Modal logics: overview

- Part I: Introduction to modal and multimodal logics
  1. Motivation and introduction
  2. The basic multimodal logic K
  3. The basic monomodal logics
  4. Completeness of $G(k, l, m, n)$ logics, and decidability of the basic modal logics
  5. Basic multimodal logics
  6. Other modal logics
- Part II: Applications
  7. Knowledge and announcements
  8. **Belief**
  9. Common knowledge and common belief
  10. Action and propositional dynamic logic
  11. Goals and intentions
  12. Ability, agency and branching time
- Part II: Proof methods
  13. Translation method
  14. Tableau method

Chapter 8.
Belief

Oct.16, 2008

Outline of chapter

- multiagent doxastic logic *KD*45
- graded belief
  - different from probability
- relation between belief and knowledge
  - knowledge reducible to belief?
  - remarks and critiques: negative introspection

Doxastic logic: language

- *doxa* = 'believe' (Greek)
- multiagent
- finite parameter set = set of agents

$$\begin{aligned} Prms &= \{i_1, i_2, \ldots, i_{\text{card}(Agts)}\} \\ &= Agts \end{aligned}$$

('*i*ndividuals')

- belief explained in terms of possible worlds:

$$\begin{aligned} \Box_i \varphi &= \text{“agent } i \text{ believes that } \varphi\text{”} \\ &= \text{“}\varphi \textit{ true in every world that is compatible with i's beliefs}\text{”} \\ &= \mathrm{B}_i\, \varphi \end{aligned}$$

[Hintikka]

Belief: logical properties

standard logic of belief = multimodal *KD*45

- principles of multimodal *K*
- consistency of belief:
  - $\mathrm{B}_i\,\varphi \to \neg \mathrm{B}_i\,\neg\varphi$ — axiom D(B_i)
- positive introspection:
  - $\mathrm{B}_i\,\varphi \to \mathrm{B}_i\,\mathrm{B}_i\,\varphi$ — axiom 4(B_i)
- negative introspection:
  - $\neg\mathrm{B}_i\,\varphi \to \mathrm{B}_i\,\neg\mathrm{B}_i\,\varphi$ — axiom 5 (B_i)
- belief does not imply truth: $\varphi \wedge \mathrm{B}_i\ \not\varphi$ consistent

Belief: omniscience

belief set of agent i = set of formulas believed by i (at a given moment)

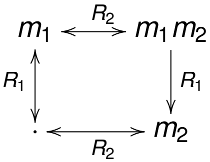
- i's belief set is. . .
  - closed under theorems:
    - ★ $\frac{\varphi}{\mathrm{B}_i\,\varphi}$ rule RN(B_i)
  - closed under logical implication:
    - ★ $\frac{\varphi \to \psi}{\mathrm{B}_i\,\varphi \to \mathrm{B}_i\,\psi}$ rule RM(B_i)
  - closed under material implication:
    - ★ $(\mathrm{B}_i\,\varphi \wedge \mathrm{B}_i\,(\varphi \to \psi)) \to \mathrm{B}_i\,\psi$ axiom K(B_i)
- *omniscience problem*
  - if I believe the axioms and inference rules of Peano Arithmetic then I either believe Goldbach's conjecture is true, or I either believe Goldbach's conjecture is false
  - *KD*45 is an idealization: rational agent, perfect reasoner
  - inadequate for human agents
  - widely accepted in AI

The logic of belief: semantics

- serial, transitive and Euclidian relation R_i

$$
\begin{aligned}
R_i(w) &= \text{"}i\text{'s alternatives to } w\text{"} \\
&= \text{"set of worlds } i \text{ cannot distinguish from } w\text{"} \\
&= \text{"set of worlds compatible with } i\text{'s belief"} \\
&= \text{"}\textit{belief state} \text{ of agent } i \text{ at } w\text{"}
\end{aligned}
$$

- variant of the muddy children puzzle: child 1 wrongly believes it is not muddy



$$M, (m_1 m_2) \Vdash m_1 \wedge m_1 \wedge \mathrm{B}_1 (\neg m_1 \wedge m_2) \wedge \neg \mathrm{B}_2 \mathrm{B}_1 \neg m_1 \wedge \neg \mathrm{B}_2 \mathrm{B}_1 m_1$$

The logic of belief: properties

- sound and complete: $\vdash_{KD45} \varphi$ iff $\models_{KD45} \varphi$
- decidable
- complexity of $KD45$-satisfiability:
  - NP-complete if $\text{card}(Agts) = 1$
  - PSPACE-complete if $\text{card}(Agts) > 1$
- normal form only for monoagent $KD45$ (but not if $\text{card}(Agts) > 1$!)
- difficult to extend by public announcements
  - what if agent i wrongly believes that p, and $\neg p$ is announced?
  - requires integration of belief revision ...

Excursion: graded belief

- language: $\mathrm{B}_i{}^{\geq d}\varphi$ = "*i* believes φ with degree at least *d*" $(d \in [0, 1])$
- semantics:

Excursion: graded belief

- language: $\mathrm{B}_i{}^{\geq d}\varphi$ = "i believes φ with degree at least d" ($d \in [0, 1]$)
- semantics: $M = \langle W, R, V \rangle$ where
  - $R : (Agts \times [0, 1]) \longrightarrow (W \times W)$ such that $R_i^{\geq d} \subseteq R_i^{\geq d+d'}$
    'system of spheres'

  $wR_i^{\geq d}v$ = "for i, at w world v has degree of possibility at least d"
- axiomatics:

Excursion: graded belief

- language: $\mathsf{B}_i{}^{\geq d}\varphi$ = "i believes φ with degree at least d" $(d \in [0,1])$
- semantics: $M = \langle W, R, V \rangle$ where
  - $R : (Agts \times [0,1]) \longrightarrow (W \times W)$ such that $R_i^{\geq d} \subseteq R_i^{\geq d+d'}$

    'system of spheres'

  $wR_i^{\geq d}v$ = "for i, at w world v has degree of possibility at least d"
- axiomatics:
  - $KD45(\mathsf{B}_i{}^{\geq d})$, for every i and d
  - $\mathsf{B}_i{}^{\geq d}\varphi \rightarrow \mathsf{B}_i{}^{\geq d'}\varphi$ if $d \geq d'$

Can knowledge be defined from belief?

[Plato, Theaetetus]

- $K_i\,\varphi \stackrel{\text{def}}{=} B_i\,\varphi \wedge \varphi$

Can knowledge be defined from belief?

[Plato, Theaetetus]

- $\mathrm{K}_i\,\varphi \stackrel{\text{def}}{=} \mathrm{B}_i\,\varphi \wedge \varphi$
  - problem: ‘knowledge by accident’
- $\mathrm{K}_i\,\varphi \stackrel{\text{def}}{=} \mathrm{B}_i\,\varphi \wedge \varphi \wedge \mathit{hasJustif}(i, \varphi)$

Can knowledge be defined from belief?

[Plato, Theaetetus]

- $K_i \varphi \stackrel{\text{def}}{=} B_i \varphi \wedge \varphi$
  - problem: 'knowledge by accident'
- $K_i \varphi \stackrel{\text{def}}{=} B_i \varphi \wedge \varphi \wedge \mathit{hasJustif}(i, \varphi)$
  - problem: what is a justification?
  - Gettier Problem [1963]:
    - ★ suppose a logic of belief and justification such that $\frac{\varphi \rightarrow \psi}{\mathit{hasJustif}(i,\varphi) \rightarrow \mathit{hasJustif}(i,\psi)}$
    - ★ suppose i wrongly believes p, but has some justification for that: $\neg p \wedge B_i\, p \wedge \mathit{hasJustif}(i, p)$
    - ★ ...hence i believes that $p \vee q$ and i believes that $p \vee \neg q$ (by axiom $M(B_i)$)
    - ★ ...and $\mathit{hasJustif}(i, (p \vee q))$ and $\mathit{hasJustif}(i, (p \vee \neg q))$ (use inference rule for *hasJustif*)
    - ★ ...**and either i knows that $p \vee q$, or i knows that $p \vee q$, for any q**: $\models B_i\, p \wedge \mathit{hasJustif}(i, p) \rightarrow (K_i (p \vee q) \vee K_i (p \vee \neg q))$

Relation between knowledge and belief?

- suppose a logic of knowledge and belief defined as:
  - $KD45(\mathrm{B}_i)$
  - $KT5(\mathrm{K}_i)$
  - $\mathrm{K}_i\,\varphi \rightarrow \mathrm{B}_i\,\varphi$
  - $\mathrm{B}_i\,\varphi \rightarrow \mathrm{B}_i\,\mathrm{K}_i\,\varphi$

Relation between knowledge and belief?

- suppose a logic of knowledge and belief defined as:
  - $KD45(\mathrm{B}_i)$
  - $KT5(\mathrm{K}_i)$
  - $\mathrm{K}_i\,\varphi \rightarrow \mathrm{B}_i\,\varphi$
  - $\mathrm{B}_i\,\varphi \rightarrow \mathrm{B}_i\,\mathrm{K}_i\,\varphi$
- ...but implies that $\mathrm{B}_i\,\varphi \leftrightarrow \mathrm{K}_i\,\varphi$!

Relation between knowledge and belief?

- suppose a logic of knowledge and belief defined as:
  - $KD45(\mathrm{B}_i)$
  - $KT5(\mathrm{K}_i)$
  - $\mathrm{K}_i\,\varphi \to \mathrm{B}_i\,\varphi$
  - $\mathrm{B}_i\,\varphi \to \mathrm{B}_i\,\mathrm{K}_i\,\varphi$
- ...but implies that $\mathrm{B}_i\,\varphi \leftrightarrow \mathrm{K}_i\,\varphi$!
  - intermediate step: $\neg\mathrm{B}_i\,\neg\mathrm{K}_i\,\varphi \to \neg\mathrm{K}_i\,\neg\mathrm{B}_i\,\varphi$

Relation between knowledge and belief?

- suppose a logic of knowledge and belief defined as:
  - $KD45(\mathrm{B}_i)$
  - $KT5(\mathrm{K}_i)$
  - $\mathrm{K}_i\,\varphi \rightarrow \mathrm{B}_i\,\varphi$
  - $\mathrm{B}_i\,\varphi \rightarrow \mathrm{B}_i\,\mathrm{K}_i\,\varphi$
- ...but implies that $\mathrm{B}_i\,\varphi \leftrightarrow \mathrm{K}_i\,\varphi$!
  - intermediate step: $\neg\mathrm{B}_i\,\neg\mathrm{K}_i\,\varphi \rightarrow \neg\mathrm{K}_i\,\neg\mathrm{B}_i\,\varphi$
- culprit: negative introspection for knowledge