Multi-Agent Only-Knowing Revisited

Multi-Agent Only-Knowing Revisited

Vaishak Belle and **Gerhard Lakemeyer**
Dept. of Computer Science
RWTH Aachen
52056 Aachen
Germany
{belle,gerhard}@cs.rwth-aachen.de

Abstract

Levesque introduced the notion of only-knowing to precisely capture the beliefs of a knowledge base. He also showed how only-knowing can be used to formalize non-monotonic behavior within a monotonic logic. Despite its appeal, all attempts to extend only-knowing to the many agent case have undesirable properties. A belief model by Halpern and Lakemeyer, for instance, appeals to proof-theoretic constructs in the semantics and needs to axiomatize validity as part of the logic. It is also not clear how to generalize their ideas to a first-order case. In this paper, we propose a new account of multi-agent only-knowing which, for the first time, has a natural possible-world semantics for a quantified language with equality. We then provide, for the propositional fragment, a sound and complete axiomatization that faithfully lifts Levesque's proof theory to the many agent case. We also discuss comparisons to the earlier approach by Halpern and Lakemeyer.

Introduction

Levesque's notion of only-knowing is a single agent monotonic logic that was proposed with the intention of capturing certain types of nonmonotonic reasoning. Levesque (1990) already showed that there is a close connection to Moore's (1985) autoepistemic logic (AEL). Recently, Lakemeyer and Levesque (2005) showed that only-knowing can be adapted to capture default logic as well. The main benefit of using Levesque's logic is that, via simple semantic arguments, nonmonotonic conclusions can be reached without the use of meta-logical notions such as fixpoints (Rosati 2000; Levesque and Lakemeyer 2001) . Only-knowing is then naturally of interest in a many agent context, since agents capable of non-trivial nonmonotonic behavior should believe other agents to also be equipped with nonmonotonic mechanisms. For instance, if all that Bob knows is that Tweety is a bird and a default that birds typically fly, then Alice, if she knows all that Bob knows, concludes that Bob believes Tweety can fly.[1] Also, the idea of only-knowing a collection of sentences is useful for modeling the beliefs of a knowledge base (KB), since sentences that are not logically entailed by the KB are taken to be precisely those not believed. If many agents are involved, and suppose Alice has some beliefs on Bob's KB, then she could capitalize on Bob's knowledge to collaborate on tasks, or plan a strategy against him.

As a logic, Levesque's construction is unique in the sense that in addition to a classical epistemic operator for belief, he introduces a modality to denote what is *at most* known. This new modality has a subtle relationship to the belief operator that makes extensions to a many agent case non-trivial. Most extensions so far make use of arbitrary Kripke structures, that already unwittingly discard the simplicity of Levesque's semantics. They also have some undesirable properties, perhaps invoking some caution in their usage. For instance, in a canonical model (Lakemeyer 1993), certain types of epistemic states cannot be constructed. In another Kripke approach (Halpern 1993), the modalities do not seem to interact in an intuitive manner. Although an approach by Halpern and Lakemeyer (2001) does indeed successfully model multi-agent only-knowing, it forces us to have the semantic notion of validity directly in the language and has proof-theoretic constructs in the semantics via maximally consistent sets. Precisely for this reason, that proposal is not natural, and it is matched with a proof theory that has a set of new axioms to deal with these new notions. It is also not clear how one can extend their semantics to the first-order case. Lastly, an approach by Waaler (2004) avoids such an axiomatization of validity, but the model theory also has problems (Waaler and Solhaug 2005). Technical discussions on their semantics are deferred to later.

The goal of this paper is to show that there is indeed a natural semantics for multi-agent only-knowing for the quantified language with equality. For the propositional subset, there is also a sound and complete axiomatization that faithfully generalizes Levesque's proof theory.[2] We also differ from Halpern and Lakemeyer in that we do not enrich the language any more than necessary (modal operators for each agent), and we do not make use of canonical Kripke models. And while canonical models, in general, are only workable

[1] We use the terms "knowledge" and "belief" interchangeably in the paper.

[2] The proof theory for a quantified language is well known to be *incomplete* for the single agent case. It is also known that any complete axiomatization cannot be *recursive* (Halpern and Lakemeyer 1995; Levesque and Lakemeyer 2001).

semantically and can not be used in practice, our proposal has a computational appeal to it. We also show that if we do enrich the language with a modal operator for *validity*, but only to establish a common language with (Halpern and Lakemeyer 2001), then we agree on the set of valid sentences. Finally, we obtain a first-order multi-agent generalization of AEL, defined solely using notions of classical logical entailment and theoremhood.

The rest of the paper is organized as follows. We review Levesque's notions,[3] and define a semantics with so-called k-*structures*. We then compare the framework to earlier attempts. Following that, we introduce a sound and complete axiomatization for the propositional fragment. In the last sections, we sketch the multi-agent (first-order) generalization of AEL, and prove that k-structures and (Halpern and Lakemeyer 2001) agree on valid sentences, for an enriched language. Then, we conclude and end.

The k-structures Approach

The non-modal part of Levesque's logic[4] $\mathcal{ONL}$ consists of standard first-order logic with $=$ and a countably infinite set of standard names $\mathcal{N}$.[5] To keep matters simple, function symbols are not considered in this language. We call a predicate other than $=$, applied to first-order variables or standard names, an *atomic* formula. We write α^x_n to mean that the variable x is substituted in α by a standard name. If all the variables in a formula α are substituted by standard names, then we call it a *ground* formula. Here, a world is simply a set of ground atoms, and the semantics is defined over the set of all possible worlds $\mathcal{W}$. The standard names are thus *rigid designators*, and denote precisely the same entities in all worlds. $\mathcal{ONL}$ also has two modal operators: $\boldsymbol{L}$ and $\boldsymbol{N}$. While $\boldsymbol{L}\alpha$ is to be read as "at least α is known", $\boldsymbol{N}\alpha$ is to be read as "at most $\neg\alpha$ is known". A set of possible worlds is referred to as the agent's *epistemic state* e. Defining a model to be the pair (e, w) for $w \in \mathcal{W}$, components of $\mathcal{ONL}$'s meaning of truth are:

1. $e, w \models p$ iff $p \in w$ and p is a ground atom,
2. $e, w \models (m = n)$ iff m and n are identical standard names,
3. $e, w \models \neg\alpha$ iff $e, w \not\models \alpha$,
4. $e, w \models \alpha \vee \beta$ iff $e, w \models \alpha$ or $e, w \models \beta$,
5. $e, w \models \forall x.\, \alpha$ iff $e, w \models \alpha^x_n$ for all standard names n,
6. $e, w \models \boldsymbol{L}\alpha$ iff for all $w' \in e$, $e, w' \models \alpha$, and
7. $e, w \models \boldsymbol{N}\alpha$ iff for all $w' \notin e$, $e, w' \models \alpha$.

The main idea is that α is (at least) believed iff it is true at all worlds considered possible, while (at most) α is believed to be false iff it is true at all worlds considered *impossible*. So, an agent is said to only-know α, syntactically expressed as $\boldsymbol{L}\alpha \wedge \boldsymbol{N}\neg\alpha$, when worlds in e are precisely those where α is true. Halpern and Lakemeyer (2001) underline three features of the semantical framework of $\mathcal{ONL}$, the intuitions of which we desire to maintain in the many agent setting:

1. Evaluating $\boldsymbol{N}\alpha$ does *not affect* the epistemic possibilities. Formally, in $\mathcal{ONL}$, after evaluating formulas of the form $\boldsymbol{N}\alpha$ the agent's epistemic state is still given by e.
2. A union of the agent's possibilities, that evaluate $\boldsymbol{L}$, and the impossible worlds that evaluate $\boldsymbol{N}$, is *fixed* and *independent* of e, and is the set of all *conceivable* states. Formally, in $\mathcal{ONL}$, $\boldsymbol{L}\alpha$ is evaluated *wrt.* worlds $w \in e$, and $\boldsymbol{N}\alpha$ is evaluated *wrt.* worlds $w \in \mathcal{W} - e$; the union of which is $\mathcal{W}$. The intuition is that the exact complement of an agent's possibilities is used in evaluating $\boldsymbol{N}$.
3. Given any set of possibilities, there is always a model where *precisely* this set is the epistemic state. Formally, in $\mathcal{ONL}$, any subset of $\mathcal{W}$ can be defined as the epistemic state.

Although these notions seem clear enough in the single agent case, generalizing them to the many agent case is non-trivial (Halpern and Lakemeyer 2001). We shall return to analyze the features shortly. Let us begin by extending the language. Let $\mathcal{ONL}_n$ be a first-order modal language that enriches the non-modal subset of $\mathcal{ONL}$ with modal operators $\boldsymbol{L}_i$ and $\boldsymbol{N}_i$ for $i = a, b$. For ease of exposition, we only have two agents a (Alice) and b (Bob). Extensions to more agents is straightforward. We freely use $\boldsymbol{O}_i$, such that $\boldsymbol{O}_i\alpha$ is an abbreviation for $\boldsymbol{L}_i\alpha \wedge \boldsymbol{N}_i\neg\alpha$, and is read as "all that i knows is α". Objective and subjective formulas are understood as follows.

Definition 1. *The i-depth of a formula α, denoted $|\alpha|_i$, is defined inductively as ($\Box_i$ denotes $\boldsymbol{L}_i$ or $\boldsymbol{N}_i$):*

1. $|\alpha|_i = 1$ *for atoms,*
2. $|\neg\alpha|_i = |\alpha|_i$,
3. $|\forall x.\, \alpha|_i = |\alpha|_i$,
4. $|\alpha \vee \beta|_i = max(|\alpha|_i, |\beta|_i)$,
5. $|\Box_i\alpha|_i = |\alpha|_i$,
6. $|\Box_j\alpha|_i = |\alpha|_j + 1$, *for* $j \neq i$

A formula has a depth k if max(a-depth,b-depth) = k. A formula is called i-objective if all epistemic operators which do not occur within the scope of another epistemic operator are of the form $\Box_j$ for $i \neq j$. A formula is called i-subjective if every atom is in the scope of an epistemic operator and all epistemic operators which do not occur within the scope of another epistemic operator are of the form $\Box_i$.

For example, a formula of the form $\boldsymbol{L}_a\boldsymbol{L}_b\boldsymbol{L}_a p \vee \boldsymbol{L}_b q$ has a depth of 4, a a-depth of 3 and a b-depth of 4. $\boldsymbol{L}_b q$ is both b-subjective and a-objective. A formula is called *objective* if it does not mention any modal operators. A formula is called *basic* if it does not mention any $\boldsymbol{N}_i$ for $i = a, b$. We now define a notion of epistemic states using k-*structures*. The main intuition is that we keep separate the worlds Alice believes from the worlds she considers Bob to believe, to depth k.

[3] There are other notions of "all I know", which will not be discussed here (Halpern and Moses 1985; Ben-David and Gafni 1989). Also see (Rosati 2000).

[4] We name the logic following (Halpern and Lakemeyer 2001) for ease of comparisons later on. It is referred to as $\mathcal{OL}$ in (Halpern and Lakemeyer 1995; Levesque and Lakemeyer 2001).

[5] More precisely, we have logical connectives $\vee$, $\forall$ and $\neg$. Other connectives are taken for their usual syntactic abbreviations.

Definition 2. *A k-structure ($k \geq 1$), say e^k, for an agent is defined inductively as:*

– $e^1 \subseteq \mathcal{W} \times \{\{\}\}$,
– $e^k \subseteq \mathcal{W} \times \mathbb{E}^{k-1}$, *where $\mathbb{E}^m$ is the set of all m-structures.*

A e^1 for Alice, denoted as e^1_a, is intended to represent a set of worlds $\{\langle w, \{\}\rangle, \ldots\}$. A e^2 is of the form $\{\langle w, e^1_b\rangle, \langle w', {e'}^1_b\rangle, \ldots\}$, and it is to be read as "at w, she believes Bob considers worlds from e^1_b possible but at w', she believes Bob to consider worlds from ${e'}^1_b$ possible". This conveys the idea that Alice has only partial information about Bob, and so at different worlds, her beliefs about what Bob knows differ. We define a e^k for Alice, a e^j for Bob and a world $w \in \mathcal{W}$ as a (k, j)-model (e^k_a, e^j_b, w). Only sentences of a maximal a-depth of k, and a maximal b-depth of j are interpreted *wrt.* a (k, j)-model. The complete semantic definition is:

1. $e^k_a, e^j_b, w \models p$ iff $p \in w$ and p is a ground atom,
2. $e^k_a, e^j_b, w \models (m = n)$ iff $m, n \in \mathcal{N}$ and are identical,
3. $e^k_a, e^j_b, w \models \neg\alpha$ iff $e^k_a, e^j_b, w \not\models \alpha$,
4. $e^k_a, e^j_b, w \models \alpha \vee \beta$ iff $e^k_a, e^j_b, w \models \alpha$ or $e^k_a, e^j_b, w \models \beta$,
5. $e^k_a, e^j_b, w \models \forall x.\, \alpha$ iff $e^k_a, e^j_b, w \models \alpha^x_n$ for all $n \in \mathcal{N}$,
6. $e^k_a, e^j_b, w \models \boldsymbol{L}_a\alpha$ iff for all $\langle w', e^{k-1}_b\rangle \in e^k_a$, $e^k_a, e^{k-1}_b, w' \models \alpha$,
7. $e^k_a, e^j_b, w \models \boldsymbol{N}_a\alpha$ iff for all $\langle w', e^{k-1}_b\rangle \notin e^k_a$, $e^k_a, e^{k-1}_b, w' \models \alpha$

And since $\boldsymbol{O}_a\alpha$ syntactically denotes $\boldsymbol{L}_a\alpha \wedge \boldsymbol{N}_a\neg\alpha$, it follows from the semantics that

8. $e^k_a, e^j_b, w \models \boldsymbol{O}_a\alpha$ iff for all worlds w', for all e^{k-1} for Bob, $\langle w', e^{k-1}_b\rangle \in e^k_a$ iff $e^k_a, e^{k-1}_b, w' \models \alpha$

(The semantics for $\boldsymbol{L}_b\alpha$ and $\boldsymbol{N}_b\alpha$ are given analogously.) A formula α (of a-depth of k and of b-depth of j) is *satisfiable* iff there is a (k, j)-model such that $e^k_a, e^j_b, w \models \alpha$. The formula is *valid* ($\models \alpha$) iff α is true at all (k, j)-models. Satisfiability is extended to a set of formulas Σ (of maximal a, b-depth of k, j) in the manner that there is a (k, j)-model e^k_a, e^j_b, w such that $e^k_a, e^j_b, w \models \alpha'$ for every $\alpha' \in \Sigma$. We write $\Sigma \models \alpha$ to mean that for every (k, j)-model e^k_a, e^j_b, w, if $e^k_a, e^j_b, w \models \alpha'$ for all $\alpha' \in \Sigma$, then $e^k_a, e^j_b, w \models \alpha$.

Validity is not affected if models of a depth greater than that needed are used. This is to say, if α is true *wrt.* all (k, j)-models, then α is true *wrt.* all (k', j')-models for $k' \geq k, j' \geq j$. We obtain this result by constructing for every $e^{k'}_a$, a k-structure $e_a{\downarrow}^{k'}_k$, such that they agree on all formulas of maximal a-depth k. Analogously for $e^{j'}_b$.

Definition 3. *Given $e^{k'}_a$, we define $e_a{\downarrow}^{k'}_k$ for $k' \geq k \geq 1$:*

1. $e_a{\downarrow}^1_1 = e^1_a$,
2. $e_a{\downarrow}^{k'}_1 = \{\langle w, \{\}\rangle \mid \langle w, e^{k'-1}_b\rangle \in e^{k'}_a\}$,
3. $e_a{\downarrow}^{k'}_k = \{\langle w, e_b{\downarrow}^{k'-1}_{k-1}\rangle \mid \langle w, e^{k'-1}_b\rangle \in e^{k'}_a\}$.

Lemma 4. *For all formulas α of maximal a, b-depth of k, j, $e^{k'}_a, e^{j'}_b, w \models \alpha$ iff $e_a{\downarrow}^{k'}_k, e_b{\downarrow}^{j'}_j, w \models \alpha$, for $k' \geq k, j' \geq j$.*

Proof. By induction on the depth of formulas. The proof immediately holds for atomic formulas, disjunctions and negations since we have the same world w. Assume that the result holds for formulas of a, b-depth 1. Let α such a formula, and suppose $e^{k'}_a, e^{j'}_b, w \models \boldsymbol{L}_a\alpha$ (where $\boldsymbol{L}_a\alpha$ has a, b-depth of $1, 2$). Then, for all $\langle w', e^{k'-1}_b\rangle \in e^{k'}_a$, $e^{k'}_a, e^{k'-1}_b, w' \models \alpha$ iff (by induction hypothesis) $e_a{\downarrow}^{k'}_1, e_b{\downarrow}^{k'-1}_1, w' \models \alpha$ iff $e_a{\downarrow}^{k'}_2, \{\}, w \models \boldsymbol{L}_a\alpha$. By construction, we also have $e_a{\downarrow}^{k'}_1, \{\}, w \models \boldsymbol{L}_a\alpha$. Lastly, since $\boldsymbol{L}_a\alpha$ is a-subjective, b's structure is irrelevant, and thus, $e_a{\downarrow}^{k'}_1, e_b{\downarrow}^{j'}_2, w \models \boldsymbol{L}_a\alpha$.

For the reverse direction, suppose $e_a{\downarrow}^{k'}_1, e_b{\downarrow}^{j'}_2, w \models \boldsymbol{L}_a\alpha$. Then for all $w' \in e_a{\downarrow}^{k'}_1$, $e_a{\downarrow}^{k'}_1, \{\}, w' \models \alpha$ iff (by construction) for all $\langle w', e^{k'-1}_b\rangle \in e^{k'}_a$, $e^{k'}_a, e^{k'-1}_b, w' \models \alpha$ iff $e^{k'}_a, \{\}, w \models \boldsymbol{L}_a\alpha$. Since b's structure is irrelevant, we have $e^{k'}_a, e^{j'}_b, w \models \boldsymbol{L}_a\alpha$. The cases for $\boldsymbol{L}_b\alpha$, $\boldsymbol{N}_a\alpha$ and $\boldsymbol{N}_b\alpha$ are completely symmetric. ∎

Theorem 5. *For all formulas α of a, b-depth of k, j, if α is true at all (k, j)-models, then α is true at all (k', j')-models with $k' \geq k$ and $j' \geq j$.*

Proof. Suppose α is true at all (k, j)-models. Given any (k', j')-model, by assumption $e_a{\downarrow}^{k'}_k, e_b{\downarrow}^{j'}_j, w \models \alpha$ and by Lemma 4, $e^{k'}_a, e^{j'}_b, w \models \alpha$. ∎

Knowledge with k-structures satisfy *weak* S5 properties, and the Barcan formula (Hughes and Cresswell 1972).

Lemma 6. *If α is a formula, the following are valid* wrt. *models of appropriate depth ($\Box_i$ denotes $\boldsymbol{L}_i$ or $\boldsymbol{N}_i$):*

1. $\Box_i\alpha \wedge \Box_i(\alpha \supset \beta) \supset \Box_i\beta$,
2. $\Box_i\alpha \supset \Box_i\Box_i\alpha$,
3. $\neg\Box_i\alpha \supset \Box_i\neg\Box_i\alpha$,
4. $\forall \boldsymbol{x}.\, \Box_i\alpha \supset \Box_i(\forall \boldsymbol{x}.\, \alpha)$.

Proof. The proofs are similar. For item 3, wlog let $\Box_i$ be $\boldsymbol{L}_a$. Suppose $e^k_a, e^j_b, w \models \neg\boldsymbol{L}_a\alpha$. There is some $\langle w', e^{k-1}_b\rangle \in e^k_a$ such that $e^k_a, e^{k-1}_b, w' \models \neg\alpha$. Let w'' be any world such that $\langle w'', {e'}^{k-1}_b\rangle \in e^k_a$. Then, $e^k_a, {e'}^{k-1}_b, w'' \models \neg\boldsymbol{L}_a\alpha$. Thus, $e^k_a, e^j_b, w \models \boldsymbol{L}_a\neg\boldsymbol{L}_a\alpha$. The case of $\boldsymbol{N}_a$ is analogous. ∎

Before moving on, let us briefly reflect on the fact that k-structures have finite depth. So suppose a only-knows KB, of depth k. Using k-structures allows us to reason about what is believed, up to depth k. Also, if we construct epistemic states from k'-structures where $k' \geq k$, then the logic correctly captures non-beliefs beyond the depth k. To illustrate, let *true* (depth 1) be all that a knows. Then, it can easily be shown that both the sentences $\boldsymbol{O}_a(\mathit{true}) \supset \neg\boldsymbol{L}_a\neg\boldsymbol{L}_b\alpha$ and $\boldsymbol{O}_a(\mathit{true}) \supset \neg\boldsymbol{L}_a\boldsymbol{L}_b\alpha$ are valid sentences in the logic, by considering any e^2 (and higher) for a. For most purposes, this restriction of having a parameter k seems harmless in the sense that agents usually have a finite knowledge base with sentences of some maximal depth k and they should not be able to conclude anything about what is known at depths higher than k, with one exception. If we were to include a notion of common knowledge (Fagin et al. 1995), then we would get entailments about what is believed at arbitrary depths. With our current model, this cannot be captured, but we are willing to pay that price because in return we get, for the first time, a very simple possible-world style account of only-knowing. Similarly, we have nothing to say about (infinite) knowledge bases with unbounded depth.

Multi-Agent Only-Knowing

In this section, we return to the features of only-knowing discussed earlier and verify that the new semantics reasonably extends them to the multi-agent case. We also briefly discuss earlier attempts at capturing these features. Halpern (1993), Lakemeyer (1993), and Halpern and Lakemeyer (2001) independently attempted to extend $\mathcal{ONL}$ to the many agent case.[6] There are some subtle differences in their approaches, but the main restriction is they only allow a propositional language. Henceforth, to make the comparison feasible, we shall also speak of the propositional subset of $\mathcal{ONL}_n$ with the understanding that the semantical framework is now defined for propositions (from an infinite set Φ) rather than ground atoms.

The main component in these features is the notion of *possibility*. In the single agent case, each world represents a possibility. Thus, from a logical viewpoint, a possibility is simply the set of objective formulas true at some world. Further, the set of epistemic possibilities is given by $\{\{\text{objective formulas true at } w\} \mid w \in e\}$. Halpern and Lakemeyer (2001) correctly argue that the appropriate generalization of the notion of possibility in the many agent case are i-objective formulas. Intuitively, a possible state of affairs according to a include the state of the world (objective formulas), as well as what b is taken to believe. The earlier attempts by Halpern and Lakemeyer use Kripke structures with accessibility relations $\mathcal{K}_i$ for each agent i. Given a Kripke structure M, the notion of possibility is defined as the set of i-objective formulas true at some Kripke world, and the set of epistemic possibilities is obtained from the i-objective formulas true at all i-accessible worlds. Formally, the set of epistemic possibilities true at (M, w), where w is a world in M, is defined as $\{obj_i^+(M, w') \mid w' \in \mathcal{K}_i(w)\}$, where $obj_i^+(M, w')$ is a set consisting of i-objective formulas true at (M, w').[7] Although intuitive, note that, even for the propositional subset of $\mathcal{ONL}$, a Kripke world is a completely different entity from what Levesque supposes. Perhaps, one consequence is that the semantic proofs in earlier approaches are very involved. In contrast, we define worlds exactly as Levesque supposes. And, our notion of possibility is obtained from the set of a-objective formulas true at each $\langle w, e_b^{k-1}\rangle$ in e_a^k.

[6]For space reasons, we do not review all aspects of these approaches.

[7]The superscript $+$ denotes that the set includes non-basic formulas. Given X^+, we let $X = \{\phi \text{ is basic } \mid \phi \in X^+\}$.

Definition 7. *Suppose $M = (e_a^k, e_b^j, w)$ is a (k, j)-model.*

1. *let $obj_i^+(M) = \{i\text{-objective } \phi \mid M \models \phi\}$,*
2. *let $Obj_a^+(e_a^k) = \{obj_a^+(\{\}, e_b^{k-1}, w) \mid \langle w, e_b^{k-1}\rangle \in e_a^k\}$,*
3. *let $Obj_b^+(e_b^j) = \{obj_b^+(e_a^{j-1}, \{\}, w) \mid \langle w, e_a^{j-1}\rangle \in e_b^j\}$.*

All the a-objective formulas true at a model M, essentially the objective formulas true *wrt.* w and the b-subjective formulas true w*rt.* e_b^j, are given by $obj_a^+(M)$. Note that these formulas do not strictly correspond to a's possibilities. Rather, we define Obj_a^+ on her epistemic state e_a^k, and this gives us all the a-objectives formulas that a considers possible. We shall now argue that the intuition of all of Levesque's properties is maintained.[8]

Property 1. In the single agent case, this property ensured that an agent's epistemic possibilities are not affected on evaluating $\boldsymbol{N}$. This is immediately the case here. Given a model, say (e_a^k, e_b^j, w), a's epistemic possibilities are determined by $Obj_a^+(e_a^k)$. To evaluate $\boldsymbol{N}_a\alpha$, we consider all models (e_a^k, e_b^{k-1}, w') such that $\langle w', e_b^{k-1}\rangle \notin e_a^k$. Again, a's possibilities are given by $Obj_a^+(e_a^k)$ for all these models, and does not change.

Property 2. In the single agent case, this property ensured that evaluating $\boldsymbol{L}\alpha$ and $\boldsymbol{N}\alpha$ is always *wrt.* the set of all possibilities, and completely independent of e. As discussed, in the many agent case, possibilities mean i-objective formulas and analogously, if α is a possibility in a's view, say an a-objective formula of maximal b-depth of k, then we should interpret $\boldsymbol{L}_a\alpha$ and $\boldsymbol{N}_a\alpha$ *wrt.* all a-objective possibilities of max. depth k: the set of $(k+1)$-structures. Clearly then, the result is fixed and independent of the corresponding e^{k+1}. The following lemma is a direct consequence of the definition of the semantics.

Lemma 8. *Let α be a i-objective formula of j-depth k, for $j \neq i$. Then, the set of $k+1$-structures that evaluate $\boldsymbol{L}_i\alpha$ and $\boldsymbol{N}_i\alpha$ is $\mathbb{E}^{k+1}$.*

Property 3. The third property ensures that one can characterize epistemic states from any set of i-objective formulas. Intuitively, given such a set, we must have a model where *precisely* this set is the epistemic state. Earlier attempts at clarifying this property involved constructing a *set* of maximally $\mathsf{K45}_n$-consistent sets of basic i-objective formulas, and showing that there exist an epistemic state that precisely corresponds to this set. But, defining possibilities via $\mathsf{K45}_n$ proof-theoretic machinery inevitably leads to some limitations, as we shall see. We instead proceed semantically, and go beyond basic formulas. Let Ω be a satisfiable set of i-objective formulas, say of maximal j-depth k, for $j \neq i$. Let Ω' be a set obtained by adding a i-objective formula γ of maximal j-depth k such that Ω' is also satisfiable. By considering all i-objective formulas of maximal j-depth k, let

[8]It is interesting to note that such a formulation of Levesque's properties is not straightforward in the first-order case. That is, for the quantified language, it is known that there are epistemic states that can not be characterized using only objective formulas (Levesque and Lakemeyer 2001). Thus, it is left open how one must correctly generalize the features of first-order $\mathcal{ONL}$.

us construct Ω', Ω'', ... by adding formulas iff the resultant set remains satisfiable. When we are done, the resulting Ω^* is what we shall call a maximally satisfiable i-objective set.[9] Naturally, there may be many such sets corresponding to Ω. We show that given a *set* of maximally satisfiable i-objective sets, there is a model where precisely this set characterizes the epistemic state.

Theorem 9. *Let S_i be a set of maximally satisfiable sets of i-objective formulas, and σ a satisfiable objective formula. Suppose S_a is of max. b-depth $k-1$ and S_b is of max. a-depth $j-1$. Then there is a model $M^* = \langle {e^*}^k_a, {e^*}^j_b, w^* \rangle$ such that $M^* \models \sigma$, $S_a = Obj^+_a({e^*}^k_a)$ and $S_b = Obj^+_b({e^*}^j_b)$.*

Proof. Consider S_a. Each $S' \in S_a$ is a maximally satisfiable a-objective set, and thus by definition, there is a k-structure $\langle w', e^{k-1}_b \rangle$ such that $\{\}, e^{k-1}_b, w' \models S'$. Define such a set of k-structures $\{\langle w', e^{k-1}_b \rangle\}$, corresponding to each $S' \in S_a$, and let this be ${e^*}^k_a$. It is immediate to verify that $Obj^+_a({e^*}^k_a) = S_a$. Analogously, for ${e^*}^j_b$ using S_b. Finally, there is clearly some world w^* where σ holds. ■

On Validity

How does the semantics compare to earlier approaches? In particular, we are interested in valid formulas. Lakemeyer (1993) proposes a semantics using $\mathsf{K45}_n$-canonical models, but he shows that the formula $\neg \boldsymbol{O}_a \neg \boldsymbol{O}_b p$ for any proposition p is valid. Intuitively, it says that all that Alice knows is that Bob does not only know p, and as Lakemeyer argues, the validity of $\neg \boldsymbol{O}_a \neg \boldsymbol{O}_b p$ is unintuitive. After all, Bob could *honestly* tell Alice that he does not only know p. The negation of this formula, on the other hand, is satisfiable in a Kripke structure approach by Halpern (1993), called the i-set approach.[10] It is also satisfiable in the k-structure semantics. Interestingly, the i-set approach and k-structures agree on one more notion. The formula $\boldsymbol{L}_a \bot \supset \neg \boldsymbol{N}_a \neg \boldsymbol{O}_b \neg \boldsymbol{O}_a p$ (ζ) is valid in both, while $\neg\zeta$ is satisfiable *wrt.* Lakemeyer (1993). (It turns out that the validity of ζ in our semantical framework is implicitly related to the satisfiability of $\boldsymbol{O}_b \neg \boldsymbol{O}_a p$, so this property is not unreasonable.)

However, we immediately remark that the i-set approach and k-structures do not share too many similarities beyond those presented above. In fact, the i-set approach does not truly satisfy Levesque's second property. For instance, $\boldsymbol{N}_a \neg \boldsymbol{O}_b p \wedge \boldsymbol{L}_a \neg \boldsymbol{O}_b p$ (λ) is satisfiable in Halpern (1993). Recall that, in this property, the union of models that evaluate $\boldsymbol{N}_i \alpha$ and $\boldsymbol{L}_i \alpha$ must lead to all conceivable states. So, the satisfiability of λ leaves open the question as to why $\boldsymbol{O}_b p$ is not considered since $\neg \boldsymbol{O}_b p$ is true at all conceivable states. We show that, in contrast, λ is not satisfiable in the k-structures approach. Lastly, (Halpern and Lakemeyer 2001) involves enriching the language, the intuitions of which are perhaps best explained after reviewing the proof theory, and so we defer discussions to later.[11]

Theorem 10. *The following are properties of the semantics:*

1. *$\boldsymbol{O}_a \neg \boldsymbol{O}_b p$, for any $p \in \Phi$, is satisfiable.*
2. *$\models \boldsymbol{L}_a \bot \supset \neg \boldsymbol{N}_a \neg \boldsymbol{O}_b \neg \boldsymbol{O}_a p$.*
3. *$\boldsymbol{N}_a \neg \boldsymbol{O}_b p \wedge \boldsymbol{L}_a \neg \boldsymbol{O}_b p$ is not satisfiable.*

Proof. **Item 1.** Let $\mathcal{W}_p = \{w \mid w \models p\}$ and let E be all subsets of $\mathcal{W}$ except the set $\mathcal{W}_p$. It is easy to see that if $e^1_b \in E$, then $\{\}, e^1_b, w \not\models \boldsymbol{O}_b p$, for any world w. Now, define a e^2 for a that has all of $\mathcal{W} \times E$. Thus, $e^2_a, \{\}, w \models \boldsymbol{O}_a \neg \boldsymbol{O}_b p$.

Item 2. Suppose $e^k_a, \{\}, w \models \boldsymbol{L}_a \bot$ for any $w \in \mathcal{W}$. Then, for all $\langle w', e^{k-1}_b \rangle \in e^k_a$, $e^k_a, e^{k-1}_b, w' \models \bot$, and thus, $e^k_a = \{\}$. Suppose now $e^k_a, \{\}, w \models \boldsymbol{N}_a \neg \boldsymbol{O}_b \neg \boldsymbol{O}_a p$. Then, *wrt.* all of $\langle w', e^{k-1}_b \rangle \notin e^k_a$ i.e. all of $\mathbb{E}^k$, $\neg \boldsymbol{O}_b \neg \boldsymbol{O}_a p$ must hold. That is, $\neg \boldsymbol{O}_b \neg \boldsymbol{O}_a p$ must be valid. From above, we know this is not the case.

Item 3. Suppose $e^k_a, \{\}, w \models \boldsymbol{L}_a \neg \boldsymbol{O}_b p$, for any w. Then, for all $\langle w', e^{k-1}_b \rangle \in e^k_a$, $e^k_a, e^{k-1}_b, w' \models \neg \boldsymbol{O}_b p$. Since $\boldsymbol{O}_b p$ is satisfiable, there is a ${e^*}^{k-1}_b$ such that $\{\}, {e^*}^{k-1}_b, w^* \models \boldsymbol{O}_b p$, and $\langle w^*, {e^*}^{k-1}_b \rangle \notin e^k_a$. Then, $e^k_a, \{\}, w \models \neg \boldsymbol{N}_a \neg \boldsymbol{O}_b p$. ■

Thus, k-structures seem to satisfy our intuitions on the behavior of only-knowing. To understand why, notice that $\neg \boldsymbol{O}_a \neg \boldsymbol{O}_b p$ and λ involve the nesting of $\boldsymbol{N}_i$ operators. Lakemeyer (1993) makes an unavoidable technical commitment. A (i-objective) possibility is formally a maximally $\mathsf{K45}_n$-consistent set of *basic* i-objective formulas. The restriction to basic formulas is an artifact of a semantics based on the canonical model. Unfortunately, there is more to agent i's possibility than just basic formulas. In the case of Halpern (1993), the problem seems to be that $\boldsymbol{N}_i$ and $\boldsymbol{L}_i$ do not interact naturally, and that the full complement of epistemic possibilities is not considered in interpreting $\boldsymbol{N}_i$. In contrast, Theorem 9 shows that we allow non-basic formulas and by using a strictly semantic notion, we avoid problems that arise from the proof-theoretic restrictions. And, since the semantics faithfully complies with the second property, λ is not satisfiable.

The natural question is if there are axioms that characterize the semantics. We begin, in the next section, with a proof theory by Lakemeyer (1993) that is known to be sound and complete for all attempts so far, but for a restricted language.

Proof Theory

In the single agent case, $\mathcal{ONL}$'s proof theory consists of axioms of propositional logic, axioms that treat $\boldsymbol{L}$ and $\boldsymbol{N}$ as a classical belief operator in $\mathsf{K45}$, an axiom that allows us to use $\boldsymbol{N}$ and $\boldsymbol{L}$ freely on subjective formulas, modus ponens (**MP**) and necessitation (**NEC**) for both $\boldsymbol{L}$ and $\boldsymbol{N}$ as inference rules, and the following axiom:[12]

[9] A maximally satisfiable set is to be understood as a semantically characterized *complete* description of a possibility, analogous to a proof theoretically characterized notion of maximally consistent set of formulas.

[10] In his original formulation, Halpern (1993) constructs *trees*. We build on discussions in (Halpern and Lakemeyer 2001).

[11] An approach by (Waaler 2004; Waaler and Solhaug 2005) is also motivated by the proof theory. Discussions are deferred.

[12] Strictly speaking, this is not the proof theory introduced in (Levesque 1990), where an axiom replaces the inference rule **NEC**. Here, we consider an equivalent formulation by Halpern and Lakemeyer (2001).

A5. $\boldsymbol{N}\alpha \supset \neg\boldsymbol{L}\alpha$ if $\neg\alpha$ is a propositionally consistent objective formula.

As we shall see, only the axiom **A5** is controversial, since extending any objective α to *any* i-objective α is problematic. Mainly, the soundness of the axiom in the single agent case relies on propositional logic. But in the multi-agent case, since we go beyond propositional formulas establishing this consistency is non-trivial, and even circular. To this end, Lakemeyer (1993) proposes to resolve this consistency by relying on the existing logic $\mathsf{K45}_n$. As a consequence, his proof theoretic formulation appropriately generalizes all of Levesque's axioms, except for **A5** where its application is restricted to only basic i-objective consistent formulas. We use $\vdash$ to denote provability.

Definition 11. *$\mathcal{ONL}_n^-$ consists of all formulas α in $\mathcal{ONL}_n$ such that no $\boldsymbol{N}_j$ may occur in the scope of a $\boldsymbol{L}_i$ or a $\boldsymbol{N}_i$, for $i \neq j$.*

The following axioms, along with **MP** and **NEC** (for $\boldsymbol{L}_i$ and $\boldsymbol{N}_i$) is an axiomatization that we refer to as AX_n. AX_n is sound and complete for the canonical model and the i-set approach for formulas in $\mathcal{ONL}_n^-$.

- $\mathbf{A1}_n$. All instances of propositional logic,
- $\mathbf{A2}_n$. $\boldsymbol{L}_i(\alpha \supset \beta) \supset (\boldsymbol{L}_i\alpha \supset \boldsymbol{L}_i\beta)$,
- $\mathbf{A3}_n$. $\boldsymbol{N}_i(\alpha \supset \beta) \supset (\boldsymbol{N}_i\alpha \supset \boldsymbol{N}_i\beta)$,
- $\mathbf{A4}_n$. $\sigma \supset \boldsymbol{L}_i\sigma \wedge \boldsymbol{N}_i\sigma$ for i-subjective σ,
- $\mathbf{A5}_n$. $\boldsymbol{N}_i\alpha \supset \neg\boldsymbol{L}_i\alpha$ if $\neg\alpha$ is a $\mathsf{K45}_n$-consistent i-objective basic formula.

Observe that, as discussed, the soundness of $\mathbf{A5}_n$ is built on $\mathsf{K45}_n$-consistency. Since our semantics is not based on Kripke structures, proving that every $\mathsf{K45}_n$-consistent formula is satisfiable in some (k, j)-model is not immediate. We propose a construction called the (k, j)-*correspondence model*. In the following, in order to disambiguate $\mathcal{W}$ from Kripke worlds, we shall refer to our worlds as propositional valuations.

Definition 12. *The $\mathsf{K45}_n$ canonical model $M^c = \langle \mathcal{W}^c, \pi^c, \mathcal{K}_a^c, \mathcal{K}_b^c \rangle$ is defined as follows:*

1. *$\mathcal{W}^c = \{w \mid w$ is a (basic) maximally consistent set $\}$*
2. *for all $p \in \Phi$ and worlds w, $\pi^c(w)(p) =$ true iff $p \in w$*
3. *$(w, w') \in \mathcal{K}_i^c$ iff $w \backslash \boldsymbol{L}_i \subseteq w'$, $w \backslash \boldsymbol{L}_i = \{\alpha \mid \boldsymbol{L}_i\alpha \in w\}$*

Definition 13. *Given M^c, define a set of propositional valuations $\mathcal{W}$ such that for each world $w \in \mathcal{W}^c$, there is a valuation $[\![w]\!] \in \mathcal{W}$, $[\![w]\!] = \{p \mid p \in w\}$.*

Definition 14. *Given M^c and a world $w \in \mathcal{W}^c$, construct a (k, j)-model $\langle e_{[\![w]\!]}{}_a^k, e_{[\![w]\!]}{}_b^j, [\![w]\!]\rangle$ from valuations $\mathcal{W}$ inductively:*

1. *$e_{[\![w]\!]}{}_a^1 = \{\langle [\![w']\!], \{\}\rangle \mid w' \in \mathcal{K}_a^c(w)\}$,*
2. *$e_{[\![w]\!]}{}_a^k = \{\langle [\![w']\!], e_{[\![w']\!]}{}_b^{k-1}\rangle \mid w' \in \mathcal{K}_a^c(w)\}$, where $e_{[\![w']\!]}{}_b^{k-1} = \{\langle [\![w'']\!], e_{[\![w'']\!]}{}_a^{k-2}\rangle \mid w'' \in \mathcal{K}_b^c(w')\}$.*

Further, $e_{[\![w]\!]}{}_b^j$ is constructed analogously. Let us refer to this model as the (k, j)-correspondence model of (M^c, w).

Roughly, Defn. 14 is a construction of a (k, j)-model that appeals to the accessibility relations in the canonical model.[13] Thus, a e_a^1 for Alice *wrt.* w has precisely the valuations of Kripke worlds $w' \in \mathcal{K}_a^c(w)$. Quite analogously, a e_a^k is a set $\{\langle [\![w']\!], e^{k-1}\rangle\}$, where $w' \in \mathcal{K}_a^c(w)$ as before, but e^{k-1} is an epistemic state for Bob and hence refers all worlds $w'' \in \mathcal{K}_b^c(w')$. By a induction on the depth of a *basic* formula α, we obtain a theorem that α of maximal a, b-depth k, j is satisfiable at (M^c, w) iff the (k, j)-correspondence model satisfies the formula.

Theorem 15. *For all basic formulas α in $\mathcal{ONL}_n^-$ and of maximal a, b-depth of k, j,*

$$M^c, w \models \alpha \text{ iff } e_{[\![w]\!]}{}_a^k, e_{[\![w]\!]}{}_b^j, [\![w]\!] \models \alpha.$$

Proof. By definition, the proof holds for propositional formulas, disjunctions and negations. So let us say the result holds for formulas of a, b-depth 1. Suppose now $M^c, w \models \boldsymbol{L}_a\alpha$, where $\boldsymbol{L}_a\alpha$ has a, b-depth of $1, 2$. Then for all $w' \in \mathcal{K}_a^c(w)$, $M^c, w' \models \alpha$ iff (by induction hypothesis) $e_{[\![w']\!]}{}_a^1, e_{[\![w']\!]}{}_b^1, [\![w']\!] \models \alpha$ iff $e_{[\![w]\!]}{}_a^2, \{\}, [\![w]\!] \models \boldsymbol{L}_a\alpha$. By construction, we also have $e_{[\![w]\!]}{}_a^1, \{\}, [\![w]\!] \models \boldsymbol{L}_a\alpha$. Since b's structure is irrelevant, we get $e_{[\![w]\!]}{}_a^1, e_{[\![w]\!]}{}_b^2, [\![w]\!] \models \boldsymbol{L}_a\alpha$ proving the hypothesis.

For the other direction, suppose $e_{[\![w]\!]}{}_a^1, e_{[\![w]\!]}{}_b^2, [\![w]\!] \models \boldsymbol{L}_a\alpha$. For all $[\![w']\!] \in e_{[\![w]\!]}{}_a^1$, $e_{[\![w]\!]}{}_a^1, \{\}, [\![w']\!] \models \alpha$ iff (by hyp.) $M^c, w' \models \alpha$ for all $w' \in \mathcal{K}_a^c(w)$ iff $M^c, w \models \boldsymbol{L}_a\alpha$. ■

Lemma 16. *Every $\mathsf{K45}_n$-consistent basic formula α is satisfiable wrt. some (k, j)-model.*

Proof. It is a property of the canonical model that every $\mathsf{K45}_n$-consistent basic formula is satisfiable *wrt.* the canonical model. Supposing that the formula has a a, b-depth of k, j then from Thm 15, we know there is at least the correspondence (k, j)-model that also satisfies the formula. ■

Theorem 17. *For all $\alpha \in \mathcal{ONL}_n^-$, if $AX_n \vdash \alpha$ then $\models \alpha$.*

Proof. The soundness is easily shown to hold for $\mathbf{A1}_n - \mathbf{A4}_n$. The soundness of $\mathbf{A5}_n$ is shown by induction on the depth. Suppose α is a propositional formula, and say $\neg\alpha$ is a consistent propositional formula (and hence $\mathsf{K45}_n$-consistent). Then there is a world w^* such that $\{\}, \{\}, w^* \models \neg\alpha$. Given a e_a^k, if $\langle w^*, e_b^{k-1}\rangle \in e_a^k$ for some e_b^{k-1}, then $e_a^k, \{\}, w \models \neg\boldsymbol{L}_a\alpha$ for any world w. If not, then $e_a^k, \{\}, w \models \neg\boldsymbol{N}_a\alpha$. Thus, $e_a^k, \{\}, w \models \boldsymbol{N}_a\alpha \supset \neg\boldsymbol{L}_a\alpha$. Wlog, assume the proof holds for a-objective formulas of max. b-depth $k-1$. Suppose now, α is such a formula, and $\neg\alpha$ is $\mathsf{K45}_n$-consistent. By Lemma 16, there is $\langle w^*, e^*{}_b^{k-1}\rangle$, such that $\{\}, e^*{}_b^{k-1}, w^* \models \neg\alpha$. Again, if $\langle w^*, e^*{}_b^{k-1}\rangle \in e_a^k$, then $e_a^k, \{\}, w \models \neg\boldsymbol{L}_a\alpha$ and if not, then $e_a^k, \{\}, w \models \neg\boldsymbol{N}_a\alpha$. ■

We proceed with the completeness over the following definition, and lemmas.

[13] The construction is somewhat similar to the notion of *generated submodels* of Kripke frames (Hughes and Cresswell 1984).

Definition 18. *A formula ψ is said to be* independent *of the formula ϕ wrt. an axiom system AX, if neither $AX \vdash \phi \supset \psi$ nor $AX \vdash \phi \supset \neg\psi$.*

Lemma 19 (Halpern and Lakemeyer, 2001). *If $\phi_1, \ldots, \phi_m$ are $K45_n$-consistent basic i-objective formulas then there exists a basic i-objective formula ψ of the form $\boldsymbol{L}_j\psi'$ ($j \neq i$) that is independent of $\phi_1, \ldots, \phi_m$ wrt. $K45_n$.*

Lemma 20. *In the lemma above, if ϕ_i are i-objective and of maximal j-depth k for $j \neq i$, then there is a ψ of j-depth $2k + 2$.*

Lemma 21 (Halpern and Lakemeyer, 2001). *If ϕ and ψ are i-objective basic formulas, and if $\boldsymbol{L}_i\phi \wedge \boldsymbol{N}_i\psi$ is AX_n-consistent, then $\phi \vee \psi$ is valid.*

Lemma 22 (Halpern and Lakemeyer, 2001). *Every formula $\alpha \in \mathcal{ONL}_n$ is provably equivalent to one in the normal form (written below for $n = \{a, b\}$):*

$$\bigvee(\sigma \wedge \boldsymbol{L}_a\varphi_{a0} \wedge \neg\boldsymbol{L}_a\varphi_{a1} \ldots \wedge \neg\boldsymbol{L}_a\varphi_{am_1} \wedge \boldsymbol{L}_b\varphi_{b0} \ldots \wedge \neg\boldsymbol{L}_b\varphi_{bm_2} \wedge \boldsymbol{N}_a\psi_{a0} \ldots \wedge \neg\boldsymbol{N}_a\psi_{an_1} \wedge \boldsymbol{N}_b\psi_{b0} \ldots \wedge \neg\boldsymbol{N}_b\psi_{bn_2})$$

where σ is a propositional formula, and φ_{im} and ψ_{in} are i-objective. If $\alpha \in \mathcal{ONL}_n^-$, φ_{im} and ψ_{in} are basic.

Theorem 23. *For all formulas $\alpha \in \mathcal{ONL}_n^-$, if $\models \alpha$ then $AX_n \vdash \alpha$.*

Proof. It is sufficient to prove that every AX_n-consistent formula ξ is satisfiable *wrt.* some (k, j)-model. If ξ is basic, then by Lemma 16, the statement holds. If ξ is not basic, then wlog, it can be considered in the normal form:

$$\bigvee(\sigma \wedge \boldsymbol{L}_a\varphi_{a0} \wedge \neg\boldsymbol{L}_a\varphi_{a1} \ldots \wedge \neg\boldsymbol{L}_a\varphi_{am_1} \wedge \boldsymbol{L}_b\varphi_{b0} \ldots \wedge \neg\boldsymbol{L}_b\varphi_{bm_2} \wedge \boldsymbol{N}_a\psi_{a0} \ldots \wedge \neg\boldsymbol{N}_a\psi_{an_1} \wedge \boldsymbol{N}_b\psi_{b0} \ldots \wedge \neg\boldsymbol{N}_b\psi_{bn_2})$$

where σ is a propositional formula, and φ_{im} and ψ_{in} are i-objective and basic. Since σ is propositional and consistent, there is clearly a world w^* such that $w^* \models \sigma$. We construct a k'-structure such that it satisfies all the a-subjective formulas in the normal form above. Following that, a j'-structure for all the b-subjective formulas is constructed identically. The resulting (k', j')-model (with w^*) satisfies ξ.

Let A be all $K45_n$-consistent formulas of the form $\varphi_{a0} \wedge \psi_{a0} \wedge \neg\varphi_{aj}$ (for $j \geq 1$) or the form $\varphi_{a0} \wedge \psi_{a0} \wedge \neg\psi_{aj}$. Let γ be independent of all formulas in A, as in Lemma 19 and 20. Note that, while we take ξ itself to be of maximal a, b-depth of k, j, the depth of $\varphi_{a0}, \ldots$ being a-objective are of maximal b-depth $k - 1$, and hence γ is of b-depth $2k$ (Lemma 20). Given a consistent set of formulas, the standard Lindenbaum construction can be used to construct a maximally consistent set of formulas, all of a maximal b-depth $k - 1$. That is, a formula is considered in the construction only if it has a maximal b-depth $k - 1$. Now, let S_a be a set of all maximally consistent sets of formulas, constructed by only considering formulas of maximal b-depth $k - 1$, and containing $\varphi_{a0} \wedge (\neg\psi_{a0} \vee (\psi_{a0} \wedge \gamma))$. Since each of these consistent sets are basic and a-objective, they are satisfiable by Lemma 16. Thus the sets $S' \in S_a$ are satisfiable *wrt.* $2k$-structures $\langle w, e_b^{2k}\rangle$. Let $k' = 2k + 1$. By constructing a k'-structure for Alice, say $e_a^{k'}$, from each $\langle w, e_b^{2k}\rangle$ for every $S' \in S_a$, we have that $Obj_a(e_a^{k'}) = S_a$. We shall show that all the a-subjective formulas in the normal form are satisfied *wrt.* $\langle e_a^{k'}, \{\}, w^*\rangle$.

Since for all $S' \in S_a$, we have $\varphi_{a0} \in S'$ we get that $e_a^{k'}, \{\}, w^* \models \boldsymbol{L}_a\varphi_{a0}$. Now, since $\boldsymbol{L}_a\varphi_{a0} \wedge \neg\boldsymbol{L}_a\varphi_{aj}$ is consistent, it must be that $\varphi_{a0} \wedge \neg\varphi_{aj}$ is consistent. For suppose not, then $\neg\varphi_{a0} \vee \varphi_{aj}$ is provable and thus, we have $\varphi_{a0} \supset \varphi_{aj}$. We then prove $\boldsymbol{L}_a\varphi_{a0} \supset \boldsymbol{L}_a\varphi_{aj}$, and since we have $\boldsymbol{L}_a\varphi_{a0}$ we prove $\boldsymbol{L}_a\varphi_{aj}$, clearly inconsistent with $\boldsymbol{L}_a\varphi_{a0} \wedge \neg\boldsymbol{L}_a\varphi_{aj}$. Now that $\varphi_{a0} \wedge \neg\varphi_{aj}$ is consistent, we either have that $\varphi_{a0} \wedge \neg\varphi_{aj} \wedge \psi_{a0}$ or $\varphi_{a0} \wedge \neg\varphi_{aj} \wedge \neg\psi_{a0}$ is consistent. With the former, we also have that $\varphi_{a0} \wedge \neg\varphi_{aj} \wedge \psi_{a0} \wedge \gamma$ is consistent. There are maximally consistent sets that contain one of them, both of which contain $\neg\varphi_{aj}$. This means that, $e_a^{k'}, \{\}, w^* \models \neg\boldsymbol{L}_a\varphi_{aj}$.

Now, consider some k'-structure $\langle w^\bullet, e^{\bullet}{}_b^{2k}\rangle \notin e_a^{k'}$. One of the following a-objective formulas must hold *wrt.* this k'-structure: (a) $\varphi_{a0} \wedge \psi_{a0}$, (b) $\varphi_{a0} \wedge \neg\psi_{a0}$, (c) $\neg\varphi_{a0} \wedge \psi_{a0}$ or (d) $\neg\varphi_{a0} \wedge \neg\psi_{a0}$. It can not be (d), since $\boldsymbol{L}_a\varphi_{a0} \wedge \boldsymbol{N}_a\psi_{a0}$ is consistent, and this implies that $\varphi_{a0} \vee \psi_{a0}$ is valid (by Lemma 21). It certainly cannot be (b), for it would be in some $S' \in S_a$. This leaves us with options (c) and (a), both of which have $\neg\psi_{a0}$. Since the k'-structure was arbitrary, we must have for all $\langle w, e_b^{2k}\rangle \notin e_a^{k'}, \{\}, e_b^{2k}, w \models \psi_{a0}$. Thus, $e_a^{k'}, \{\}, w^* \models \boldsymbol{N}_a\psi_{a0}$.

Finally, since $\boldsymbol{N}_a\psi_{a0} \wedge \neg\boldsymbol{N}_a\psi_{aj}$ is consistent, it must be that $\psi_{a0} \wedge \neg\psi_{aj}$ is consistent. Further, either $\psi_{a0} \wedge \neg\psi_{aj} \wedge \varphi_{a0}$ or $\psi_{a0} \wedge \neg\psi_{aj} \wedge \neg\varphi_{a0}$ is consistent. If the former, then $\psi_{a0} \wedge \neg\psi_{aj} \wedge \varphi_{a0} \wedge \neg\gamma$ is also consistent. Let β be that which is consistent. Note that $\neg\beta \wedge (\varphi_{a0} \wedge (\neg\psi_{a0} \vee (\psi_{a0} \wedge \gamma)))$ is consistent, and hence part of all $S' \in S_a$. This means that $e_a^{k'}, \{\}, w^* \models \boldsymbol{L}_a(\neg\beta)$. But since β itself is consistent, there is a k'-structure such that $\{\}, e^\bullet_{(b,2k)}, w^\bullet \models \beta$. And this k'-structure can not be in $e_a^{k'}$. This means that $e_a^{k'}, \{\}, w^* \models \neg\boldsymbol{N}_a\psi_{aj}$. Thus, all the a-subjective formulas in the normal form above are satisfiable *wrt.* $e_a^{k'}$. ■

Now, observe that, although $\boldsymbol{L}_a\bot \supset \neg\boldsymbol{N}_a\neg\boldsymbol{O}_b\neg\boldsymbol{O}_a p$ (ζ) from Theorem 10 is valid, yet it is not derivable from AX_n. In fact, the soundness result is easily extended to the full language $\mathcal{ONL}_n$. Then, the proof theory cannot be complete for the full language since there is $\zeta \in \mathcal{ONL}_n$ such that $\not\vdash \zeta$ and $\models \zeta$. Similarly, the validity of non-provable formulas $\neg\boldsymbol{O}_a\neg\boldsymbol{O}_b p$ and ζ *wrt.* the canonical model and the i-set approach respectively, show that although AX_n is also sound for the full language in these approaches, it cannot be compelete. Mainly, axiom $\mathbf{A5}_n$ has to somehow go beyond basic formulas. As Halpern and Lakemeyer (2001) discuss, the problem is one of circularity. We would like the axiom to hold for any α such that it is a consistent i-objective formula, but to deal with consistency we have to clarify what the axiom system looks like.

The approach taken by Halpern and Lakemeyer is to introduce *validity* (and its dual satisfiability) directly into the language. Formulas in the new language, $\mathcal{ONL}_n^+$, are shown to be provably equivalent to $\mathcal{ONL}_n$. Some new axioms involving validity and satisfiability are added to the axiom system, and the resultant proof theory AX_n^+ is shown to be sound and complete for formulas in $\mathcal{ONL}_n^+$, *wrt.* an *extended* canonical model. (An extended canonical model follows the spirit of the canonical model construction but by considering maximally AX_n^+-consistent sets, and treat $\boldsymbol{L}_i$ and $\boldsymbol{N}_i$ as two independent modal operators.) So, one approach is to show that for formulas in the extended language the set of valid formulas overlap in the extended canonical model and k-structures. But then, as we argued, axiomatizing validity is not natural. Also, the proof theory is difficult to use. And in the end, we would still understand the axioms to characterize a semantics bridged on proof-theoretic elements.

Again, what is desired is a generalization of Levesque's axiom $\mathbf{A5}$, and nothing more. To this end, we propose a new axiom system, that is subtly related to the structure of formulas as are parameters k and j. The axiom system has an additional t-axioms, and is to correspond to a sequence of languages $\mathcal{ONL}_n^t$.[14]

Definition 24. *Let $\mathcal{ONL}_n^1 = \mathcal{ONL}_n^-$. Let $\mathcal{ONL}_n^{t+1}$ be all Boolean combinations of formulas of $\mathcal{ONL}_n^t$ and formulas of the form $\boldsymbol{L}_i\alpha$ and $\boldsymbol{N}_i\alpha$ for $\alpha \in \mathcal{ONL}_n^t$.*

It is not hard to see that $\mathcal{ONL}_n^{t+1} \supseteq \mathcal{ONL}_n^t$. Note that t here does not correspond to the depth of formulas. Indeed, a formula of the form $(\boldsymbol{L}_b\boldsymbol{L}_a)^{k+1}p$ is already in $\mathcal{ONL}_n^-$. Let AX_n^{t+1} be an axiom system consisting of $\mathbf{A1}_n - \mathbf{A4}_n$, $\mathbf{MP}$, $\mathbf{NEC}$ and $\mathbf{A5}_n^1 - \mathbf{A5}_n^{t+1}$ defined inductively as:

$\mathbf{A5}_n^1$. $\boldsymbol{N}_i\alpha \supset \neg\boldsymbol{L}_i\alpha$, if $\neg\alpha$ is a $\mathsf{K45}_n$-consistent i-objective basic formula.

$\mathbf{A5}_n^{t+1}$. $\boldsymbol{N}_i\alpha \supset \neg\boldsymbol{L}_i\alpha$, if $\neg\alpha \in \mathcal{ONL}_n^t$, is i-objective, and consistent *wrt.* $\mathbf{A1}_n - \mathbf{A4}_n$, $\mathbf{A5}_n^1 - \mathbf{A5}_n^t$.

Theorem 25. *For all $\alpha \in \mathcal{ONL}_n^t$, if $AX_n^t \vdash \alpha$ then $\models \alpha$.*

Proof. We prove by induction on t. The case of AX_n^1 is identical to Theorem 17. So, for the induction hypothesis, let us assume that *wrt.* AX_n^t, if $AX_n^t \vdash \beta$ for $\beta \in \mathcal{ONL}_n^t$ then $\models \beta$. Now, suppose that $\neg\alpha$ is consistent *wrt.* AX_n^t and is a-objective. This implies that $\not\models \alpha$. Thus, there is some k-structure $\langle w^*, e{^*}_b^k\rangle$ such that $\{\}, e{^*}_b^k, w^* \models \neg\alpha$. Suppose now $\langle w^*, e{^*}_b^k\rangle \in e_a^{k+1}$ then $e_a^{k+1}, \{\}, w' \models \neg\boldsymbol{L}_a\alpha$ and if not then $e_a^{k+1}, \{\}, w' \models \neg\boldsymbol{N}_a\alpha$. Thus, $e_a^{k+1}, \{\}, w' \models \boldsymbol{N}_a\alpha \supset \neg\boldsymbol{L}_a\alpha$, demonstrating the soundness of AX_n^{t+1}. ■

We establish completeness in a manner identical to Theorem 23, and thus it necessary to ensure that Lemma 19, 20 and 21 hold for non-basic formulas.

Lemma 26. *If $\phi_1, \ldots, \phi_m$ are AX_n^t-consistent i-objective formulas, then there is a basic formula ψ of the form $\boldsymbol{L}_j\psi$ $(j \neq i)$ that is independent of $\phi_1, \ldots, \phi_m$ wrt. AX_n^t.*

Proof. Suppose that ϕ_i are a-objective and of maximal b-depth k. A formula ψ of the form $(\boldsymbol{L}_b\boldsymbol{L}_a)^{k+1}p$ (where $p \in \Phi$ is in the scope of $k+1$ $\boldsymbol{L}_b\boldsymbol{L}_a$) is shown to be independent of $\phi_1, \ldots, \phi_m$. Let us suppose we can derive a γ of the form $\boldsymbol{L}_b\boldsymbol{L}_a\boldsymbol{L}_b\boldsymbol{L}_a \ldots p$ of maximal depth k, to show that neither $\vdash \gamma \supset \psi$ nor $\vdash \gamma \supset \neg\psi$. Given any formula, the only axioms in AX_n^t that can introduce γ in the scope of modal operators is $\mathbf{A4}_n$ and $\mathbf{A5}_n^t$. Applying $\mathbf{A4}_n$ gives $\boldsymbol{L}_b\gamma$ or $\boldsymbol{N}_b\gamma$, and then using the axiom again we have $\boldsymbol{L}_b\boldsymbol{L}_b\gamma$ or $\boldsymbol{L}_b\boldsymbol{N}_b\gamma$. It is easy to see that the resulting formulas are clearly independent from ψ. Applying $\mathbf{A5}_n^t$ on the other hand, allows us to derive $\vdash \gamma \supset \boldsymbol{N}_a\gamma$ or $\vdash \gamma \supset \neg\boldsymbol{L}_a\gamma$ (γ is consistent *wrt.* AX_n^t and hence also *wrt.* $\mathbf{A5}_n^{t-1}$). Again, we could show $\vdash \gamma \supset \neg\boldsymbol{L}_b\neg\boldsymbol{L}_a\gamma$. Continuing this way, it might only be possible to derive $\neg\boldsymbol{L}_b\neg\boldsymbol{L}_a \ldots \boldsymbol{L}_b\boldsymbol{L}_a \ldots p$ of depth $2k+2$, that is indeed independent of ψ. ■

Lemma 27. *If ϕ and ψ are i-objective formulas, $\phi, \psi \in \mathcal{ONL}_n^t$ and $\boldsymbol{L}_i\phi \wedge \boldsymbol{N}_i\psi$ is AX_n^{t+1}-consistent then $\models \phi \vee \psi$.*

Proof. Suppose not. Then $\neg\phi \wedge \neg\psi$ is AX_n^t-consistent, and by $\mathbf{A5}_n^{t+1}$ we prove $\boldsymbol{N}_a(\phi \vee \psi) \supset \neg\boldsymbol{L}_a(\phi \vee \psi)$, and thus, $\boldsymbol{N}_a\psi \supset \neg\boldsymbol{L}_a\phi$, and this is not AX_n^{t+1}-consistent with $\boldsymbol{L}_a\phi \wedge \boldsymbol{N}_a\psi$. ■

Theorem 28. *For all $\alpha \in \mathcal{ONL}_n^t$, if $\models \alpha$ then $AX_n^t \vdash \alpha$.*

Proof. Proof by induction on t. It is sufficient to show that if a formula $\beta \in \mathcal{ONL}_n^{t+1}$ is AX_n^{t+1}-consistent then it is satisfiable *wrt.* some model. We already have the proof for $\mathcal{ONL}_n^1$ (see Theorem 23). Let us assume the proof holds for all formulas $\alpha \in \mathcal{ONL}_n^t$. Particularly, this means that any formula that is AX_n^t-consistent is satisfiable *wrt.* some (k', j')-model. Let $\alpha \in \mathcal{ONL}_n^{t+1}$ (say of maximal a, b-depth of $k+1, j+1$), and suppose that α is consistent *wrt.* AX_n^{t+1}. It is sufficient to show that α is satisfiable. Wlog, we take it in the normal form:

$$\bigvee(\sigma \wedge \boldsymbol{L}_a\varphi_{a0} \wedge \neg\boldsymbol{L}_a\varphi_{a1} \ldots \wedge \neg\boldsymbol{L}_a\varphi_{am_1} \wedge \boldsymbol{L}_b\varphi_{b0} \ldots \wedge \neg\boldsymbol{L}_b\varphi_{bm_2} \wedge \boldsymbol{N}_a\psi_{a0} \ldots \wedge \neg\boldsymbol{N}_a\psi_{an_1} \wedge \boldsymbol{N}_b\psi_{b0} \ldots \wedge \neg\boldsymbol{N}_b\psi_{bn_2}).$$

Note that, by definition, it must be that all of φ_{im}, ψ_{in} are at most in $\mathcal{ONL}_n^t$ (i.e. they may also be in $\mathcal{ONL}_n^{t-1}, \ldots$), and i-objective. We proceed as we did for Theorem 23 but without restricting to basic formulas. Let A be all AX_n^t-consistent formulas of the form $\varphi_{a0} \wedge \psi_{a0} \wedge \neg\varphi_{aj}$ or $\varphi_{a0} \wedge \psi_{a0} \wedge \neg\psi_{aj}$ (they are of maximal b-depth k). Let γ be independent of all formulas in A. Let S_a be the set of all (AX_n^t-) maximally consistent sets of formulas, constructed from formulas of maximal b-depth k, and containing $\varphi_{a0} \wedge (\neg\psi_{a0} \vee (\psi_{a0} \wedge \gamma))$, and hence by induction hypothesis they are satisfiable in some model. Note that all formulas in S_a are in $\mathcal{ONL}_n^t$. The b-depth is maximally $2k+2$. Letting $k'' = 2k+2$, we have that for all $S' \in S_a$, there is a $\langle w, e_b^{k''}\rangle$ such that $\{\}, e_b^{k''}, w \models S'$. Let $k' = k''+1$. Letting $e_a^{k'}$ be all such k'-structures $\langle w, e_b^{k''}\rangle$ for each $S' \in S_a$ makes $Obj_a^+(e_a^{k'}) = S_a$ (in contrast, for Thereom 23 we dealt with Obj_a). We claim that this k'-structure for Alice, a j'-structure for Bob constructed similarly, and a world where σ holds (there is such a world since σ is propositional and consistent) is a model where α is satisfied. The proof proceeds as in Theorem 23. We show the case of $\neg\boldsymbol{L}_a\varphi_{aj}$.

[14] The idea was also suggested by a reviewer in (Halpern and Lakemeyer 2001) for an axiomatic characterization of the extended canonical model, although its completeness was left open.

Since $\boldsymbol{L}_a\varphi_{a0} \wedge \neg\boldsymbol{L}_a\varphi_{aj}$ is consistent *wrt.* AX_n^{t+1}, it must be that $\varphi_{a0} \wedge \neg\varphi_{aj}$ is consistent *wrt.* AX_n^{t+1}. Further, since $\varphi_{a0}, \varphi_{aj} \in \mathcal{ONL}_n^t$, they must consistent be *wrt.* AX_n^t (for if not, they cannot by definition be consistent *wrt.* AX_n^{t+1}). This means that either $\varphi_{a0}\wedge\neg\varphi_{aj}\wedge\psi_{a0}$ or $\varphi_{a0}\wedge\neg\varphi_{aj}\wedge\neg\psi_{a0}$ is consistent. If the former is, then so is $\varphi_{a0}\wedge\neg\varphi_{aj}\wedge\psi_{a0}\wedge\gamma$. Since S_a consist of all AX_n^t-consistent formulas containing $\varphi_{a0}\wedge(\neg\psi_{a0}\wedge(\psi_{a0}\wedge\gamma))$, there is clearly a $S' \in S_a$ such that $\neg\varphi_{aj} \in S'$. Consequently, it can not be that $e_a^{k'}, \{\}, w' \models \boldsymbol{L}_a\varphi_{aj}$. Thus, $e_a^{k'}, \{\}, w' \models \neg\boldsymbol{L}_a\varphi_{aj}$. ■

Thus, we have a sound and complete axiomatization for the propositional fragment of $\mathcal{ONL}_n$. In comparison to Lakemeyer (1993), the axiomatization goes beyond a language that restricts the nesting of $\boldsymbol{N}_i$. In contrast to Halpern and Lakemeyer (2001), the axiomatization does not necessitate the use of semantic notions in the proof theory. A third axiomatization by (Waaler 2004; Waaler and Solhaug 2005) proposes an interesting alternative to deal with the circularity in a generalized $\mathbf{A5}$. The idea is to first define consistency by formulating a fragment of the axiom system in the sequent calculus. Quite analogous to having t-axioms, they allow us to apply $\mathbf{A5}_n$ on i-objective formulas of a lower depth, thus avoiding circularity without the need to appeal to satisfiability as in (Halpern and Lakemeyer 2001). Waaler and Solhaug (2005) also define a semantics for multi-agent only-knowing which does not appeal to canonical models. Instead, they define a class of Kripke structures which need to satisfy certain constraints. Unfortunately, these constraints are quite involved and, as the authors admit, the nature of these models "is complex and hard to penetrate."

To get a feel of the axiomatization, let us consider a well studied example from (Halpern and Lakemeyer 2001) to see where we differ. Suppose Alice assumes the following default: unless I know that Bob knows my secret then he does not know it. If the default is all that she knows, then she *nonmonotonically* comes to believe that Bob does not know her secret. Let γ be a proposition that denotes Alice's secret, and we want to show that $\vdash \boldsymbol{O}_a(\delta) \supset \boldsymbol{L}_a\neg\boldsymbol{L}_b\gamma$, where $\delta = \neg\boldsymbol{L}_a\boldsymbol{L}_b\gamma \supset \neg\boldsymbol{L}_b\gamma$. We write (Def.) to mean $\boldsymbol{O}_a\alpha \equiv \boldsymbol{L}_a\alpha \wedge \boldsymbol{N}_a\neg\alpha$, and we freely reason with propositional logic (PL) or $\mathsf{K45}_n$.

1. $\boldsymbol{O}_a(\delta) \supset \boldsymbol{L}_a\neg\boldsymbol{L}_a\boldsymbol{L}_b\gamma \supset \boldsymbol{L}_a\neg\boldsymbol{L}_b\gamma$ — Def.,PL,$\mathbf{A2}_n$
2. $\boldsymbol{O}_a(\delta) \supset \boldsymbol{N}_a\neg\boldsymbol{L}_a\boldsymbol{L}_b\gamma \wedge \boldsymbol{N}_a\boldsymbol{L}_b\gamma$ — Def.,PL,$\mathsf{K45}_n$
3. $\boldsymbol{N}_a\boldsymbol{L}_b\gamma \supset \neg\boldsymbol{L}_a\boldsymbol{L}_b\gamma$ — $\mathbf{A5}_n^1$
4. $\neg\boldsymbol{L}_a\boldsymbol{L}_b\gamma \supset \boldsymbol{L}_a\neg\boldsymbol{L}_a\boldsymbol{L}_b\gamma$ — $\mathbf{A4}_n$
5. $\boldsymbol{O}_a(\delta) \supset \boldsymbol{L}_a\neg\boldsymbol{L}_a\boldsymbol{L}_b\gamma$ — 2,3,4,PL
6. $\boldsymbol{O}_a(\delta) \supset \boldsymbol{L}_a\neg\boldsymbol{L}_b\gamma$ — 1,5,PL

We use $\mathbf{A5}_n^1$, and it is applicable because $\neg\boldsymbol{L}_b\gamma$ is a-objective and $\mathsf{K45}_n$-consistent. Now, suppose Alice is cautious. She changes her default to assume that if she does not believe Bob to only-know some set of facts $\theta \in \Phi$, then θ is not all that he knows. We would like to show

$$\vdash \boldsymbol{O}_a(\neg\boldsymbol{L}_a\boldsymbol{O}_b\theta \supset \neg\boldsymbol{O}_b\theta) \supset \boldsymbol{L}_a\neg\boldsymbol{O}_b\theta$$

Of course, this default is different from δ in containing $\boldsymbol{O}_b\theta$ rather than $\boldsymbol{L}_b\gamma$. The proof is identical, except that we use $\mathbf{A5}_n^2$, since $\neg\boldsymbol{O}_b\theta \in \mathcal{ONL}_n^1$ is a-objective and AX_n^1-consistent. The latter proof requires reasoning with the *satisfiability* modal operator in Halpern and Lakemeyer (2001), and is not provable with the axioms of Lakemeyer (1993).

Autoepistemic Logic

Having examined the properties of multi-agent only-knowing, in terms of a semantics for both the first-order and propositional case, and an axiomatization for the propositional case, in the current section we discuss how the semantics also captures autoepistemic logic (AEL). AEL, as originally developed by Moore (1985), intends to allow agents to draw conclusions, by making observations of their own epistemic states. For instance, Alice concludes that she has no brother because if she did have one then she would have known about it, and she does not know about it (Moore 1985). The characterization of such beliefs are defined using fixpoints called *stable expansions*. In the single agent case, Levesque (1990) showed that the beliefs of an agent who only-knows α is *precisely* the stable expansion of α. Of course, the leverage with the former is that it is specified using regular entailments. In Lakemeyer (1993), and Halpern and Lakemeyer (2001), a many agent generalization of AEL is considered in the sense of a stable expansion for every agent, and relating this to what the agent only-knows. But their generalizations are only for the propositional fragment, while Levesque's definitions involved first-order entailments. In contrast, we obtain the corresponding quantificational multi-agent generalization of AEL. We state the main theorems below. The proofs are omitted since they follow very closely from the ideas for the single agent case (Levesque and Lakemeyer 2001).

Definition 29. *Let A be a set of formulas, and Γ is the i-stable expansion of A iff it the set of first-order implications of $A \cup \{\boldsymbol{L}_i\beta \mid \beta \in \Gamma\} \cup \{\neg\boldsymbol{L}_i\beta \mid \beta \notin \Gamma\}$.*

Definition 30 (Maximal structure). *If e_a^k is a k-structure, let $\boldsymbol{e}_a^+$ be a k-structure with the addition of all $\langle w', e_b^{k-1}\rangle \notin e_a^k$ such that for every $\alpha \in \mathcal{ONL}_n^-$ of maximal a,b-depth $k, k-1$, if $e_a^k, \{\}, w \models \boldsymbol{L}_a\alpha$ for any world w then $e_a^k, e_b^{k-1}, w' \models \alpha$. Define $\Gamma = \{\beta \mid \beta$ is basic and $\boldsymbol{e}_a^+, \{\}, w \models \boldsymbol{L}_a\beta\}$ as the belief set of $\boldsymbol{e}_a^+$.*

Theorem 31. *Let $M = \langle \boldsymbol{e}_a^+, e_b^j, w\rangle$ be a model, where $\boldsymbol{e}_a^+$ is a maximal structure for a. Let Γ be the belief set of $\boldsymbol{e}_a^+$, and suppose $\alpha \in \mathcal{ONL}_n^-$ is of maximal a,b-depth $k, k-1$. Then, $M \models \boldsymbol{O}_a\alpha$ iff Γ is the a-stable expansion of α.*

Theorem 31 essentially says that the complete set of basic beliefs at a *maximal* epistemic state where α is all that i knows, precisely coincides with the i-stable expansion of α.

Axiomatizing Validity

Extending the work in (Lakemeyer 1993) and (Halpern 1993), which was only restricted to formulas in $\mathcal{ONL}_n^-$, Halpern and Lakemeyer (2001) proposed a multi-agent only-knowing logic that handles the nesting of $\boldsymbol{N}_i$ operators. But as discussed, there are two undesirable features. The first is a semantics based on canonical models, and the second is a proof theory that axiomatizes validity. Although such a construction is far from natural, we show in this section that they do indeed capture the desired properties of only-knowing. This also instructs us that our axiomatization avoids such problems in a reasonable manner.

Recall that the language of (Halpern and Lakemeyer 2001) is $\mathcal{ONL}_n^+$, which is $\mathcal{ONL}_n$ and a modal operator for validity, *Val*. A modal operator *Sat*, for satisfiability, is used freely such that $Val(\alpha)$ is syntactically equivalent to $\neg Sat(\neg\alpha)$. To enable comparisons, we present a variant of our logic, that has all its main features, but has additional notions to handle the extended language. We then show that this logic and (Halpern and Lakemeyer 2001) agree on the set of valid sentences from $\mathcal{ONL}_n^+$ (and also $\mathcal{ONL}_n$).

The main feature of (Halpern and Lakemeyer 2001) is the proof theory AX'_n, and a semantics that is sound and complete for AX'_n via the extended canonical model. AX'_n consists of $\mathbf{A1}_n - \mathbf{A4}_n$, $\mathbf{MP}$, $\mathbf{NEC}$ and the following:

$\mathbf{A5}'_n$. $Sat(\neg\alpha) \supset (\boldsymbol{N}_i\alpha \supset \neg\boldsymbol{L}_a\alpha)$, if α is i-objective.

$\mathbf{V1}$. $Val(\alpha) \wedge Val(\alpha \supset \beta) \supset Val(\beta)$.

$\mathbf{V2}$. $Sat(p_1 \wedge \ldots p_n)$, if p_i's are literals and $p_1 \wedge \ldots p_n$ is propositionally consistent.

$\mathbf{V3}$. $Sat(\alpha \wedge \beta_1) \wedge \ldots Sat(\alpha \wedge \beta_k) \wedge Sat(\gamma \wedge \delta_1) \ldots \wedge Sat(\gamma \wedge \delta_m) \wedge Val(\alpha \vee \gamma) \supset Sat(\boldsymbol{L}_i\alpha \wedge \neg\boldsymbol{L}_i\neg\beta_1 \ldots \wedge \boldsymbol{N}_i\gamma \wedge \neg\boldsymbol{N}_i\neg\delta_1 \ldots)$, if $\alpha, \beta_i, \gamma, \delta_i$ are i-objective.

$\mathbf{V4}$. $Sat(\alpha) \wedge Sat(\beta) \supset Sat(\alpha \wedge \beta)$, if α is i-objective and β is i-subjective.

$\mathbf{NEC}_{Val}$. From α infer $Val(\alpha)$.

The essence of our new logic, in terms of a notion of depth (with $|Val(\alpha)|_i = |\alpha|_i$) and a semantical account over possible worlds, is as before. The complete semantic definition for formulas in $\mathcal{ONL}_n^+$ of maximal a, b-depth of k, j is:

1. -8. as before,
9. $e_a^k, e_b^j, w \models Val(\alpha)$ if $e_a^k, e_b^j, w \models \alpha$ for all e_a^k, e_b^j, w.

Satisfiability and validity ($\models$) are understood analogously.[15] Let $\mathcal{ONL}_n^{+1}, \ldots \mathcal{ONL}_n^{+t}$ be also defined analogously. Further, let axioms $\mathbf{A1}_n - \mathbf{A5}_n^t$ be defined for $\mathcal{ONL}_n^{+t}$. For instance, $\mathbf{A5}_n^t$ is defined for any i-objective $\neg\alpha \in \mathcal{ONL}_n^{+t-1}$ that is consistent with $\mathbf{A1}_n - \mathbf{A5}_n^{t-1}$. Then, the semantics above is characterized by the proof theory AX_n^{+t} defined (inductively) for $\mathcal{ONL}_n^{+t}$, consisting of AX_n^t ($\mathbf{A1}_n - \mathbf{A5}_n^t$, $\mathbf{MP}$, $\mathbf{NEC}$) with $\mathbf{NEC}_{Val}$ as an additional inference rule.

Lemma 32. *For all $\alpha \in \mathcal{ONL}_n^{+t}$, $AX_n^{+t} \vdash \alpha$ iff $\models \alpha$.*

The proof of this lemma, and those of the following theorems are given in the appendix. We proceed to show that $Sat(\alpha)$ is provable from AX'_n iff α is AX_n^{+t}-consistent.

Theorem 33. *For all $\alpha \in \mathcal{ONL}_n^{+t}$, $AX'_n \vdash Sat(\alpha)$ iff α is AX_n^{+t}-consistent.*

This allows us to show that AX'_n and AX_n^{+t} agree on provable sentences.

[15]Note that *Val* corresponds precisely to how validity is defined.

Theorem 34. *For all $\alpha \in \mathcal{ONL}_n^{+t}$, $AX'_n \vdash \alpha$ iff $AX_n^{+t} \vdash \alpha$.*

Lemma 35. *For all $\alpha \in \mathcal{ONL}_n^{+t}$, $\models \alpha$ iff α is valid in (Halpern and Lakemeyer 2001).*

Proof. AX'_n is sound and complete for (Halpern and Lakemeyer 2001), and AX_n^{+t} is sound and complete for $\models$. ■

Since it can be shown that every $\alpha \in \mathcal{ONL}_n^+$ is provably equivalent to some $\alpha' \in \mathcal{ONL}_n$ (Halpern and Lakemeyer 2001), we also obtain the following corollary.

Corollary 36. *For all $\alpha \in \mathcal{ONL}_n^t$, $\models \alpha$ iff α is valid in (Halpern and Lakemeyer 2001).*

Conclusions

This paper has the following new results. We have a first-order modal logic for multi-agent only-knowing that we show, for the first time, generalizes Levesque's semantics. Unlike all attempts so far, we neither make use of proof-theoretic notions of maximal consistency nor Kripke structures (Waaler and Solhaug 2005). The benefit is that the semantic proofs are straightforward, and we understand possible worlds precisely as Levesque meant. We then analyzed a propositional subset, and showed first that the axiom system from Lakemeyer (1993) is sound and complete for a restricted language. We used this result to devise a new proof theory that does not require us axiomatize any semantic notions (Halpern and Lakemeyer 2001). Our axiomatization was shown to be sound and complete for the semantics, and its use is straightforward on formulas involving the nesting of *at most* operators. In the process, we revisited the features of only-knowing and compared the semantical framework to other approaches. Its behavior seems to coincide with our intuitions, and it also captures a multi-agent generalization of Moore's AEL. Finally, although the axiomatization of Halpern and Lakemeyer (2001) is not natural, we showed that they essentially capture the desired properties of multi-agent only-knowing, but at much expense.

Acknowledgements

The authors would like to thank the reviewers for helpful suggestions and comments. The first author is supported by a DFG scholarship from the graduate school GK 643.

References

Ben-David, S., and Gafni, Y. 1989. All we believe fails in impossible worlds. *Manuscript*.

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

Halpern, J. Y., and Lakemeyer, G. 1995. Levesque's axiomatization of only knowing is incomplete. *Artif. Intell.* 74(2):381–387.

Halpern, J. Y., and Lakemeyer, G. 2001. Multi-agent only knowing. *J. Log. Comput.* 11(1):41–70.

Halpern, J., and Moses, Y. 1985. Towards a theory of knowledge and ignorance: preliminary report. *Logics and models of concurrent systems* 459–476.

Halpern, J. Y. 1993. Reasoning about only knowing with many agents. In *AAAI*, 655–661.

Hughes, G. E., and Cresswell, M. J. 1972. *An introduction to modal logic*. Methuen London.

Hughes, G. E., and Cresswell, M. J. 1984. *A companion to modal logic*. Routledge.

Lakemeyer, G., and Levesque, H. J. 2005. Only-knowing: taking it beyond autoepistemic reasoning. In *AAAI'05*, 633–638. AAAI Press.

Lakemeyer, G. 1993. All they know: A study in multi-agent autoepistemic reasoning. In *IJCAI-93*.

Levesque, H., and Lakemeyer, G. 2001. *The logic of knowledge bases*. The MIT Press.

Levesque, H. J. 1990. All I know: a study in autoepistemic logic. *Artif. Intell.* 42(2-3):263–309.

Moore, R. C. 1985. Semantical considerations on nonmonotonic logic. *Artif. Intell.* 25(1):75–94.

Rosati, R. 2000. On the decidability and complexity of reasoning about only knowing. *Artif. Intell.* 116(1-2):193–215.

Waaler, A., and Solhaug, B. 2005. Semantics for multi-agent only knowing: extended abstract. In *TARK*, 109–125.

Waaler, A. 2004. Consistency proofs for systems of multi-agent only knowing. In *Advances in Modal Logic*, 347–366.

Appendix

Lemma 32. *For all $\alpha \in \mathcal{ONL}_n^{+t}$, $AX_n^{+t} \vdash \alpha$ iff $\models \alpha$.*

Proof. The proof is via induction. Using Theorems 25 and 28 as the base cases in the induction, there is one additional step on the structure of formulas.

Soundness: The base case holds for formulas $\alpha \in \mathcal{ONL}_n^t$ for AX_n^t. Suppose now if $AX_n^t \vdash \alpha$, then $AX_n^{+t} \vdash \mathit{Val}(\alpha)$. But if $AX_n^t \vdash \alpha$ then (by induction hypothesis) at all models $e_a^k, e_b^j, w \models \alpha$, and so by the definition at all models $e_a^k, e_b^j, w \models \mathit{Val}(\alpha)$ or $\models \mathit{Val}(\alpha)$.

Completeness. For the base case, we know that if for all models $e_a^k, e_b^j, w \models \alpha$ then $AX_n^t \vdash \alpha$. Suppose $\models \mathit{Val}(\alpha)$, then by definition, for all models $e_a^k, e_b^j, w \models \alpha$ iff (by hypothesis) $AX_n^t \vdash \alpha$. So, $AX_n^{+t} \vdash \mathit{Val}(\alpha)$. ■

Theorem 33. *For all $\alpha \in \mathcal{ONL}_n^{+t}$, $AX'_n \vdash \mathit{Sat}(\alpha)$ iff α is AX_n^{+t}-consistent.*

Proof. It is helpful to have the following variant of Lemma 27 at hand, and a corollary thereof.

Lemma 37. *Suppose $\phi, \psi \in \mathcal{ONL}_n^{+t-1}$ are i-objective AX_n^{+t}-consistent formulas, and $\models \phi \vee \psi$. Then $\boldsymbol{L}_i\phi \wedge \boldsymbol{N}_i\psi$ is AX_n^{+t}-consistent.*

Proof. Suppose not. Then $AX_n^{+t} \vdash \neg(\boldsymbol{L}_a\phi \wedge \boldsymbol{N}_a\psi)$, that is $AX_n^{+t} \vdash \neg\boldsymbol{L}_a\phi \vee \neg\boldsymbol{N}_a\psi$. Then, by Lemma 32, $\models \neg\boldsymbol{L}_a\phi \vee \neg\boldsymbol{N}_a\psi$. Let $\mathcal{W}_\phi = \{w \mid w \models \phi\}$. Let $e_a^k = \mathcal{W}_\phi \times \mathbb{E}^{k-1}$ be a e^k for Alice. Then clearly, $e_a^k, \{\}, w \not\models \neg\boldsymbol{L}_a\phi$. It must be then that $e_a^k, \{\}, w \models \neg\boldsymbol{N}_a\psi$. Then there is some $\langle w', e_b^{k-1}\rangle \notin e_a^k$ such that $e_a^k, e_b^{k-1}, w' \models \neg\psi$. And clearly, for all $\langle w', e_b^{k-1}\rangle \notin e_a^k$, $e_a^k, e_b^{k-1}, w' \models \neg\phi$ (by construction). It follows that there is a $\langle w', e_b^{k-1}\rangle \notin e_a^k$ where e_a^k, $e_b^{k-1}, w' \models \neg(\phi \vee \psi)$, contradicting the validity of $\phi \vee \psi$. ■

Corollary 38. *Suppose $\alpha, \beta_1, \ldots \beta_k, \gamma, \delta_1, \ldots \delta_m \in \mathcal{ONL}_n^{+t-1}$, are i-objective AX_n^{+t}-consistent formulas, and $\models \alpha \vee \gamma$. Then $\boldsymbol{L}_i\alpha \wedge \neg\boldsymbol{L}_i\neg\beta_1 \ldots \wedge \neg\boldsymbol{L}_i\neg\beta_k \wedge \boldsymbol{N}_i\gamma \wedge \neg\boldsymbol{N}_i\delta_1 \ldots \wedge \neg\boldsymbol{N}_i\neg\delta_m$ is AX_n^{+t}-consistent.*

Returning to Theorem 33: Proof on the *length* of the derivative, using induction on t. Let α be a consistent propositional formula. Then, by $\mathbf{V2}$, $AX'_n \vdash \mathit{Sat}(\alpha)$. Since it is a consistent propositional formula, it is also AX_n^{+t}-consistent. Assume theorem holds for $\alpha \in \mathcal{ONL}_n^{+t-1}$. Suppose we have $\mathit{Sat}(\alpha \wedge \beta_k), \mathit{Sat}(\gamma \wedge \delta_m), \neg\mathit{Sat}(\neg(\alpha \vee \gamma)) \in \mathcal{ONL}_n^{+t-1}$ then by $\mathbf{V3}$, $AX'_n \vdash \mathit{Sat}(\boldsymbol{L}_i\alpha \wedge \neg\boldsymbol{L}_i\neg\beta_k \wedge \boldsymbol{N}_i\gamma \wedge \neg\boldsymbol{N}_i\neg\delta_m)$. By hypothesis $\alpha \wedge \beta_k$, $\gamma \wedge \delta_m$ are AX_n^{+t}-consistent. And $\neg(\alpha \vee \gamma)$ is not AX_n^{+t}-consistent, and so $AX_n^{+t} \vdash \alpha \vee \gamma$. By Lemma 32, $\models \alpha \vee \gamma$. Clearly, by Corollary 38, $\boldsymbol{L}_i\alpha \wedge \neg\boldsymbol{L}_i\neg\beta_k \wedge \boldsymbol{N}_i\gamma \wedge \neg\boldsymbol{N}_i\neg\delta_m$ is AX_n^{+t}-consistent. Finally, suppose that you have $\mathit{Sat}(\alpha)$ for some i-objective α and $\mathit{Sat}(\beta)$ for some i-subjective β, then by $\mathbf{V4}$, $AX'_n \vdash \mathit{Sat}(\alpha \wedge \beta)$. By induction hypothesis, α and β are AX_n^{+t}-consistent. By Lemma 32, α is satisfiable and β is satisfiable, and so is $\alpha \wedge \beta$. By Lemma 32, $\alpha \wedge \beta$ is AX_n^{+t}-consistent. The other direction is symmetric. ■

Theorem 34. $AX'_n \vdash \alpha$ *iff* $AX_n^{+t} \vdash \alpha$*, for $\alpha \in \mathcal{ONL}_n^{+t}$.*

Proof. Since axioms $\mathbf{A1}_n - \mathbf{A4}_n$, $\mathbf{MP}$, $\mathbf{NEC}$, $\mathbf{NEC}_{\mathit{Val}}$ are common to both, their use is not discussed. To show that $AX'_n \vdash \alpha \Rightarrow AX_n^{+t} \vdash \alpha$, suppose you had $\mathit{Sat}(\neg\alpha)$ for some i-objective $\alpha \in \mathcal{ONL}_n^{+t-1}$ then using $\mathbf{A5}'_n$, one could show that $\boldsymbol{N}_i\alpha \supset \neg\boldsymbol{L}_i\alpha$. From Theorem 33, we also know $\neg\alpha$ is AX_n^{+t-1}-consistent. Then, we can show $\boldsymbol{N}_i\alpha \supset \neg\boldsymbol{L}_i\alpha$ as well using $\mathbf{A5}_n^t$. $\mathbf{V2}, \mathbf{V3}, \mathbf{V4}$ follow immediately from Theorem 33. Assuming now that the proof holds for base cases, using $\mathbf{V1}$, if $AX'_n \vdash \mathit{Val}(\alpha)$ and $AX'_n \vdash \mathit{Val}(\alpha \supset \beta)$ then $AX'_n \vdash \mathit{Val}(\beta)$. Now, by induction hypothesis, $AX_n^{+t} \vdash \mathit{Val}(\alpha)$ iff by Lemma 32 $\models \mathit{Val}(\alpha)$, and so $\models \alpha$. Similarly, $\models \alpha \supset \beta$, and thus, $\models \beta$ and $\models \mathit{Val}(\beta)$ by the semantics. By Lemma 32, $AX_n^{+t} \vdash \mathit{Val}(\beta)$.

To show that $AX_n^{+t} \vdash \alpha \Rightarrow AX'_n \vdash \alpha$, suppose $\neg\alpha \in \mathcal{ONL}_n^{+t-1}$ is i-objective and AX_n^{t-1}-consistent, then one can prove $\boldsymbol{N}_i\alpha \supset \neg\boldsymbol{L}_i\alpha$. Now, $\neg\alpha$ is also AX_n^{+t}-consistent and by Theorem 33, $AX'_n \vdash \mathit{Sat}(\neg\alpha)$. Then we can prove $\boldsymbol{N}_i\alpha \supset \neg\boldsymbol{L}_i\alpha$, as desired. ■