Only-Knowing: Taking It Beyond Autoepistemic Reasoning

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

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

Abstract

The idea of only-knowing a collection of sentences has been previously shown to have a close connection with autoepistemic logic. Here we propose a more general account of only-knowing that captures not only autoepistemic logic but default logic as well. This allows us not only to study the properties of default logic in terms of an underlying model of belief, but also the relationship among different forms of nonmonotonic reasoning, all within a classical monotonic logic characterized semantically in terms of possible worlds.

Introduction

When considering a knowledge-based agent, it seems natural to think of the beliefs[1] of the agent to be those that follow from the assumption that its knowledge base (KB) is *all* that is believed. Levesque (1990) was the first to capture this notion explicitly in his logic of *only-knowing*. One of the advantages of this approach is that beliefs can be analyzed in terms of the valid sentences of a logic without requiring additional meta-logical notions like fixed points or partial orders. This is done by using two modal operators in the language, $\boldsymbol{K}$ for belief, and $\boldsymbol{O}$ for only knowing. For example, in the logic proposed by Levesque, the sentence

$$\boldsymbol{O}(P(a) \vee P(b)) \;\supset\; \boldsymbol{K}(\exists x.P(x) \wedge \neg\boldsymbol{K}P(x))$$

is valid, which can be read as "if we only know that $P(a)$ or $P(b)$, then we know that something is a P, but not what."

Levesque also showed that, when the KB itself is allowed to mention $\boldsymbol{K}$, then $\boldsymbol{O}$ captures the autoepistemic logic (AEL) proposed by Moore (1985), in the sense that the beliefs entailed by only-knowing KB are precisely those which are in all stable expansions of KB. This connection made it possible to study autoepistemic reasoning within a classical monotonic logic, leading, among other things, to an axiomatic characterization of the logic in the propositional case, and a first-order account that handles quantifying-in. Only-knowing has been studied and extended in various ways over the years (Levesque and Lakemeyer 2001; Halpern and Lakemeyer 2001; Waaler 2004), but in terms of nonmonotonic reasoning it has never gone beyond AEL.

[1] In this paper, we use the terms "knowledge" and "belief" interchangeably to mean belief.

In this paper, we broaden the scope of only-knowing to capture other forms of nonmonotonic reasoning, and in particular, the default logic (DL) proposed by Reiter (1980). We will propose a new modal logic which simultaneously captures DL, AEL, and a variant of AEL due to Konolige (1988). This will allow us not only to study the properties of DL in terms of an underlying model of belief, but also the relationship among these three different forms of nonmonotonic reasoning, all within a classical monotonic logic characterized semantically in terms of possible worlds.

This is not the first time AEL and DL have been compared in a modal setting. The first such attempt was made by Konolige (1988). As described by Reiter, a default rule $\alpha : \beta \,/\, \gamma$ has an intuitive reading of "if α is believed and it is consistent to believe β then infer that γ is true." Hence Konolige proposed translating the default rule into a sentence of AEL of the form

$$\boldsymbol{K}\alpha \wedge \boldsymbol{M}\beta \;\supset\; \gamma.$$

Here $\boldsymbol{M}$ is understood as the dual of $\boldsymbol{K}$ in the sense that $\boldsymbol{M}\beta$ stands for $\neg\boldsymbol{K}\neg\beta$. It turns out that this translation is not always faithful as there are cases where the AEL default admits extensions which are not DL extensions. The problem is, roughly, that in DL what is believed and what is consistent to believe are not duals. To account for this asymmetry, Lin and Shoham (1988) and later Lifschitz (1994) proposed nonmonotonic bimodal logics of belief with operators similar to $\boldsymbol{K}$ and $\boldsymbol{M}$, where the duality is given up. While the semantics of their logics is quite simple, the translation of an AEL default into their logic differs from the translation above, which complicates a direct comparison. Recently, Denecker *et al.* (2003) proposed a more uniform representation of defaults for both AEL and DL. Unfortunately, their semantics relies on meta-logical notions like fixed points in a certain lattice structure. Moreover, these approaches all require non-classical (that is, nonmonotonic) logics. Finally, while Amati et al. (1997) are able to capture DL within a modal logic, this comes at the price of defining fixed points in the language and a more complicated translation of defaults.

Here we propose a logic where we can always use the direct Konolige translation of a default rule, but where different forms of default reasoning correspond to different variants of only-knowing. To this end, instead of a single operator $\boldsymbol{O}$, we imagine three operators, $\boldsymbol{O}_M$ (for Moore), $\boldsymbol{O}_K$

(for Konolige) and $\boldsymbol{O_R}$ (for Reiter), all coexisting in a single first-order quantified language, which we call $\mathcal{O}_3\mathcal{L}$, together with the operators $\boldsymbol{K}$ and $\boldsymbol{M}$ (used primarily within beliefs). The $\boldsymbol{O_M}$ operator is precisely the $\boldsymbol{O}$ operator of Lakemeyer and Levesque and will turn out to be the most basic of the three. The other two are new. The semantics of $\mathcal{O}_3\mathcal{L}$ will be in terms of possible worlds and takes inspiration from the semantics in (Lin and Shoham 1988; Lifschitz 1994). Its simplicity will allow us to prove non-trivial properties beyond the handling of defaults such as the non-reducibility of belief in the first-order case.

The rest of the paper is organized as follows. In the next section, we present the formal details of the logic $\mathcal{O}_3\mathcal{L}$, its syntax and semantics. Then we discuss some of the properties of $\mathcal{O}_3\mathcal{L}$, and in particular the relationships among the three only knowing operators. Next, we turn our attention to defaults and state the main result of the paper, which is that $\mathcal{O}_3\mathcal{L}$ correctly captures the three forms of default reasoning. For reasons of simplicity and space, we restrict our attention to propositional defaults and leave a discussion of the first-order case to the conclusions.

Syntax and semantics

The symbols of the $\mathcal{O}_3\mathcal{L}$ language are the usual logical connectives, quantifiers, punctuation, variables, the equality symbol, predicates (of every arity), a countably infinite set of *standard names*, and the modal operators $\boldsymbol{M}$, $\boldsymbol{K}$, $\boldsymbol{O_M}$, $\boldsymbol{O_K}$, and $\boldsymbol{O_R}$. For simplicity, we omit constants and function symbols.[2] The *terms* of $\mathcal{O}_3\mathcal{L}$ are the variables and standard names. The *formulas* of $\mathcal{O}_3\mathcal{L}$ are defined by the following:

1. if $t_1, \ldots, t_k$ are terms and P is a predicate of arity k, then $P(t_1, \ldots, t_k)$ is an (atomic) formula;
2. if t_1 and t_2 are terms, then $(t_1 = t_2)$ is a formula;
3. if α and β are formulas and x is any variable, then $\neg\alpha$, $(\alpha\wedge\beta)$, and $\forall x.\alpha$ are formulas, as are the *modal* formulas, $\boldsymbol{M}\alpha$, $\boldsymbol{K}\alpha$, $\boldsymbol{O_M}\alpha$, $\boldsymbol{O_K}\alpha$, and $\boldsymbol{O_R}\alpha$.

As usual, we treat $(\alpha \vee \beta)$, $(\alpha \supset \beta)$, $(\alpha \equiv \beta)$, and $\exists x.\alpha$ as abbreviations. The notion of a free and bound variable is defined in the usual way, and α^x_n means α with all free occurrences of x replaced by n. We call a formula without free variables a *sentence*, and a formula of the form $P(n_1, \ldots, n_k)$, where the n_i are standard names, a *primitive sentence*. Formulas without modal operators are called *objective*, and those where all the predicates appear in the scope of a modal operator are called *subjective*. Formulas without $\boldsymbol{O_M}$, $\boldsymbol{O_K}$ and $\boldsymbol{O_R}$ are called *basic*. Basic formulas that only mention $\boldsymbol{K}$ are called $\boldsymbol{K}$*-basic* and those that mention only $\boldsymbol{M}$ are called $\boldsymbol{M}$*-basic*. In most of what we do below, the modal operators will be applied to basic sentences only.

The semantics of $\mathcal{O}_3\mathcal{L}$ builds on the semantics of $\mathcal{OL}$ from (Levesque and Lakemeyer 2001). The starting point is the notion of a *world* (or world state) which is a function from the primitive sentences to $\{0, 1\}$. We let W be the set of all worlds. An epistemic state in $\mathcal{OL}$ is any set of worlds. What we will do differently in $\mathcal{O}_3\mathcal{L}$ is to use *two* epistemic states, one to interpret formulas with $\boldsymbol{K}$, and one to interpret formulas with $\boldsymbol{M}$ (since, as we noted, there will be contexts where the two operators are not duals).

[2] The standard names can be thought of as constants that satisfy the unique name assumption and an infinitary version of domain closure.

Let w be a world, and e_1 and e_2 be epistemic states. We can define when a basic sentence α is *true* wrt e_1, e_2, and w, which we write as $e_1, e_2, w \models \alpha$, as follows:

1. $e_1, e_2, w \models P(n_1, \ldots, n_k)$ iff $w[P(n_1, \ldots, n_k)] = 1$;
2. $e_1, e_2, w \models (n_1 = n_2)$ iff n_1 and n_2 are the same standard name;
3. $e_1, e_2, w \models \neg\alpha$ iff $e_1, e_2, w \not\models \alpha$;
4. $e_1, e_2, w \models (\alpha\wedge\beta)$ iff $e_1, e_2, w \models \alpha$ and $e_1, e_2, w \models \beta$;
5. $e_1, e_2, w \models \forall x.\alpha$ iff $e_1, e_2, w \models \alpha^x_n$ for every standard name n;
6. $e_1, e_2, w \models \boldsymbol{K}\alpha$ iff $e_1, e_2, w' \models \alpha$ for every $w' \in e_1$;
7. $e_1, e_2, w \models \boldsymbol{M}\alpha$ iff $e_1, e_2, w' \models \alpha$ for some $w' \in e_2$.

Observe that when $e_1 = e_2$, $\boldsymbol{K}$ and $\boldsymbol{M}$ will behave like the usual duals. Next we define $\boldsymbol{O_M}$ to coincide with the $\boldsymbol{O}$ of (Levesque and Lakemeyer 2001):

8. $e_1, e_2, w \models \boldsymbol{O_M}\alpha$ iff for every $w' \in W$, $e_1, e_2, w' \models \alpha$ iff $w' \in e_1$.

This has the effect of replacing an "if" in the clause for $\boldsymbol{K}$ by an "iff". Finally, the definitions of $\boldsymbol{O_K}$ and $\boldsymbol{O_R}$ use this one:

9. $e_1, e_2, w \models \boldsymbol{O_K}\alpha$ iff for every e' such that $e_1 \subseteq e'$, $e', e', w \models \boldsymbol{O_M}\alpha$ iff $e' = e_1$;
10. $e_1, e_2, w \models \boldsymbol{O_R}\alpha$ iff for every e' such that $e_1 \subseteq e'$, $e', e_1, w \models \boldsymbol{O_M}\alpha$ iff $e' = e_1$;

Note that the definition of $\boldsymbol{O_K}$ and $\boldsymbol{O_R}$ differ only in one place: where $\boldsymbol{O_K}$ uses the e' for its second epistemic argument (thus keeping the two arguments identical), $\boldsymbol{O_R}$ uses the given e_1.

To complete the specification of the logic, we define $e, w \models \alpha$ to mean $e, e, w \models \alpha$, and we say that a sentence α is *valid* (which we write as $\models \alpha$) iff $e, w \models \alpha$ for every e and w. If α is objective, we often omit the e and write $w \models \alpha$; if α is subjective, we write $e \models \alpha$.

Properties of $\mathcal{O}_3\mathcal{L}$

It is not too hard to see that the subset of the language that mentions only the modal operators $\boldsymbol{K}$ and $\boldsymbol{O_M}$ coincides with (that is, has exactly the same valid sentences as) $\mathcal{OL}$ from (Levesque and Lakemeyer 2001). In particular, the $\boldsymbol{K}$-basic subset of $\mathcal{O}_3\mathcal{L}$ has precisely the valid sentences of the modal system $K45$ or *weak S5* (Chellas 1980). Since we have a fixed universe of discourse (the standard names), the Barcan Formula $\forall x.\boldsymbol{K}\alpha \supset \boldsymbol{K}\forall x.\alpha$ holds as well.

Moreover, if a sentence does not mention $\boldsymbol{O_R}$, then $\boldsymbol{M}$ and $\boldsymbol{K}$ are duals in the sense that we can replace $\boldsymbol{M}$ by $\neg\boldsymbol{K}\neg$.

Theorem 1: *Let α be a sentence not mentioning $\boldsymbol{O_R}$. Let α' be α with any number of occurrences of $\boldsymbol{M}$ replaced by $\neg\boldsymbol{K}\neg$. Then $\models \alpha \equiv \alpha'$.*

The proof is a simple induction argument on the structure of α using the fact that without occurrences of $\boldsymbol{O_R}$, the two epistemic states can be assumed to be identical in all cases and that $e, e, w \models \boldsymbol{M}\alpha$ iff $e, e, w \models \neg\boldsymbol{K}\neg\alpha$. Indeed, because of the way we have defined validity, $\boldsymbol{M}\alpha$ is logically equivalent to $\neg\boldsymbol{K}\neg\alpha$ as would normally be expected.

As the following example demonstrates, $\boldsymbol{K}$ and $\neg\boldsymbol{M}\neg$ behave differently in the context of an $\boldsymbol{O}_R$. Here and in other examples below, we often refer to the epistemic states $e_p = \{w \mid w \models p\}$ and $e_0 = W$, the set of all worlds.

Example 1:
$e_p \models \boldsymbol{O}_R(\neg\boldsymbol{M}\neg p \supset p)$, but $e_p \not\models \boldsymbol{O}_R(\boldsymbol{K}p \supset p)$.

Proof:
To prove that $e_p \models \boldsymbol{O}_R(\neg\boldsymbol{M}\neg p \supset p)$, it suffices to show that (1) $e_p, e_p \models \boldsymbol{O}_M(\neg\boldsymbol{M}\neg p \supset p)$ and that (2) for no proper superset e' of e_p, $e', e_p \models \boldsymbol{O}_M(\neg\boldsymbol{M}\neg p \supset p)$ holds. To see why (1) holds, let $w \in e_p$. Then $w \models p$ and hence $e_p, e_p, w \models \neg\boldsymbol{M}\neg p \supset p$. Now suppose $w \notin e_p$. Then $w \not\models p$ and, since $e_p, e_p \models \neg\boldsymbol{M}\neg p$, we have that $e_p, e_p, w \not\models \neg\boldsymbol{M}\neg p \supset p$. Therefore, $e_p, e_p \models \boldsymbol{O}_M(\neg\boldsymbol{M}\neg p \supset p)$. To show (2), let $e' \supsetneq e$. Then there is a $w' \in e'$ such that $w' \not\models p$. Then $e', e_p, w' \not\models \neg\boldsymbol{M}\neg p \supset p$ and, hence, $e', e_p \not\models \boldsymbol{O}_M(\neg\boldsymbol{M}\neg p \supset p)$.

Furthermore $e_p \not\models \boldsymbol{O}_R(\boldsymbol{K}p \supset p)$ follows from the fact that $e_0, e_p \models \boldsymbol{O}_M(\boldsymbol{K}p \supset p)$. (Note that $e_0 \models \neg\boldsymbol{K}p$.) ∎

Now let us consider the relationships among the three forms of only-knowing. It is easy to see that all three coincide on objective sentences.

Theorem 2: *If ϕ is an objective sentence, then we have that $\models \boldsymbol{O}_M\phi \equiv \boldsymbol{O}_K\phi$ and $\models \boldsymbol{O}_K\phi \equiv \boldsymbol{O}_R\phi$.*

This is so because there is a unique e such that $e \models \boldsymbol{O}_M\phi$, namely $e = \{w \mid w \models \phi\}$, and the second epistemic state is irrelevant for objective sentences.

From the definition of the $\boldsymbol{O}$-operators it also follows immediately that $\boldsymbol{O}_M$ is the the most basic of the three in the following sense:

Theorem 3: *For all sentences α,*
$\models (\boldsymbol{O}_K\alpha \supset \boldsymbol{O}_M\alpha)$ *and* $\models (\boldsymbol{O}_R\alpha \supset \boldsymbol{O}_M\alpha)$.

The converse fails in both cases. For let $\alpha = (\boldsymbol{K}p \supset p)$. Then $e_p \models \boldsymbol{O}_M\alpha$, yet $e_p \not\models \boldsymbol{O}_K\alpha$ because $e_0 \models \boldsymbol{O}_M\alpha$ and $e_0 \supsetneq e_p$. We already showed above that $e_p \not\models \boldsymbol{O}_R\alpha$.

From the definition of $\boldsymbol{O}_M$ and $\boldsymbol{O}_K$ it is not difficult to derive necessary and sufficient conditions for when the two modalities coincide:

Lemma 1: *For any epistemic state e and basic sentence α,*
$e \models \boldsymbol{O}_K\alpha$ *iff* $e \models \boldsymbol{O}_M\alpha$ *and for all* $e' \supsetneq e$, $e' \not\models \boldsymbol{O}_M\alpha$.

Theorem 4: $\models \boldsymbol{O}_M\alpha \equiv \boldsymbol{O}_K\alpha$ *iff*
for all e, if $e \models \boldsymbol{O}_M\alpha$ then for all $e' \supsetneq e$, $e' \not\models \boldsymbol{O}_M\alpha$.

As a special case of this theorem we get that $\boldsymbol{O}_M$ and $\boldsymbol{O}_K$ agree on α if there is a *unique* epistemic state that only-knows it. Note that α does not need to be objective. Formally:

Corollary 1: *Suppose there is a unique epistemic state e such that $e \models \boldsymbol{O}_M\alpha$. Then $\models \boldsymbol{O}_M\alpha \equiv \boldsymbol{O}_K\alpha$.*

In (Levesque and Lakemeyer 2001) such sentences are called *definite*. They show, for example, that for any standard name n, the sentence

$$P(n) \wedge \forall x.\neg\boldsymbol{K}P(x) \supset \neg P(x)$$

is definite. Furthermore, if α is this sentence, then

$$\models \boldsymbol{O}_M\alpha \supset \boldsymbol{K}(\forall x.P(x) \equiv (x = n)).$$

In other words, only-knowing α amounts to making the closed-world assumption for P.

Theorem 4 fails when we use $\boldsymbol{O}_R$ instead of $\boldsymbol{O}_K$. Indeed, it does not even hold in the case where there is a unique e such that $e \models \boldsymbol{O}_M\alpha$, as the following example demonstrates. Let α be the sentence

$$(\boldsymbol{K}p \supset p) \wedge (\boldsymbol{M}\neg p \supset p).$$

Since $\boldsymbol{O}_M$ treats $\boldsymbol{K}$ and $\boldsymbol{M}$ as duals, $\boldsymbol{O}_M\alpha$ is logically equivalent to $\boldsymbol{O}_M p$, and so e_p is the only epistemic state e such that $e \models \boldsymbol{O}_M\alpha$. But $\boldsymbol{O}_R$ treats $\boldsymbol{K}$ and $\boldsymbol{M}$ differently and in particular, $e_p \not\models \boldsymbol{O}_R\alpha$ because $e_0, e_p \models \boldsymbol{O}_M\alpha$.

However, $\boldsymbol{O}_R$ does reduce to $\boldsymbol{O}_K$ when $\boldsymbol{K}$ is the only modality in α, and to $\boldsymbol{O}_M$ when $\boldsymbol{M}$ is the only modality:[3]

Theorem 5: *For any $\boldsymbol{K}$-basic sentence α, $\models \boldsymbol{O}_R\alpha \equiv \boldsymbol{O}_K\alpha$, and for any $\boldsymbol{M}$-basic sentence α, $\models \boldsymbol{O}_R\alpha \equiv \boldsymbol{O}_M\alpha$.*

Proof: For the first part, recall that $\boldsymbol{O}_K$ and $\boldsymbol{O}_R$ differ only in one place, which concerns the interpretation of $\boldsymbol{M}$. A simple induction on α shows that for any e_1, e_2, e_3, w, and a $\boldsymbol{K}$-basic α, $e_1, e_2, w \models \alpha$ iff $e_1, e_3, w \models \alpha$. The equivalence of $\boldsymbol{O}_R\alpha$ and $\boldsymbol{O}_K\alpha$ then follows immediately from the definitions of the two operators. Turning now to the second part, the only-if direction is immediate because of Theorem 3. For the converse, suppose $e \models \boldsymbol{O}_M\alpha$. Then it suffices to show that for all $e' \supsetneq e$, $e', e \not\models \boldsymbol{O}_M\alpha$. As α does not mention $\boldsymbol{K}$, a simple induction shows that for any e_1, e_2, e_3, w, $e_1, e_2, w \models \alpha$ iff $e_3, e_2, w \models \alpha$. Now let $w \in e' - e$. By assumption, $e, e, w \not\models \alpha$. Therefore, $e', e, w \not\models \alpha$, from which $e', e \not\models \boldsymbol{O}_M\alpha$ follows. ∎

As a corollary of the last theorem we get that, in general, $\models \boldsymbol{O}_R\alpha \supset \boldsymbol{O}_K\alpha$ does not hold. For example, let α be the sentence $(\neg\boldsymbol{M}\neg p \supset p)$. Then $e_p \models \boldsymbol{O}_R\alpha$ while $e_p \not\models \boldsymbol{O}_K\alpha$.

In the propositional case it is well known that a basic sentence has finitely many stable expansions and that each stable expansion is uniquely characterized by its objective sentences. This result was recast in (Levesque and Lakemeyer 2001) in terms of only knowing:

Theorem 6: *[Levesque and Lakemeyer] Let α be a basic sentence without quantifiers. Then there is a set of objective sentences $\Phi = \{\phi_1, \ldots, \phi_n\}$ such that*

$$\models \boldsymbol{O}_M\alpha \equiv (\boldsymbol{O}_M\phi_1 \vee \ldots \vee \boldsymbol{O}_M\phi_n).$$

Together with Theorem 3 we immediately obtain

Theorem 7: *There are sets $\{\phi^R_1, \ldots, \phi^R_k\} \subseteq \Phi$ and $\{\phi^K_1, \ldots, \phi^K_m\} \subseteq \Phi$ such that*

1. $\models \boldsymbol{O}_R\alpha \equiv (\boldsymbol{O}_R\phi^R_1 \vee \ldots \vee \boldsymbol{O}_M\phi^R_k)$;
2. $\models \boldsymbol{O}_K\alpha \equiv (\boldsymbol{O}_K\phi^K_1 \vee \ldots \vee \boldsymbol{O}_M\phi^K_m)$.

[3] It is a happy coincidence that the first letters of the two names align with the two basic modalities.

This means that, in the propositional case, only knowing an arbitrary basic α is reducible to a sentence without nested modalities. It turns out that, in the first-order case, none of these reductions hold. This was already shown for $\boldsymbol{O}_M$ by Levesque and Lakemeyer using the following example.

Let ζ be the conjunction of the following eight sentences:

1. $\forall xyz.[R(x,y) \land R(y,z) \supset R(x,z)]$;
2. $\forall x.\neg R(x,x)$;
3. $\forall x.[\boldsymbol{K}P(x) \supset \exists y.R(x,y) \land \boldsymbol{K}P(y)]$;
4. $\forall x.[\boldsymbol{K}\neg P(x) \supset \exists y.R(x,y) \land \boldsymbol{K}\neg P(y)]$;
5. $\exists x.\boldsymbol{K}P(x) \land \exists x\boldsymbol{K}\neg P(x)$;
6. $\exists x.\neg\boldsymbol{K}P(x)$;
7. $\forall x.\boldsymbol{K}P(x) \supset P(x)$;
8. $\forall x.\boldsymbol{K}\neg P(x) \supset \neg P(x)$.

(1) and (2) say that R is transitive and irreflexive, respectively; (3) and (4) say that for every known instance (non-instance) of P, there is another one that is R related to it; (5) says that there are at least one known instance and non-instance of P; (6) says that there is something that is not known to be an instance of P; (7) and (8) say that every known instance of P is a P and every known non-instance is not.

Theorem 8: *[Levesque and Lakemeyer] $\boldsymbol{O}_M\zeta$ is not equivalent to any sentence without nested modalities.*

We remark that their proof (Levesque and Lakemeyer 2001), pp. 156–158, generalizes to the language of $\mathcal{O}_3\mathcal{L}$ since any occurrence of only knowing within α would be restricted to an objective formula, where the three versions of only knowing coincide by Theorem 2.

Theorem 9: *Neither $\boldsymbol{O}_K\zeta$ nor $\boldsymbol{O}_R\zeta$ is equivalent to a sentence without nested modalities.*

Proof: The proof for $\boldsymbol{O}_K\zeta$ is exactly the same as the one by Levesque and Lakemeyer for $\boldsymbol{O}_M$, except for one thing. We need to establish that there is an e such that $e \models \boldsymbol{O}_K\zeta$. For that we consider the following e. Let Ω be the set $\{\#1, \#3, \#5, \ldots\}$ and let us call a standard name *odd* if it is in Ω and *even* otherwise. Let e be the set of world states w which satisfy the following conditions:

a) w satisfies all of the following objective sentences:

$$\{P(\#1), \neg P(\#2), P(\#3), \neg P(\#4), \ldots.\};$$

b) w satisfies conjuncts (1) and (2) of ζ stating that R is transitive and irreflexive;

c) for every even n there are infinitely many even standard names m which are R-related to n, that is, for which $w \models R(n,m)$;

d) for every odd n there are infinitely many odd standard names which are R-related to n.

Levesque and Lakemeyer showed that $e \models \boldsymbol{O}_M\zeta$. It is not hard to show that for any $e' \supsetneq e$, $e' \not\models \boldsymbol{O}_M\zeta$. Hence $e \models \boldsymbol{O}_K\zeta$. Finally, turning to $\boldsymbol{O}_R\zeta$, since ζ is $\boldsymbol{K}$-basic, by Theorem 5, $\models \boldsymbol{O}_R\zeta \equiv \boldsymbol{O}_K\zeta$, and so the result is immediate. ∎

Defaults

Having examined the general properties of the logic $\mathcal{O}_3\mathcal{L}$, we now investigate how the logic behaves with respect to defaults and, in particular, how the characterizations of default reasoning due to Moore, Konolige, and Reiter are similar and different.

We begin with the notion of a default. For Reiter, these have the form $\alpha : \beta_1, \ldots, \beta_k \,/\, \gamma$ where, of course, the α, β_i and γ are all objective. We follow Konolige and represent a default like this with a basic formula

$$\boldsymbol{K}\alpha \land \boldsymbol{M}\beta_1 \land \ldots \land \boldsymbol{M}\beta_k \supset \gamma.$$

Reiter allowed open defaults, treated them as standing for all their ground instances, and then defined extensions only for ground formulas. Moore and Konolige only considered propositional formulas. We will follow them here and consider propositional theories (and hence closed defaults) only. So for the rest of this section, we restrict our attention to the propositional subset of $\mathcal{O}_3\mathcal{L}$. We will get back to first-order defaults at the end of the paper.

So given a default theory $\langle F, D\rangle$, where F is a finite set of objective sentences and D is a finite set of closed defaults, our representation of the default theory is the sentence $(\phi \land \delta)$ where ϕ is the conjunction over F, and δ is the conjunction of the basic sentences representing the defaults in D.

We assume that the notion of a *Reiter extension* of a default theory is the standard one (Reiter 1980). By a *Moore extension* of a default theory we mean a stable expansion as defined in (Moore 1985) of the formula $(\phi \land \delta)$ that represents the theory. By a *Konolige extension* of a default theory we mean a moderately grounded stable expansion as defined in (Konolige 1988) of the same formula. What we will prove is a correspondence between these notions and the sentences believed at an epistemic state. By the basic beliefs of e we mean the sentences believed at e that are basic, and the objective beliefs of e are those that are objective.

The main result of the paper is this:

Theorem 10: *Let α be $(\phi \land \delta)$, the representation of a default theory $\langle F, D\rangle$. Then*

1. *Γ is a Moore extension of $\langle F, D\rangle$ iff there is an e such that $e \models \boldsymbol{O}_M\alpha$ and Γ is the set of basic beliefs of e;*
2. *Γ is a Konolige extension of $\langle F, D\rangle$ iff there is an e such that $e \models \boldsymbol{O}_K\alpha$ and Γ is the set of basic beliefs of e;*
3. *Γ is a Reiter extension of $\langle F, D\rangle$ iff there is an e such that $e \models \boldsymbol{O}_R\alpha$ and Γ is the set of objective beliefs of e.*

Corollary 2: *Let α be the representation of a default theory $\langle F, D\rangle$, and let ψ be any objective sentence. Then*

1. *$\models (\boldsymbol{O}_M\alpha \supset \boldsymbol{K}\psi)$ iff ψ is an element of every Moore extension of $\langle F, D\rangle$;*
2. *$\models (\boldsymbol{O}_K\alpha \supset \boldsymbol{K}\psi)$ iff ψ is an element of every Konolige extension of $\langle F, D\rangle$;*
3. *$\models (\boldsymbol{O}_R\alpha \supset \boldsymbol{K}\psi)$ iff ψ is an element of every Reiter extension of $\langle F, D\rangle$.*

This shows that models of only knowing as we have defined them are in 1–1 correspondence with the syntactically defined extensions of default theories, for Moore, Konolige, and Reiter. Moreover, we have captured the three varieties of extensions within a single possible-world framework.

Before considering the proof of this theorem, observe that by using the theorem, many properties of default theories will now follow directly from general properties of belief. For example, the fact that every Reiter extension is also a Moore extension follows from our Theorem 3. The fact that the converse does not hold in general, but holds in the case of prerequisite-free default theories follows from Theorem 5.

To prove the theorem, we consider the three cases in turn. First, for $\boldsymbol{O}_M$, we use an existing result by Levesque (1990), who proves an exact correspondence between Moore's autoepistemic logic and $\boldsymbol{O}_M$.

Theorem 11: *Γ is a stable expansion of a basic α iff for some e, $e \models \boldsymbol{O}_M\alpha$ and Γ is the set of basic beliefs of e.*

The first part of the theorem then follows from the observation that $(\phi \wedge \delta)$ is basic. Next, to prove the correspondence with $\boldsymbol{O}_K$, we first observe that Γ is defined to be a moderately grounded stable expansion iff it is a stable expansion with a minimal set of objective facts. Then using Theorem 11, the second part of the main theorem follows from the fact that the objective beliefs of an epistemic state e are minimal when e is as large as possible, and then apply Lemma 1.

Finally, to prove the correspondence with $\boldsymbol{O}_R$, we first turn to the logic MBNF proposed by Lifschitz.

MBNF

Extending work by (Lin and Shoham 1988), Lifschitz proposed a bimodal nonmonotonic logic of minimal belief and negation as failure MBNF (Lifschitz 1994) whose semantics is closely related to $\boldsymbol{O}_R$. Here we focus on the propositional subset of the language and leave a discussion of the first-order version and its connection to $\mathcal{O}_3\mathcal{L}$ to an extended paper. The language of MBNF can be taken as the objective part of propositional $\mathcal{O}_3\mathcal{L}$ together with two modal operators $\boldsymbol{B}$ (for belief) and not (for negation as failure).

As in $\mathcal{O}_3\mathcal{L}$, the semantics is defined with respect to a world w and two sets of worlds e_1, e_2 for the operators $\boldsymbol{B}$ and not, respectively:

1. $e_1, e_2, w \models_M p$ iff $w[p] = 1$;
2. $e_1, e_2, w \models_M \neg\alpha$ iff $e_1, e_2, w \not\models_M \alpha$
3. $e_1, e_2, w \models_M (\alpha \wedge \beta)$ iff
   $e_1, e_2, w \models_M \alpha$ and $e_1, e_2, w \models_M \beta$;
4. $e_1, e_2, w \models_M \boldsymbol{B}\alpha$ iff $e_1, e_2, w' \models_M \alpha$ for all $w' \in e_1$;
5. $e_1, e_2, w \models_M \mathsf{not}\,\alpha$ iff
   $e_1, e_2, w' \not\models_M \alpha$ for some $w' \in e_2$.

A pair (e, w) is called an MBNF-model of a sentence α if $e, e, w \models_M \alpha$ and for all w' and e', if $e' \supsetneq e$ then $e', e, w' \not\models_M \alpha$. When α is subjective, we often omit the w and write $e_1, e_2 \models_M \alpha$ instead of $e_1, e_2, w \models_M \alpha$ and e instead of (e, w).

The definitions of $\boldsymbol{O}_R$ and MBNF-models seem very similar, especially when we consider MBNF-models of formulas of the form $\boldsymbol{B}\alpha$. In both cases, one of the epistemic states is allowed to vary while the other remains fixed. The main difference seems to be that $\boldsymbol{O}_R$ minimizes what is only-known and MBNF-models minimize what is believed. In the following we show that there are at least two important cases where the notions coincide, but that they differ in general.

To formally address this issue, we begin by defining a mapping * from basic $\mathcal{O}_3\mathcal{L}$ formulas into corresponding formulas of MBNF. In particular, let α^* be α with all occurrences of $\boldsymbol{K}$ replaced by $\boldsymbol{B}$ and $\boldsymbol{M}$ replaced by $\mathsf{not}\neg$.

The question we want to answer then is the following: given a basic α and an epistemic state e, is it the case that $e \models \boldsymbol{O}_R\alpha$ iff e is an MBNF-model of $\boldsymbol{B}\alpha^*$? The following lemma, which is easily proven by induction, is helpful in answering this question.

Lemma 2: $e_1, e_2, w \models \alpha$ *iff* $e_1, e_2, w \models_M \alpha^*$.

The first result then is that the answer to our question is *yes* if we restrict ourselves to $\boldsymbol{M}$-basic sentences:

Theorem 12: *Let α be $\boldsymbol{M}$-basic and e an epistemic state. Then $e \models \boldsymbol{O}_R\alpha$ iff e is an* MBNF-*model of $\boldsymbol{B}\alpha^*$.*

The proof actually follows from a result by Rosati (1999), who showed, roughly, that the beliefs of MBNF-models of $\boldsymbol{B}\alpha^*$ correspond exactly to the Moore extensions of α (where $\boldsymbol{M}$ is interpreted as the dual of belief). With this, the theorem follows as a corollary to Theorems 5 and 11.

For sentences that represent default theories we obtain an exact correspondence as well:

Theorem 13: *Let $(\phi \wedge \delta)$ be the representation of a default theory, and let e be any epistemic state. Then*
$e \models \boldsymbol{O}_R(\phi \wedge \delta)$ *iff e is an* MBNF-*model of $\boldsymbol{B}(\phi \wedge \delta^*)$.*

Proof: To prove the if direction, let e be an MBNF-model of $\boldsymbol{B}(\phi \wedge \delta^*)$. For simplicity we assume that every default δ_i^* has the form $\boldsymbol{B}\alpha_i \wedge \mathsf{not}\neg\beta_i \supset \gamma_i$. Suppose, without loss of generality, that $e, e \models_M \boldsymbol{B}\alpha_i \wedge \mathsf{not}\neg\beta_i$ for $1 \le i \le k$ and $e, e \not\models_M \boldsymbol{B}\alpha_j \wedge \mathsf{not}\neg\beta_j$ for $k+1 \le j \le n$ with $0 \le k \le n$. Then $e \models_M \boldsymbol{B}(\phi \wedge \gamma_1 \wedge \ldots \wedge \gamma_k)$. Let $e' = \{w \mid w \models \phi \wedge \gamma_1 \wedge \ldots \wedge \gamma_k\}$. Note that $e \subseteq e'$. We claim that $e', e \models_M \boldsymbol{B}(\phi \wedge \delta^*)$. Clearly, $e', e \models_M \boldsymbol{B}\phi$ and $e', e \models_M \boldsymbol{B}\gamma_i$ for $1 \le i \le k$. Hence $e', e \models_M \boldsymbol{B}(\phi \wedge \bigwedge_i^k \delta_i^*)$. By assumption, $e, e \not\models_M \boldsymbol{B}\alpha_j$ or $e, e \not\models_M \mathsf{not}\neg\beta_j$ for $k+1 \le j \le n$. If $e, e \not\models_M \boldsymbol{B}\alpha_j$ then $e', e \not\models_M \boldsymbol{B}\alpha_j$ since e' is a superset of e. If $e, e \not\models_M \mathsf{not}\neg\beta_j$ then $e', e \not\models_M \mathsf{not}\neg\beta_j$ as e' is irrelevant here. Therefore, $e', e \not\models_M \boldsymbol{B}\alpha_j \wedge \mathsf{not}\neg\beta_j$ for $k+1 \le j \le n$ and, hence, $e', e \models_M \boldsymbol{B}(\phi \wedge \delta^*)$. Since e is an MBNF-model of $\boldsymbol{B}(\phi \wedge \delta^*)$, we have $e' = e$. With that and Lemma 2 we also get that $e \models \boldsymbol{O}_M(\phi \wedge \delta)$. Since e is an MBNF-model of $\boldsymbol{B}(\phi \wedge \delta^*)$, we have that for all $e' \supsetneq e$, $e', e \not\models_M \boldsymbol{B}(\phi \wedge \delta^*)$. Hence $e', e \not\models \boldsymbol{K}(\phi \wedge \delta)$ by Lemma 2 and $e', e \not\models \boldsymbol{O}_M(\phi \wedge \delta)$, from which $e \models \boldsymbol{O}_R(\phi \wedge \delta)$ follows.

Conversely, suppose $e \models \boldsymbol{O}_R(\phi \wedge \delta)$. Then $e \models \boldsymbol{O}_M(\phi \wedge \delta)$ and, hence, $e \models \boldsymbol{K}(\phi \wedge \delta)$. By Lemma 2, $e, e \models_M \boldsymbol{B}(\phi \wedge \delta^*)$. Suppose e is not an MBNF-model of $\boldsymbol{B}(\phi \wedge \delta^*)$. Then there is an $e' \supsetneq e$ such that $e', e \models_M (\boldsymbol{B}\phi \wedge \delta^*)$. Then, by Lemma 2, $e', e \models \boldsymbol{K}(\phi \wedge \delta)$. Wlog, Let $e, e \models \boldsymbol{K}\alpha_i \wedge \boldsymbol{M}\beta_i$ for $1 \le i \le k$ and $e, e \not\models \boldsymbol{K}\alpha_j \wedge \boldsymbol{M}\beta_j$ for $k+1 \le j \le n$ with $0 \le k \le n$. Since $e, e \not\models \boldsymbol{K}\alpha_j \wedge \boldsymbol{M}\beta_j$ for $k+1 \le j \le n$, we have $e', e \not\models \boldsymbol{K}\alpha_j \wedge \boldsymbol{M}\beta_j$. Suppose, wlog, that $e', e \models \boldsymbol{K}\alpha_i \wedge \boldsymbol{M}\beta_i$ for $1 \le i \le l$ and $e', e \not\models \boldsymbol{K}\alpha_j \wedge \boldsymbol{M}\beta_j$ for $l+1 \le j \le k$ with $0 \le l \le k$. Then $e' \subseteq \{w \mid w \models \phi \wedge \gamma_1 \wedge \ldots \wedge \gamma_l\}$. Now we extend e' to a set e^* such that

$e^*, e \models \boldsymbol{O}_M(\phi \wedge \delta)$ to get a contradiction. Let $e^* = \{w \mid w \models \phi \wedge \gamma_1 \wedge \ldots \wedge \gamma_l\}$. Then for all $w \in e^*$, $e^*, e, w \models \delta_i$ for $1 \leq i \leq l$, $e^*, e \not\models \boldsymbol{K}\alpha_j \wedge \boldsymbol{M}\beta_j$ for $l+1 \leq j \leq n$. Hence $e^*, e, w \models \phi \wedge \delta$. Conversely, suppose $e^*, e, w \models \phi \wedge \delta$. Then $w \models \phi \wedge \gamma_1 \wedge \ldots \wedge \gamma_l$ and, hence, $w \in e$, from which $e^*, e \models \boldsymbol{O}_M(\phi \wedge \delta)$ follows, contradicting our assumption that $e \models \boldsymbol{O}_R(\phi \wedge \delta)$. ∎

In general, however, models of $\boldsymbol{O}_R$ and MBNF disagree. To see why, let α be $[(\boldsymbol{K}p \supset p) \wedge \boldsymbol{K}(p \vee q)]$. It is easy to show that $e \models \boldsymbol{O}_R\alpha$ iff $e = e_p$. On the other hand, the only epistemic state which is an MBNF-model of $\boldsymbol{B}\alpha^*$ is $e = \{w \mid w \models (p \vee q)\}$.

One way to interpret this discrepancy is to note that minimizing *beliefs*, as done in MBNF, differs from minimizing *only-knowing*. Our counterexample above mentions only $\boldsymbol{K}$, and so the translation to MBNF does not mention not. Such sentences in MBNF are called *positive*. It has been known since (Kaminski 1991) that MBNF restricted to positive subjective sentences coincides with the logic of minimal knowledge of (Halpern and Moses 1985). Hence the counterexample shows that $\boldsymbol{O}_R$ differs from that logic as well.

Turning to defaults specifically, for a default of the form $\alpha : \beta_1, \ldots, \beta_k \,/\, \gamma$, Lifschitz proposes the following translation into MBNF, adapted from (Lin and Shoham 1988):

$$\boldsymbol{B}\alpha \wedge \mathsf{not}\neg\beta_1 \wedge \ldots \wedge \mathsf{not}\neg\beta_k \supset \boldsymbol{B}\gamma.$$

Note the modal operator applied to the conclusion γ. Given a set of defaults D, let us denote by δ_M the conjunction of their translations of default rules. A default theory $\langle F, D\rangle$ is then represented as $\boldsymbol{B}\phi \wedge \delta_M$ in MBNF. Adapting a result by Lin and Shoham (1988), Lifschitz proved that this gives us a faithful embedding of default theories into MBNF:

Theorem 14: *[Lifschitz] Γ is a Reiter extension of $\langle F, D\rangle$ iff there is an e such that e is an* MBNF*-model of $\boldsymbol{B}\phi \wedge \delta_M$ and Γ is the set of objective beliefs of e.*

Let us call two formulas α and β equivalent in MBNF if for all e_1, e_2, w, we have that $e_1, e_2, w \models \alpha$ iff $e_1, e_2, w \models \beta$. It is easy to see that $\boldsymbol{B}\phi \wedge \delta_M$ is equivalent to $\boldsymbol{B}(\phi \wedge \delta^*)$ in MBNF. So Theorems 13 and 14 together prove the third and final part of our main Theorem 10, thus establishing that our embedding of default logic into $\mathcal{O}_3\mathcal{L}$ is indeed correct.

Conclusions

In this paper, we showed how DL and two variants of AEL could be embedded in a classical logic of only-knowing. While we considered propositional defaults only, $\mathcal{O}_3\mathcal{L}$ is clearly not limited to that. In fact, Theorem 11, which established the correspondence between $\boldsymbol{O}_M$ and AEL, was already proven for a generalized version of AEL with quantifying-in. Similarly, $\boldsymbol{O}_K$ applied to arbitrary basic sentences provides us with a natural generalization of moderately grounded AEL to the first-order case. So what about first-order default logic? As we remarked earlier, Reiter allowed open defaults such as $\mathit{Bird}(x) : \mathit{Fly}(x) \,/\, \mathit{Fly}(x)$, which are meant as shorthand for the set of all its ground instances. With $\mathcal{O}_3\mathcal{L}$ we have the further option of considering the universal closure of open defaults such as $\forall x. \boldsymbol{K}\mathit{Bird}(x) \wedge \boldsymbol{M}\mathit{Fly}(x) \supset \mathit{Fly}(x)$. Note that the conclusions from the set of all ground instances of a default are generally weaker than those from the universal closure, as there may be domain elements which are not referred to by any term. For this reason Lifschitz introduced a variant of Reiter's default logic with a fixed universe and with *names* in the language for each domain element (Lifschitz 1990). He then showed that this variant of default logic can be correctly embedded in a first-order extension of MBNF using the universal closure of open defaults as above. With that and the results of the previous section, it is not difficult to obtain a corresponding result for $\mathcal{O}_3\mathcal{L}$. The details are left to an extended version of this paper.

As for future work, given that we are working within a standard monotonic logic, there now seems to be a reasonable chance to arrive at an *axiomatic* characterization of DL and moderately grounded AEL, at least in the propositional case, as was done for AEL by Levesque.

References

G. Amati, L. C. Aiello, and F. Pirri, Definability and commonsense reasoning. *Artif. Intell.* **93**(1–2), 169–199, 1997.

B. Chellas, *Modal Logic: An Introduction*. Cambridge University Press, Cambridge, 1980.

M. Denecker, V. W. Marek, and M. Truszczynski, Uniform semantic treatment of default and autoepistemic logic. *Artificial Intelligence* **143**(1), 79–122, 2003.

Y. Moses and J. Y. Halpern, Towards a theory of knowledge and ignorance: preliminary report. In Apt, K. (Ed.), *Logic and Models of Concurrent Systems*, Springer-Verlag, 1985.

J. Y. Halpern and G. Lakemeyer, Multi-Agent Only Knowing. *Journal of Logic and Computation* **11**(1), 41-70, 2001.

M. Kaminski, Embedding a default system into nonmonotonic logics. *Fundamenta Informaticae* **14**, 345–354, 1991.

K. Konolige, On the relation between default logic and autoepistemic theories. *Artif. Intell.* **35**(3), 343–382, 1988.

H. J. Levesque, All I Know: A Study in Autoepistemic Logic. *Artificial Intelligence*, **42**, 263–309, 1990.

H. J. Levesque and G. Lakemeyer, *The Logic of Knowledge Bases*. MIT Press, 2001.

V. Lifschitz, On open defaults, in J. Lloyd (Ed.), *Proc. of the Symp. on Computational Logic*, Springer, 80–95, 1990.

V. Lifschitz, Minimal Belief and Negation as Failure, *Artificial Intelligence*, **70**(1–2), 53–72, 1994.

F. Lin and Y. Shoham, Epistemic Semantics for Fixpoint Nonmonotonic Logics. *Proceedings of TARK 1990,* Morgan Kaufmann, 1990.

R. C. Moore, Semantical considerations on nonmonotonic logic. *Artificial Intelligence* **25**, 75–94, 1985.

R. Reiter, A logic for default reasoning. *Artificial Intelligence* **13**(1–2), 81–132, 1980.

R. Rosati, Reasoning about minimal belief and negation as failure. *J. of Art. Intelligence Res.*, **11**, 277–300, 1999.

A. Waaler, Consistency proofs for systems of multi-agent only knowing. *Proc. of Advances in Modal Logic (AiML-2004)*, 2004.