Towards Tractable Inference for Resource-Bounded Agents

Toryn Q. Klassen and **Sheila A. McIlraith** and **Hector J. Levesque**
Department of Computer Science
University of Toronto
Toronto, Ontario, Canada
{toryn,sheila,hector}@cs.toronto.edu

Abstract

For a machine to act with common sense, it is not enough that information about commonsense things be written down in a formal language. What *actual* knowledge—i.e. conclusions available for informing actions—a formalization is meant to provide cannot be determined without some specification of what sort of reasoning is expected. The traditional view in epistemic logic says that agents see all logical consequences of the information they have, but that would give agents capabilities far beyond common sense or what is physically realizable. To work towards addressing this issue, we introduce a new epistemic logic, based on a three-valued version of neighborhood semantics, which allows for talking about the effort used in making inferences. We discuss the advantages and limitations of this approach and suggest that the ideas used in it could also find a role in autoepistemic reasoning.

1 Introduction

Commonsense reasoning is easy for people. This easiness is not an incidental property, but rather what makes such reasoning widespread and useful in everyday life. Part of understanding common sense is knowing what it does *not* encompass, e.g. being able to solve complicated puzzles that merely happen to mention commonplace objects like piggy banks or broken eggs. Therefore, elaborate formalizations of commonsense knowledge that are made without regard for how reasoning will be done in them may miss the point.

We would like, then, to have a formal system that tells us which inferences are reasonably easy for an agent to make and which are hard—an epistemic logic in which we can say that things are obvious (or a doxastic logic, but we will not be distinguishing between belief and knowledge). Let us note that the standard approach in epistemic logic, following (Hintikka 1962), does not fit our purpose at all. In the standard approach, an agent's uncertainty about the world is modeled with a set of "possible worlds", each of which is associated with a truth assignment that describes one of the different ways the agent thinks reality might be. The agent believes whatever is true in all the worlds. If all sentences in a set Γ are true in every world, then so is any sentence α that is a logical consequence of Γ. Hence the agent believes all logical consequences of its beliefs.

This unrealistic property, termed "logical omniscience" (or "deductive omniscience"), has inspired considerable discussion in the literature. That knowledge and belief in (Hintikka 1962) were modeled as being "much too strong" was pointed out in (Castañeda 1964), an otherwise largely positive book review. Hocutt (1972) wrote that real people are "logically obtuse" rather than omniscient. Stalnaker (1991) wrote that

> [A]ny kind of information processing or computation is unintelligible as an activity of a deductively omniscient agent. It is hard to see what a logic of knowledge could be for if it were a harmless simplification for it to ignore these activities that are so essential to rationality and cognition.

There have been numerous proposals for epistemic logics that avoid logical omniscience, and we will not have space to discuss all of them in this paper. For more information, the reader is referred to the survey papers (McArthur 1988; Sim 1997; Moreno 1998; Whitsey 2003).

Hintikka (1962, Section 2.10) suggested an alternative interpretation (albeit an interpretation he did not favor) of his modal "knowledge" operator as indicating what follows from the agent's knowledge, rather than the knowledge itself. This interpretation was taken by Levesque (1984), who wrote that logical omniscience characterizes "implicit belief", but that "explicit" (i.e real) beliefs should be described by a different logic.

The distinction between implicit and explicit belief is useful, and one that we will frequently refer to. However, a single type of explicit belief is not really enough. The amount of effort that an agent is expected to apply to a problem will depend on context. Therefore, we will be following the line of research developed in (Liu, Lakemeyer, and Levesque 2004; Liu 2006; Lakemeyer and Levesque 2013; 2014) in which there is an infinite family of "levels of belief".

An intuitive understanding of a level of belief is that a sentence is in a level if it can be concluded with an amount of effort that is bounded by the number of that level. So, instead of talking about what agents do believe, the logics in these papers describe what agents would believe, conditioned on their spending a given amount of effort. An autonomous agent would need to have some mechanism to determine how much effort was appropriate to spend in a given

situation. The logics are not meant to define such a mechanism, but to specify the behavior of a *reasoning service* that an agent with such a mechanism could make use of. For an idea as to how such a service could be used, see (Kowalski 1995), which suggested defining agents to cyclically consume inputs, reason for a bounded amount of time, and take action. The bound would provide a compromise between operating in a deliberative and reactive manner.

The logic we will be developing in this paper is inspired by (the propositional fragment of) the logic $\mathcal{ESL}$ from (Lakemeyer and Levesque 2014). However, the semantics of $\mathcal{ESL}$ are defined in a rather ad hoc and syntactic way, whereas ours will be based on neighborhood semantics. The outline of this paper is as follows: In section 2, we provide background information and introduce notation that will be needed. Section 3 shows how a simple explicit belief operator can be defined in a three-valued neighborhood semantics. Then, in section 4, we also define a system of levels of belief. In section 5 we compare our logic with $\mathcal{ESL}$ and consider ways in which it would be interesting to extend our work, and in section 6 we point out some related work.

2 Background

Suppose S and T are sets. We will write $S \to T$ to denote the set of all functions from S to T. The power set of S will be denoted by $\mathscr{P}(S)$. If $\prec$ is a partial order, then $\min_{\prec}(S)$ is the set of minimal elements of S according to $\prec$.

A propositional *logic* is defined by three things: a language (set of sentences), a set of semantic objects, and a satisfaction relation between semantic objects and sentences (which indicates which semantic objects make which sentences true). The language determines the *syntax* of the logic, while the semantic objects and satisfaction relation together determine the *semantics*.

Let us assume that we have some non-empty set (possibly infinite) Φ of *atomic* symbols. The propositional language $\mathscr{L}(\Phi)$ is defined by the grammar

$$\alpha ::= p \mid (\alpha \wedge \alpha) \mid \neg\alpha$$

where $p \in \Phi$. As is conventional, $\wedge$ is meant to be understood as a conjunction operator and $\neg$ as a negation operator. We can define other operators like disjunction, the material conditional, and equivalence as the usual abbreviations: $(\alpha \vee \beta) := \neg(\neg\alpha \wedge \neg\beta)$, $(\alpha \supset \beta) := (\neg\alpha \vee \beta)$, and $(\alpha \equiv \beta) := ((\alpha \supset \beta) \wedge (\beta \supset \alpha))$

Lowercase Latin characters $p, q, r, \ldots$ will typically be used to denote atoms in Φ, and Greek letters $\alpha, \beta, \gamma, \ldots$ to denote sentences of $\mathscr{L}(\Phi)$. We may use subscripts on letters. The length of a sentence α, written $\mathsf{len}(\alpha)$, is defined inductively as follows:

$$\begin{aligned} \mathsf{len}(p) &= 1 \\ \mathsf{len}(\neg\alpha) &= 1 + \mathsf{len}(\alpha) \\ \mathsf{len}((\alpha \wedge \beta)) &= 3 + \mathsf{len}(\alpha) + \mathsf{len}(\beta) \end{aligned}$$

We will be considering two different logics using the language $\mathscr{L}(\Phi)$, the classical two-valued logic and a three-valued logic (specifically, Kleene's three-valued logic from (Kleene 1938)). The two classical *truth values* are T ("true") and F ("false"), and for three-valued logic there is a third truth value that we will call N ("neither"). Let $\mathbb{C} := \{\mathsf{T}, \mathsf{F}\}$ and $\mathbb{K} := \{\mathsf{T}, \mathsf{F}, \mathsf{N}\}$.

The semantic objects of three-valued logic are functions, called *truth assignments*, from the set $\Phi \to \mathbb{K}$. We will denote the satisfaction relation of three-valued logic by $\models_{\mathbb{K}}$. For $v \in \Phi \to \mathbb{K}$ and $\alpha \in \mathscr{L}(\Phi)$, $v \models_{\mathbb{K}} \alpha$ iff $v'(\alpha) = \mathsf{T}$, where $v' \in \mathscr{L}(\Phi) \to \mathbb{K}$ is defined in terms of v as follows:

$$v'(p) = v(p) \text{ for } p \in \Phi$$

$$v'(\neg\alpha) = \begin{cases} \mathsf{T} & \text{if } v'(\alpha) = \mathsf{F} \\ \mathsf{F} & \text{if } v'(\alpha) = \mathsf{T} \\ \mathsf{N} & \text{if } v'(\alpha) = \mathsf{N} \end{cases}$$

$$v'(\alpha \wedge \beta) = \begin{cases} \mathsf{T} & \text{if } v'(\alpha) = \mathsf{T} \text{ and } v'(\beta) = \mathsf{T} \\ \mathsf{F} & \text{if } v'(\alpha) = \mathsf{F} \text{ or } v'(\beta) = \mathsf{F} \\ \mathsf{N} & \text{otherwise} \end{cases}$$

We can identify an element of $\Phi \to \mathbb{K}$ with the set of *literals* it makes true (a literal is an atom or the negation of an atom). This enables us to compare elements of $\Phi \to \mathbb{K}$ with the subset relation, to talk of them being finite or infinite, and to take intersections and (sometimes) unions. Note that if $u \in \Phi \to \mathbb{K}$ and $v \in \Phi \to \mathbb{K}$, $u \cup v$ might not be an element of $\Phi \to \mathbb{K}$, because there might be some $p \in \Phi$ such that both $p \in u \cup v$ and $\neg p \in u \cup v$.

Definition 1 (compatibility)**.** $u \in \Phi \to \mathbb{K}$ and $v \in \Phi \to \mathbb{K}$ are compatible, written $u \diamond v$, if $u \cup v \in \Phi \to \mathbb{K}$.

Given the three-valued logic we have described, we can think of classical two-valued logic as a restriction of it, which differs only in that truth functions cannot map any atom to N. That is, the semantic objects in classical logic are elements of $\Phi \to \mathbb{C}$ (we can view $\Phi \to \mathbb{C}$ as a subset of $\Phi \to \mathbb{K}$ by identifying functions with their graphs), and the satisfaction relation $\models_{\mathbb{C}}$ in classical logic coincides with $\models_{\mathbb{K}}$ for elements of $\Phi \to \mathbb{C}$.

The *valid* sentences or *tautologies* of a logic are those which are satisfied by every semantic object. If α is a sentence in the language of a logic with satisfaction relation $\models$, then we will write $\models \alpha$ to indicate that α is valid in that logic. Note that there are no valid sentences in Kleene's three-valued logic, since $\emptyset \in \Phi \to \mathbb{K}$ and $\emptyset$ does not make any sentence true. For a set of sentences Γ, we will write $\Gamma \models \alpha$ if every semantic object that makes every $\gamma \in \Gamma$ true also makes α true. We may write $\gamma \models \alpha$ to mean $\{\gamma\} \models \alpha$.

In any logic, the *proposition* expressed by a sentence is the set of semantic objects that satisfy that sentence. For classical and three-valued logics, let us introduce some notation to denote the propositions expressed by sentences:

$$[\![\alpha]\!]^{\mathbb{C}} := \{v \in \Phi \to \mathbb{C} : v \models_{\mathbb{C}} \alpha\}$$

$$[\![\alpha]\!]^{\mathbb{K}} := \{v \in \Phi \to \mathbb{K} : v \models_{\mathbb{K}} \alpha\}$$

We will also find the following definition useful:

$$|\!|\alpha|\!| := \min_{\subset}([\![\alpha]\!]^{\mathbb{K}})$$

If there are an infinite number of atoms, $[\![\alpha]\!]^{\mathbb{K}}$ is always either infinite or empty, while $|\!|\alpha|\!|$ is always finite and all elements of it are finite. Note that for any $\alpha, \beta \in \mathscr{L}(\Phi)$, $v \models_{\mathbb{K}} \beta$ for all $v \in [\![\alpha]\!]^{\mathbb{K}}$ if and only if $v \models_{\mathbb{K}} \beta$ for all $v \in |\!|\alpha|\!|$.

2.1 Neighborhood semantics

Neighborhood semantics (sometimes called Montague-Scott semantics) for modal logic were suggested by (Montague 1968) and (Scott 1970). Various forms of these semantics have been used in AI for modeling belief; a survey can be found in (Sim 1997, Section IV-B).

In this section we will sketch descriptions of a simple form of neighborhood semantics with only one agent and no support for nested beliefs.

The semantic objects are (two-valued) epistemic states, defined below:

Definition 2. A (two-valued) epistemic state is an element of $\mathscr{P}(\mathscr{P}(\Phi \to \mathbb{C}))$, i.e. a set of sets of truth assignments from $\Phi \to \mathbb{C}$.

The intuition is that if $\mathfrak{M}$ is an agent's epistemic state, then for each $V \in \mathfrak{M}$, the agent thinks the world is described by one of the truth assignments in V. If the agent were logically omniscient, it would therefore think that the real world corresponded to one of the truth assignments in $\bigcap \mathfrak{M}$. However, the point of the semantics is that agents do not have to be logically omniscient, i.e. explicit belief can be modeled.

There are two established ways in which we might go about defining how the satisfaction relation treats explicit belief, namely, the strict and loose neighborhood semantics. Let us introduce two modal operators, $[=]$ and $[\subseteq]$, one for each type of explicit belief. The strict neighborhood semantics defines explicit belief by

$$\mathfrak{M} \vDash [=]\alpha \text{ if there exists } V \in \mathfrak{M} \text{ such that } V = [\![\alpha]\!]^{\mathbb{C}}$$

while the loose neighborhood semantics defines it by

$$\mathfrak{M} \vDash [\subseteq]\alpha \text{ if there exists } V \in \mathfrak{M} \text{ such that } V \subseteq [\![\alpha]\!]^{\mathbb{C}}$$

Note that $V \subseteq [\![\alpha]\!]^{\mathbb{C}}$ iff $v \vDash_{\mathbb{C}} \alpha$ for every $v \in V$.

The "strict" and "loose" terminology and the $[=]$ and $[\subseteq]$ notation are from (Areces and Figueira 2009). Both types of semantics have long been considered in AI research; (Vardi 1986) used strict neighborhood semantics, and loose neighborhood semantics were used by the "logic of local reasoning" from (Fagin and Halpern 1988, Section 6).

The intuitive way to understand the strict neighborhood semantics is to view an epistemic state as simply a set of every proposition that the agent explicitly believes. A problem with this semantics, noted by (Vardi 1986) and others, is that if an agent believes a sentence α, then the agent believes any β equivalent to α (so, for example, if the agent believes any one tautology, then the agent believes all tautologies).

From the point of view of the loose neighborhood semantics, an epistemic state is not the set of everything believed, because inferences can be made from each proposition in the epistemic state. Note that if $\alpha \vDash_{\mathbb{C}} \beta$ and $\mathfrak{M} \vDash [\subseteq]\alpha$, then $\mathfrak{M} \vDash [\subseteq]\beta$. This is closer to logical omniscience (some authors have defined logical omniscience as being exactly this), but the agent still cannot bring together information from separate propositions. For example, $\{[\![p]\!]^{\mathbb{C}}, [\![q]\!]^{\mathbb{C}}\} \nvDash [\subseteq](p \wedge q)$ and $\{[\![p]\!]^{\mathbb{C}}, [\![(p \supset q)]\!]^{\mathbb{C}}\} \nvDash [\subseteq]q$.

3 Three-valued neighborhood semantics

Both the strict and weak neighborhood semantics are in a sense too strong, as exemplified in the way they treat belief in tautologies. By basing neighborhood semantics on Kleene's three-valued logic, which has no tautologies, we can go some way towards improving matters.

We will relax the definition of an epistemic state to allow it to involve three-valued truth assignments.

Definition 3 (epistemic state)**.** An epistemic state is an element of $\mathscr{P}(\mathscr{P}(\Phi \to \mathbb{K}))$, i.e. a set of sets of truth assignments from $\Phi \to \mathbb{K}$.

We can view two-valued epistemic states as a special case, in which none of the functions involved has N in its image.

The point of three-valued epistemic states is not that the agent thinks that the world is really three-valued. Rather, a three-valued truth assignment provides a *partial* description of the world. Let us make a definition:

Definition 4 (compatibility with an epistemic state)**.** For $\mathfrak{M}$ an epistemic state and $u \in \Phi \to \mathbb{K}$, u is compatible with $\mathfrak{M}$ if for each $V \in \mathfrak{M}$ there is some $v \in V$ such that $v \diamond u$.

If an agent's epistemic state is $\mathfrak{M}$, then the agent (implicitly) thinks that the two-valued truth assignment that corresponds to the real world is compatible with $\mathfrak{M}$.

Definition 5 ($\Subset$)**.** Let $U \subseteq \Phi \to \mathbb{K}$ and $V \subseteq \Phi \to \mathbb{K}$. Then $V \Subset U$ if for every $v \in V$, there is some $u \in U$ such that $u \subseteq v$.

Note that if $U \subseteq \Phi \to \mathbb{C}$ and $V \subseteq \Phi \to \mathbb{C}$, then $V \Subset U$ iff $V \subseteq U$. Also note that, for any $\alpha \in \mathscr{L}(\Phi)$, $V \Subset [\![\alpha]\!]$ iff $v \vDash_{\mathbb{K}} \alpha$ for all $v \in V$. Therefore, we can define a new modal operator $[\Subset]$ which can be thought of as the three-valued analogue to $[\subseteq]$ as follows:

$$\mathfrak{M} \vDash [\Subset]\alpha \text{ if there exists } V \in \mathfrak{M} \text{ such that } V \Subset [\![\alpha]\!]$$

How does $[\Subset]$ compare with $[\subseteq]$ as an explicit belief operator? It is true that if $\alpha \vDash_{\mathbb{K}} \beta$ and $\mathfrak{M} \vDash [\Subset]\alpha$, then $\mathfrak{M} \vDash [\Subset]\beta$. However, this is often a much less onerous requirement for the agent to fulfill than the classical version of that. Consider that to decide whether $\{[\![(p \vee q)]\!]^{\mathbb{C}}\} \vDash [\subseteq]\alpha$ holds an agent may have to reflect not just on the truth values of p and q, but also on the atoms in α (since, for example, α might be a tautology). On the other hand, to determine if $\{[\![(p \vee q)]\!]\} \vDash [\Subset]\alpha$ all that has to be done is check whether both of the two truth assignments in $[\![(p \vee q)]\!]$ make α true. This is easy, especially since each truth assignment in $[\![(p \vee q)]\!]$ is undefined on every atom but one.

We could also create a three-valued version of the strict neighborhood semantics, but we will not look into that here. We would like epistemic states to be, instead of enumerations of everything believed (which would often be infinite), reasonably compact objects which could be physically realized in a relatively straightforward way.

Note that three-valued neighborhood semantics of the sort we have described are essentially the same as the semantics based on "belief cells" that (McArthur 1988, Section 4.2) recounted from an unpublished paper by Levesque.

3.1 On reasonable closure properties

We would like for an agent to have *some* ability to combine information from different elements of its epistemic state, without requiring unrealistic reasoning powers. For example, we might like for explicit belief in α and in β to make $(\alpha \wedge \beta)$ also explicitly believed. How shall we achieve this?

One obvious approach (followed in e.g. (Vardi 1986, Section 4)) is to impose a restriction on the set of semantic objects. Let us first make a definition:

Definition 6 ($\Cap$)**.** For $U, V \in \mathscr{P}(\Phi \to \mathbb{K})$, let $U \Cap V := \{u \cup v : u \in U, v \in V, \text{ and } u \diamond v\}$.

The intuition behind $\Cap$ is that it is the semantic version of the $\wedge$ operator. Note that $\llbracket\alpha\rrbracket \Cap \llbracket\beta\rrbracket = \llbracket\alpha \wedge \beta\rrbracket$. Also, if none of the functions in U or V assign the value N to any atom, then $U \Cap V = U \cap V$.

Now, a restriction on semantic objects could be to require an epistemic state $\mathfrak{M}$ to satisfy that if $U \in \mathfrak{M}$ and $V \in \mathfrak{M}$, then $U \Cap V \in \mathfrak{M}$. Unfortunately, this sort of approach makes epistemic states much too strong. To illustrate, suppose that $\mathfrak{M}$ is such that $\{\llbracket\alpha_i\rrbracket : 1 \leq i \leq n\} \subseteq \mathfrak{M}$ for some sentences $\alpha_1, \alpha_2, \ldots, \alpha_n$. Then, in order to fulfill the requirement it must be that $\llbracket\bigwedge_{1 \leq i \leq n} \alpha_i\rrbracket \in \mathfrak{M}$. That means that the agent explicitly believes *all* the logical consequences (in Kleene's logic) of $\{\alpha_i : 1 \leq i \leq n\}$.

In the next section, we will consider an alternative way to get closure under conjunction, a way that leaves the semantic objects alone and instead expands the satisfaction relation by providing additional conditions under which explicit belief exists. We will thereby avoid requiring so much power.

However, this does not mean that restricting the set of semantic objects may not sometimes be useful. We below introduce a more reasonable restriction on semantic objects, that of being harmonized.

Definition 7 (harmonization)**.** Let $\mathfrak{M}$ be an epistemic state. The harmonization of $\mathfrak{M}$, written $\mathcal{H}(\mathfrak{M})$, is the least superset of $\mathfrak{M}$ satisfying the following condition: if V and $\{u\}$ are elements, then so is $\{v \in V : v \diamond u\}$.

An epistemic state $\mathfrak{M}$ is said to be harmonized if $\mathfrak{M} = \mathcal{H}(\mathfrak{M})$. The motivation behind harmonization is to give a semantic generalization of the proof-theoretic notion of unit propagation. Harmonizing an epistemic state does not confer anything like logical omniscience; for example, if $\mathfrak{M} = \{\llbracket(p \vee q)\rrbracket, \llbracket((p \vee q) \supset r)\rrbracket\}$ then $\mathfrak{M}$ is already harmonized and yet $\mathfrak{M} \nvDash [\Subset]r$.

4 A logic with levels of belief

We are now almost ready to formally define a logic based on three-valued neighborhood semantics that has a version of levels of belief. First, though, let us make a definition.

Definition 8 (expansion)**.** Let $\mathfrak{M}$ be an epistemic state and $\alpha \in \mathscr{L}(\Phi)$. Then the expansion of $\mathfrak{M}$ by α, written $\mathfrak{M}[\alpha]$, is the epistemic state $\mathcal{H}(\mathfrak{M} \cup \{\llbracket\alpha\rrbracket\})$.

$\mathfrak{M}[\alpha]$ could be thought of as the epistemic state that results from the agent learning or being told α. $\mathfrak{M}[\alpha]$ might also be a state temporarily entered when the agent assumes α for the sake of argument. A major reason for our using harmonization is so that, if α "obviously" conflicts with the information in $\mathfrak{M}$, $\mathfrak{M}[\alpha]$ will include the empty set (and so make every level of belief contain every sentence). This allows for reasoning by contradiction.

Our logic will use the modal language $\mathscr{M}(\Phi)$, which is defined by the grammar below, in which $\alpha \in \mathscr{L}(\Phi)$, and k is any nonnegative integer.

$$\varphi ::= \mathsf{B}\alpha \mid \mathsf{B}_k\alpha \mid [\Subset]\alpha \mid [\alpha]\varphi \mid (\varphi \wedge \varphi) \mid \neg\varphi$$

Note that sentences of $\mathscr{L}(\Phi)$ cannot appear outside the scope of a modal operator. Also, $[\alpha]$ is the only sort of modal operator for which other modal operators can be in its scope.

The semantic objects of our logic are harmonized epistemic states.

We will next define a satisfaction relation inductively. To make the induction well-founded we will have to use a slightly more complex order on sentences than just length. In preparation for defining this order, we will inductively define two functions f, g mapping $\mathscr{M}(\Phi)$ to integers.

$$\begin{aligned} f(\mathsf{B}\alpha) = f([\Subset]\alpha) &= -1 \\ f(\mathsf{B}_k\alpha) &= k \\ f([\alpha]\varphi) = f(\neg\varphi) &= f(\varphi) \\ f((\varphi \wedge \psi)) &= \max(f(\varphi), f(\psi)) \end{aligned}$$

Note that $f(\varphi)$ is the value of the highest subscript in φ if there is one, and -1 otherwise. We next define g:

$$\begin{aligned} g(\mathsf{B}\alpha) = g(\mathsf{B}_k\alpha) = g([\Subset]\alpha) &= 1 + \mathsf{len}(\alpha) \\ g([\alpha]\varphi) &= 2 + \mathsf{len}(\alpha) + g(\varphi_1) \\ g((\varphi \wedge \psi)) &= 3 + g(\varphi) + g(\psi) \\ g(\neg\varphi) &= 1 + g(\varphi) \end{aligned}$$

If we consider the B_k and $[\Subset]$ operators to have length 1, then $g(\varphi)$ is the length of φ.

Recall the purpose of f and g is to define a partial order on sentences. Let us say that $\varphi \prec \psi$ if $\langle f(\varphi), g(\varphi)\rangle$ lexicographically precedes $\langle f(\psi), g(\psi)\rangle$, i.e. if $f(\varphi) < f(\psi)$ or if both $f(\varphi) = f(\psi)$ and $g(\varphi) < g(\psi)$.

Now we can say that the satisfaction relation, $\vDash$, is defined by induction on the order $\prec$ as follows:

1. $\mathfrak{M} \vDash \mathsf{B}\alpha$ iff, for each $w \in \Phi \to \mathbb{C}$ that is compatible with $\mathfrak{M}$, $w \vDash_{\mathbb{C}} \alpha$
2. $\mathfrak{M} \vDash [\Subset]\alpha$ iff there exists $V \in \mathfrak{M}$ such that $V \Subset \llbracket\alpha\rrbracket$
3. $\mathfrak{M} \vDash [\alpha]\varphi$ iff $\mathfrak{M}[\alpha] \vDash \varphi$
4. $\mathfrak{M} \vDash (\varphi \wedge \psi)$ iff $\mathfrak{M} \vDash \varphi$ and $\mathfrak{M} \vDash \psi$
5. $\mathfrak{M} \vDash \neg\varphi$ iff $\mathfrak{M} \nvDash \varphi$
6. $\mathfrak{M} \vDash \mathsf{B}_k\alpha$, where k is a nonnegative integer, iff at least one of the following is true:
   (a) $k = 0$ and $\mathfrak{M} \vDash [\Subset]\alpha$
   (b) $\alpha = (\alpha_1 \wedge \alpha_2)$, and $\mathfrak{M} \vDash \mathsf{B}_k\alpha_1$ and $\mathfrak{M} \vDash \mathsf{B}_k\alpha_2$
   (c) $\alpha = \neg(\alpha_1 \wedge \alpha_2)$, and $\mathfrak{M} \vDash \mathsf{B}_k\neg\alpha_1$ or $\mathfrak{M} \vDash \mathsf{B}_k\neg\alpha_2$
   (d) $\alpha = \neg\neg\alpha_1$ and $\mathfrak{M} \vDash \mathsf{B}_k\alpha_1$
   (e) $k > 0$ and there exists $p \in \Phi$ such that both $\mathfrak{M} \vDash [p]\mathsf{B}_{k-1}\alpha$ and $\mathfrak{M} \vDash [\neg p]\mathsf{B}_{k-1}\alpha$

Rule (1) defines B as an implicit belief operator, which is easily seen to be characterized by logical omniscience. Suppose $\Gamma \models_{\mathbb{C}} \alpha$. If $\mathfrak{M} \models \mathsf{B}\gamma$ for every $\gamma \in \Gamma$, then each $w \in \Phi \to \mathbb{C}$ compatible with $\mathfrak{M}$ makes every element of Γ true, and so must make α true as well. Hence $\mathfrak{M} \models \mathsf{B}\alpha$.

The definition of $[\Subset]$ that we have seen before is repeated by rule (2). Rule (3) defines an operator for expansion by α, which could be compared to a public announcement of α in dynamic epistemic logic (van Ditmarsch, van der Hoek, and Kooi 2007). Rules (4) and (5) make connectives outside the scope of modal operators behave in their traditional ways.

The various parts of rule (6) define the infinite family of operators $\{\mathsf{B}_k : k \geq 0\}$. An intuitive reading of $\mathsf{B}_k\alpha$ is that "upon being queried about the truth of α, confirmation takes at most k effort", though for brevity we suggest the conventional reading "α is in level k". Though we still have the operator $[\Subset]$, we will think of B_0 as indicating a form of explicit belief.

Rule (6a) ensures that level 0 contains every α for which $[\Subset]\alpha$ is true. Rule (6b) allows for forming conjunctions from conjuncts that are separately believed, and constitutes our alternative to the idea (mentioned in the last section) of accomplishing that by imposing constraints on the semantic objects. Note that this alternative means that whether a sentence is in a level depends not just on what proposition it expresses, but also on its syntactic form. For example, that $((\alpha \vee \neg\beta) \wedge \beta)$ was in a level would not necessarily mean that $(\alpha \wedge \beta)$ was also. The rules (6c) and (6d) allow other simple ways of syntactically building up beliefs.

Rule (6e) describes how the higher levels of belief are formed from the lower ones. The idea is that when k effort is allowed, reasoning by cases (i.e. considering what would be the case if p were true, and what would be the case if instead $\neg p$ were true) can be done, nested up to a depth of k. This is the same way higher levels of belief are defined in (Lakemeyer and Levesque 2014). The idea of using the depth of case-splitting allowed as a measure of effort can also be found in more proof-theory oriented papers like (Finger 2004; D'Agostino and Floridi 2009; D'Agostino, Finger, and Gabbay 2013).

4.1 Properties

We will now present various properties of our logic, mostly without proof. The first shows how implicit belief works:

Proposition 1. *Let $\Gamma \subseteq \mathscr{L}(\Phi)$ and suppose that $\mathfrak{M} = \mathcal{H}(\{[\![\gamma]\!] : \gamma \in \Gamma\})$. Then $\mathfrak{M} \models \mathsf{B}\alpha$ if and only if $\Gamma \models_{\mathbb{C}} \alpha$.*

Lemma 1. *If $\mathfrak{M} \models \mathsf{B}_k\alpha$ and $\mathfrak{M}'$ is a harmonized epistemic state such that $\mathfrak{M} \subseteq \mathfrak{M}'$, then $\mathfrak{M}' \models \mathsf{B}_k\alpha$.*

Lemma 2 (monotonicity of expansions)**.** *Let $\alpha, \beta \in \mathscr{L}(\Phi)$. If $\mathfrak{M} \models \mathsf{B}_k\alpha$, then $\mathfrak{M}[\beta] \models \mathsf{B}_k\alpha$ (and so $\mathfrak{M} \models [\beta]\mathsf{B}_k\alpha$).*

Proposition 2 (levels are cumulative)**.** $\models \mathsf{B}_k\alpha \supset \mathsf{B}_{k+1}\alpha$

Proof. Suppose that $\mathfrak{M} \models \mathsf{B}_k\alpha$. Pick any $p \in \Phi$. By Lemma 2, $\mathfrak{M} \models [p]\mathsf{B}_k\alpha$ and $\mathfrak{M} \models [\neg p]\mathsf{B}_k\alpha$. □

Proposition 3 (level soundness)**.** $\models \mathsf{B}_k\alpha \supset \mathsf{B}\alpha$.

Proposition 4 (eventual completeness)**.** *Suppose that $\mathfrak{M}$ is finite, and for each $V \in \mathfrak{M}$, V is finite and each $v \in V$ is finite. If $\mathfrak{M} \models \mathsf{B}\alpha$, then there is some k such that $\mathfrak{M} \models \mathsf{B}_k\alpha$.*

Proof sketch. Suppose that $\mathfrak{M} \models \mathsf{B}\alpha$. Let n be the number of atoms that are mentioned in $\mathfrak{M}$ or α. Let m be the number of atoms p such that either $[\![p]\!] \in \mathfrak{M}$ or $[\![\neg p]\!] \in \mathfrak{M}$. We will prove that $\mathfrak{M} \models \mathsf{B}_{n-m}\alpha$ by induction on $n - m$.

The base case, where $n - m = 0$, is straightforward. For the inductive step, suppose that $n - m > 0$. Let $p \in \Phi$ be such that neither $[\![p]\!] \in \mathfrak{M}$ nor $[\![\neg p]\!] \in \mathfrak{M}$. Since $\mathfrak{M} \models \mathsf{B}\alpha$, it is also the case that $\mathfrak{M}[p] \models \mathsf{B}\alpha$ and $\mathfrak{M}[\neg p] \models \mathsf{B}\alpha$. Therefore, by the inductive hypothesis, $\mathfrak{M}[p] \models \mathsf{B}_{n-m-1}\alpha$ and $\mathfrak{M}[\neg p] \models \mathsf{B}_{n-m-1}\alpha$. Hence $\mathfrak{M} \models \mathsf{B}_{n-m}\alpha$ by rule (6e). □

Note that the proof of (Lakemeyer and Levesque 2014, Theorem 3) is similar.

Proposition 5 (miscellaneous properties of levels)**.**

$$\models \mathsf{B}_k(\alpha \wedge \beta) \equiv (\mathsf{B}_k\alpha \wedge \mathsf{B}_k\beta) \quad (1)$$

$$\models \mathsf{B}_k\neg\neg\alpha \equiv \mathsf{B}_k\alpha \quad (2)$$

$$\models \mathsf{B}_k(\alpha \vee (\beta \vee \gamma)) \equiv \mathsf{B}_k((\alpha \vee \beta) \vee \gamma) \quad (3)$$

$$\models \mathsf{B}_k((\alpha \wedge \beta) \vee (\alpha \wedge \gamma)) \supset \mathsf{B}_k(\alpha \wedge (\beta \vee \gamma)) \quad (4)$$

$$\models \mathsf{B}_k(\alpha \vee (\beta \wedge \gamma)) \supset \mathsf{B}_k((\alpha \vee \beta) \wedge (\alpha \vee \gamma)) \quad (5)$$

The converses of (4) and (5) are not valid in general.

4.2 A reasoning service

Our logic can be used to specify a reasoning service in the following way: after being told the sequence of sentences $\alpha_1, \alpha_2, \ldots, \alpha_n$ (and nothing else), the service will, given an input sentence β and integer k, return "Yes" if $\emptyset[\alpha_1][\alpha_2]\cdots[\alpha_n] \models \mathsf{B}_k\beta$, and "No" otherwise. Equivalently (because of Lemma 1), we could phrase the question as determining whether $\models [\alpha_1][\alpha_2]\cdots[\alpha_n]\mathsf{B}_k\beta$.

The following complexity result can be shown:

Proposition 6. *For $\alpha_1, \ldots, \alpha_n$ in* disjunctive normal form (DNF)*, any sentence β, and k a fixed constant, whether $\models [\alpha_1][\alpha_2]\cdots[\alpha_n]\mathsf{B}_k\beta$ can be computed in polynomial time.*

A sentence is in DNF if it is a disjunction of conjunctions of literals. The requirement that $\alpha_1, \ldots, \alpha_n$ be in DNF may seem like a serious constraint, since (as is well-known) converting a sentence into DNF may take exponential time. However, for knowledge representation purposes, we may often be dealing with large collections of facts which are individually simple—i.e. the number n of sentences may grow to be very large, but each sentence typically remains small. In such a case it could be practical to convert each of the sentences into DNF.

5 Discussion

5.1 Comparison with $\mathcal{ESL}$

Our logic is to some extent patterned after the propositional fragment of the logic $\mathcal{ESL}$ from (Lakemeyer and Levesque 2014). So, how does it compare?

The analogues in $\mathcal{ESL}$ of our epistemic states are called "setups" and they are sets of clauses (a clause is a disjunction of literals). Setups are a much less expressive class of

objects than epistemic states. To make an epistemic state $\mathfrak{M}$ restricted in an analogous way, we would have to require that every $V \in \mathfrak{M}$ be finite (we typically would want that anyway) and, more seriously, that for each $v \in V$, exactly one atom is mapped to a non-N value by v. As a consequence, in $\mathcal{ESL}$ the following sentence is valid:

$$\mathsf{B}_0((p \wedge q) \vee (r \wedge s)) \supset (\mathsf{B}_0(p \wedge q) \vee \mathsf{B}_0(r \wedge s))$$

That seems undesirable. Of course, that sentence is not valid in our logic. The idea that at heart an agent's knowledge consists of a set of disjunctions of literals seems to be without philosophical or psychological motivation.

Furthermore, the satisfaction relation in $\mathcal{ESL}$ is defined in a more syntactic way, resulting in many sentences not being in a level even when intuitively they seem to follow—without the need to reason by cases—from other sentences in the level. For example, $\mathsf{B}_0(p \vee q) \supset \mathsf{B}_0(p \vee \neg\neg q)$ (a validity in our logic) is not valid in $\mathcal{ESL}$.

5.2 Limitations and possible extensions

Parikh (1987) defined a *knowledge algorithm* as consisting of a database and a procedure that, given an input question and a resource bound, works up to the bound, and then either answers the question or says "I don't know". Also, in a feature that might be interesting to extend our logic with, the database may be updated as a result of the query. This could be used to model Socratic questioning, where a series of well-chosen questions make the agent realize what it (implicitly) knew all along (see (Crawford and Kuipers 1989) for an existing approach to formalizing this).

Parikh also suggested that, in some cases, the agent may know that an implicit belief does not exist. McCarthy (1977) gave the example of being sure that you will not be able to, by reasoning alone, determine whether the president is currently standing. Our levels of belief can be thought of as approximations of implicit belief from below; it would be interesting to have approximations from above, that would identify sentences that were obviously neither believed nor disbelieved. See (Schaerf and Cadoli 1995) and (Finger and Wassermann 2007) for existing approaches at this.

One of the advantages some authors have found with neighborhood semantics is that agents can have conflicting beliefs without believing everything. Unfortunately, our eventual completeness result means that in our logic agents can ultimately derive anything from contradictory beliefs.

Our logic also does not feature any introspection, but we expect that an extension incorporating that would be interesting. In particular, incorporating a notion of effort into autoepistemic reasoning would be useful. Recall (Moore 1985)'s well-known example of autoepistemic reasoning:

> [I]f I did have an older brother I would know about it; therefore, since I don't know of any older brothers, I must not have any.

If we formalize "know" in the brother example as implicit knowledge, then determining whether "I don't know of any older brothers" may be very difficult. A better formalization might capture the following idea, which is probably what we normally really mean if we say that we would know if we had an older brother:

> If I had an older brother, it would be *obvious* to me that I did.

Some notion of effort would clearly be relevant to this. Moore's brother problem was formalized in (Elgot-Drapkin and Perlis 1990, Section 6.2) by equating knowing α with having already drawn the conclusion that α. This may suffice for the brother problem in particular, but is not very flexible.

For a more complicated example of agents reasoning about their own knowledge, consider the following problem:

> A classroom is full of students, about to write an exam. The instructor announces (truthfully) that the exam only requires material from up to chapter five in the textbook, and that she expects the exam to be easy.
>
> Formalize how the instructor's announcement might help the students.

Unlike the contrived puzzles often considered in epistemic logic (e.g. the "muddy children" problem discussed at length in (Fagin et al. 1995)), this problem actually describes something that could plausibly occur in everyday life. Furthermore, the machinery provided by standard epistemic logics (including dynamic epistemic logic) is not of much help in capturing the important aspects of this problem.

6 Related work

Duc (2001, Chapter 5) also presents a logic with numbered knowledge operators, where the numbers are meant to indicate bounds on how much time it would take to verify that the sentences in question are true. However, most of the work in defining what those bounds would be for particular sentences is not done by the logic, but by cost functions that a user of the logic would have to provide.

Two-valued strict neighborhood semantics are the basis for the "active logic" described in (Nirkhe, Kraus, and Perlis 1995, Section 4). In this logic, time is represented, and epistemic states expand over time, which is meant to model an agent reasoning. However, as was criticized by (Jago 2006, Section 4.4.2), all tautologies are believed from time 0 on.

Some other papers that we have not yet mentioned which involve something like a notion of effort are (Crawford and Kuipers 1991; Dalal 1996; Crawford and Etherington 1998).

7 Conclusion

Following (McCarthy and Hayes 1969)'s epistemological-heuristic division, some researchers have approached the problem of formalizing common sense without regard for the complexity of reasoning. In contrast, we have described a logic in which the effort involved in reasoning can, to some extent, be described. Further evaluation will be needed to compare against actual human performance. What is more important than our particular formalism, though, is the idea that what it is formalizing is something that needs to be formalized and is relevant to research into common sense.

Acknowledgements

This work was funded by the Natural Sciences and Engineering Research Council of Canada, and by an Ontario Graduate Scholarship.

References

Areces, C., and Figueira, D. 2009. Which Semantics for Neighbourhood Semantics? In *Proceedings of the 21st International Joint Conference on Artificial Intelligence*, 671–676.

Castañeda, H. 1964. Review: Jaakko Hintikka, Knowledge and Belief. An Introduction to the Logic of the Two Notions. *Journal of Symbolic Logic* 29(3):132–134.

Crawford, J. M., and Etherington, D. W. 1998. A Non-Deterministic Semantics for Tractable Inference. In *Proceedings of the Fifteenth National Conference on Artificial Intelligence*, 286–291.

Crawford, J. M., and Kuipers, B. 1989. Towards a Theory of Access-limited Logic for Knowledge Representation. In *Proceedings of the First International Conference on Principles of Knowledge Representation and Reasoning*, 67–78.

Crawford, J. M., and Kuipers, B. 1991. Negation and Proof by Contradiction in Access-Limited Logic. In *Proceedings of the 9th National Conference on Artificial Intelligence*, 897–903.

D'Agostino, M., and Floridi, L. 2009. The enduring scandal of deduction. *Synthese* 167(2):271–315.

D'Agostino, M.; Finger, M.; and Gabbay, D. 2013. Semantics and proof-theory of depth bounded Boolean logics. *Theoretical Computer Science* 480(0):43–68.

Dalal, M. 1996. Semantics of an Anytime Family of Reasoners. In *Proceedings of the 12th European Conference on Artificial Intelligence*, 360–364.

Duc, H. N. 2001. *Resource-Bounded Reasoning about Knowledge*. Ph.D. Dissertation, Faculty of Mathematics and Informatics, University of Leipzig.

Elgot-Drapkin, J. J., and Perlis, D. 1990. Reasoning situated in time I: basic concepts. *Journal of Experimental & Theoretical Artificial Intelligence* 2(1):75–98.

Fagin, R., and Halpern, J. Y. 1988. Belief, Awareness, and Limited Reasoning. *Artificial Intelligence* 34:39–76.

Fagin, R.; Halpern, J. Y.; Moses, Y.; and Vardi, M. Y. 1995. *Reasoning about Knowledge*. MIT Press.

Finger, M., and Wassermann, R. 2007. Anytime Approximations of Classical Logic from Above. *Journal of Logic and Computation* 17(1):53–82.

Finger, M. 2004. Polynomial Approximations of Full Propositional Logic via Limited Bivalence. In *Logics in Artificial Intelligence, 9th European Conference, JELIA 2004*, volume 3229 of *Lecture Notes in Computer Science*. Springer Berlin Heidelberg. 526–538.

Hintikka, J. 1962. *Knowledge and Belief. An Introduction to the Logic of the Two Notions*. Ithaca, New York: Cornell University Press.

Hocutt, M. O. 1972. Is epistemic logic possible? *Notre Dame Journal of Formal Logic* 13(4):433–453.

Jago, M. 2006. *Logics for Resource-Bounded Agents*. Ph.D. Dissertation, University of Nottingham.

Kleene, S. C. 1938. On Notation for Ordinal Numbers. *The Journal of Symbolic Logic* 3(4):150–155.

Kowalski, R. 1995. Using Meta-Logic to Reconcile Reactive With Rational Agents. In Apt, K. R., and Turini, F., eds., *Metalogics and Logic Programming*, 227–242. MIT Press. Updated version at http://www.doc.ic.ac.uk/%7Erak/recon-abst.pdf.

Lakemeyer, G., and Levesque, H. J. 2013. Decidable Reasoning in a Logic of Limited Belief with Introspection and Unknown Individuals. In *Proceedings of the 23rd International Joint Conference on Artificial Intelligence*, 969–975.

Lakemeyer, G., and Levesque, H. J. 2014. Decidable Reasoning in a Fragment of the Epistemic Situation Calculus. In *Principles of Knowledge Representation and Reasoning: Proceedings of the Fourteenth International Conference*.

Levesque, H. J. 1984. A Logic of Implicit and Explicit Belief. In *Proceedings of the Fourth National Conference on Artificial Intelligence*, 198–202.

Liu, Y.; Lakemeyer, G.; and Levesque, H. J. 2004. A Logic of Limited Belief for Reasoning with Disjunctive Information. In *Principles of Knowledge Representation and Reasoning: Proceedings of the Ninth International Conference*, 587–597.

Liu, Y. 2006. *Tractable Reasoning in Incomplete First-order Knowledge Bases*. Ph.D. Dissertation, University of Toronto.

McArthur, G. L. 1988. Reasoning about knowledge and belief: a survey. *Computational Intelligence* 4(3):223–243.

McCarthy, J., and Hayes, P. J. 1969. Some philosophical problems from the standpoint of artificial intelligence. In *Machine Intelligence 4*, 463–502. Edinburgh University Press.

McCarthy, J. 1977. Epistemological Problems of Artificial Intelligence. In *Proceedings of the 5th International Joint Conference on Artificial Intelligence*, 1038–1044.

Montague, R. 1968. Pragmatics. In Klibansky, R., ed., *Contemporary Philosophy*. Firenze: La Nuova Italia Editrice. 102–122.

Moore, R. C. 1985. Semantical Considerations on Nonmonotonic Logic. *Artificial Intelligence* 25(1):75–94.

Moreno, A. 1998. Avoiding logical omniscience and perfect reasoning: a survey. *AI Communications* 11(2):101.

Nirkhe, M.; Kraus, S.; and Perlis, D. 1995. Thinking takes time: A modal active-logic for reasoning in time. In *Proceedings of the Fourth Bar Ilan Symposium on Foundations of Artificial Intelligence*.

Parikh, R. 1987. Knowledge and the problem of logical omniscience. In *Methodologies for Intelligent Systems, Proceedings of the Second International Symposium*, 432–439.

Schaerf, M., and Cadoli, M. 1995. Tractable reasoning via approximation. *Artificial Intelligence* 74(2):249 – 310.

Scott, D. 1970. Advice on Modal Logic. In Lambert, K., ed., *Philosophical Problems in Logic*. Dordrecht, Holland: D. Reidel.

Sim, K. M. 1997. Epistemic Logic and Logical Omniscience: A Survey. *International Journal of Intelligent Systems* 12(1):57–81.

Stalnaker, R. 1991. The Problem of Logical Omniscience, I. *Synthese* 89(3):425–440.

van Ditmarsch, H.; van der Hoek, W.; and Kooi, B. 2007. *Dynamic Epistemic Logic*. Dordrecht, The Netherlands: Springer.

Vardi, M. Y. 1986. On Epistemic Logic and Logical Omniscience. In *Proceedings of the 1986 Conference on Theoretical Aspects of Reasoning About Knowledge*, 293–305.

Whitsey, M. 2003. Logical Omniscience: A Survey. Technical Report NOTTCS-WP-2003-2, School of Computer Science and Information Technology, University of Nottingham.