MULTI-AGENT ONLY-KNOWING

Gerhard Lakemeyer
Computer Science,
RWTH Aachen University
Germany

AI, Logic, and Epistemic Planning, Copenhagen
October 3, 2013

Joint work with Vaishak Belle
Contents of this talk

- Only-knowing
- The single-agent case
- The multi-agent case
Levesque (1990) introduced the logic of only-knowing to capture the beliefs of a knowledge-base. (Other variants such as Halpern & Moses, Ben-David & Gafni, Waaler not discussed here.)

**EXAMPLE**

If *all I know* is that the father of George is a teacher, then
- I know that someone is a teacher;
- but not who the teacher is.

**Note:** Does not work if *all I know* is replaced by *I know*
- Need to express that nothing else is known.
The Single-Agent Case

Levesque considered only the single-agent case.

- Compelling (i.e. simple) model-theoretic account
  - A possible-worlds framework for a first-order language
  - An epistemic state is simply a set of worlds
Levesque considered only the single-agent case.
- Compelling (i.e. simple) model-theoretic account
  - A possible-worlds framework for a first-order language
  - An epistemic state is simply a set of worlds
- Compelling proof-theoretic account
  (for the propositional fragment)
  - K45 + 2 extra axioms
Levesque showed only-knowing captures Moore’s Autoepistemic Logic:

- beliefs that follow from only-knowing facts and defaults are precisely those contained in all autoepistemic expansions

**Example**

If *all I know* is that Tweety is a bird and that birds fly unless known otherwise, then I believe that Tweety flies.
The Language $\text{ONLY}$

$\text{ONLY}$ is a first-order language with $=$

- infinitely many standard names $n_1, n_2, \ldots$
syntactically like constants, serve as the fixed domain
of discourse (rigid designators);
- variables $x, y, z, \ldots$
- predicate symbols of every arity;
- the usual logical connectives and quantifiers: $\land, \neg, \forall$;
- modal operators:
  - $K\alpha$ "at least" $\alpha$ is believed.
  - $N\alpha$ "at most" $\alpha$ is believed to be false;
- Only-knowing:
  \[ O\alpha \equiv K\alpha \land N\neg\alpha \]
Levesque’s Semantics

Primitive Formula = predicate with standard names as args.

A world $w$ is set of primitive formulas
An epistemic state $e$ is a set of worlds.

- $e, w \models P(\bar{n})$ iff $P(\bar{n}) \in w$;
- $e, w \models \forall x. \alpha$ iff $e, w \models \alpha^x_n$ for all $n$
- $e, w \models K\alpha$ iff for all $w' \in e$, $e, w' \models \alpha$;
- $e, w \models N\alpha$ iff for all $w' \not\in e$, $e, w' \models \alpha$
Levesque’s Semantics

Primitive Formula = predicate with standard names as args.

A world $w$ is set of primitive formulas
An epistemic state $e$ is a set of worlds.

$\triangleright e, w \models P(\vec{n})$ iff $P(\vec{n}) \in w$;
$\triangleright e, w \models \forall x.\alpha$ iff $e, w \models \alpha^x_n$ for all $n$
$\triangleright e, w \models K\alpha$ iff for all $w' \in e$, $e, w' \models \alpha$;
$\triangleright e, w \models N\alpha$ iff for all $w' \notin e$, $e, w' \models \alpha$
$\triangleright (e, w \models O\alpha$ iff for all $w'$, $w' \in e$ iff $e, w' \models \alpha)$
Axioms (propositional)

objective: non-modal formulas
subjective: all predicates within a modal

Let $L$ stand for both $K$ and $N$:

1. Axioms of propositional logic.
2. $L(\alpha \supset \beta) \supset (L\alpha \supset L\beta)$.
3. $\sigma \supset L\sigma$, where $\sigma$ is subjective.
4. The $N$ vs. $K$ axiom: $(N\phi \supset \neg K\phi)$, where $\neg \phi$ is consistent and objective.
5. $O\alpha \equiv (K\alpha \land N\neg \alpha)$.
Axioms (propositional)

objective: non-modal formulas
subjective: all predicates within a modal

Let $L$ stand for both $K$ and $N$:

1. Axioms of propositional logic.
2. $L(\alpha \supset \beta) \supset L\alpha \supset L\beta$.
3. $\sigma \supset L\sigma$, where $\sigma$ is subjective.
4. The $N$ vs. $K$ axiom:
   $(N\phi \supset \neg K\phi)$, where $\neg \phi$ is consistent and objective;
5. $O\alpha \equiv (K\alpha \land N\neg\alpha)$.
6. Inference rules:
   Modus ponens and Necessitation (for $K$ and $N$)
Intuitively, only-knowing for many agents seems easy:

**EXAMPLE**

If Alice believes that all that Bob knows is that birds normally fly and that Tweety is a bird, then Alice believes that Bob believes that Tweety flies.
Intuitively, only-knowing for many agents seems easy:

**EXAMPLE**

If Alice believes that all that Bob knows is that birds normally fly and that Tweety is a bird, then Alice believes that Bob believes that Tweety flies.

But technically things were surprisingly cumbersome! The problem lies in the complexity in what agents consider possible:

- For a single agent possibilities are just worlds.
- For many agents possibilities include other agents beliefs.
- The problem is that it is not so clear how to come up with models that contain all possibilities.
Some Previous Attempts

▶ (Lakemeyer 1993) uses the K45\textsubscript{n}-canonical model
  ▶ Yet, certain types of epistemic states cannot be constructed
    \( \neg O_a \neg O_b p \) is valid

▶ (Halpern 1993) proposes a tree approach
  ▶ Modalities do not interact in an intuitive manner

▶ In (Halpern and Lakemeyer 2001), a solution is proposed. But
  ▶ again uses canonical models
  ▶ proof theory needs to axiomatize validity
The Logic $\mathcal{ONL}_n$

- $\mathcal{ONL}_n \doteq \text{multi-agent version of } \mathcal{ONL}$
  - Here, only for $a$ and $b$ ($K_a$, $K_b$, $N_a$, $N_b$)
- **depth**: alternating nesting of modalities
  - a notion of $a$-depth and $b$-depth

**Example ($a$-depth)**

- $p$: 1
- $K_a p$: 1
- $K_b p$: 2
- $K_a K_b p$: 2

- $a$-objective: formulas not in scope of $K_a$ or $N_a$:
  - $p \land K_b p$ is $a$-objective,
  - $p \land K_a p$ is not.
Beyond Sets of Worlds

- Alice’s epistemic state is again a set of states of affairs, but where a state of affairs consists of a world and Bob’s epistemic state.
- Similarly, Bob’s epistemic state is again a set of affairs where a state of affairs consists of a world and Alice’s epistemic state (that determines her beliefs at this state).
- To be well-defined, we can do this only to some finite depth.
- For formulas of $a$-depth $k$ and $b$-depth $j$, it is sufficient to look at an epistemic state for Alice of depth $k$ and an epistemic state for Bob of depth $j$. 
Define an epistemic state for Alice as a set of pairs
- $e^1_a = \{\langle w, \{\} \rangle, \langle w', \{\} \rangle, \ldots \}$ (for formulas of $a$-depth 1)
- $e^k_a = \{\langle w, e^{k-1}_b \rangle, \ldots \}$

Similarly, an epistemic state for Bob
- $e^1_b = \{\langle w, \{\} \rangle, \ldots \}$
- $e^j_b = \{\langle w, e^{j-1}_a \rangle, \ldots \}$

$(k, j)$-model $= \langle e^k_a, e^j_b, w \rangle$.

Given a formula of $a$-depth $k$ and $b$-depth $j$
- $e^k_a, e^j_b, w \models K_a \alpha$ iff for all
  $\langle w', e^{k-1}_b \rangle \in e^k_a, e^k_a, e^{k-1}_b, w' \models \alpha$
- $e^k_a, e^j_b, w \models N_a \alpha$ iff for all
  $\langle w', e^{k-1}_b \rangle \not\in e^k_a, e^k_a, e^{k-1}_b, w' \models \alpha$

$O_a \alpha \equiv K_a \alpha \land N_a \neg \alpha$. 
A formula of \( a \)-depth \( k \) and \( b \)-depth \( j \) is valid if it is true at all \((k', j')\)-models for \( k' \geq k, j' \geq j \).

- K45\(_n\) (for \( K_i \) and \( N_i \))
  - \( K_a(\alpha \supset \beta) \supset K_a\alpha \supset K_a\beta \)
  - \( K_a\alpha \supset K_aK_a\alpha \)
  - \( \neg K_a\alpha \supset K_a\neg K_a\alpha \)

- Mutual introspection:
  - \( K_a\alpha \supset N_aK_a\alpha \)

- Barcan formula
  - \( \forall x \ K_a\alpha \supset K_a\forall x \alpha \)
Examples of Only-Knowing

\[ O_a(p \land O_b p) \] entails

\[ K_a p \]
\[ K_a K_b p \]
\[ K_a \neg K_b K_a p \]
\[ \text{but not } K_a K_b \neg K_a p \]
Examples of Only-Knowing

- $O_a(p \land O_bp)$ entails
  - $K_a p$
  - $K_a K_b p$
  - $K_a \neg K_b K_a p$
  - but not $K_a K_b \neg K_a p$

- $K_a O_b \exists x \ T(x)$ entails
  - $K_a K_b (\exists x \ T(x) \land \neg K_b T(x))$
Examples of Only-Knowing

- $O_a(p \land O_b p)$ entails
  - $K_a p$
  - $K_a K_b p$
  - $K_a \neg K_b K_a p$
  - but not $K_a K_b \neg K_a p$

- $K_a O_b \exists x \ T(x)$ entails
  - $K_a K_b (\exists x \ T(x) \land \neg K_b T(x))$

- Let $\delta_b = \forall x. K_b Bird(x) \land \neg K_b \neg Fly(x) \supset Fly(x)$
  - $K_a O_b (\delta_b \land Bird(tweety))$ entails
    - $K_a K_b Fly(tweety)$. 
Proof Theory for Many Agents

- K45_n + Axiom defining $O_i$
  + new version of the $N$ vs $K$ axiom.
- Recall, in the single-agent case:
  $(N\phi \supset \neg K\phi)$, where $\neg \phi$ is consistent and objective.
- Two things to generalize here:
  - Objective formulas to $i$-objective formulas
  - But generalizing the notion of consistency of $i$-objective formulas is circular!
**$N_i$ vs. $K_i$ Axioms**

**Idea:** break the circularity by considering a hierarchy of sub-languages based on the nesting of $N_i$.

- Define a family of languages
  - Let $\mathcal{ONL}_n^1 \equiv$ no $N_j$ in the scope of $K_i, N_i (i \neq j)$
  - Let $\mathcal{ONL}_n^{t+1}$ formed from $\mathcal{ONL}_n^t, K_i\alpha$ and $N_i\alpha$ for all $\alpha \in \mathcal{ONL}_n^t$
**N_i vs. K_i Axioms**

**Idea:** break the circularity by considering a hierarchy of sub-languages based on the nesting of N_i.

- Define a family of languages
  - Let ONL_n^1 = no N_j in the scope of K_i, N_i (i ≠ j)
  - Let ONL_n^{t+1} formed from ONL_n^t, K_iα and N_iα for all α ∈ ONL_n^t

- Proof theory is K45_n + Def. of O_i +
  - A_n^1 \( N_iα \supset \neg K_iα \) if \( \neg α \) is a K45_n-consistent i-objective formula
  - A_n^{t+1} \( N_iα \supset \neg K_iα, \) if \( \neg α \in ONL_n^t, \) is i-objective and consistent wrt. K45_n, A_n^1 − A_n^t

**Theorem**

For all \( α \in ONL_n^t, \) \( \models α \) iff Axioms^t ⊢ α
Conclusions

- First-order modal logic for multi-agent only-knowing
- Faithfully generalizes intuitions of Levesque’s logic
- Semantics not based on Kripke structures or canonical models, and thus avoids some problems of previous approaches
- In other work, we incorporated this notion of only-knowing into a mult-agent variant of the situation calculus for reasoning about knowledge and action.