is a free variable in ϕ, then RES$[k, \phi, KB]$ is
$$((x = n_1) \wedge \text{RES}[k, \phi^x_{n_1}, KB]) \vee \ldots$$
$$((x = n_k) \wedge \text{RES}[k, \phi^x_{n_k}, KB]) \vee$$
$$((x \neq n_1) \wedge \ldots \wedge (x \neq n_k) \wedge \text{RES}[k, \phi^x_{n'}, KB]^{n'}_x).$$

For $KB = (\mathit{Teacher}(\mathit{sue}) \vee \mathit{Teacher}(\mathit{sam}))$ and $\phi = \mathit{Teacher}(x)$ we obtain RES$[1, \phi, KB] = (x = \mathit{sam} \wedge \text{FALSE}) \vee (x = \mathit{sue} \wedge \text{FALSE}) \vee (x \neq \mathit{sam} \wedge x \neq \mathit{sam} \wedge \text{FALSE})$, which simplifies to FALSE. If instead $KB = \mathit{Teacher}(\mathit{sue})$ then RES$[0, \phi, KB] = (x = \mathit{sue} \wedge \text{TRUE}) \vee (x \neq \mathit{sue} \wedge \text{FALSE})$, which simplifies to $x = \mathit{sue}$.

The correctness of our definition of RES is reflected in the following result.

Lemma 3 *For any proper$^+$ KB, any objective ϕ with free variables $x_1, \ldots, x_l$, and standard names $n_1, \ldots, n_l$,* $\models \boldsymbol{B}_0KB \supset \boldsymbol{B}_k\phi^{x_1}_{n_1}\cdots^{x_l}_{n_l}$ *iff* $\models \mathit{RES}[k, \phi, KB]^{x_1}_{n_1}\cdots^{x_l}_{n_l}$.

The proof is by induction on l and closely follows a similar argument in [Levesque, 1984a].

The following definition gives us the means to deal with arbitrary nestings of beliefs.

[6] We write TRUE as shorthand for $\forall x.(x = x)$ and FALSE as shorthand for $\neg$TRUE.

Given a proper$^+$ *KB* and a basic formula α, $\|\alpha\|_{\text{KB}}$ is the objective formula defined by

$$\|\alpha\|_{\text{KB}} = \alpha, \quad \text{when } \alpha \text{ is objective;}$$

$$\|\neg\alpha\|_{\text{KB}} = \neg\|\alpha\|_{\text{KB}};$$

$$\|(\alpha \vee \beta)\|_{\text{KB}} = (\|\alpha\|_{\text{KB}} \vee \|\beta\|_{\text{KB}});$$

$$\|\exists x.\alpha\|_{\text{KB}} = \exists x.\|\alpha\|_{\text{KB}};$$

$$\|\boldsymbol{B}_k\alpha\|_{\text{KB}} = \text{RES}[k, \|\alpha\|_{\text{KB}}, \textit{KB}].$$

Thus, given our example *KB* about Sue and Sam, $\|\exists x.\textit{Teacher}(x) \wedge \neg\boldsymbol{B}_1\textit{Teacher}(x)\|_{\text{KB}}$ results in $\exists x.\textit{Teacher}(x) \wedge \neg\text{FALSE}$ (after simplification), which further simplifies to $\exists x.\textit{Teacher}(x)$.

Lemma 4 *Let $s = \textit{gnd}(\textit{KB})$, α any basic formula with free variables $x_1, \ldots, x_l$, and $n_1, \ldots, n_l$ standard names. Then $s, s, k \models \alpha^{x_1}_{n_1}\cdots^{x_l}_{n_l}$ iff $s, s, k \models \|\alpha\|_{\text{KB}}{}^{x_1}_{n_1}\cdots^{x_l}_{n_l}$*

With this lemma it is easy to show that

Theorem 5 $\models \textit{OKB} \supset \boldsymbol{B}_k\alpha$ *iff* $\models \textit{OKB} \supset \boldsymbol{B}_k\|\alpha\|_{\text{KB}}$.

Since $\|\alpha\|_{\text{KB}}$ is objective, and together with Theorem 4, we obtain immediately:

Corollary 2 $\models \textit{OKB} \supset \boldsymbol{B}_k\alpha$ *iff* $\models \boldsymbol{B}_0\textit{KB} \supset \boldsymbol{B}_k\|\alpha\|_{\text{KB}}$.

Finally, given the decidability result for objective beliefs (Corollary 1) of proper$^+$ *KB*s and the fact that RES appeals to this decision problem only a finite number of times, we obtain

Corollary 3 *For any proper$^+$ KB and basic α, the validity of $\textit{OKB} \supset \boldsymbol{B}_k\alpha$ is decidable in $\mathcal{LB}$.*

5 Unknown individuals

Let us now turn to the issue of dealing with unknown individuals such as *bestFriendOfSue*. For that we augment the language with an infinite set of constants, which are treated as additional terms. The difference between constants and std. names will be that constants can denote different std. names. To simplify matters, we restrict the language for this section to basic formulas without nested beliefs (i.e. similar to $\mathcal{SL}$) and call the new logic $\mathcal{LB}^u$.

The semantics needs to be extended to account for the new terms. In the case of a world w, we interpret constants by mapping them into the standard names, e.g., $w[a] = {}^{\#}1$. We also need to extend clauses and setups. An e-clause is like a clause except that it may contain an ewff e as an additional disjunct. A ground e-clause is an e-clause without free variables, but it may contain constants. An e-setup is a possibly infinite set of ground e-clauses. This allows us, for example, to model knowledge of the kind $\forall x.a \neq x \supset T(x)$ as the infinite e-setup $\{(a = n) \vee T(n) \mid \text{for all std. names } n\}$. Let a *c-map* ν be a mapping from constants into standard names. We write $\alpha\nu$ to mean α with every constant a replaced by $\nu(a)$. Given an e-setup s, let $s\nu = \{c\nu \mid \{e\} \cup c \in s, \text{ where } \models \neg e\nu\}$. In words, $s\nu$ is a setup containing the clause part of those e-clauses in s with constants replaced by std. names, where e comes out false. For the above example, if $\nu(a) = {}^{\#}1$ then $s\nu = \{T({}^{\#}2), T({}^{\#}3), \ldots\}$.

The truth of a formula α is now defined wrt. a world w and an e-setup s. Compared to the previous semantics of $\mathcal{LB}$, only Rules 1,2, and 6 need to be changed. Rules 1 and 2 need to deal, in the obvious way, with the denotation of constants as specified by the world w (details omitted). Rule 6 is replaced by the following:

6′. $s, w \models \boldsymbol{B}_k\alpha$ iff for all c-maps ν, $s\nu, s\nu, k \models \alpha\nu$.

Note that the RHS talks only about setups in the old sense and formulas without constants, that is, we can use the existing semantic rules 8–15 of $\mathcal{LB}$ to evaluate its truth value. Validity in $\mathcal{LB}^u$ is defined as truth in all worlds and e-setups.

We now address the problem of deciding the beliefs of a proper$^+$ *KB*, where constants are allowed in $\forall$-clauses. Since we do not deal with nested beliefs, it suffices to consider the validity of formulas of the form $\boldsymbol{B}_0\textit{KB} \supset \boldsymbol{B}_k\phi$ for objective ϕ. E.g., let $\textit{KB} = \{(P(a) \vee Q(a), \forall x.x \neq {}^{\#}1 \supset \neg P(x)\}$, where a is a const. Then $\models \boldsymbol{B}_0\textit{KB} \supset \boldsymbol{B}_0(a \neq {}^{\#}1 \supset Q(a))$.

Let $\text{gnd}^e(\textit{KB}) = \{\{\neg e\theta\} \cup c\theta \mid \forall(e \supset c) \in \textit{KB}\}$, that is, $\text{gnd}^e(\textit{KB})$ is a grounding of *KB* where the ewffs are kept as part of the resulting e-clauses. We then have

Lemma 5 $\models \boldsymbol{B}_0\textit{KB} \supset \boldsymbol{B}_k\phi$ *iff* $\textit{gnd}^e(\textit{KB}) \models \boldsymbol{B}_k\phi$.

(The proof is similar to an analogous result by LLL for $\mathcal{SL}$.) We then have a result similar to one by De Giacomo et al. [2011] (Theorem 5) for proper *KB*s with unknowns:

Theorem 6 *Let $s = \textit{gnd}^e(\textit{KB})$. Let $a_1, \ldots, a_m$ be the constants in KB and ϕ. Let H be the set of those c-maps ν s.t. $\nu(a_1)$ ranges over all std. names in KB and ϕ plus one more, and for all $1 \leq i < m$, $\nu(a_{i+1})$ ranges over all names that $\nu(a_i)$ ranges over plus one more. Then*
for all ν, $s\nu, s\nu, k \models \phi\nu$ iff for all $\nu \in H$, $s\nu, s\nu, k \models \phi\nu$.

This result establishes that finitely many substitutions of constants by std. names suffice to determine the beliefs of a *KB*. Using Lemma 5 and the fact that $\text{gnd}^e(\textit{KB})\nu = \text{gnd}(\textit{KB}\nu)$ we then obtain

Theorem 7 *Let KB, ϕ, H be as in the previous theorem.*
$\models \boldsymbol{B}_0 \supset \boldsymbol{B}_k\phi$ *iff* $\models_{SL} \boldsymbol{B}_0\textit{KB}\nu \supset \boldsymbol{B}_k\phi\nu$ *for all $\nu \in H$.*

Note that this reduces the problem of determining the beliefs of a *KB* with unknown individuals to a finite number of decidable validity problems in $\mathcal{SL}$. Hence

Corollary 4 *For any proper$^+$ KB and objective ϕ, the validity of $\boldsymbol{B}_0\textit{KB} \supset \boldsymbol{B}_k\phi$ is decidable in $\mathcal{LB}^u$.*

6 Conclusions

The paper has established that the decidable reasoning service for proper$^+$ *KB*s proposed by LLL can be faithfully extended to deal with introspection. Moreover, unknown individuals can also be accommodated without sacrificing decidability. As for future work, an immediate task is to combine the results on introspection and unknown individuals. Besides decidability it will also be interesting to identify special cases where reasoning remains tractable for a fixed k. Finally, going beyond constants and allowing function symbols in some way is yet another challenge.

References

[Anderson and Belnap, 1975] A.R. Anderson and N.D. Belnap. *Entailment: The Logic of Relevance and Necessity*. Princeton University Press, 1975.

[Cadoli and Schaerf, 1992] M. Cadoli and M. Schaerf. Approximate reasoning and non-omniscient agents. In *Proc. of the 4th Conference on Theoretical Aspects of Reasoning about Knowledge (TARK-92)*, pages 159–183, 1992.

[De Giacomo et al., 2011] G. De Giacomo, Y. Lesprance, and H.J. Levesque. Efficient Reasoning in Proper Knowledge Bases with Unknown Individuals. IJCAI 2011, 827–832, 2011.

[Delgrande, 1995] J.P. Delgrande. A framework for logics of explicit belief. *Computational Intelligence*, 11(1):47–88, 1995.

[Dunn, 1976] J.M. Dunn. Intuitive semantics for first-degree entailments and coupled trees. *Philosophical Studies*, 29:149–168, 1976.

[Fagin and Halpern, 1988] R. Fagin and J.Y. Halpern. Belief, awareness, and limited reasoning. *Artificial Intelligence*, 34(1):39–76, 1988.

[Fagin *et al.*, 1990] R. Fagin, J.Y. Halpern, and M.Y. Vardi. A nonstandard approach to the logical omniscience problem. In *Proc. of the 3rd Conference on Theoretical Aspects of Reasoning about Knowledge (TARK-90)*, pages 41–55, 1990.

[Halpern and Moses, 1992] J.Y. Halpern and Y. Moses. A guide to completeness and complexity for modal logics of knowledge and belief. *Artificial Intelligence*, 54(3):319–379, 1992.

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

[Hintikka, 1975] J. Hintikka. Impossible Possible Worlds Vindicated. *Journal of Philosophical Logic,* 4, 475–484, 1975.

[Konolige, 1986] K. Konolige. *A Deduction Model of Belief*. Brown University Press, 1986.

[Kripke, 1976] S. Kripke. Is there a problem with substitutional quantification? In G. Evans and J. McDowell, editors, *Truth and Meaning*, pages 325–419. Clarendon Press, Oxford, 1976.

[Lakemeyer, 1996] G. Lakemeyer. *Limited Reasoning in First-Order Knowledge Bases with Full Introspection Artif. Intell.* 84(1-2): 209–255, 1996.

[Lakemeyer and Levesque, 2002] G. Lakemeyer and H.J. Levesque. Evaluation-based reasoning with disjunctive information in first-order knowledge bases. In *Proc. of KR-02*, pages 73–81, 2002.

[Levesque, 1984a] H. J. Levesque. Foundations of a Functional Approach to Knowledge Representation. *Artif. Intell.* 23(2), 155-212, 1984.

[Levesque, 1984b] H.J. Levesque. A logic of implicit and explicit belief. In *Proc. of AAAI-84*, pages 198–202, 1984.

[Levesque, 1998] H.J. Levesque. A completeness result for reasoning with incomplete first-order knowledge bases. In *Proc. KR-98*, pages 14–23, 1998.

[Liu et al., 2004] Y. Liu, G. Lakemeyer, H.J. Levesque. A Logic of Limited Belief for Reasoning with Disjunctive Information. *Int. Conf. on the Principles of Knowledge Representation and Reasoning (KR),* 587–597, 2004

[Patel-Schneider, 1985] P. Patel-Schneider. A decidable first-order logic for knowledge representation. In *Proc. of IJCAI-85*, pages 455–458, 1985.

[Vardi, 1986] M.Y. Vardi. On epistemic logic and logical omniscience. In *Proc. of the 1st Conference on Theoretical Aspects of Reasoning about Knowledge (TARK-86)*, pages 293–305, 1986.