The Truth about Defaults

Hector J. Levesque
Dept. of Computer Science
University of Toronto

August 2006

This is joint work with Gerhard Lakemeyer

Where's the truth?

> "To know p is to take p to be true"
>
> — *A well known KR textbook*

Surprisingly, much of the work in Knowledge Representation does not appeal directly to a notion of *truth*.

- in some cases, this is because the object of study is below the level of sentences, *e.g.* concepts in description logics
- in some cases, this is because the object of study involves sentences with a well accepted notion of truth, *e.g.* classical logic in commonsense reasoning, constraint satisfaction, planning, *etc.*

But new formalisms and reasoning methods are often proposed that bypass truth, and jump directly to *logical entailment*.

Default reasoning

We consider the sort of defaults first defined in [Reiter 80].

Formally: A default theory $T = \langle F, D \rangle$ where F is a set of ordinary sentences, and D is a set of (closed) *defaults* of the form

$$\frac{\alpha : \beta_1, \ldots, \beta_n}{\gamma}.$$

Informally: If α is believed, and each β_i can be consistently believed, then assume that γ is true.

We specify how to reason by specifying the *extensions* of T: the sets of sentences considered to be reasonable sets of beliefs, given T.

Credulous reasoning: be content with any extension.

Skeptical reasoning: find what is common to all extensions.

What are the extensions?

Two major definitions have been studied:

- Reiter extensions from [Reiter 80]

  Informally: these are minimal sets of sentences that contain the given facts, are closed under logical entailment, and have applied the defaults as much as possible.

- Moore extensions from [Moore 85] (aka "stable expansions")

  Informally: these are sets that start with the facts and the defaults represented as modal sentences, and are closed under logical entailment, as well as positive and negative introspection.

Later we will also consider a third variant from [Konolige 88].

Reasoning without truth

Neither definition appeals to the truth of sentences.

Although both definitions make reference to logical entailment,
this is only one part of a complex minimization.

So all analysis of default reasoning is done using extra-logical notions such as fixpoints, partial orders, closure operations, stable sets, *etc*.

We cannot look at a *semantic model* of a default theory and ask

- what is true,
- what is believed to be true,
- what is all that is believed to be true.

In this talk, we propose to remedy this.

Overview

- we observe that classical notions of truth and belief are too cumbersome to be used for this purpose
- we propose a first-order language called $\mathcal{O}_3\mathcal{L}$ that has a simpler notion of truth, belief, and all that is believed
- we show how to reconstruct within $\mathcal{O}_3\mathcal{L}$ the default reasoning of Moore and Reiter in truth-theoretic terms
- we show that the variant proposed by Konolige is a bridge between the two forms of default reasoning
- we present first steps towards an axiom system for this logic, but argue that it is unlikely to work in the first-order case
- we conclude with directions for future research

Truth vs. semantics

Q: Is it not enough to have a semantics?

A: In a sense, default logic already has a “semantics,” where the models are the extensions.

For KR purposes, we want a semantics that is truth-theoretic.

Q: So when is semantics truth-theoretic?

A: When it is compositional and stipulates a correspondence between

$\lor$ and “or”

$\neg$ and “not”

$\exists$ and “some” *etc.*

Acid test: do $\land$ and $\lor$ appear as duals? $\begin{cases} \text{yes, for truth} \\ \text{no, for entailment} \end{cases}$

Didn't Tarski and Kripke do this already?

The classical definition of truth [Tarski 35] and its possible-world extension for belief [Hintikka 62] (via Kripke) are too cumbersome!

Observe that the use of *Kripke structures* in the literature is almost always propositional, and that *Tarski structures* are rarely employed.

e.g. see [Lakemeyer & Levesque 04] for some of the difficulties in using standard Tarski semantics for the *situation calculus*.

To be workable, we need a definition of truth that supports

- mathematical induction over *sentences*, without having to deal with open formulas and elements of the domain;
- the ability to easily construct models that are *combinations* of other models, without having to deal with differing domains

The proposal

Following [Levesque & Lakemeyer 01], we will

1. use a first-order language with a simplified notion of truth: $\mathcal{L}$;
2. extend it with a simplified notion of belief: $\mathcal{KL}$;
3. further extend it to talk about all that is believed: $\mathcal{OL}$ and $\mathcal{O}_3\mathcal{L}$.

The languages use a fixed set of *standard names* as terms.

(For simplicity here, we omit any other function or constant symbols.)

Main idea of the semantics:

- $\forall x \alpha$ is true iff α^x_n is true for all standard names n.
- The models are *truth assignments*: $W = [Atoms \rightarrow \{0,1\}]$.

The semantics of $\mathcal{L}$

Let w be any element of W. We define what it means for an objective sentence α to be **true** wrt w, which we write as $w \models \alpha$ as follows:

1. $w \models P(n_1, \ldots, n_k)$ iff $w[P(n_1, \ldots, n_k)] = 1$;
2. $w \models (n_1 = n_2)$ iff n_1 and n_2 are the same standard name;
3. $w \models \neg\alpha$ iff $w \not\models \alpha$;
4. $w \models (\alpha \wedge \beta)$ iff $w \models \alpha$ and $w \models \beta$;
   $w \models (\alpha \vee \beta)$ iff $w \models \alpha$ or $w \models \beta$;
5. $w \models \forall x.\alpha$ iff $w \models \alpha^x_n$ for every standard name n;
   $w \models \exists x.\alpha$ iff $w \models \alpha^x_n$ for some standard name n;

We say that the sentence α is **valid**, which we write $\models \alpha$, when $w \models \alpha$ for every $w \in W$.

Is this the same as classical first-order logic?

Almost. In the Tarski definition of truth,

1. the $=$ symbol is just another binary predicate;
2. the domain of quantification is any non-empty set;
3. elements of the domain may be unnamed (no standard names).

The main difference is (2). For example, in $\mathcal{L}$ we have that

$$\models \exists x \exists y \exists z\,(x \neq y \wedge x \neq z \wedge y \neq z).$$

To deal with a finite domain in $\mathcal{L}$, *e.g.* to say that there are at most two "objects," we need to use a predicate:

$$\exists x \exists y \forall z\,(Obj(z) \supset z = x \vee z = y)$$

Truth and belief

The notion of a default from Reiter appeals to a concept of *belief*.

> Suppose all we are told about Tweety is that she is a bird.
>
> Then we believe $Bird(tweety)$, but we do not believe $\neg Fly(tweety)$ (nor do we believe $Fly(tweety)$).
>
> It will therefore be consistent to believe $Fly(tweety)$. **. . .**

We add to the language: $\mathbf{K}\alpha$, read "α is believed"
$\mathbf{M}\alpha$, read "it is consistent to believe α"

We characterize a state of belief with a *set of truth assignments* $e \subseteq W$:

- $e \models \mathbf{K}Bird(tweety)$ iff $w \models Bird(tweety)$ for every $w \in e$;
- $e \models \mathbf{M}Fly(tweety)$ iff $w \models Fly(tweety)$ for some $w \in e$.

The semantics of $\mathcal{KL}$

Let $w \in W$ and $e \subseteq W$. We define what it means for a basic sentence α to be **true** wrt e and w, which we write as $e, w \models \alpha$, as follows:

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

We say that α is **valid**, written $\models \alpha$, when $e, w \models \alpha$ for every e and w.

Is this the same as classical modal logic?

Almost. In the Kripke definition of modal truth,

1. each world gets to have its own set of *accessible* worlds;
2. each world has its own *domain* of quantification.

Our definition leads to the following *introspection* properties:

$$\models \mathbf{K}\alpha \supset \mathbf{K}\mathbf{K}\alpha$$

$$\models \neg\mathbf{K}\alpha \supset \mathbf{K}\neg\mathbf{K}\alpha$$

We also have the property of *belief generalization*:

$$\models \mathbf{K}\forall x\alpha \equiv \forall x\mathbf{K}\alpha.$$

We handle more / fewer objects with predicates: $\exists x.Obj(x) \wedge \mathbf{M}\neg Obj(x)$.

Defaults as sentences

We follow [Konolige 88] and interpret a (possibly open) Reiter default

$$\frac{\alpha : \beta_1, \ldots, \beta_n}{\gamma}$$

as the following basic sentence:

$$\forall \vec{x}.\ \mathbf{K}\alpha \wedge \mathbf{M}\beta_1 \wedge \cdots \wedge \mathbf{M}\beta_n \supset \gamma.$$

As an example, we will use T_0 to mean the conjunction of the following:

$$\left.\begin{array}{r} Bird(tweety) \\ Bird(chilly) \\ \neg Fly(chilly) \end{array}\right\} \quad \text{some objective facts}$$

$$\forall x. \mathbf{K} Bird(x) \wedge \mathbf{M} Fly(x) \supset Fly(x) \qquad \text{a default}$$

Getting Tweety off the ground

- Observe that T_0 can be *true* without $Fly(tweety)$ being true.

  Let e be W and w be such that $w[Bird(n)] = 1$ and $w[Fly(n)] = 0$.

- Observe that T_0 can be *believed* without $Fly(tweety)$ being believed.

  Let $e = \{w : w[Bird(n)] = 1$ and $w[Fly(n)] = 0$, for all $n\}$.
  Then $e \models \mathbf{K}\forall x Bird(x) \wedge \mathbf{K}\forall x \neg Fly(x)$.

- However, suppose that T_0 is *all that is believed*.

  Then intuitively, $e \not\models \mathbf{K}\neg Fly(tweety)$, and so $e \models \mathbf{KM}Fly(tweety)$.
  Also, $e \models \mathbf{K}Bird(tweety)$, and so $e \models \mathbf{KK}Bird(tweety)$.
  Since $e \models \mathbf{K}(\forall x.\mathbf{K}Bird(x) \wedge \mathbf{M}Fly(x) \supset Fly(x))$, the default,
  we have that $e \models \forall x.\mathbf{KK}Bird(x) \wedge \mathbf{KM}Fly(x) \supset \mathbf{K}Fly(x)$.
  So $e \models \mathbf{K}Fly(tweety)$.

In this case, it seems that $Fly(tweety)$ will be believed.

All that is believed

We add to the language: $\mathbf{O}\alpha$, read “α is all that is believed.”

Depending on how we characterize the truth of these sentences, we will obtain different treatments of defaults.

The simplest:

Suppose that ϕ is an objective sentence.

We have that $e \models \mathbf{K}\phi$ iff $e \subseteq \{w : w \models \phi\}$.

With additional information, we move to an $e' \subset e \subseteq \{w : w \models \phi\}$.

We say that ϕ is *all that is believed* in e iff $e = \{w : w \models \phi\}$.

More generally, $e \models \mathbf{O}\alpha$ iff

for every w', if $w' \in e$, then $e, w' \models \alpha$, and (α is believed)

for every w', if $e, w' \models \alpha$, then $w' \in e$. (nothing else is)

The semantics of $\mathcal{OL}$

Let $w \in W$ and $e \subseteq W$. We define when a sentence α is **true** wrt e and w, which we write as $e, w \models \alpha$, as follows:

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

Tweety reconsidered

We already saw that $\not\models T_0 \supset Fly(tweety)$,

and $\not\models \mathbf{K}T_0 \supset \mathbf{K}Fly(tweety)$.

We can now prove that $\models \mathbf{O}T_0 \supset \mathbf{K}Fly(tweety)$.

Proof: Suppose $e \models \mathbf{O}T_0$. We prove that $e \models \mathbf{M}Fly(tweety)$.
(The rest is then as before, leading to $e \models \mathbf{K}Fly(tweety)$.)

We have that if $e, w \models T_0$, then $w \in e$.

Let $w^*[\rho] = 1$, for all ρ, except $\rho = Fly(chilly)$.

Then $w^* \models Bird(tweety) \wedge Bird(chilly) \wedge \neg Fly(chilly)$.
Moreover, $e \models \neg\mathbf{M}Fly(chilly)$, so $e, w^* \models \forall x. \mathbf{K}Bird(x) \wedge \mathbf{M}Fly(x) \supset Fly(x)$.
Therefore, $e, w^* \models T_0$. So $w^* \in e$.

Since $w^*[Fly(tweety)] = 1$ and $w^* \in e$, we have $e \models \mathbf{M}Fly(tweety)$. ■

Default reasoning reconstructed

So the pattern for *skeptical* default reasoning is this:

> Given a default theory T, believe any α such that $\models \mathbf{O}T \supset \mathbf{K}\alpha$.

The pattern for *credulous* default reasoning is this:

> Given a default theory T, select an e such that $e \models \mathbf{O}T$, and believe any α such that $e \models \mathbf{K}\alpha$.

This version of $\mathbf{O}$ corresponds precisely to Moore’s autoepistemic logic:

> **Theorem [Levesque 90]:** E is a Moore extension of T iff for some e such that $e \models \mathbf{O}T$, $E = \{\alpha : e \models \mathbf{K}\alpha\}$.

What about the other forms of default reasoning?

$$\text{Replace } \mathbf{O} \text{ by } \begin{cases} \mathbf{O}_M & \text{Moore (= } \mathbf{O}\text{, as above)} \\ \mathbf{O}_K & \text{Konolige} \\ \mathbf{O}_R & \text{Reiter} \end{cases} \qquad \Rightarrow \qquad \mathcal{O}_3\mathcal{L}$$

From Moore to Konolige

Does $\mathbf{O}\alpha$ really capture the fact that α is all that is believed?

Consider $(\neg\mathbf{K}p \vee p)$. It has two Moore extensions:

Let $e_t = \{w \in W\}$. Then $e_t \models \mathbf{O}_M(\neg\mathbf{K}p \vee p)$.

Let $e_p = \{w : w \models p\}$. Then $e_p \models \mathbf{O}_M(\neg\mathbf{K}p \vee p)$.

But $e_p \subset e_t$, so e_p is not a *minimal* belief state.

E is a *Konolige extension* of T iff E is a Moore extension with a minimal set of objective beliefs [Konolige 88].

$e \models \mathbf{O}_K\alpha$ iff $e \models \mathbf{O}_M\alpha$ and
for all supersets e' of e, $e' \not\models \mathbf{O}_M\alpha$.

We will get that $e_t \models \mathbf{O}_K(\neg\mathbf{K}p \vee p)$, but $e_p \not\models \mathbf{O}_K(\neg\mathbf{K}p \vee p)$.

From Moore to Reiter

Consider δ which is the conjunction of the following two defaults:

$$\mathbf{K}p \wedge \mathbf{M}\text{TRUE} \supset p$$

$$\mathbf{K}\text{TRUE} \wedge \mathbf{M}\neg p \supset p.$$

Because $\mathbf{K}$ is equivalent to $\neg\mathbf{M}\neg$, we have that $\models \mathbf{O}_M\delta \equiv \mathbf{O}_M p$.

But in Reiter extensions, $\mathbf{K}$ and $\mathbf{M}$ are not duals [Lin and Shoham 90].

minimize beliefs $\Gamma(S)$, checking consistency *wrt a fixed S*.

$e \models \mathbf{O}_R\alpha$ iff $e \models \mathbf{O}_M\alpha$ and

for all supersets e' of e, $e', e \not\models \mathbf{O}_M\alpha$.

for $\mathbf{K}$ (e')

for $\mathbf{M}$ (e)

We will get that $\models \neg\mathbf{O}_R\delta$ (no Reiter extensions)

The semantics of $\mathcal{O}_3\mathcal{L}$ (1)

Let $w \in W$, $e \subseteq W$. We say that α is **true** wrt e and w, which we write as $e, w \models \alpha$, when $e, e, w \models \alpha$, where $e_1, e_2, w \models \alpha$ is defined by:

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. $e_1, e_2, w \models \mathbf{K}\alpha$ iff $e_1, e_2, w' \models \alpha$ for every $w' \in e_1$;
   $e_1, e_2, w \models \mathbf{M}\alpha$ iff $e_1, e_2, w' \models \alpha$ for some $w' \in e_2$.

Note that when $e_1 = e_2$, the $\mathbf{K}$ and $\mathbf{M}$ operators are the usual duals.

continued …

The semantics of $\mathcal{O}_3\mathcal{L}$ (2)

We continue the definition of truth for non-basic sentences:

7. $e_1, e_2, w \models \mathbf{O}_M\alpha$ iff
   for every $w' \in W$, $e_1, e_2, w' \models \alpha$ iff $w' \in e_1$.

8. $e_1, e_2, w \models \mathbf{O}_K\alpha$ iff for every e' such that $e_1 \subseteq e'$,
   $e', e', w \models \mathbf{O}_M\alpha$ iff $e' = e_1$;

9. $e_1, e_2, w \models \mathbf{O}_R\alpha$ iff for every e' such that $e_1 \subseteq e'$,
   $e', e_2, w \models \mathbf{O}_M\alpha$ iff $e' = e_1$;

Note that e_1 will differ from e_2 only in the context of an $\mathbf{O}_R$ operator, and that $\mathbf{O}_R$ differs from $\mathbf{O}_K$ only in this one small detail.

We say that the sentence α of $\mathcal{O}_3\mathcal{L}$ is **valid**, which we write $\models \alpha$, when $e, w \models \alpha$ for every $w \in W$ and $e \subseteq W$.

The main theorem

Theorem [Lakemeyer & Levesque 05]:

Let T be a closed default theory. Then

1. E is a Moore extension of T iff there is an e such that $e \models \mathbf{O}_M T$ and $E = \{$basic $\alpha : e \models \mathbf{K}\alpha\}$;
2. E is a Konolige extension of T iff there is an e such that $e \models \mathbf{O}_K T$ and $E = \{$basic $\alpha : e \models \mathbf{K}\alpha\}$;
3. E is a Reiter extension of T iff there is an e such that $e \models \mathbf{O}_R T$ and $E = \{$objective $\alpha : e \models \mathbf{K}\alpha\}$.

Corollary: Let T be a default theory and ψ be an objective sentence. Then ψ is an element of every Moore / Konolige / Reiter extension of T iff $\mathbf{K}\psi$ is logically entailed by $\mathbf{O}_M T$ / $\mathbf{O}_K T$ / $\mathbf{O}_R T$.

Is this the right account of defaults?

$\mathcal{O}_3\mathcal{L}$ captures the existing accounts of Moore, Konolige, and Reiter in a uniform truth-theoretic setting.

But each of these has faults and limitations.

Consider, for example, the treatment of *open defaults* by Reiter:

An open default stands for its set of ground instances.

From $Bird(favouritePet(oldestFriend(george)))$, we would derive by default $Fly(favouritePet(oldestFriend(george)))$.

So we do not need to know the identity of a bird to use the default.

But from $\exists x\,(OnBranch(x) \wedge Bird(x))$, we cannot use the default.
We could if had we Skolemized it.

However, if we are willing to Skolemize, then $(Bird(tweety) \vee Bird(spike))$ and $\exists x(Bird(x) \wedge (x = tweety \vee x = spike))$ will behave differently!

Towards an axiom system

Given that $\mathcal{O}_3\mathcal{L}$ is a *classical* logic, we can consider looking for a set axioms and rules of inference that generate the valid sentences.

$\Rightarrow$ step-by-step *derivations* of skeptical default reasoning

Instead of: $T_0 \longrightarrow \ldots$ nonmonotonic steps $\ldots \longrightarrow Fly(tweety)$

we have: $\mathbf{O}T_0 \longrightarrow \ldots$ classical monotonic steps $\ldots \longrightarrow \mathbf{K}Fly(tweety)$

We will develop a proof theory for $\mathcal{O}_3\mathcal{L}$ with the following restrictions:

1. only the propositional subset
2. no **O** operators within **K** or **M**
3. no nested **K** or **M** within **O**

} enough to express default theories

This will be done by handling $\mathbf{O}_M$, $\mathbf{O}_K$, and $\mathbf{O}_R$, in turn.

Axioms: Moore

A proof theory for $\mathbf{O}_M$ already exists, using another modal operator $\mathbf{N}$, where $e \models \mathbf{N}\alpha$ iff $e, w \models \alpha$ for all $w \notin e$. [Levesque 90]

Inference Rule: From α and $\alpha \supset \beta$, derive β.

Axioms: /* Let $\mathbf{L}$ stand for $\mathbf{K}$ or $\mathbf{N}$ */

1. The axioms of propositional logic
2. $\mathbf{L}\alpha$, where α is an instance of an axiom (1)
3. $\mathbf{L}(\alpha \supset \beta) \supset \mathbf{L}\alpha \supset \mathbf{L}\beta$
4. $\sigma \supset \mathbf{L}\sigma$, where σ is subjective
5. $\mathbf{M}\alpha \equiv \neg\mathbf{K}\neg\alpha$

(modal logic K45)

(two new axioms)

6. $\mathbf{O}_M\alpha \equiv (\mathbf{K}\alpha \wedge \mathbf{N}\neg\alpha)$
7. $(\mathbf{N}\neg\phi \supset \mathbf{M}\phi)$, where ϕ is any objective sentence such that $\not\models \neg\phi$.

Derivations for default reasoning

The heart of default reasoning is arriving at the conclusion $\mathbf{M}\phi$, for some objective ϕ, and then using ordinary modal logic from there.

For example, with Tweety, we start with $(F \wedge D)$, where F is the objective facts and D is the ground instances of the default:

$$\mathbf{K}Bird(chilly) \wedge \mathbf{M}Fly(chilly) \supset Fly(chilly) \quad \wedge$$
$$\mathbf{K}Bird(tweety) \wedge \mathbf{M}Fly(tweety) \supset Fly(tweety)$$

The form of the derivation starting from $\mathbf{O}_M(F \wedge D)$ is then:

- use Axiom 6 to get $\mathbf{N}\neg(F \wedge D)$;
- use ordinary modal logic (K45) to derive $\mathbf{N}\neg(F \wedge Fly(tweety))$;
- use Axiom 7 to get $\mathbf{M}(F \wedge Fly(tweety))$, since $F \not\models \neg Fly(tweety)$;
- use ordinary modal logic to derive $\mathbf{M}Fly(tweety)$.

Tweety flies again

1. $\mathbf{O}_M(F \wedge D)$	Assumption.
2. $\mathbf{K}(F \wedge D)$	1; defn. of $\mathbf{O}_M$ (Axiom 6).
3. $\mathbf{K}\neg Fly(tweety) \vee \mathbf{K}Fly(tweety)$	2; *K45*.
4. $\mathbf{N}\neg\mathbf{M}Fly(chilly)$	2; *K45*.
5. $\mathbf{N}\neg(F \wedge D)$	1; defn. of $\mathbf{O}_M$ (Axiom 6).
6. $\mathbf{N}\neg(F \wedge Fly(tweety))$	$4, 5$; *K45*.
7. $\mathbf{M}(F \wedge Fly(tweety))$	6; **N** vs. **M** (Axiom 7).
8. $\mathbf{M}Fly(tweety)$	7; *K45*.
9. $\mathbf{K}Fly(tweety)$	$3, 8$; *K45*.

Axioms: Konolige

We use the following result from [Levesque & Lakemeyer 01]:

> **Theorem:** Let α be a basic sentence without quantifiers. Then there is a set of objective sentences $\{\phi_1, \ldots, \phi_n\}$ such that
>
> $$\models \mathbf{O}_M\alpha \equiv (\mathbf{O}_M\phi_1 \vee \ldots \vee \mathbf{O}_M\phi_n).$$

Then $\mathbf{O}_K\alpha$ is equivalent to the disjunction of the *minimal* of the $\mathbf{O}_M\phi_i$.

Axiom: $\mathbf{O}_K\alpha \supset \mathbf{O}_M\alpha$

Inference Rules:

From: $(\mathbf{O}_M\psi \supset \mathbf{O}_M\alpha)$, $(\mathbf{O}_M\phi \supset \mathbf{O}_M\alpha)$, $(\mathbf{O}_M\psi \supset \mathbf{K}\phi)$, $(\mathbf{O}_M\phi \supset \neg\mathbf{K}\psi)$,
derive: $(\mathbf{K}\psi \supset \neg\mathbf{O}_K\alpha)$. ($\psi$ is not minimal)

From: $(\mathbf{O}_M\alpha \supset \mathbf{O}_M\psi \vee \bigvee \mathbf{O}_M\phi_i)$, $(\mathbf{O}_M\psi \supset \mathbf{O}_M\alpha)$, $(\mathbf{O}_M\psi \supset \bigwedge \neg\mathbf{K}\phi_i)$,
derive: $(\mathbf{O}_M\psi \supset \mathbf{O}_K\alpha)$. ($\psi$ is minimal)

Axioms: Reiter

Using ideas from [Lifschitz 94] and [Denecker *et al* 03], we observe that $\mathbf{O}_R$ behaves like $\mathbf{O}_K$ except that it holds the $\mathbf{M}$ fixed.

Axioms:

1. $\mathbf{O}_R\alpha \equiv \mathbf{O}_K\alpha,$ if α has no $\mathbf{M}$ operators
2. $\mathbf{M}\phi \supset (\mathbf{O}_R\alpha \equiv \mathbf{O}_R\alpha'),$ where α' is α with $\mathbf{M}\phi$ replaced by TRUE
3. $\neg\mathbf{M}\phi \supset (\mathbf{O}_R\alpha \equiv \mathbf{O}_R\alpha'),$ where α' is α with $\mathbf{M}\phi$ replaced by FALSE

These axioms allow us to replace every $\mathbf{M}\phi$ in α either by TRUE or by FALSE as appropriate and then to use the Konolige version.

Theorem [Lakemeyer & Levesque 06]:

Let α be any sentence of $\mathcal{O}_3\mathcal{L}$, subject to the restrictions noted.

Then α is valid iff α is derivable.

The first-order case

Our *semantic theory* of default reasoning works even when the defaults are quantified, *e.g.* $\models \mathbf{O}_R T_0 \supset \mathbf{K}Fly(tweety)$.

However, it is unlikely that any *proof theory* will work in this case.

Consider the following example:

$D =$ one normal, prerequisite-free default: $\forall x. \mathbf{M}\neg Ab(x) \supset \neg Ab(x)$.

$F =$ the following objective facts:

$$\left.\begin{array}{l} \forall x. \neg R(x,x) \\ \forall x,y,z. R(x,y) \wedge R(y,z) \supset R(x,z) \\ \exists x. Ab(x) \\ \forall x. Ab(x) \supset \exists y. R(x,y) \wedge Ab(y) \end{array}\right\} \begin{array}{l}\text{there are}\\ \text{infinitely}\\ \text{many}\\ \text{abnormalities}\end{array}$$

Then $\models \neg\mathbf{O}_M(F \wedge D)$, and so $\models \neg\mathbf{O}_R(F \wedge D)$: no extensions.

Summary

It is possible to consider default reasoning from the standpoint of *truth*:

we can look at a model of a default theory T and ask

- what is true,
- what is believed to be true,
- what is all that is believed to be true.

With a few minor adjustments, the classical versions of truth and belief can be made workable, even in the quantified case.

The exercise reveals interesting connections among the versions of default reasoning proposed by Moore, Konolige, and Reiter.

By formulating these three accounts within a monotonic logic of belief, we can also get sentence-by-sentence derivations that correspond precisely to each form of default reasoning.

Future work

- better axioms for Konolige
- use of quantified defaults
- relationship to circumscription
- getting defaults right!
- other areas where truth might help
  - answer set programming?

THE END

References:

1. [Denecker, Marek & Truszczynski 03] in *Artificial Intelligence* **143**, 2003.
2. [Halpern & Lakemeyer 95] in *Artificial Intelligence* **74**, 1995.
3. [Hintikka 62] *Knowledge and Belief*, Cornell University Press, 1962.
4. [Konolige 88] in *Artificial Intelligence* **35**, 1988.
5. [Lakemeyer & Levesque 04] in *Proc. KR-04*, 2004.
6. [Lakemeyer & Levesque 05] in *Proc. AAAI-05*, 2005.
7. [Lakemeyer & Levesque 06] in *Proc. AAAI-06*, 2006.
8. [Levesque 90] in *Artificial Intelligence* **42**, 1990.
9. [Levesque & Lakemeyer 01] *The Logic of Knowledge Bases*, MIT Press, 2001.
10. [Lifschitz 94] in *Artificial Intelligence* **70**, 1994.
11. [Lin & Shoham 90] in *Proc. TARK-1990*, 1990.
12. [Moore 85] in *Artificial Intelligence* **25**, 1985.
13. [Reiter 80] in *Artificial Intelligence* **13**, 1980.
14. [Tarski 35] in *Studia Philosophica* **1**, 1935.