A Representation Theorem for Reasoning in First-Order Multi-Agent Knowledge Bases

Christoph Schwering Maurice Pagnucco

UNSW Sydney, Australia

A Representation Theorem for Reasoning in First-Order Multi-Agent Knowledge Bases

Reasoning in multi-agent epistemic knowledge bases

reduces to

classical validity

A Representation Theorem for Reasoning in First-Order Multi-Agent Knowledge Bases

Reasoning in multi-agent epistemic knowledge bases

reduces to

classical validity

- Logical framework: Levesque's logic of only-knowing $\mathbf{K}_A\,\alpha$ $\mathbf{O}_A\,\alpha$

A Representation Theorem for Reasoning in First-Order Multi-Agent Knowledge Bases

Reasoning in multi-agent epistemic knowledge bases

Turing-reduces to

classical validity

- Logical framework: Levesque's logic of only-knowing $\mathbf{K}_A\,\alpha$ $\mathbf{O}_A\,\alpha$

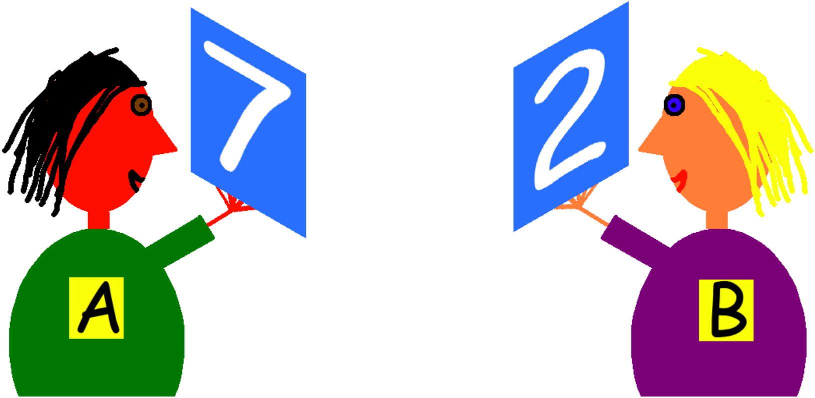
A Representation Theorem for Reasoning in First-Order Multi-Agent Knowledge Bases

Reasoning in multi-agent epistemic knowledge bases

Turing-reduces to

classical validity

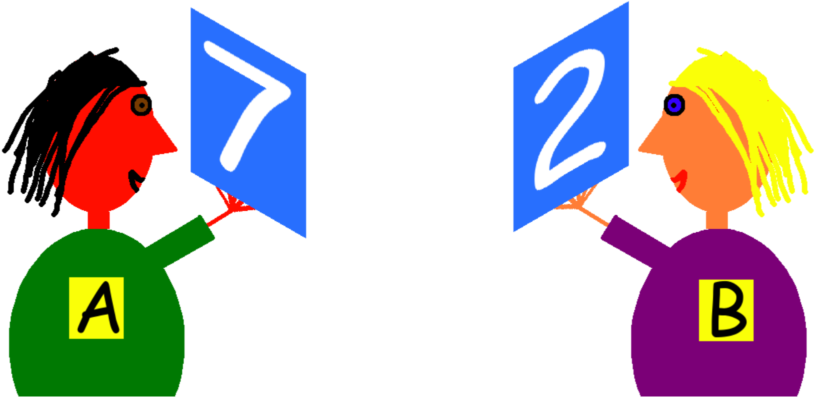
- Logical framework: Levesque's logic of only-knowing $\mathbf{K}_A\,\alpha$ $\mathbf{O}_A\,\alpha$
- Could implement reasoning service with off-the-shelf theorem prover



- A knows $\boxed{A}$
- A knows that

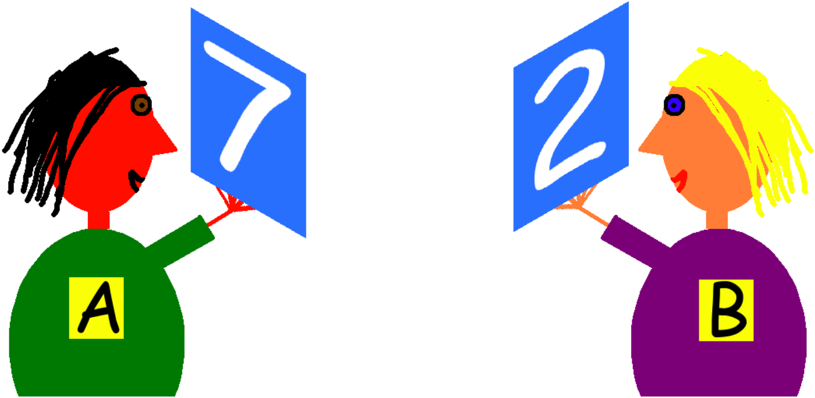
  but A doesn't know $\boxed{B}$

- B knows $\boxed{B}$



KB: $\mathbf{O}_A\left(\boxed{A} = 7 \wedge \forall x\left(\boxed{B} = x \rightarrow \mathbf{O}_B \boxed{B} = x\right)\right)$

A only knows that $\boxed{A} = 7$ and that if $\boxed{B} = x$, then B only knows that $\boxed{B} = x$



$$\text{KB:} \quad \mathbf{O}_A \left(\boxed{A} = 7 \land \forall x \left(\boxed{B} = x \to \mathbf{O}_B \boxed{B} = x \right) \right)$$

entails

$$\text{Query:} \quad \mathbf{K}_A \exists z \left(\underbrace{\boxed{B} = z}_{\text{de dicto}} \land \neg \mathbf{K}_A \underbrace{\boxed{B} = z}_{\text{de re}} \land \mathbf{K}_B \underbrace{\boxed{B} = z}_{\text{de re}} \right)$$

A knows that some number is equal to $\boxed{B}$, but A doesn't know what the number is, and B does know it

Reduction: Eliminate Modal Operators

KB: $\mathbf{O}_A\left(\boxed{A} = 7 \land \forall x\left(\boxed{B} = x \to \mathbf{O}_B\,\boxed{B} = x\right)\right)$

entails

Query: $\mathbf{K}_A\,\exists z\left(\boxed{B} = z \land \neg\mathbf{K}_A\,\boxed{B} = z \land \mathbf{K}_B\,\boxed{B} = z\right)$

Reduction: Eliminate Modal Operators

KB: $\mathbf{O}_A \left(A = 7 \land \forall x \left(B = x \rightarrow \mathbf{O}_B B = x \right) \right)$

entails

Query: $\mathbf{K}_A \exists z \left(B = z \land \neg \mathbf{K}_A B = z \land \mathbf{K}_B B = z \right)$

Reduction: Eliminate Modal Operators

KB: $\mathbf{O}_A \left(A = 7 \wedge \forall x \left(B = x \rightarrow \mathbf{O}_B\, B = x \right) \right)$

entails

Query: $\mathbf{K}_A \exists z \left(B = z \wedge \neg \mathbf{K}_A\, B = z \wedge \mathbf{K}_B\, B = z \right)$

KB′: $\mathbf{O}_A \left(A = 7 \wedge \forall x \left(B = x \rightarrow P(x) \right) \right)$

Reduction: Eliminate Modal Operators

KB: $\mathbf{O}_A\left(A = 7 \land \forall x\left(B = x \rightarrow \mathbf{O}_B B = x\right)\right)$

entails

Query: $\mathbf{K}_A \exists z\left(B = z \land \neg\mathbf{K}_A B = z \land \mathbf{K}_B B = z\right)$

KB’: $\mathbf{O}_A\left(A = 7 \land \forall x\left(B = x \rightarrow P(x)\right)\right)$

entails

Query’: $\mathbf{K}_A \exists z\left(B = z \land \neg\mathbf{K}_A B = z \land \exists x\left(P(x) \land \quad\right)\right)$

“For which x, z does $\mathbf{O}_B B = x$ entail $\mathbf{K}_B B = z$?”

Reduction: Eliminate Modal Operators

KB: $\mathbf{O}_A \big(\boxed{A} = 7 \land \forall x \big(\boxed{B} = x \to \mathbf{O}_B \boxed{B} = x \big)\big)$

entails

Query: $\mathbf{K}_A \exists z \big(\boxed{B} = z \land \neg \mathbf{K}_A \boxed{B} = z \land \mathbf{K}_B \boxed{B} = z \big)$

KB': $\mathbf{O}_A \big(\boxed{A} = 7 \land \forall x \big(\boxed{B} = x \to P(x) \big)\big)$

entails

Query': $\mathbf{K}_A \exists z \big(\boxed{B} = z \land \neg \mathbf{K}_A \boxed{B} = z \land \exists x (P(x) \land \quad) \big)$

"For which x, z is $\boxed{B} = x \to \boxed{B} = z$ valid?"

Call validity oracle! [Levesque '84]

Reduction: Eliminate Modal Operators

KB: $\mathbf{O}_A \left(\boxed{A} = 7 \land \forall x \left(\boxed{B} = x \rightarrow \mathbf{O}_B \boxed{B} = x \right) \right)$

entails

Query: $\mathbf{K}_A \exists z \left(\boxed{B} = z \land \neg \mathbf{K}_A \boxed{B} = z \land \mathbf{K}_B \boxed{B} = z \right)$

KB': $\mathbf{O}_A \left(\boxed{A} = 7 \land \forall x \left(\boxed{B} = x \rightarrow P(x) \right) \right)$

entails

Query': $\mathbf{K}_A \exists z \left(\boxed{B} = z \land \neg \mathbf{K}_A \boxed{B} = z \land \exists x (P(x) \land \underbrace{x = z}) \right)$

"For which x, z is $\boxed{B} = x \rightarrow \boxed{B} = z$ valid?"

Call validity oracle! [Levesque '84]

Summary

Assumption: agents always only-know something about each other.

✗ $\mathbf{O}_A\,(P \to \mathbf{O}_B\,\alpha)$

✓ $\mathbf{O}_A\,((P \to \mathbf{O}_B\,\alpha) \wedge (\neg P \to \mathbf{O}_B\,\beta))$

✓ $\mathbf{O}_A\,(\boxed{A} = 7 \wedge \forall x\,(\boxed{B} = x \to \mathbf{O}_B\,\boxed{B} = x))$

Then:

Reasoning in multi-agent epistemic knowledge bases

Turing-reduces to

classical validity

Appendix

Multi-Agent Knowledge Bases

- $\mathbf{O}_A\,\alpha \;=\; A$ only-knows α [Levesque '84]
  - A considers <u>all</u> models of α possible

Multi-Agent Knowledge Bases

- $\mathbf{O}_A\,\alpha \quad = A$ only-knows α [Levesque '84]
  - A considers all models of α possible
- $\mathbf{O}_A\,\phi$ entails $\mathbf{K}_A\,\psi \iff \phi \to \psi$ is valid
  - provided that ϕ, ψ are objective!

Multi-Agent Knowledge Bases

- $\mathbf{O}_A\,\alpha \;=\; A$ only-knows α [Levesque '84]
  - A considers all models of α possible
- $\mathbf{O}_A\,\phi$ entails $\mathbf{K}_A\,\psi \iff \phi \rightarrow \psi$ is valid
  - provided that ϕ, ψ are objective!
- $\mathbf{O}_A\,\alpha$ implies $\mathbf{O}_A\,\beta \iff \alpha$ and β are equivalent
  - A can only-know at most one formula

Multi-Agent Knowledge Bases

- $\mathbf{O}_A\,\alpha \;=\; A$ only-knows α [Levesque '84]
  - A considers all models of α possible
- $\mathbf{O}_A\,\phi$ entails $\mathbf{K}_A\,\psi \iff \phi \rightarrow \psi$ is valid
  - provided that ϕ, ψ are objective!
- $\mathbf{O}_A\,\alpha$ implies $\mathbf{O}_A\,\beta \iff \alpha$ and β are equivalent
  - A can only-know at most one formula
- $\mathbf{O}_A\,\alpha$ is a multi-agent KB $\iff$ every model of α satisfies some $\mathbf{O}_B\,\beta$
  - ✗ $\mathbf{O}_A\left(P \rightarrow \mathbf{O}_B\,\alpha\right)$
  - ✓ $\mathbf{O}_A\left((P \rightarrow \mathbf{O}_B\,\alpha) \land (\neg P \rightarrow \mathbf{O}_B\,\beta)\right)$
  - ✓ $\mathbf{O}_A \forall x \left(f = x \rightarrow \mathbf{O}_B\,\alpha(x)\right)$

Reduction

- Replace each $\mathbf{O}_A\,\alpha(\vec{x})$ with a fresh atom $P_\alpha(\vec{x})$

Reduction

- Replace each $\mathbf{O}_A\,\alpha(\vec{x})$ with a fresh atom $P_\alpha(\vec{x})$
- Replace each $\mathbf{K}_A\gamma(\vec{z})$ with a disjunction of

  $$\exists\vec{x}\big(P_\alpha(\vec{x}) \wedge \text{“for which } \vec{x},\vec{z} \text{ is } \alpha(\vec{x}) \rightarrow \gamma(\vec{z}) \text{ is valid?”}\big)$$

  over all $\mathbf{O}_A\,\alpha(\vec{x})$ at the same modal nesting level

Reduction

- Replace each $\mathbf{O}_A\,\alpha(\vec{x})$ with a fresh atom $P_\alpha(\vec{x})$
- Replace each $\mathbf{K}_A\gamma(\vec{z})$ with a disjunction of

$$\exists\vec{x}\big(P_\alpha(\vec{x}) \wedge \text{“for which } \vec{x},\vec{z} \text{ is } \alpha(\vec{x}) \rightarrow \gamma(\vec{z}) \text{ is valid?”}\big)$$

  over all $\mathbf{O}_A\,\alpha(\vec{x})$ at the same modal nesting level
- Axiomatise that $P_\alpha(\vec{x}), P_\beta(\vec{y})$ introduced for $\mathbf{O}_A\,\alpha(\vec{x}), \mathbf{O}_A\,\beta(\vec{y})$

$$P_\alpha(\vec{x}) \rightarrow \big(P_\beta(\vec{y}) \leftrightarrow \text{“for which } \vec{x},\vec{y} \text{ is } \alpha(\vec{x}) \rightarrow \beta(\vec{y}) \text{ is valid?”}\big)$$

Summary

Multi-agent KB:

- Based on Levesque's logic of only-knowing
- Every model of a multi-agent KB must satisfy some $\mathbf{O}_B\,\beta$
- Allows for incomplete knowledge about other agent's knowledge

Summary

Multi-agent KB:

- Based on Levesque's logic of only-knowing
- Every model of a multi-agent KB must satisfy some $\mathbf{O}_B\,\beta$
- Allows for incomplete knowledge about other agent's knowledge

Reduction to classical reasoning:

- Oracle for FOL validity
- Turing reduction: calls oracle many times
- Would need FO-K45 oracle if it weren't for

Summary

Multi-agent KB:

- Based on Levesque's logic of only-knowing
- Every model of a multi-agent KB must satisfy some $\mathbf{O}_B\,\beta$
- Allows for incomplete knowledge about other agent's knowledge

Reduction to classical reasoning:

- Oracle for FOL validity
- Turing reduction: calls oracle many times
- Would need FO-K45 oracle if it weren't for

Implementation options:

- FOL theorem prover (e.g., Vampire)
- Limited belief logic