Formalizing Epistemic Logic

Completeness of Modal Logics

Asta Halkjær From, Alexander Birch Jensen and Jørgen Villadsen
Technical University of Denmark

LAMAS & SR 2021-05-04

Outline

- Possible worlds
- Syntax and semantics
- Normal modal logics
- Soundness
- Completeness-via-canonicity
- Systems K, T, KB, K4, S4, S5
- Technicality
- Takeaways

Possible Worlds

Worlds model situations
Relations model uncertainty

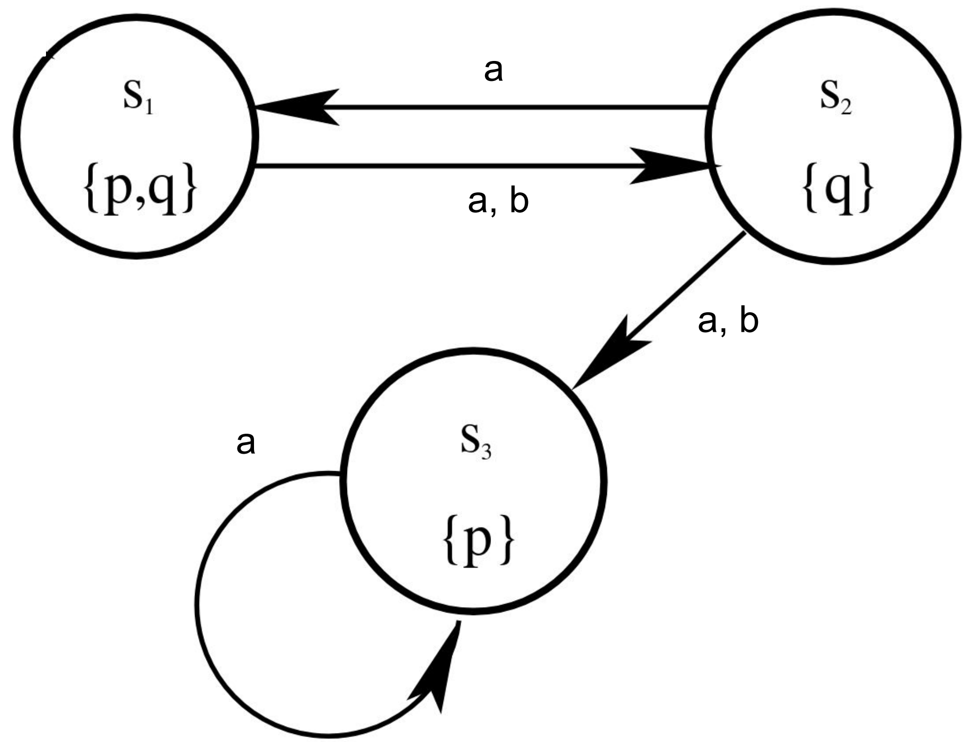
Agent *i* **knows** φ (K_i φ) at a world
if φ holds at all *i*-related worlds

At S_2 we have

- K_a p and K_b p
- Not K_a q

Also at S_s:

- K_b K_a p
- Not K_a K_b p



Syntax and Semantics

We use *x* for propositional symbols and *i* for agent labels:

$$\phi, \psi ::= \bot \mid x \mid \phi \vee \psi \mid \phi \wedge \psi \mid \phi \to \psi \mid K_i \phi$$

The language is interpreted on Kripke models $M = ((W, R_1, R_2, \ldots), V)$:

worlds W, relations R_i, valuation V

$$\begin{array}{lll}
\mathfrak{M}, w \not\models \bot & & \\
\mathfrak{M}, w \models x & \text{iff} & w \in V(x) \\
\mathfrak{M}, w \models \phi \vee \psi & \text{iff} & \mathfrak{M}, w \models \phi \text{ or } \mathfrak{M}, w \models \psi \\
\mathfrak{M}, w \models \phi \wedge \psi & \text{iff} & \mathfrak{M}, w \models \phi \text{ and } \mathfrak{M}, w \models \psi \\
\mathfrak{M}, w \models \phi \to \psi & \text{iff} & \mathfrak{M}, w \not\models \phi \text{ or } \mathfrak{M}, w \models \psi \\
\mathfrak{M}, w \models K_i \phi & \text{iff} & w R_i w' \text{ implies } \mathfrak{M}, w' \models \phi \text{ for all } w' \in W
\end{array}$$

Syntax REDUX

Deep embedding in Isabelle/HOL

Model syntax as an object in the higher-order logic:

```
type_synonym id = string

datatype 'i fm
  = FF ("⊥")
  | Pro id
  | Dis ‹'i fm› ‹'i fm› (infixr "∨" 30)
  | Con ‹'i fm› ‹'i fm› (infixr "∧" 35)
  | Imp ‹'i fm› ‹'i fm› (infixr "⟶" 25)
  | K 'i ‹'i fm›
```

Abbreviations as usual (“considers possible”):

```
abbreviation ‹L i p ≡ ¬ K i (¬ p)›
```

Semantics REDUX

Kripke models as another datatype (type variable ‘w models W):

```
datatype ('i, 'w) kripke = Kripke (π: ‹'w ⇒ id ⇒ bool›) (𝒦: ‹'i ⇒ 'w ⇒ 'w set›)
```

Interpret syntax into the higher-order logic:

```
primrec semantics :: ‹('i, 'w) kripke ⇒ 'w ⇒ 'i fm ⇒ bool›
  ("_, _ ⊨ _" [50, 50] 50) where
  ‹(M, w ⊨ ⊥) = False›
| ‹(M, w ⊨ Pro x) = π M w x›
| ‹(M, w ⊨ (p ∨ q)) = ((M, w ⊨ p) ∨ (M, w ⊨ q))›
| ‹(M, w ⊨ (p ∧ q)) = ((M, w ⊨ p) ∧ (M, w ⊨ q))›
| ‹(M, w ⊨ (p ⟶ q)) = ((M, w ⊨ p) ⟶ (M, w ⊨ q))›
| ‹(M, w ⊨ K i p) = (∀v ∈ 𝒦 M i w. M, v ⊨ p)›
```

Epistemic Principles

At S_3 we have K_b q vacuously

We may want only *true knowledge*

Reflexive relations

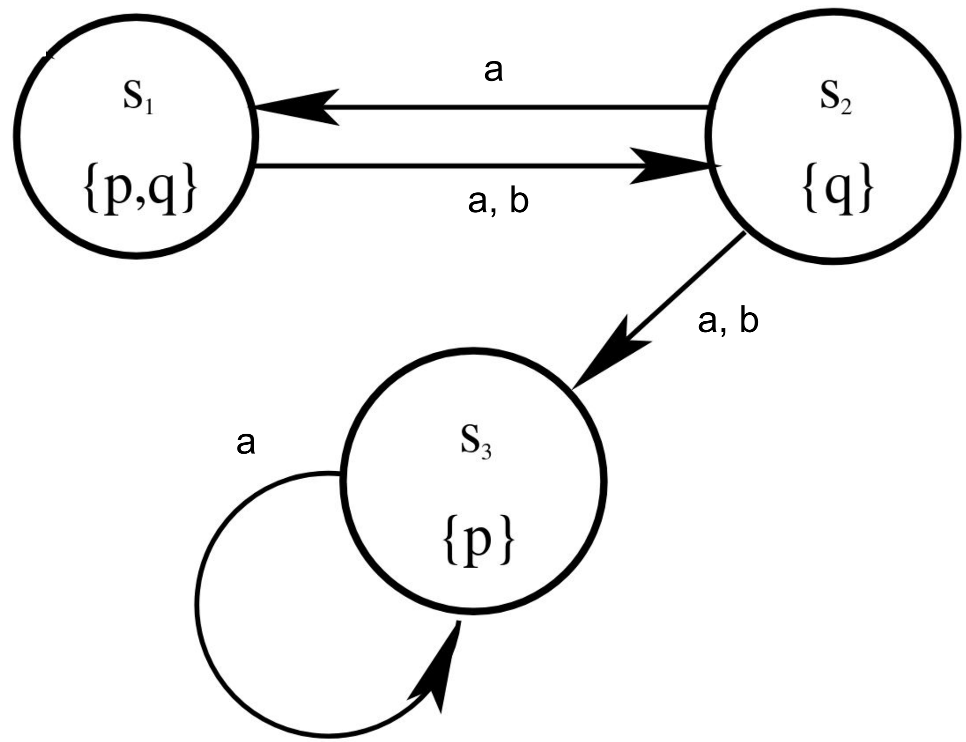
K_i p implies p

We may want *positive introspection*

Transitive relations

K_i p implies K_i K_i p

And so on



Normal Modal Logics

Consider a *family* of proof systems for epistemic reasoning:

```
inductive AK :: ‹('i fm ⇒ bool) ⇒ 'i fm ⇒ bool› ("_ ⊢ _" [50, 50] 50)
  for A :: ‹'i fm ⇒ bool› where
    A1: ‹tautology p ⟹ A ⊢ p›
  | A2: ‹A ⊢ (K i p ∧ K i (p ⟶ q) ⟶ K i q)›
  | Ax: ‹A p ⟹ A ⊢ p›
  | R1: ‹A ⊢ p ⟹ A ⊢ (p ⟶ q) ⟹ A ⊢ q›
  | R2: ‹A ⊢ p ⟹ A ⊢ K i p›
```

A1: all propositional tautologies

A2: distribution axiom

R1: modus ponens

R2: necessitation

Ax: *any epistemic principles we want (as admitted by A)*

Soundness

Generalized soundness result for any normal modal logic

If all extra axioms are sound on models admitted by *P*,
then the resulting logic is sound on *P*-models:

```
theorem soundness:
  fixes M :: ‹('i, 'w) kripke›
  assumes ‹⋀(M :: ('i, 'w) kripke) w p. A p ⟹ P M ⟹ M, w ⊨ p›
  shows ‹A ⊢ p ⟹ P M ⟹ M, w ⊨ p›
```

Completeness-via-Canonicity I

Following proofs by Fagin et al. and Blackburn et al.

- Assume φ has no derivation
- Then {¬φ} is consistent (no finite subset implies ⊥)
- Extend to a maximal consistent set V (Lindenbaum's lemma)
- Canonical model satisfies ¬φ at V (truth lemma)
- So φ could not have been valid

For completeness over a class of frames:

show that the canonical frame belongs to that class

Completeness-via-Canonicity II

Fagin et al. prove completeness for K and write for T:

> “A proof identical to that of Theorem 3.1.3 can now be used.”

We do not want to *copy/paste* our efforts for each logic.

Blackburn et al. write (emphasis ours):

> “The canonical frame of any normal logic containing T is reflexive, the canonical frame of any normal logic containing B is symmetric, and the canonical frame of any normal logic containing D is right unbounded. *This allows us to ‘add together’ our results.*”

Let’s aim for such *compositionality*!

Maximal Consistent Sets wrt. A (A-MCSs)

A set of formulas is *A-consistent* if no finite subset implies ⊥ (using A-axioms)

```
definition consistent :: ‹('i fm ⇒ bool) ⇒ 'i fm set ⇒ bool› where
  ‹consistent A S ≡ ∄S'. set S' ⊆ S ∧ A ⊢ imply S' ⊥›
```

And A-maximal if any proper extension destroys A-consistency:

```
definition maximal :: ‹('i fm ⇒ bool) ⇒ 'i fm set ⇒ bool› where
  ‹maximal A S ≡ ∀p. p ∉ S ⟶ ¬ consistent A ({p} ∪ S)›
```

The usual properties hold:

```
theorem mcs_properties:
  assumes ‹consistent A V› ‹maximal A V›
  shows ‹A ⊢ p ⟹ p ∈ V›
    and ‹p ∈ V ⟷ (¬ p) ∉ V›
    and ‹p ∈ V ⟹ (p ⟶ q) ∈ V ⟹ q ∈ V›
```

Lindenbaum's Lemma

In Isabelle, *extend A S f n* computes S_n from $S_0 = S$ and enumeration *f*:

$$S_{n+1} = \begin{cases} S_n & \text{if } \{\phi_n\} \cup S_n \text{ is not } A\text{-consistent} \\ \{\phi_n\} \cup S_n & \text{otherwise} \end{cases}$$

Extend A S f is the infinite union. We have:

```
lemma consistent_Extend:
  assumes ‹consistent A S›
  shows ‹consistent A (Extend A S f)›
```

```
lemma maximal_Extend:
  assumes ‹surj f›
  shows ‹maximal A (Extend A S f)›
```

Canonical Model

As worlds W take all sets of formulas*

```
abbreviation pi :: ‹'i fm set ⇒ id ⇒ bool› where
  ‹pi V x ≡ Pro x ∈ V›

abbreviation known :: ‹'i fm set ⇒ 'i ⇒ 'i fm set› where
  ‹known V i ≡ {p. K i p ∈ V}›

abbreviation reach :: ‹('i fm ⇒ bool) ⇒ 'i ⇒ 'i fm set ⇒ 'i fm set set› where
  ‹reach A i V ≡ {W. known V i ⊆ W ∧ consistent A W ∧ maximal A W}›
```

reach ensures that we stay in A-MCS worlds

* Preferably only A-MCSs but Isabelle/HOL's logic only supports this if we fix A

Truth Lemma

Following Fagin et al. (860 lines up to and including this result):

```
lemma truth_lemma:
  fixes A and p :: ‹('i :: countable) fm›
  defines ‹M ≡ Kripke pi (reach A)›
  assumes ‹consistent A V› ‹maximal A V›
  shows ‹(p ∈ V ⟷ M, V ⊨ p) ∧ ((¬ p) ∈ V ⟷ M, V ⊨ ¬ p)›
```

Useful abstraction:

```
lemma canonical_model:
  assumes ‹consistent A S› ‹p ∈ S›
  defines ‹V ≡ Extend A S from_nat› and ‹M ≡ Kripke pi (reach A)›
  shows ‹M, V ⊨ p› ‹consistent A V› ‹maximal A V›
```

System K

No extra axioms (A admits nothing):

```
abbreviation SystemK :: ‹'i fm ⇒ bool› ("⊢K _" [50] 50) where
  ‹⊢K p ≡ (λ_. False) ⊢ p›

lemma soundnessK: ‹⊢K p ⟹ M, w ⊨ p›
  using soundness by metis

abbreviation ‹validK p ≡ ∀(M :: (nat, nat fm set) kripke) w. M, w ⊨ p›

theorem mainK: ‹validK p ⟷ ⊢K p›
```

Validity in our universe implies validity in any other:

```
corollary ‹validK p ⟶ M, w ⊨ p›
```

Extra Axioms

Axiom	Formula	Frame condition	Principle
T	$K_i\varphi \to \varphi$	Reflexive	True knowledge
B	$\varphi \to K_i L_i \varphi$	Symmetric	Knowledge of consistency of truths[a]
4	$K_i\varphi \to K_i K_i \varphi$	Transitive	Positive introspection
5	$\neg K_i\varphi \to K_i \neg K_i \varphi$	Euclidean[b]	Negative introspection

```
inductive AxT :: ‹'i fm ⇒ bool› where
  ‹AxT (K i p ⟶ p)›
```

```
lemma AxT_reflexive:
  assumes ‹⋀p. AxT p ⟹ A p› ‹consistent A V› ‹maximal A V›
  shows ‹V ∈ reach A i V›
```

```
lemma Ax4_transitive:
  assumes ‹⋀p. Ax4 p ⟹ A p› ‹consistent A V› ‹maximal A V›
    and ‹W ∈ reach A i V› ‹U ∈ reach A i W›
  shows ‹U ∈ reach A i V›
```

Compositionality

System	Axioms	Class
K		All frames
T	T	Reflexive frames
KB	B	Symmetric frames
K4	4	Transitive frames
S4	T, 4	Reflexive and transitive frames
S5	T, B 4 or T, 5	Frames with equivalence relations

```
abbreviation SystemS4 :: ‹'i fm ⇒ bool› ("⊢S4 _" [50] 50) where
  ‹⊢S4 p ≡ AxT ⊕ Ax4 ⊢ p›
```

```
abbreviation ‹validS4 p ≡ ∀(M :: (nat, nat mcsS4) kripke) w.
    reflexive M ⟶ transitive M ⟶ M, w ⊨ p›
```

```
theorem mainS4: ‹validS4 p ⟷ ⊢S4 p›
```

Behind the Curtain

Recall how our canonical model had too many worlds?

For each system we need to define the type of their worlds:

```
typedef 'i mcsS4 =
  ‹{V :: ('i :: countable) fm set. consistent (AxT ⊕ Ax4) V ∧ maximal (AxT ⊕ Ax4) V}›

abbreviation ‹piS4 ≡ pi o Rep_mcsS4›
abbreviation ‹reachS4 i V ≡ Abs_mcsS4 ` (reach (AxT ⊕ Ax4) i (Rep_mcsS4 V))›
```

We show this submodel reflexive etc.:

```
lemma mcsS4_reflexive: ‹reflexive (Kripke piS4 reachS4)›
```

To reuse the truth lemma, show the models satisfy the same formulas:

```
lemma mcsS4_equiv:
  assumes ‹consistent (AxT ⊕ Ax4) V› ‹maximal (AxT ⊕ Ax4) V›
  shows ‹(Kripke pi (reach (AxT ⊕ Ax4)), V ⊨ p) = (Kripke piS4 reachS4, Abs_mcsS4 V ⊨ p)›
```

Takeaways

- Epistemic logic models the knowledge of agents
- Different epistemic principles give rise to different logics
- Using Isabelle/HOL we have given a disciplined treatment of
  - Normal modal logics ranging from K to S5
  - Completeness-via-canonicity arguments
  - The compositional nature of this method
- Modeling worlds as types gives slight complications
- Soundness and completeness for 7 systems in just over 1600 lines
  - A clear recipe for adding more

References

Fagin, R., Halpern, J.Y., Moses, Y., Vardi, M.Y.: Reasoning About Knowledge. MIT Press (1995).

Blackburn, P., de Rijke, M., Venema, Y.: Modal Logic, Cambridge Tracts in Theoretical Computer Science, vol. 53. Cambridge University Press (2001).

From, A.H.: Epistemic logic: Completeness of modal logics. Archive of Formal Proofs (2018), https://devel.isa-afp.org/entries/Epistemic_Logic.html, Formal proof development

See also four formalizations by Bentzen, Li, Neeley and Wu & Gore in Lean and by Hagemeier in Coq.