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

Only-Knowing

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

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
- Compelling proof-theoretic account (for the propositional fragment)
  - K45 + 2 extra axioms

Connection with Default Reasoning

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 $\mathcal{ONL}$

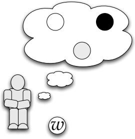
$\mathcal{ONL}$ is a first-order language with $=$

- 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: $\wedge, \neg, \forall$;
- modal operators:
  - $\boldsymbol{K}\alpha$ "at least" α is believed.
  - $\boldsymbol{N}\alpha$ "at most" α is believed to be false;
  - Only-knowing:
    $\boldsymbol{O}\alpha \doteq \boldsymbol{K}\alpha \wedge \boldsymbol{N}\neg\alpha$

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.

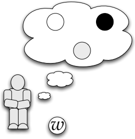


- $e, w \models P(\vec{n})$ iff $P(\vec{n}) \in w$;
- $e, w \models \forall x.\alpha$ iff $e, w \models \alpha^x_n$ for all n
- $e, w \models \boldsymbol{K}\alpha$ iff for all $w' \in e$, $e, w' \models \alpha$;
- $e, w \models \boldsymbol{N}\alpha$ iff for all $w' \notin e$, $e, w' \models \alpha$

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.



- $e, w \models P(\vec{n})$ iff $P(\vec{n}) \in w$;
- $e, w \models \forall x.\alpha$ iff $e, w \models \alpha^x_n$ for all n
- $e, w \models \boldsymbol{K}\alpha$ iff for all $w' \in e$, $e, w' \models \alpha$;
- $e, w \models \boldsymbol{N}\alpha$ iff for all $w' \notin e$, $e, w' \models \alpha$
- ($e, w \models \boldsymbol{O}\alpha$ 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 $\boldsymbol{L}$ stand for both $\boldsymbol{K}$ and $\boldsymbol{N}$:

Axioms (propositional)

objective: non-modal formulas
subjective: all predicates within a modal
Let $\boldsymbol{L}$ stand for both $\boldsymbol{K}$ and $\boldsymbol{N}$:

1. Axioms of propositional logic.
2. $\boldsymbol{L}(\alpha \supset \beta) \; \supset \; \boldsymbol{L}\alpha \supset \boldsymbol{L}\beta$.
3. $\sigma \; \supset \; \boldsymbol{L}\sigma$, where σ is subjective.
4. The $\boldsymbol{N}$ vs. $\boldsymbol{K}$ axiom:
   $(\boldsymbol{N}\phi \supset \neg\boldsymbol{K}\phi)$, where $\neg\phi$ is consistent and objective;
5. $\boldsymbol{O}\alpha \equiv (\boldsymbol{K}\alpha \wedge \boldsymbol{N}\neg\alpha)$.
6. Inference rules:
   Modus ponens and Necessitation (for $\boldsymbol{K}$ and $\boldsymbol{N}$)

Many Agents

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.

Many Agents

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 is not so clear how to come up with models that contain all possibilities.

Some Previous Attempts

- (Lakemeyer 1993) uses the $K45_n$-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 $\mathcal{ONL}_n$

- $\mathcal{ONL}_n \doteq$ multi-agent version of $\mathcal{ONL}$
  - Here, only for *a* and *b* ($\boldsymbol{K}_a, \boldsymbol{K}_b, \boldsymbol{N}_a, \boldsymbol{N}_b$)
- depth: alternating nesting of modalities
  - a notion of *a*-depth and *b*-depth

Example (*a*-depth)

- p: 1
- $\boldsymbol{K}_a p$: 1
- $\boldsymbol{K}_b p$: 2
- $\boldsymbol{K}_a \boldsymbol{K}_b p$: 2

- *a*-objective: formulas not in scope of $\boldsymbol{K}_a$ or $\boldsymbol{N}_a$:
  $p \wedge \boldsymbol{K}_b p$ is *a*-objective,
  $p \wedge \boldsymbol{K}_a p$ 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.

Formal Semantics

- Define an epistemic state for Alice as a set of pairs
  - $e_a^1 = \{\langle w, \{\}\rangle, \langle w', \{\}\rangle \ldots\}$ (for formulas of a-depth 1)
  - $e_a^k = \{\langle w, e_b^{k-1}\rangle, \ldots\}$
- Similarly, an epistemic state for Bob
  - $e_b^1 = \{\langle w, \{\}\rangle \ldots\}$
  - $e_b^j = \{\langle w, e_a^{j-1}\rangle, \ldots\}$
- (k, j)-model $\doteq \langle e_a^k, e_b^j, w\rangle$.

  Given a formula of a-depth k and b-depth j
  - $e_a^k, e_b^j, w \models \boldsymbol{K}_a\alpha$ iff for all $\langle w', e_b^{k-1}\rangle \in e_a^k$, $e_a^k, e_b^{k-1}, w' \models \alpha$
  - $e_a^k, e_b^j, w \models \boldsymbol{N}_a\alpha$ iff for all $\langle w', e_b^{k-1}\rangle \notin e_a^k$, $e_a^k, e_b^{k-1}, w' \models \alpha$
  - $\boldsymbol{O}_a\alpha \equiv \boldsymbol{K}_a\alpha \wedge \boldsymbol{N}_a\neg\alpha$.

Some Properties

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 $\boldsymbol{K}_i$ and $\boldsymbol{N}_i$)
  - $\boldsymbol{K}_a(\alpha \supset \beta) \supset \boldsymbol{K}_a\alpha \supset \boldsymbol{K}_a\beta$
  - $\boldsymbol{K}_a\alpha \supset \boldsymbol{K}_a\boldsymbol{K}_a\alpha$
  - $\neg\boldsymbol{K}_a\alpha \supset \boldsymbol{K}_a\neg\boldsymbol{K}_a\alpha$
- Mutual introspection:
  - $\boldsymbol{K}_a\alpha \supset \boldsymbol{N}_a\boldsymbol{K}_a\alpha$
- Barcan formula
  - $\forall x\, \boldsymbol{K}_a\alpha \supset \boldsymbol{K}_a\forall x\, \alpha$

Examples of Only-Knowing

- $\boldsymbol{O}_a(p \wedge \boldsymbol{O}_b p)$ entails
  - $\boldsymbol{K}_a p$
  - $\boldsymbol{K}_a \boldsymbol{K}_b p$
  - $\boldsymbol{K}_a \neg \boldsymbol{K}_b \boldsymbol{K}_a p$
  - but not $\boldsymbol{K}_a \boldsymbol{K}_b \neg \boldsymbol{K}_a p$

Examples of Only-Knowing

- $\boldsymbol{O}_a(p \wedge \boldsymbol{O}_b p)$ entails
  - $\boldsymbol{K}_a p$
  - $\boldsymbol{K}_a \boldsymbol{K}_b p$
  - $\boldsymbol{K}_a \neg \boldsymbol{K}_b \boldsymbol{K}_a p$
  - but not $\boldsymbol{K}_a \boldsymbol{K}_b \neg \boldsymbol{K}_a p$
- $\boldsymbol{K}_a \boldsymbol{O}_b \exists x\ T(x)$ entails
  - $\boldsymbol{K}_a \boldsymbol{K}_b (\exists x\ T(x) \wedge \neg \boldsymbol{K}_b T(x))$

Examples of Only-Knowing

- $\boldsymbol{O}_a(p \wedge \boldsymbol{O}_b p)$ entails
  - $\boldsymbol{K}_a p$
  - $\boldsymbol{K}_a \boldsymbol{K}_b p$
  - $\boldsymbol{K}_a \neg \boldsymbol{K}_b \boldsymbol{K}_a p$
  - but not $\boldsymbol{K}_a \boldsymbol{K}_b \neg \boldsymbol{K}_a p$
- $\boldsymbol{K}_a \boldsymbol{O}_b \exists x\; T(x)$ entails
  - $\boldsymbol{K}_a \boldsymbol{K}_b(\exists x\; T(x) \wedge \neg \boldsymbol{K}_b T(x))$
- Let $\delta_b = \forall x. \boldsymbol{K}_b Bird(x) \wedge \neg \boldsymbol{K}_b \neg Fly(x) \supset Fly(x)$

  $\boldsymbol{K}_a \boldsymbol{O}_b(\delta_b \wedge Bird(tweety))$ entails
  - $\boldsymbol{K}_a \boldsymbol{K}_b Fly(tweety)$.

Proof Theory for Many Agents

- $\mathsf{K45}_n$ + Axiom defining $\boldsymbol{O}_i$
  + new version of the $\boldsymbol{N}$ vs $\boldsymbol{K}$ axiom.
- Recall, in the single-agent case:
  $(\boldsymbol{N}\phi \supset \neg\boldsymbol{K}\phi)$, 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!

$\boldsymbol{N}_i$ vs. $\boldsymbol{K}_i$ Axioms

Idea: break the circularity by considering a hierarchy of sub-languages based on the nesting of $\boldsymbol{N}_i$.

- Define a family of languages
  - Let $\mathcal{ONL}_n^1 \doteq$ no $\boldsymbol{N}_j$ in the scope of $\boldsymbol{K}_i$, $\boldsymbol{N}_i$ ($i \neq j$)
  - Let $\mathcal{ONL}_n^{t+1}$ formed from $\mathcal{ONL}_n^t$, $\boldsymbol{K}_i\alpha$ and $\boldsymbol{N}_i\alpha$ for all $\alpha \in \mathcal{ONL}_n^t$

$\boldsymbol{N}_i$ vs. $\boldsymbol{K}_i$ Axioms

Idea: break the circularity by considering a hierarchy of sub-languages based on the nesting of $\boldsymbol{N}_i$.

- Define a family of languages
  - Let $\mathcal{ONL}_n^1 \doteq$ no $\boldsymbol{N}_j$ in the scope of $\boldsymbol{K}_i$, $\boldsymbol{N}_i$ ($i \neq j$)
  - Let $\mathcal{ONL}_n^{t+1}$ formed from $\mathcal{ONL}_n^t$, $\boldsymbol{K}_i\alpha$ and $\boldsymbol{N}_i\alpha$ for all $\alpha \in \mathcal{ONL}_n^t$
- Proof theory is $K45_n$ + Def. of $\boldsymbol{O}_i$ +

$\mathbf{A}_n^1$ $\boldsymbol{N}_i\alpha \supset \neg\boldsymbol{K}_i\alpha$ if $\neg\alpha$ is a $K45_n$-consistent *i*-objective formula

$\mathbf{A}_n^{t+1}$ $\boldsymbol{N}_i\alpha \supset \neg\boldsymbol{K}_i\alpha$, if $\neg\alpha \in \mathcal{ONL}_n^t$, is *i*-objective and consistent *wrt*. $K45_n$, $\mathbf{A}_n^1 - \mathbf{A}_n^t$

Theorem

For all $\alpha \in \mathcal{ONL}_n^t$, $\models \alpha$ iff Axioms$^t \vdash \alpha$

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.