Possible Worlds Semantics and Algorithmic Knowledge of Mathematics

Zeynep Soysal

University of Rochester

Possible Worlds Semantics

Possible worlds semantics models meanings as constructions from possible objects and possible worlds.

Possible worlds account of propositions (henceforth ‘PW-account’):
The proposition that a declarative sentence ϕ expresses is modeled as some $f : W \to \{\top, \bot\}$, or equivalently, as some $P \subseteq W$ (the proposition expressed by ϕ is the set of possible worlds in which ϕ is true).

Big problem for the PW-account

There is only one necessary proposition, $W = \{w \mid w \text{ is a possible world}\}$.

- All true mathematical sentences mean the same thing;
- whoever knows any necessary proposition knows them all, and, thus, in particular, is mathematically omniscient.

Plan:

1. The metalinguistic strategy: mathematical propositions aren't W or $\emptyset$, but propositions about the relation between some mathematical sentence and W;

Plan:

1. The metalinguistic strategy: mathematical propositions aren't W or $\emptyset$, but propositions about the relation between some mathematical sentence and W;
2. the closure problem;
3. the fragmentation strategy;
4. the computation strategy.

The Metalinguistic Strategy

- Mathematical propositions are propositions about the relationship between some mathematical sentence and the one necessarily true proposition, W.

The Metalinguistic Strategy

- Mathematical propositions are propositions about the relationship between some mathematical sentence and the one necessarily true proposition, W.
- Ola seems ignorant that 5801 is prime: Ola isn't ignorant that W, but that '5801 is prime' expresses W.

The Metalinguistic Strategy

- Mathematical propositions are propositions about the relationship between some mathematical sentence and the one necessarily true proposition, W.
- Ola seems ignorant that 5801 is prime: Ola isn't ignorant that W, but that '5801 is prime' expresses W.
- Mathematical propositions are propositions of form $\{w \mid \phi$ expresses W at $w\}$, for mathematical sentence ϕ.

The Metalinguistic Strategy

- Mathematical propositions are propositions about the relationship between some mathematical sentence and the one necessarily true proposition, W.
- Ola seems ignorant that 5801 is prime: Ola isn't ignorant that W, but that '5801 is prime' expresses W.
- Mathematical propositions are propositions of form $\{w \mid \phi$ expresses W at $w\}$, for mathematical sentence ϕ.
- Since for any distinct φ and ψ, $\{w \mid \varphi \text{ expresses } \mathscr{W} \text{ at } w\} \neq \{w \mid \psi \text{ expresses } \mathscr{W} \text{ at } w\}$, there are as many distinct mathematical propositions as there are distinct mathematical sentences.

The Closure Problem

On the standard model traditionally associated with the PW-account, knowledge and belief are closed under entailment:

- If Π entails Q, and if $B_S(P)$ for all $P \in \Pi$, then $B_S(Q)$. [Closure Principle]

The Closure Problem

1. Ola knows that the axioms of PA are true.
2. Ola knows that the inference rules preserve truth.
3. For any given theorem of arithmetic, ϕ, the axioms being true and the rules of inference being truth-preserving entails that ϕ is true.
4. Ola knows that ϕ is true [by Closure Principle].

The Closure Problem

1. Ola knows that the axioms of PA are true.
2. Ola knows that the inference rules preserve truth.
3. For any given theorem of arithmetic, ϕ, the axioms being true and the rules of inference being truth-preserving entails that ϕ is true.
4. Ola knows that ϕ is true [by Closure Principle].

The Standard Model

Knowledge and belief is "truth in all accessible worlds":

- $\mathcal{I}_S$: the set of worlds that are epistemically accessible to S, S's "information state."
- $K_S(P)$ iff $\mathcal{I}_S \subseteq P$.
- If $P \subseteq Q$, and $K_S(P)$, then $\mathcal{I}_S \subseteq Q$, i.e., $K_S(Q)$.
- $\mathcal{B}_S$: the set of worlds that are doxastically accessible to S, S's "belief state."

Stalnaker's causal-pragmatic account of belief

Stalnaker's causal-pragmatic account of belief

> *To believe that p is to be disposed to act in ways that would tend to satisfy one's desires, whatever they are, in a world in which p (together with one's other beliefs) were true. (Stalnaker 1984, 15)*

Stalnaker's causal-pragmatic account of belief

Necessarily, an agent believes P iff:

1. the agent is in a state she would only be in if P were the case, and,
2. the agent is disposed to act in ways that would tend to satisfy her desires in a world in which P together with her other beliefs is true.

Stalnaker's causal-pragmatic account of belief

Necessarily, an agent believes P iff:

1. the agent is in a state she would only be in if P were the case, and,
   - Closed under entailment.
2. the agent is disposed to act in ways that would tend to satisfy her desires in a world in which P together with her other beliefs is true.

Stalnaker's causal-pragmatic account of belief

Necessarily, an agent believes P iff:

1. the agent is in a state she would only be in if P were the case, and,
   - Closed under entailment.
2. the agent is disposed to act in ways that would tend to satisfy her desires in a world in which P together with her other beliefs is true.
   - If S is disposed to act in ways that would tend to satisfy her desires in P-worlds, and if $P \subseteq Q$, then she is also disposed to act in ways that would satisfy her desires in $P \cap Q$;
   - S believes $\{w \mid P$ is true at w and Q is true at $w\}$;
   - S believes Q (by distribution) [?];

Stalnaker's causal-pragmatic account of belief

Necessarily, an agent believes P iff:

1. the agent is in a state she would only be in if P were the case, and,
   - Closed under entailment.
2. the agent is disposed to act in ways that would tend to satisfy her desires in a world in which P together with her other beliefs is true.
   - In any case, $\{w$ |PA is true and rules are truth-preserving at $w\} = \{w$ |PA is true and rules are truth-preserving and FLT is a theorem and FLT is true at $w\}$, so, if Ola believes the former she believes the latter.

The Fragmentation Strategy

- Agents can be "fragmented" in the sense of having more than one belief state at the same time.
- Each belief state corresponds to a context the agent is in, or a task that the agent is engaged in.
- $\rightarrow$ Information can be accessible for some purposes or in some contexts, but inaccessible for others.

The Fragmentation Strategy

- Agents can be "fragmented" in the sense of having more than one belief state at the same time.
- Each belief state corresponds to a context the agent is in, or a task that the agent is engaged in.
- $\rightarrow$ Information can be accessible for some purposes or in some contexts, but inaccessible for others.
- E.g.: Information that 'dreamt' is a word of English with 6 letters ending with 'mt':
  - Inaccessible to S for purpose of solving cross-word puzzle;
  - Accessible to S for purpose of answering "Is 'dreamt' a word of English with six letters and ending in 'mt'?"

Cases that can't be explained by fragmentation:

1. $K_O(\{w \mid a_1 \text{ expresses } W \text{ at } w\})$ [*assumption*].
2. $K_O(\{w \mid a_2 \text{ expresses } W \text{ at } w\})$ [*assumption*].
3. $K_O(\{w \mid\vdash_{\langle A,R\rangle} \psi \text{ at } w\})$ [*since* $\{w \mid\vdash_{\langle A,R\rangle} \psi$ *at* $w\} = W$].
4. $K_O(\{w \mid$At w, for any φ, if a_1 expresses W and a_2 expresses W and $\vdash_{\langle A,R\rangle} \varphi$, then φ expresses $W\})$ [*assumption*].
5. $K_O(\{w \mid \psi \text{ expresses } W \text{ at } w\})$ [*by* 1.–4. *and closure under entailment*].

Cases that can't be explained by fragmentation:

1. $K_O(\{w \mid a_1 \text{ expresses } W \text{ at } w\})$ [*assumption*].
2. $K_O(\{w \mid a_2 \text{ expresses } W \text{ at } w\})$ [*assumption*].
3. $K_O(\{w \mid \vdash_{\langle A,R\rangle} \psi \text{ at } w\})$ [*since* $\{w \mid \vdash_{\langle A,R\rangle} \psi$ *at* $w\} = W$].
4. $K_O(\{w \mid$At w, for any φ, if a_1 expresses W and a_2 expresses W and $\vdash_{\langle A,R\rangle} \varphi$, then φ expresses $W\})$ [*assumption*].
5. $K_O(\{w \mid \psi \text{ expresses } W \text{ at } w\})$ [*by* 1.–4. *and closure under entailment*].

The Computation Strategy

- What Ola doesn't have is the ability to prove ϕ from $\langle A, R \rangle$.
- What Watson lacks and Holmes has is the ability to compute who the culprit is based on the information they each have.

Algorithmic Knowledge Models

Developed by Halpern, Moses, Vardi, Konolige, Parikh, Pucella, ...:

- Supplements the standard model with algorithms;
- Agent has a knowledge algorithm that returns 'Yes', 'No', or '?' given a formula ϕ;
- An agent then (algorithmically) knows ϕ iff her knowledge algorithm returns 'Yes' on input ϕ.

[Simple case: An *algorithmic structure* is a tuple $M = \langle W, W', \pi, A \rangle$ where $\langle W, W', \pi \rangle$ is a K45 Kripke structure, and A a knowledge algorithm that returns 'Yes', 'No', or '?' given ϕ.
$(M, w) \models K\phi$ iff $A(\phi) =$ 'Yes'.]

Simple Proposal

S knows $\{w \mid \phi$ expresses W at $w\}$ iff:

1. S is in a state that indicates (carries the information) $\{w \mid \phi$ expresses W at $w\}$, and,
2. S has an algorithm that reliably outputs 'True' if asked to determine ϕ's truth-value, using at most resources R.

Simple Proposal

S knows $\{w \mid \phi$ expresses W at $w\}$ iff:

1. S is in a state that indicates (carries the information) $\{w \mid \phi$ expresses W at $w\}$, and,
2. S has an algorithm that reliably outputs 'True' if asked to determine ϕ's truth-value, using at most resources R.

- Doesn't face closure problem.

Simple Proposal

S knows $\{w \mid \phi$ expresses W at $w\}$ iff:

1. S is in a state that indicates (carries the information) $\{w \mid \phi$ expresses W at $w\}$, and,
2. S has an algorithm that reliably outputs 'True' if asked to determine ϕ's truth-value, using at most resources R.

- Doesn't face closure problem.
- Mathematical knowledge plausibly requires more than 2.
- Ignores relevance of desires to action.
- Only applies to metalinguistic propositions.
- Could be made to fit more explicitly with functionalism.

Improved Proposal, 1

S knows P iff:

1*. S is in a state that indicates (carries the information) P, and,

2*. S has an algorithm that reliably produces desire-satisfying behavior in $w \in P$, using at most resources R.

Improved Proposal, 1

S knows P iff:

1*. S is in a state that indicates (carries the information) P, and,

2*. S has an algorithm that reliably produces desire-satisfying behavior in $w \in P$, using at most resources R.

- Faces closure problem:

Improved Proposal, 1

S knows P iff:

1*. S is in a state that indicates (carries the information) P, and,

2*. S has an algorithm that reliably produces desire-satisfying behavior in $w \in P$, using at most resources R.

- Faces closure problem:
- Assume you know axioms and rules of inference, you are able to exhibit desire-satisfying behaviors in $w \in \{w \mid$ PA is true and rules are truth-preserving$\}$.
- But $\{w \mid$ PA is true and rules are truth-preserving$\} = \{w \mid$ PA is true and rules are truth-preserving and FLT is a theorem and FLT is true at $w\}$.

Simple Proposal:

S knows $\{w \mid \phi \text{ expresses } W \text{ at } w\}$ iff:

1. S is in a state that indicates (carries the information) $\{w \mid \phi \text{ expresses } W \text{ at } w\}$, and,
2. S has an algorithm that reliably outputs 'True' if asked to determine ϕ's truth-value, using at most resources R.

Improved Proposal, 1:

S knows P iff:

1*. S is in a state that indicates (carries the information) P, and,
2*. S has an algorithm that reliably produces desire-satisfying behavior in $w \in P$, using at most resources R.

Improved Proposal, 2

S knows P iff:

1**. S is in a state that indicates (carries the information) P, and,

2**. S has an algorithm that reliably produces desire-satisfying behavior in $w \in P$, using at most resources R, with respect to tasks T.

Improved Proposal, 2

S knows P iff:

1**. S is in a state that indicates (carries the information) P, and,

2**. S has an algorithm that reliably produces desire-satisfying behavior in $w \in P$, using at most resources R, with respect to tasks T.

- Even if $P = Q$, having an algorithm that produces desire-satisfying behavior in $w \in P$ with respect to task T_1 might not entail having an algorithm that produces desire-satisfying behavior in $w \in Q$ with respect to task T_2.

Improved Proposal, 2

S knows P iff:

1**. S is in a state that indicates (carries the information) P, and,

2**. S has an algorithm that reliably produces desire-satisfying behavior in $w \in P$, using at most resources R, with respect to tasks T.

Improved Proposal, 2

S knows $\{w \mid \phi$ expresses W at $w\}$ iff:

1**. S is in a state that indicates (carries the information) $\{w \mid \phi$ expresses W at $w\}$, and,

2**. S has an algorithm that reliably produces desire-satisfying behavior in $w \in \{w \mid \phi$ expresses W at $w\}$, using at most resources R, with respect to ϕ-related tasks.

Improved Proposal, 2

S knows P iff:

1**. S is in a state that indicates (carries the information) P, and,

2**. S has an algorithm that reliably produces desire-satisfying behavior in $w \in P$, using at most resources R, with respect to tasks T.

$P = \{w \mid$ PA is true and rules are truth-preserving and FLT is a theorem and FLT is true at w$\}$?

Other Ps?

Improved Proposal, 2

S knows P iff:

1**. S is in a state that indicates (carries the information) P, and,

2**. S has an algorithm that reliably produces desire-satisfying behavior in $w \in P$, using at most resources R, with respect to tasks T.

- Even if P entails Q, the tasks relative to which one attributes 'knowledge that P' can be different from the tasks relative to which one attributes 'knowledge that Q'.

Improved Proposal, 2

S knows $\{w \mid \phi \text{ expresses } W \text{ at } w\}$ iff:

1**. S is in a state that indicates (carries the information) $\{w \mid \phi \text{ expresses } W \text{ at } w\}$, and,

2**. S has an algorithm that reliably produces desire-satisfying behavior in $w \in \{w \mid \phi \text{ expresses } W \text{ at } w\}$, using at most resources R, with respect to ϕ-related tasks.

Conclusion

1. The metalinguistic strategy;
2. the closure problem;
3. the fragmentation strategy;
4. the computation strategy.

Conclusion

Thank you!