

A Tractable, Expressive, and Eventually Complete First-Order Logic of Limited Belief

Gerhard Lakemeyer¹ and Hector J. Levesque²

¹ Dept. of Computer Science, RWTH Aachen University, Germany

² Dept. of Computer Science, University of Toronto, Canada
gerhard@cs.rwth-aachen.de, hector@cs.toronto.edu

Abstract

In knowledge representation, obtaining a notion of belief which is tractable, expressive, and eventually complete has been a somewhat elusive goal. Expressivity here means that an agent should be able to hold arbitrary beliefs in a very expressive language like that of first-order logic, but without being required to perform full logical reasoning on those beliefs. Eventual completeness means that any logical consequence of what is believed will eventually come to be believed, given enough reasoning effort. Tractability in a first-order setting has been a research topic for many years, but in most cases limitations were needed on the form of what was believed, and eventual completeness was so far restricted to the propositional case. In this paper, we propose a novel logic of limited belief, which has all three desired properties.

1 Introduction

The whole idea of a knowledge-based system, going all the way back to [McCarthy, 1963], is to provide a system with a knowledge base (KB) of facts expressed in a logical language that the system would then draw conclusions from. But in trying to characterize just what conclusions the system should be able to draw, it has been difficult to come up with a notion of belief that allows for full-first order logic on the one hand, and is not overly demanding computationally on the other.

There has been considerable research on this topic (since our first efforts 35 years ago in [Levesque, 1984]), and we will review this research in a later section. In this paper, we present a new model of belief that is the culmination of our thoughts on this subject. Following [Liu et al., 2004], it is based on a notion of *mental effort*, characterized by a single parameter k . The idea is that when $k = 0$, the only sentences believed are those that are in the given KB (or “easy” logical consequences of them). At higher values of k , more and more logical consequences of the KB are revealed. The main properties we are claiming for this new model are the following:

Expressiveness: Any sentence of a first-order logic with equality and standard names can be a sentence in the KB and therefore believed at $k = 0$. In particular, it will

be possible to believe (at $k = 0$) a disjunction without having to believe either disjunct, and to believe an existential without having to believe any of its instances.

Eventual completeness: Every sentence believed (at any level k) will be a logical consequence of the KB. In addition, if a sentence is a logical consequence of the KB, there will be a level k at which the sentence will be believed. So with enough mental effort it will be possible to do sound and complete first-order logical reasoning.

Tractability: The reasoning required to determine what is believed will be tractable in the following specific sense. Imagine that the KB consists of a potentially large number N of sentences each of which is bounded in size by a constant c . Suppose that we are interested in determining whether some other sentence (of size also bounded by c) is believed at some fixed level k . We will show that there is procedure that is polynomial in N for doing this.

(Note that although the elements of the KB can be of arbitrary logical form, belief is guaranteed to be tractable only when each individual sentence in the KB is small.) As far as we know, no existing model of belief, including our own earlier proposals, satisfy these properties. For example, cases where belief is guaranteed to be tractable might end up being less expressive than full first-order logic with equality, as in [Crawford and Etherington, 1998; Liu and Levesque, 2003; Liu et al., 2004], for example. Cases where all of first-order logic is allowed may fail to be eventually complete, that is, there may be logical consequences of the KB that are outside the reach of what can be believed [Patel-Schneider, 1985; Lakemeyer, 1996; Lakemeyer and Levesque, 2016].

The organization of the paper is as follows. In Section 2 we sketch the main ideas of our approach. Section 3 introduces the first-order logic with equality that our work is based on, and other logical preliminaries. Section 4 formalizes the notions of belief and only-believing that we are after, followed by a description of the full logic. In Section 6 we show that the logic indeed has all the desired properties and more. The paper ends with a discussion of related work and conclusions.

2 The Approach

In order to talk about what the system is considered to believe at any point, it will be convenient to use a logic of belief, where formulas $B_k\phi$ for $k = 0, 1, 2 \dots$ are used to say that

the system believes a first-order sentence ϕ at level k . (When we say that a logic of belief is tractable, expressive, and eventually complete, we mean that the notion of belief it characterizes has these properties.) In order to talk about the use of a knowledge base, we also use a form of *only-believing* by way of another modal operator O where $O\phi$ means that ϕ is believed (at level 0), but nothing stronger [Levesque, 1990]. This will allow us to write sentences like $(OKB \supset B_k\phi)$, which can be read as “if the given knowledge base is KB, then ϕ is believed at level k .” (When KB is used in a formula, we mean the conjunction of its elements.) The limited beliefs of a knowledge base can then be studied in terms of the valid sentences of this form.

With this logic of belief in mind, the properties of limited reasoning we are looking for are these:

- expressiveness: for any first-order ϕ , the sentence $O\phi$ will be satisfiable, and moreover $\models (O\phi \supset B_0\phi)$.
- cumulativity: for any k and any ϕ , $\models (B_k\phi \supset B_{k+1}\phi)$.
- soundness: for any k , any KB and ϕ , if $\models (OKB \supset B_k\phi)$, then $\models (KB \supset \phi)$.
- eventual completeness: for any KB and ϕ , if $\models (KB \supset \phi)$, then there will be some k such that $\models (OKB \supset B_k\phi)$.
- tractability: for any KB and ϕ , the question as to whether $\models (OKB \supset B_k\phi)$ for fixed k will be decidable with polynomial data complexity as described above.

So while it will be computationally feasible to determine if $B_k\phi$ is true (with an effort that depends on the k), if an agent really needs to determine whether or not ϕ is true, it may have to look at higher and higher values of k . Because of the undecidability of first-order logic, it will be undecidable to determine whether there exists a k such that $B_k\phi$ is true.

2.1 Reduction to Clauses

One possible approach to satisfying the above requirements is to start with a sound and complete logical reasoning procedure (like Resolution, say), but cut it off after k steps. In other words, we might have arranged the semantics so that $B_k\phi$ is true iff ϕ can be derived from the KB by the reasoning procedure in k or fewer steps.

However, such a straightforward interpretation of belief levels is problematic, as it would likely mean that many desirable properties of belief are lost. For example, $B_k(\alpha \wedge \beta)$ and $B_k(\beta \wedge \alpha)$ may not be logically equivalent, since the reasoning procedure might need some extra steps to go from one conjunction to the other. As we will see in Section 6, our notion of belief will preserve these and other equivalences.

In the semantics of O and B_k to follow starting in the next section, we obtain these desirable equivalences by reducing the general definition of belief over arbitrary sentences to belief over ground *clauses*, that is, disjunctions of literals. We will be making use of reductions of the following forms:

1. $B_k\phi$ will hold iff $B_k\exists\vec{x}.\psi$ holds, where the formula ψ is a Skolemized version of ϕ with no quantifiers;
2. $B_k\exists x.\psi$ will hold iff there are terms t_0, \dots, t_k such that $B_k(\psi_{t_0}^x \vee \dots \vee \psi_{t_k}^x)$ holds;

3. finally, when ϕ is quantifier-free, $B_k\phi$ will hold iff for every ground clause c in the *CNF* of ϕ , $B_k(c)$ holds.

The reduction in item (1) involves a conversion to prenex form and the dual form of Skolemization (sometimes called Herbrandization) where universally quantified variables are eliminated using Skolem functions; item (2) is what amounts to an application of a bounded version of Herbrand’s Theorem [Buss, 1998]; finally, item (3) involves the usual conversion to *CNF*.

This of course still leaves us with the problem of specifying which ground clauses c are to be believed given a KB. We need to be careful here, since the problem of determining whether KB logically entails a ground clause c (even without equality) is already undecidable. In fact, it is undecidable even when the KB is a single universally quantified clause.

2.2 Epistemic States

In possible-world semantics [Kripke, 1959; Hintikka, 1962; Chellas, 1980], perhaps the simplest model of the epistemic state of a KB is the set of all worlds satisfying the KB, given a fixed universe of discourse [Levesque and Lakemeyer, 2001]. For tractable reasoning however, this notion of epistemic state is too coarse, as it lumps together all logically equivalent knowledge bases. For example, for $KB = \{p, (p \supset q)\}$, we might want an epistemic state where $B_0(q)$ is false, but for the logically equivalent $KB = \{p, (p \supset q), q\}$, we want a different epistemic state where $B_0(q)$ would be true.

So in this paper, we will be using a much finer-grained notion of epistemic state. There may be many ways of doing this (and see future work), but we will show here that it is sufficient to think of an epistemic state not as a set of worlds, but as a finite set of clauses (which may contain variables). Intuitively, these clauses are the ones obtained by Skolemizing the KB and converting it to *CNF*. This solves the problem above since logically equivalent knowledge bases need not correspond to the same epistemic state when they do not have the same *CNF* (like the two examples above). (However, if two of them have exactly the same *CNF*, only believing one will end up being equivalent to only believing the other.)

The use of clauses as epistemic primitive is not new and dates back to [Liu et al., 2004], where epistemic states were possibly infinite sets of ground clauses (see also Section 7). While we consider finite sets of clauses with variables, these can be thought of as representing the (infinite) set of all ground instances.

2.3 When Are Clauses Believed?

What we will be proposing for the belief of clauses is easiest to understand in the propositional case (without the complications of quantifiers or equality). First, define a clause to be trivial iff it contains complementary literals. Then define

$$RP(C) = C \cup \{(a \cup b) \mid \{\rho\} \cup a \in C, \{\bar{\rho}\} \cup b \in C\}.$$

So $RP(C)$ performs one step of Resolution on the elements of C , $RP(RP(C))$ performs two steps, and so on. Then finally, define a clause b to be believed at level k in an epistemic state defined by a set of clauses C iff b is trivial or there is a clause $a \in RP^k(C)$ such that $a \subseteq b$. For example, if the epistemic

state is defined by $\{(p \vee q), (\neg p \vee r), (\neg q \vee r)\}$, then the clauses $(p \vee q)$ and $(s \vee \neg s)$ and their supersets are believed at levels 0 and above, the clause $(p \vee r)$ and its supersets are believed at level 1 and above, and the clause r and its supersets are believed at level 2 and above. The rest of the paper will make all these ideas precise for a full first-order logic with equality and standard names.

3 First-Order Preliminaries

We begin by introducing the logic \mathcal{L} from [Levesque and Lakemeyer, 2001] as our base logic, followed by various other notions we need from first-order logic.

3.1 The Base Logic \mathcal{L}

The language of \mathcal{L} is a first-order dialect with $=$ and an infinite supply of function and predicate symbols of every arity. In addition, the language also features a set \mathcal{N} of standard names $\#1, \#2, \#3, \dots$, which are syntactically treated like constants but which are intended to be isomorphic to the (fixed) domain of discourse. In other words, standard names can be thought of as constants that satisfy the unique name assumption and an infinitary version of domain closure. Among other things, standard names allow for a very simple, substitutional account of quantifiers. See [Levesque and Lakemeyer, 2001] for more discussion on why standard names are useful. In the following we often simply write “name” instead of “standard name.” Terms and atomic formulas are defined in the usual way, and so are formulas using the connectives \neg and \wedge and the quantifier \forall . Other connectives like \vee, \supset, \equiv and the quantifier \exists are freely used as syntactic abbreviations. Any formula from \mathcal{L} is also called an *objective* formula. A *sentence* is a formula without free variables.

Function symbols with names as arguments are called *primitive terms*, and predicate symbols with names as arguments are called *primitive atoms*.

The semantics is defined in terms of *worlds*, which are mappings from the primitive terms into \mathcal{N} and from the primitive atoms into $\{0, 1\}$.

The meaning of an arbitrary ground term is given in terms of its *coreferring* standard name. Formally, given a ground term t and a world w we define $|t|_w$ (read: the coreferring standard name for t given w) by:

1. If $t \in \mathcal{N}$, then $|t|_w = t$;
2. $|h(t_1, \dots, t_k)|_w = w[h(n_1, \dots, n_k)]$,
where $n_i = |t_i|_w$.

The truth of a sentence wrt world w (written as $w \models \phi$) is defined inductively as follows:

1. $w \models p(t_1, \dots, t_k)$ iff $w[p(n_1, \dots, n_k)] = 1$
where $|t_i|_w = n_i$;
2. $w \models (t_1 = t_2)$ iff $|t_1|_w$ and $|t_2|_w$ are the same names;
3. $w \models \neg\phi$ iff $w \not\models \phi$;
4. $w \models (\phi \wedge \psi)$ iff $w \models \phi$ and $w \models \psi$;
5. $w \models \forall x. \phi$ iff $w \models \phi_n^x$ for all names n ;

Here ϕ_n^x stands for ϕ with every free occurrence of x replaced by n . A sentence ϕ is valid ($\models \phi$) iff for all worlds w , $w \models \phi$.

Apart from standard names and equality, \mathcal{L} behaves exactly like classical first-order logic: it is shown in [Levesque and Lakemeyer, 2001] that a sentence without standard names and equality is valid iff it is valid in classical logic. Standard names are, of course, special in that sentences like $(\#6 \neq \#7)$ are valid, for example. Also, since the domain of discourse is infinite, sentences like $\exists x \forall y (x = y)$ are unsatisfiable in \mathcal{L} .

3.2 Other Preliminaries

The logic of belief here uses the language \mathcal{L} but incorporates some ideas from standard first-order logic that we review here. First some notation. We use the following metavariables throughout: x, y for variables, t, u, v for terms, n for standard names, f, g for function symbols, p, q for predicate symbols, a, b, c, d for clauses (defined below), θ for substitutions, ρ, τ, μ for literals, $\bar{\rho}$ for the complement of a literal, ϕ, ψ, χ for objective formulas, i, j, k, m, r, s for non-negative integers. A capitalized version of any of these will be used for sets. We assume throughout that all formulas are *rectified*, that is, that no variable appears bound by two quantifiers, one within the scope the other, as in something like $\exists x(\phi \wedge \forall x \psi)$. For any substitution θ , and any literal, clause, formula, or set of clauses z , $z\theta$ is like z but with free variables x in z replaced by $\theta(x)$. Two literals ρ and τ with distinct variables unify if there is a θ such that $\rho\theta = \tau\theta$. A most general unifier of two literals is one where any other unifier can be obtained by applying further substitutions.

A *clause* is a finite set of literals (including equalities and inequalities) which we understand intuitively as a disjunction. (The empty clause is understood as a formula that is always false.) For any formula $\phi \in \mathcal{L}$ without quantifiers, $CNF(\phi)$ is the finite set of clauses that results from putting ϕ into conjunctive normal form, defined as follows:

Definition 1 Assume that ϕ has no quantifiers and has been rewritten so that it does not use $\vee, \supset, \text{ or } \equiv$. Then $CNF(\phi)$ is defined inductively by:

1. $CNF(\phi) = \{\{\phi\}\}$, when ϕ is a literal;
2. $CNF(\phi \wedge \psi) = CNF(\phi) \cup CNF(\psi)$;
3. $CNF(\neg\neg\phi) = CNF(\phi)$;
4. $CNF(\neg(\phi \wedge \psi)) =$
 $\{a \cup b \mid a \in CNF(\neg\phi), b \in CNF(\neg\psi)\}$.

We obtain the following property of CNF :

Proposition 1 Let ϕ be any formula with no quantifiers.

If $CNF(\phi) = \{c_1, \dots, c_k\}$,
then $CNF(\neg\phi) = \{\{\bar{\rho}_1, \dots, \bar{\rho}_k\} \mid \bar{\rho}_i \in c_i\}$.

For any formula $\phi \in \mathcal{L}$, $SKO(\phi)$ is the formula without quantifiers that results from Skolemizing ϕ , defined as follows:

Definition 2 Assume that ϕ has been rewritten so that it does not use $\exists, \vee, \supset, \text{ or } \equiv$. Then let $SKO(\phi) = SKO'(\phi, \{\})$, which is defined inductively by:

1. $SKO'(\phi, X) = \phi$, when ϕ is a literal;
2. $SKO'((\phi \wedge \psi), X) = (SKO'(\phi, X) \wedge SKO'(\psi, X))$;
3. $SKO'(\neg\neg\phi), X) = SKO'(\phi, X)$;
4. $SKO'(\neg(\phi \wedge \psi), X) = (SKO'(\neg\phi, X) \vee SKO'(\neg\psi, X))$;

5. $SKO'(\forall x\phi, X) = SKO'(\phi, X \cup \{x\})$;
6. $SKO'(\neg\forall x\phi, X) = SKO'(\neg\phi_t^x, X)$,
 where $t = f(x_1, \dots, x_n)$, where $X = \{x_1, \dots, x_n\}$ and
 f is a function symbol (a so-called Skolem function) that
 occurs nowhere else.

In a nutshell, Skolemization involves replacing each existentially quantified variable in a formula by a new function applied to all the universally quantified variables that it appears within the scope of. We can get a version of Skolemization that replaces universally quantified variables instead of existential ones by using $\neg SKO(\neg\phi)$, which is sometimes referred to as *Herbrandization*. This is made precise in the following proposition:

Proposition 2 *For any objective ϕ , let ψ be the formula that results from removing all quantifiers from ϕ . Then there are substitutions θ and θ' such that $SKO(\phi) = \psi\theta$ and $SKO(\neg\phi) = \neg\psi\theta'$. Moreover, θ and θ' can be unified: there is a θ^* such that $\theta\theta^* = \theta'\theta^*$.*

Definition 3 *For any two literals ρ and τ , $MGU[\rho, \tau]$ is the set of most general unifiers of ρ and τ (empty if the two literals do not unify).*

Definition 4 $UNA = \{\{\neg(n = n')\} \mid n \text{ and } n' \text{ are distinct standard names}\}$.

Definition 5 *For function symbols F and predicate symbols P , $EQ_{F,P}$ is the union of the following sets of clauses:*

1. $\{(x = x)\}$,
2. $\{\neg(x = y), (y = x)\}$,
3. $\{\neg(x = y), \neg(y = z), (x = z)\}$,
4. $\{\neg(x_1 = y_1), \dots, \neg(x_k = y_k),$
 $(f(x_1, \dots, x_k) = f(y_1, \dots, y_k))\}$, for every k -ary
 function symbol $f \in F$,
5. $\{\neg(x_1 = y_1), \dots, \neg(x_k = y_k),$
 $\neg p(x_1, \dots, x_k), p(y_1, \dots, y_k)\}$,
 for every k -ary predicate symbol $p \in P$.

Note that these clauses are equivalent to the standard axioms of equality restricted to the symbols in F and P . When the context is clear, we omit the F and P subscripts.

Finally, we say that a set of clauses (possibly with variables) is *first-order satisfiable* (or fo-satisfiable) if it is satisfiable in classical first-order logic. There is a close connection between these notions and \mathcal{L} : a sentence ϕ is valid in \mathcal{L} , that is, $\models \phi$ iff the set of clauses $EQ \cup UNA \cup CNF(SK O(\neg\phi))$ is fo-unsatisfiable.

4 Believing and Only-Believing

To define the logic of belief in the next section, we begin by defining what it means to believe a clause and a sentence, following the approach outlined above in Section 2.

Definition 6 *An epistemic state is any finite set of clauses, where the variables in each clause are distinct.*

Definition 7 *A clause c is trivial iff it contains complementary literals or an element of UNA .*

Note that $(n = n)$ is not a trivial clause, but will be handled as a consequence of the first item of EQ defined above.

Definition 8 *For any set of clauses C let $F(C) = \bigcup_{c \in C} F(c)$, where $F(c) = \{c\} \cup F(\{c\theta \mid \{\rho, \tau\} \subseteq c, \rho \neq \tau, \theta \in MGU[\rho, \tau]\})$.*

Definition 9 *For any set of clauses C , $RP(C) = C \cup \{(a \cup b)\theta \mid \{\rho\} \cup a \in F(C), \{\tau\} \cup b \in F(C), \theta \in MGU[\bar{\rho}, \tau]\}$.*

In these definitions, we assume the clauses in C use distinct variables, and that just one θ is chosen (if one exists) so that the new clauses $c\theta$ and $(a \cup b)\theta$ also have distinct variables. Definition 8 realizes what is known as *Factoring*, that is, the unification of literals within the same clause, which is needed for Resolution to be complete. For example, if $[P(x), P(a), Q(x)]$ is in C , then $[P(a), Q(a)]$ is in $F(C)$. Note that $F(C)$, while recursive, is well defined, as in each step newly added clauses have fewer literals than those from which they were generated. Definition 9 then performs one step of resolution on all clauses in C extended by those obtainable by Factoring.

Definition 10 *For any integer $k \geq 0$, clause b , and set of clauses C , we say that b is believed at level k in epistemic state C (written $C, k \models b$) iff b is trivial or there exist θ and $a \in RP^k(C \cup EQ_{F,P})$ such that for every $\rho \in a\theta$, $\bar{\rho} \in UNA$ or $\rho \in b$. (The F and P here are sets that contain the function and predicate symbols in C or b .)*

This definition goes well beyond the propositional version given in Section 2. It handles equality by making EQ be part of the argument to RP , and it handles standard names by first, incorporating UNA in the definition of trivial, and second, by allowing $a \in RP$ to contain literals that are not in b as long they are guaranteed to be false (like $\#5 = \#7$).

We can now define what it means to believe an arbitrary objective sentence of \mathcal{L} :

Definition 11 *For any integer $k \geq 0$, objective sentence ψ , and set of clauses C , we say that ψ is believed at level k in epistemic state C iff there are $k + 1$ substitutions, $\theta_0, \dots, \theta_k$, such that for all clauses $b \in CNF(\psi'\theta_0 \vee \dots \vee \psi'\theta_k)$, where $\psi' = \neg SKO(\neg\psi)$, we have that $C, k \models b$.*

Note that believing ψ involves Skolemizing $\neg\psi$ (and then negating the result) so that existentially quantified variables remain but universally quantified ones disappear. So, for example, $\forall x.q(x)$ is believed at level k iff the clause $\{q(u)\}$ is believed at level k , where u is a new Skolem constant appearing nowhere else. (Since the KB in question cannot mention u , the sentence will only be believed if the KB entails a suitable universal.) But $\exists x.q(x)$ will be believed at level k iff there are terms t_0, \dots, t_k (not necessarily new ones) such that the clause $\{q(t_0), \dots, q(t_k)\}$ is believed at level k . For a concrete example, consider the sentence $\psi = \exists x\forall y(\neg P(y) \vee P(x))$, which is valid and equivalent to $\forall y\neg P(y) \vee \exists xP(x)$. After Skolemizing the universal we obtain $\psi' = (\neg P(f(x)) \vee P(x))$. Now consider two substitutions θ_0 and θ_1 , where θ_0 maps x to itself and θ_1 maps x to $f(x)$. Then $CNF(\psi'\theta_0 \vee \psi'\theta_1)$ is the trivial

clause $[\neg P(f(x)), P(x), \neg P(f(f(x))), P(f(x))]$. Hence, if C is empty then ψ is believed at level 1 but not at level 0.

Finally, let us now consider what it means to say that ϕ is all that is believed:

Definition 12 For any objective sentence ϕ and set of clauses C , we say that ϕ is only-believed in epistemic state C iff there is a relabelling of the variables and Skolem functions in C such that $C = \text{CNF}(\text{SKO}(\phi))$.

As mentioned above, having ϕ logically equivalent to ψ does not make only-believing ϕ the same as only-believing ψ . For example, let $\phi = (p \wedge (\neg p \vee q))$ and let $\psi = (\phi \wedge q)$. Then only-believing ϕ requires an epistemic state with exactly two clauses, while only-believing ψ requires an epistemic state with three clauses. These two epistemic states will have different beliefs: the sentence q will not be believed at level 0 in the first one, but it will in the second. (They will agree, however, on beliefs at level 1 and beyond.)

Note also that only-believing ϕ involves Skolemizing ϕ itself (and not its negation), so that universally quantified variables remain but existentially quantified ones now disappear. So, for example, $\forall x.q(x)$ is only-believed in C iff C consists of the single clause $\{q(x)\}$ (or a variant with some other variable). But $\exists x.q(x)$ is only-believed in C iff C consists of the single clause $\{q(v)\}$, where v is a Skolem constant.

5 A Logic of Limited Reasoning

Our logic of belief extends \mathcal{L} by adding belief operators. In particular, for any objective formula ϕ and any non-negative integer k , $B_k\phi$ and $O\phi$ are also formulas. Note that we do not consider nested beliefs here. A formula is called *subjective* if all predicate and function symbols appear within the scope of a modal operator.

The semantics of sentences of the extended language is defined with respect to worlds and epistemic states, that is, sets of clauses. For the objective part, the semantics is identical to that of \mathcal{L} .

Definition 13 For any set of clauses C , any world w , and any sentence α , we say that α is true wrt C and w (written $C, w \models \alpha$) according to the following definition:

1. $C, w \models p(t_1, \dots, t_m)$ iff $w[p(|t_1|_w, \dots, |t_m|_w)] = 1$;
2. $C, w \models (t_1 = t_2)$ iff $|t_1|_w$ and $|t_2|_w$ are ident. names;
3. $C, w \models \neg\alpha$ iff $C, w \not\models \alpha$;
4. $C, w \models (\alpha \wedge \beta)$ iff $C, w \models \alpha$ and $C, w \models \beta$;
5. $C, w \models \forall x.\alpha$ iff $C, w \models \alpha_n^x$ for all standard names n ;
6. $C, w \models B_k\phi$ iff ϕ is believed at level k in ep. state C ;
7. $C, w \models O\phi$ iff ϕ is only-believed in epistemic state C .

When α is subjective, we leave out the w and write $C \models \alpha$. We say that α is valid in this logic (written $\models \alpha$) iff $C, w \models \alpha$ for every finite set of clauses C and world w .

Before looking at the general properties of this logic, let us consider a simple example:

We show that $\models (O\forall x(x \neq \#5 \supset p(x)) \supset B_0p(\#7))$. Suppose that $C \models O\forall x(x \neq \#5 \supset p(x))$. Then C consists of one clause, $\{x = \#5, p(x)\}$, modulo renaming of x . So that

clause is an element of $RP^0(C \cup EQ)$. Now let θ_0 be the empty substitution. Let $\psi' = \neg\text{SKO}(\neg p(\#7)) = \neg\neg p(\#7)$, so that $\text{CNF}(\psi'\theta_0) = \{\{p(\#7)\}\}$. So for every $b \in \text{CNF}(\psi'\theta_0)$ there is indeed a clause $a \in RP^0(C \cup EQ)$ (namely $\{x = \#5, p(x)\}$) and a θ (namely $[x/\#7]$), such that for each literal $\rho \in a\theta$, either $\bar{\rho} \in \text{UNA}$ (when ρ is $\#7 = \#5$) or $\rho \in b$ (when ρ is $p(\#7)$). Therefore, $C \models B_0p(\#7)$.

6 Properties of the Logic

In this section, we confirm that the logic of belief defined above behaves appropriately. We prove that it has the desired properties listed in Section 2.

6.1 Satisfying the Desiderata

Theorem 1 (Cumulativity) $\models (B_k\psi \supset B_{k+1}\psi)$.

Proof: The theorem follows easily from the fact that for any a, k , and C , if $C, k \models a$ then $C, k + 1 \models a$. ■

For expressiveness, we need this lemma:

Lemma 1 ϕ is believed at level 0 in the epistemic state $\text{CNF}(\text{SKO}(\phi))$.

Proof: Let $C = \text{CNF}(\text{SKO}(\phi))$ and $D = \text{CNF}(\neg\text{SKO}(\neg\phi))$. We need to show that there is a θ_0 such that for every $b \in \text{CNF}(D\theta_0)$, $C, 0 \models b$. Let θ, θ' , and θ^* be as in Proposition 2. So $C = \text{CNF}(\psi)\theta$ and $D = \text{CNF}(\psi)\theta'$. Let $\theta_0 = \theta^*$ and suppose b is any element of $D\theta_0 = \text{CNF}(\psi)\theta'\theta^*$. So there is a $b' \in \text{CNF}(\psi)$ such that $b = b'\theta'\theta^*$. But $b'\theta'\theta^* = b'\theta\theta^*$. So $b'\theta \in C$ and $b'\theta\theta^* = b$. Therefore, $C, 0 \models b$. ■

Theorem 2 (Expressiveness)

For any objective sentence ϕ , the sentence $O\phi$ is satisfiable, and moreover $\models (O\phi \supset B_0\phi)$.

Proof: Let $C = \text{CNF}(\text{SKO}(\phi))$. Then $C \models O\phi$. Now assume that C' is any finite set of clauses. If $C' \models O\phi$, then C' is like C except for a possible relabelling. By Lemma 1, ϕ is believed at level 0 in C , and so $C' \models B_0\phi$. ■

Turning now to tractability, we have the following:

Lemma 2 There is a procedure which given b, k and C can decide whether or not $C, k \models b$.

Proof: Here is a sketch of the procedure: First calculate $C' = RP^k(C \cup EQ)$ in the obvious way. (This C' will be finite.) Then confirm that either b is trivial or there is a θ and a $c \in C'$ such that for every $\rho \in c\theta$, $\bar{\rho} \in \text{UNA}$ or $\rho \in b$. (This is a small tweak to a procedure known to be decidable called theta-subsumption.) ■

Theorem 3 (Tractability) Suppose that $\phi_1, \dots, \phi_N, \psi$ are sentences such that $|\phi_i| < c$ and $|\psi| < c$, where c is some constant. Let $\phi = (\phi_1 \wedge \dots \wedge \phi_N)$. Then for any $k \geq 0$, the validity of $(O\phi \supset B_k\psi)$ can be determined in time polynomial in N .

Proof: Here is a sketch of a decision procedure which runs in time polynomial in N : First, calculate $C = CNF(SKO(\phi))$, $\psi' = \neg SKO(\neg\psi)$, and guess at a $\Theta = \{\theta_0, \dots, \theta_k\}$. Then calculate $B = CNF(\psi'\theta_0 \vee \dots \vee \psi'\theta_k)$. Finally, use the procedure of Lemma 2 to confirm for each $b \in B$ that $C, k \models b$. The θ_i in question will be substitutions that map the variables in ψ' to the (finitely many) terms that appear in $SKO(\phi)$ or $SKO(\neg\psi)$. ■

Note that this procedure requires converting the entire knowledge base ϕ to CNF , which can be exponential in general. However, it is polynomial here since ϕ is assumed to be a conjunction of small sentences. For any fixed k , the calculation of $C' = RP^k(C \cup EQ)$ required by Lemma 2 will be polynomial in $|C|$, but of course, this will grow exponentially as k increases. A similar argument applies when converting $(\psi'\theta_0 \vee \dots \vee \psi'\theta_k)$ into CNF , since the query is also assumed to be bounded in size. Note that, in the case of Factoring (F), which is part of RP , for each clause at most a polynomial number of new clauses is added. (Note also that the RP^k calculation only needs to be done once, however, and can even be done offline before considering any queries.) The “guessing” of an appropriate Θ in the proof can be made determinate by trying all potential MGUs between terms in ψ' and terms in C , of which there are only polynomially many. Overall, under the given assumptions and for any fixed k , the procedure will be polynomial in the size of the knowledge base.

To obtain soundness and eventual completeness, we need the following lemmas:

Lemma 3 *Let C be a finite set of clauses, and ϕ an objective formula. If for all $d \in CNF(\phi)$, $C, k \models d$, then $C \cup EQ \cup UNA \cup \{\neg\phi\}$ is fo-unsatisfiable.*

Proof: Suppose M is a first-order interpretation and M satisfies $C \cup EQ \cup UNA$. Since M satisfies UNA , M satisfies any trivial clause and M also satisfies $RP^k(C \cup EQ)$ (by induction on k). So for any d , if $C, k \models d$, then M satisfies d . Therefore, M satisfies all the clauses in $CNF(\phi)$. So M satisfies ϕ and does not satisfy $\neg\phi$. ■

Lemma 4 *Let C be a finite set of clauses, and ϕ a formula. (Both can have variables.) Then the set of clauses $C \cup EQ \cup UNA \cup CNF(\phi)$ is fo-unsatisfiable iff there is a k and substitutions $\theta_0, \dots, \theta_k$ such that for all $d \in CNF(\neg\phi\theta_0 \vee \dots \vee \neg\phi\theta_k)$, $C, k \models d$.*

Proof: (Sketch) (\Leftarrow) Suppose that for all $d \in CNF(\neg\phi\theta_0 \vee \dots \vee \neg\phi\theta_k)$, $C, k \models d$. By Lemma 3, $C \cup EQ \cup UNA \cup \{\phi\theta_0 \wedge \dots \wedge \phi\theta_k\}$ is fo-unsatisfiable, and so the set $C \cup EQ \cup UNA \cup CNF(\phi)$ must be fo-unsatisfiable also.

(\Rightarrow) Now suppose $C \cup EQ \cup UNA \cup CNF(\phi)$ is fo-unsatisfiable. Let $D = CNF(\phi)$. By the refutation completeness of Resolution, there is a resolution derivation of the empty clause from the clauses in $C \cup EQ \cup UNA \cup D$. Say the derivation is b_0, \dots, b_r where $b_r = \{\}$. We prove the following claim: For every $0 \leq j \leq r$, there is a k such that for every θ , there are substitutions $\theta_0, \dots, \theta_k$ such that for all clauses $d \in CNF(\neg\phi\theta_0 \vee \dots \vee \neg\phi\theta_k)$, $C, k \models b_j\theta \cup d$. The theorem will then follow since $b_r = \{\}$.

The proof is by induction on j . For the base case ($j = 0$), we have $b_0 \in C \cup EQ \cup UNA \cup D$, and it can be shown that the lemma holds for $k = 0$. In the induction step, b_j can be either the result of a resolution step or factoring. In both cases, there is a k such that $C, k \models b_j\theta \cup d$. ■

Theorem 4 (Soundness and Eventual Completeness) *Let ϕ and ψ be objective sentences. Then $\models (\phi \supset \psi)$ iff for some k , $\models (O\phi \supset B_k\psi)$.*

Proof: We have $\models (\phi \supset \psi)$ iff $EQ \cup UNA \cup CNF(SKO(\phi \wedge \neg\psi))$ is fo-unsatisfiable iff $CNF(SKO(\phi)) \cup EQ \cup UNA \cup CNF(SKO(\neg\psi))$ is fo-unsatisfiable iff (by Lemma 4) for some k , there are substitutions $\theta_0, \dots, \theta_k$ such that for all $d \in CNF(\psi'\theta_0 \vee \dots \vee \psi'\theta_k)$, $C, k \models d$, where $\psi' = \neg SKO(\neg\psi)$ iff for some k , ψ is believed at level k in epistemic state $CNF(SKO(\phi))$ iff for some k , $\models (O\phi \supset B_k\psi)$. ■

6.2 The Logical Properties of Belief

Theorem 5 (Equivalent beliefs) *For any k and any objective sentences ϕ , ψ , and χ , the following sentences are valid:*

$$\begin{aligned} B_k\phi &\equiv B_k(\phi \wedge \phi), \\ B_k\phi &\equiv B_k(\phi \vee \phi), \\ B_k\phi &\equiv B_k\neg\neg\phi, \\ B_k(\phi \wedge \psi) &\equiv B_k(\psi \wedge \phi), \\ B_k(\phi \vee \psi) &\equiv B_k(\psi \vee \phi), \\ B_k(\neg(\phi \wedge \psi)) &\equiv B_k(\neg\phi \vee \neg\psi), \\ B_k(\neg(\phi \vee \psi)) &\equiv B_k(\neg\phi \wedge \neg\psi), \\ B_k(\phi \wedge (\psi \wedge \chi)) &\equiv B_k((\phi \wedge \psi) \wedge \chi), \\ B_k(\phi \vee (\psi \vee \chi)) &\equiv B_k((\phi \vee \psi) \vee \chi), \\ B_k(\phi \wedge (\psi \vee \chi)) &\equiv B_k((\phi \wedge \psi) \vee (\phi \wedge \chi)), \\ B_k(\phi \vee (\psi \wedge \chi)) &\equiv B_k((\phi \vee \psi) \wedge (\phi \vee \chi)), \end{aligned}$$

These follow from the fact that $CNF(\neg SKO(\neg\phi)) = CNF(\neg SKO(\neg\psi))$ implies $\models (B_k\phi \equiv B_k\psi)$.

Theorem 6 (Non-equivalent beliefs) *There are obj. sentences ϕ and ψ such that $\models (\phi \equiv \psi)$, but $\not\models (B_k\phi \equiv B_k\psi)$.*

Proof: Let p and q be distinct atomic sentences. Let $\phi = (p \wedge (\neg p \vee q))$ and $\psi = (\phi \wedge q)$. Then $\models (\phi \equiv \psi)$, but $\not\models (B_0\phi \equiv B_0\psi)$. ■

The final thing to show is the relationship between beliefs and logical combinations of other beliefs.

Theorem 7 *For any sentences ϕ and ψ and any formula χ with a single-free variable x :*

1. $\models (B_k\phi \vee B_k\psi) \supset B_k(\phi \vee \psi)$.
2. $\models \exists x B_k\chi \supset B_k\exists x\chi$.
3. $\models B_k(\phi \wedge \psi) \supset (B_k\phi \wedge B_k\psi)$, but the converse fails.
4. $\models B_k\forall x\chi \supset \forall x B_k\chi$, but the converse fails.

Proof: Here we only prove (3.): The formula $\neg SKO(\neg(\phi \wedge \psi))$ is $(\neg SKO(\neg\phi) \wedge \neg SKO(\neg\psi))$. So $CNF(\neg SKO(\neg(\phi \wedge \psi))) = CNF(\neg SKO(\neg\phi)) \cup CNF(\neg SKO(\neg\psi))$. So if $C \models B_k(\phi \wedge \psi)$, then $C \models B_k\phi$ and $C \models B_k\psi$. Regarding the converse, let $\phi = \exists x.p(x)$, $\psi = \exists y.q(y)$, and $C = \{\{p(\#1), p(\#2)\}, \{q(\#1), q(\#2)\}\}$. Then $C \models B_1\phi$

and $C \models B_1\psi$, but $C \not\models B_1\exists x\exists y.p(x) \wedge q(y)$ and so $C \not\models B_1(\phi \wedge \psi)$. ■

We do however get “eventual” versions of the closure of belief under conjunction and universal generalization:

Theorem 8 *For any ϕ, ψ, χ as above, any k and finite C :*

1. *If $C \models (B_k\phi \wedge B_k\psi)$, then $\exists k', C \models B_{k'}(\phi \wedge \psi)$.*
2. *If $C \models \forall x B_k\chi$, then $\exists k', C \models B_{k'}\forall x\chi$.*

7 Related Work

Perhaps the best understood formal models of belief are those based on possible-world semantics [Kripke, 1959; Hintikka, 1962; Halpern and Moses, 1992]. However, it was recognized early on that belief under this framework suffers from *logical omniscience* [Hintikka, 1975], that is, beliefs are closed under (classical) logical entailment and hence reasoning becomes intractable in the propositional and even undecidable in the first-order case.

Later, weaker logics of belief started to emerge with a focus on tractability, but initially restricted to the propositional case. These include [Levesque, 1984; Cadoli and Schaerf, 1996; Frisch, 1987; Fagin *et al.*, 1990; Delgrande, 1995; Cadoli and Schaerf, 1996] and either provided an explicit model of limited belief or a limited entailment relation, often based on *tautological entailment* [Dunn, 1976], a fragment of relevance logic [Anderson and Belnap, 1975]. There were also proposals for tractable entailment relations of increasing complexity [Dalal, 1996; Crawford and Etherington, 1998; D’Agostino, 2015], again limited to the propositional case. In a first-order setting, tractable reasoning was investigated either in terms of limited inference [Levesque, 1998; Liu and Levesque, 2003] or in terms of belief levels as in this paper [Liu *et al.*, 2004; Liu and Levesque, 2005; Lakemeyer and Levesque, 2013; 2014; 2016; Schwering, 2017]. In this line of work, epistemic states, which they call *setups*, are also sets of clauses, but in contrast to our approach they need to be ground. Belief levels are defined in terms of splitting, either on a clause in the epistemic state [Liu *et al.*, 2004; Liu and Levesque, 2005; Lakemeyer and Levesque, 2013], an arbitrary ground literal [Lakemeyer and Levesque, 2014], or on the possible denotations of a term [Lakemeyer and Levesque, 2016; Schwering, 2017]. For example, splitting on a literal l at belief level k means adding l to the epistemic state and then checking whether the belief in question obtains at level $k-1$, and doing the same for the complement of l . These approaches, as well as [Crawford and Etherington, 1998; D’Agostino, 2015] discussed next, go further than us in allowing the use of *unit propagation*, that is, resolution with a clause containing a single literal, already at belief level 0. In all cases reasoning is tractable for every belief level k . While the knowledge base needs to be in CNF, query evaluation does not require a transformation into CNF. Schwering [2017] also provides an implementation for his logic with interesting results for games like Sudoku or Minesweeper.

Closely related to ours is the work by D’Agostino [2015], who proposes a three-valued nondeterministic semantics which was also considered in [Crawford and Etherington,

1998], together with a k -consequence relation, where splitting is allowed on arbitrary formulas at levels greater than 0. Again, tractability results are obtained for every level k , and here neither the KB nor the query need to be transformed into CNF. The author also provides a sound and complete proof-theoretic account of the k -consequence relation. But again, this is all for a propositional language only.

There are very few expressive models of limited reasoning handling full first-order logic such as [Patel-Schneider, 1985; Lakemeyer, 1996; Lakemeyer and Levesque, 2016]. None of the approaches are eventually complete for full first-order logic, although a few proposals, such as [Lakemeyer and Levesque, 2014; D’Agostino, 2015; Klassen *et al.*, 2015; Lakemeyer and Levesque, 2016], do exhibit eventually completeness for the propositional case.

Finally, returning to Hintikka, he considered ideas related to ours a long time ago [Hintikka, 1973]. His proposal deals with full first-order logic and distinguishes notions such as trivial inconsistencies vs. those that need to be discovered. In contrast to us, he uses a special disjunctive normal form, where each disjunct represents possible worlds. There is no discussion of tractability, however. We leave it to future work to investigate what the exact connections are.

8 Conclusions

In this paper we proposed, for the first time, a first-order logic of limited belief which is at once tractable, expressive, and eventually complete. One possible direction for future work involves simplifying the semantics by replacing the use of clauses in the definition of belief by three-valued interpretations. (Interestingly, this seems to work fine in the propositional case, but in the first-order case, the use of Skolemization and what amounts to the Herbrand Theorem still appears to be needed.) There are at least two directions in which the logic could be extended. For one, we so far completely ignored nested beliefs, which are necessary when an agent needs to introspect on what is and is not believed. Previous approaches to limited belief that considered introspection include [Lakemeyer, 1996; Lakemeyer and Levesque, 2013]. For another, there are additional easy forms of reasoning one may want to include even at belief level 0. A simple example would be *unit propagation* which, as noted in the related-work section, was included in the proposals presented in [Crawford and Etherington, 1998] and later in [D’Agostino, 2015] as well as [Liu *et al.*, 2004] and its descendants. In the case of a very expressive language like ours, one must be careful, however, as an unrestricted use of unit propagation is already undecidable. Possible restrictions could be to restrict unit propagation to when the arguments of the predicates are standard names, or to when the nesting of function symbols does not exceed a certain depth.

Acknowledgments

We thank the reviewers for their thoughtful comments, which helped improve the presentation of the paper. Gerhard Lakemeyer also acknowledges his affiliation with the University of Toronto while this work was conducted.

References

- [Anderson and Belnap, 1975] A.R. Anderson and N.D. Belnap. *Entailment: The Logic of Relevance and Necessity*. Princeton University Press, 1975.
- [Buss, 1998] S.R. Buss. An Introduction to Proof Theory. In S.R. Buss (Ed.), *Handbook of Proof Theory*, Elsevier, 1998.
- [Chellas, 1980] B. Chellas, *Modal Logic: An Introduction*. Cambridge University Press, Cambridge, 1980.
- [Cadoli and Schaerf, 1996] M. Cadoli and M. Schaerf. On the complexity of entailment in propositional multivalued logics. *Annals of Mathematics and Artificial Intelligence*, 18(1):29–50, 1996.
- [Crawford and Etherington, 1998] J.M. Crawford and D. Etherington. A non-deterministic semantics for tractable inference. In *Proc. of AAAI-98*, 286–291, 1998.
- [Dalal, 1996] M. Dalal. Semantics of an anytime family of reasoners. In *Proc. of the 12th European Conference on Artificial Intelligence (ECAI-96)*, pages 360–364, 1996.
- [D’Agostino, 2015] M. D’Agostino. An informational view of classical logic. *Theor. Comput. Sci.* 606, 79–97, 2015.
- [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 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.
- [Frisch, 1987] A. Frisch. Inference without chaining. In *Proc. of IJCAI-87*, pages 515–519, 1987.
- [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, 1973] J. Hintikka. *Logic, Language Games and Information. Kantian themes in the philosophy of logic*. Clarendon Press, Oxford, 1973.
- [Hintikka, 1975] J. Hintikka. Impossible Possible Worlds Vindicated. *Journal of Philosophical Logic*, 4, 475–484, 1975.
- [Klassen et al., 2015] T.Q. Klassen, S.A. McIlraith, and H.J. Levesque. Towards Tractable Inference for Resource-Bounded Agents. AAAI Spring Symposium on Logical Formalizations of Commonsense Reasoning, 89–95, AAAI Press, 2015.
- [Kripke, 1959] S. Kripke. A completeness theorem in modal logic. *Journal of Symbolic Logic*, 24:1–14, 1959.
- [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, 2013] G. Lakemeyer and H.J. Levesque. Decidable Reasoning in a Logic of Limited Belief with Introspection and Unknown Individuals. *23rd International Joint Conference on Artificial Intelligence (IJCAI)*, 2013.
- [Lakemeyer and Levesque, 2014] G. Lakemeyer and H.J. Levesque. Decidable Reasoning in a Fragment of the Epistemic Situation Calculus. *Int. Conf. on the Principles of Knowledge Representation and Reasoning (KR-14)*, Vienna, 2014.
- [Lakemeyer and Levesque, 2016] G. Lakemeyer and H.J. Levesque. Decidable Reasoning in a Logic of Limited Belief with Function Symbols. In *Proc of KR-16*, pages 288–297, 2016.
- [Levesque, 1984] H.J. Levesque. A logic of implicit and explicit belief. In *Proc. of AAAI-84*, pages 198–202, 1984.
- [Levesque, 1990] H.J. Levesque. All I know: a study in autoepistemic logic. *Artificial Intelligence* 42, 263–309, 1990.
- [Levesque, 1998] H.J. Levesque. A completeness result for reasoning with incomplete first-order knowledge bases. In *Proc. KR-98*, pages 14–23, 1998.
- [Levesque and Lakemeyer, 2001] H.J. Levesque and G. Lakemeyer. *The Logic of Knowledge Bases*. MIT Press, 2001.
- [Liu et al., 2004] Y. Liu, G. Lakemeyer, and 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.
- [Liu and Levesque, 2003] Y. Liu and H.J. Levesque. A tractability result for reasoning with incomplete first-order knowledge bases. In *Proc. of IJCAI-03*, pages 83–88, 2003.
- [Liu and Levesque, 2005] Y. Liu and H.J. Levesque. Tractable reasoning in first-order knowledge bases with disjunctive information. In *Proc. of the Twentieth National Conference on Artificial Intelligence (AAAI-05)*, 2005.
- [McCarthy, 1963] J. McCarthy. Programs with common sense. M. Minsky (ed.), *Semantic Information Processing*, MIT Press, Cambridge, 403–418, 1963.
- [Patel-Schneider, 1985] P. Patel-Schneider. A decidable first-order logic for knowledge representation. In *Proc. of IJCAI-85*, pages 455–458, 1985.
- [Schwering, 2017] C. Schwering. A Reasoning System for a First-Order Logic of Limited Belief. In *Proc. of IJCAI-17*, pages 1247–1253, 2017.