"Individual and collective intentionality" Introductory course @ ESSLLI'09

Andreas Herzig and Emiliano Lorini
IRIT-CNRS, University of Toulouse, France
www.irit.fr/~Andreas.Herzig/Esslli09course
herzig@irit.fr,lorini@irit.fr

Bordeaux, July 2009

Introduction to the course

Agents and mental states

- agents in interaction: physical vs. *mental* fact [Brentano]
- concept of *mental state* of an agent
  - philosophy (philosophy of mind, epistemology)
  - psychology
  - economics
  - computer science (MAS, AI, distributed systems)
- many kinds of mental states of an individual i:
  - i is angry; is sad; ...
    - no argument (moods, needs)
  - i loves individual j; hates individual k; ...
    - argument = object
  - i is angry *that* p; believes that p; has goal that $\neg p$; admires j for having brought about p;...
    - argument = proposition

Intentionality

- intentional mental state = mental states that are *about* something [Brentano, Searle]
  - about an object
  - *about a proposition* [Anscombe]
- two kinds of ‘being about a proposition’
  - informational mental states:
    - knowledge
    - beliefs
    - acceptance ($\neq$ belief)
  - motivational (proactive, teleological) mental states:
    - desires
    - preferences
    - goals
    - standards, values (internalized norms)
    - future-directed intentions
    - present-directed intentions (plans)

Collective intentionality

- examples:
  - "The team believes it will win today's game."
  - "The British believe that the Euro will eventually be introduced in the UK."
  - "The United States believe that those responsible for these dreadful acts must be punished."
  - "The Communist Party of Ruritania believes that capitalist countries will soon perish (but none of its members really believes so)."
  - "Microsoft intends to strengthen its position in the market."
- collective mind? collective consciousness?
  - metaphorical
  - 'intentional stance': *ascribe* mental attitudes to groups [Davidson, Dennett]
    - Microsoft's attempt to take over Yahoo can be explained by Microsoft's desire to strengthen its position in the market, and Microsoft's belief that it is able to buy Yahoo

Individual attitudes: the issues

- how represented in the agent's mind?
  - language of thought [Fodor]
- which informational attitudes?
  - knowledge, belief, acceptance
    - ★ knowledge implies truth, belief doesn't
    - ★ knowledge that φ = (belief that φ) $\wedge \varphi \wedge \ldots$
  - which logical principles?
    - ★ omniscience problem
- which proactive attitudes ?
  - desires, preferences, goals, intentions
    - ★ desires primitive?
    - ★ conscious?
  - can be formalized in logic?
- which dynamics?
  - mental attitudes trigger actions
  - (perception of) events triggers change of mental attitudes

Collective attitudes: the issues

- status?
  - exist, or just ascribed?
  - for any set of agents, or just for *constituted groups*?
  - reducible to individual attitudes?
- which informational attitudes?
  - shared knowledge, shared belief
  - distributed knowledge, distributed belief
  - common knowledge, common belief ('mutual belief')
  - *group* acceptance
    - group: more than a set of agents
- which proactive attitudes?
  - collective actions, plans, strategies
    - parallel (independence) vs. joint (moving a piano)
  - collective goals and intentions ('we-intentions', 'teamwork')
- which dynamics?
  - results from dynamics of individual attitudes?

Related @ ESSLLI 2009

- Week 1
  - "Logical Methods for Social Concepts"
    - ★ workshop, A. Herzig and E. Lorini
  - "Games, Actions and Social Software"
    - ★ introductory course, R. Verbrugge and J. van Eijck
- Week 2
  - "Logics of Rational Agency"
    - ★ foundational course, E. Pacuit
  - "Logic and Agent Programming Languages"
    - ★ introductory course, N. Alechina and B. Logan
  - "Dynamic Logics for Interactive Belief Revision"
    - ★ advanced course, A. Baltag and S. Smets

Course overview

Monday	**epistemic logic and its dynamics**
Tuesday	doxastic logic and its dynamics
Wednesday	logic of goals and intentions
Thursday	common belief, group belief and group acceptance
Friday	group action, group intention

Monday: Epistemic logics and the dynamics of knowledge

Plan

Reasoning about knowledge: *de dicto* vs. *de re*

(1) "*there are* irrational x and y such that x^y is rational"
(2) "Hilbert *knows that there are* irrational x, y such that x^y is rational"
(3) "*there are* irrational x, y such that Hilbert *knows* that x^y is rational"

- write these statements in the language of logic
  - abbreviate $\neg Rat(x) \wedge \neg Rat(y) \wedge Rat(x^y)$ by $P(x, y)$

Reasoning about knowledge: *de dicto* vs. *de re*

(1) "*there are* irrational x and y such that x^y is rational"
(2) "Hilbert *knows that there are* irrational x, y such that x^y is rational"
(3) "*there are* irrational x, y such that Hilbert *knows* that x^y is rational"

- write these statements in the language of logic
  - abbreviate $\neg Rat(x) \wedge \neg Rat(y) \wedge Rat(x^y)$ by $P(x, y)$
- it follows from the axioms of Peano Arithmetic that $\exists x \exists y P(x, y)$
  - non-constructive proof (5 lines)
- Hilbert knew Peano Arithmetic
- Hilbert knew that $\exists x \exists y P(x, y)$
- there are no x, y of which Hilbert knew that $P(x, y)$
  - although there is a constructive proof ($\sim$20 pages, $\sim$1950)
  - Hilbert was not a perfect, 'omniscient' reasoner

Reasoning about knowledge: muddy children

a famous puzzle:

1. two children come back from the garden, both with mud on their forehead; their father looks at them and says:
   "at least one of you has mud on his forehead"
   then he asks:
   "those who know whether they are dirty, step forward!"
2. nobody steps forward
3. the father asks again:
   "those who know whether they are dirty, step forward!"
4. both simultaneously answer: *"I know!"*

N.B.: can be generalized to an arbitrary number $n \geq 2$ of children

Reasoning about knowledge: muddy children

- use (second-order) predicate $Knows(i, \varphi)$, where $i \in \{1, 2\}$
  - $Knows(i, \varphi)$ = "agent i knows that φ"
- some of child 2's knowledge at the different stages:

(S0) background knowledge:
$Knows(2,\ Knows(1, m_2) \vee Knows(1, \neg m_2))$
equivalently:
$Knows(2,\ \neg Knows(1, \neg m_2) \rightarrow Knows(1, m_2))$

(S1) learns that at least one of them has mud on his forehead:
$Knows(2, Knows(1, (m_1 \vee m_2)))$

(S2) child 2 does not respond:
$Knows(2, \neg Knows(1, m_1))$

(S3) should follow from (S0)-(S2):
$Knows(2, m_2)$

- proof?

Reasoning about knowledge: muddy children

deduction of (S3) from (S0), (S1), (S2):

1. $Knows(2,\ Knows(1, (m_1 \vee m_2)))$ — hyp. (S1)
2. $Knows(2,\ Knows(1, \neg m_2) \to Knows(1, m_1))$ — conseq. of 1.
3. $Knows(2,\ \neg Knows(1, m_1) \to \neg Knows(1, \neg m_2))$ — equiv. to 2.
4. $Knows(2,\ \neg Knows(1, m_1))$ — hyp. (S2)
5. $Knows(2,\ \neg Knows(1, \neg m_2))$ — from 3. and 4.
6. $Knows(2,\ \neg Knows(1, \neg m_2) \to Knows(1, m_2))$ — equiv. to hyp. (S0)
7. $Knows(2,\ Knows(1, m_2))$ — from 5. and 6.
8. $Knows(2,\ m_2)$ — from 7., bec. $Knows(1, m_2) \to m_2$ ('knowledge implies truth')

informal deduction $\Rightarrow$ formal rules? $\Rightarrow$ deduction in a formal logic?

A second-order theory of the $Knows$ predicate

- desirable principles:
  - $\forall i \forall p \, (Knows(i,p) \to p)$
    - used in step 8.
  - $\forall i \forall p \forall q \, ((Knows(i, p \vee q) \wedge Knows(i, \neg p)) \to Knows(i,q))$
    - used in step 2.
  - …
- make up theory of knowledge $\mathcal{T}_{Knows}$
  - second-order formulas: "$\forall p$" quantifies over propositions
- reasoning about knowledge in second-order logic (SOL):
  - $T_{Knows} \vdash_{SOL} ((S0) \wedge (S1) \wedge (S2)) \to (S3)$
  - SOL consequence problem
    - undecidable …

$Knows$: from second-order to first-order logic

idea [Hin62, FHMV95]:

$Knows(i, \varphi)$ = "φ true in all worlds that are possible for i"

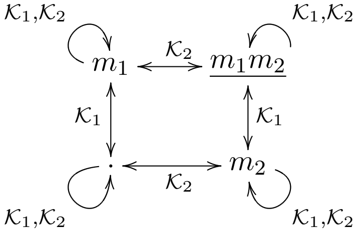
- set of possible worlds W
- ternary 'accessibility' relation $\mathcal{K}(i, w_1, w_2)$
  - i = agent
  - w_1 = actual world
  - w_2 = world that i *cannot distinguish* from w_1
- in first-order logic:

$$Knows(i, \varphi, w) \quad = \quad \text{"at } w, i \text{ knows that } \varphi\text{"}$$
$$\stackrel{\text{def}}{=} \quad \forall w' \; (R(i, w, w') \to \varphi[w'])$$

$Knows$: from second-order to first-order logic, ctd.

- muddy children:
  - $Knows(1, m_2, w) \;=\; \forall w' \; (R(1, w, w') \to m_2(w'))$
  - $\neg Knows(1, m_1, w) \;=\; \exists w' \; (R(1, w, w') \land \neg m_1(w'))$
- exercise: draw the set of possible worlds and the accessibility relation in the initial situation

- muddy children:
  - $Knows(1, m_2, w) \ = \ \forall w' \ (R(1, w, w') \rightarrow m_2(w'))$
  - $\neg Knows(1, m_1, w) \ = \ \exists w' \ (R(1, w, w') \wedge \neg m_1(w'))$
- exercise: draw the set of possible worlds and the accessibility relation in the initial situation



$Knows$: from second-order to first-order logic, ctd.

- desirable principles for knowledge $\Rightarrow$ properties of $\mathcal{K}$
  - $\forall i \forall p \, (Knows(i,p) \to p)$ corresponds to: $\forall i \forall w \, \mathcal{K}(i,w,w)$
  - ...
- make up first-order theory $\mathcal{T}_{Knows}$
- reasoning about knowledge:
  - $T_{Knows} \vdash_{FOL} \forall w \, (((S0) \wedge (S1) \wedge (S2)) \to (S3))[w]$
  - consequence problem in first-order logic (FOL)
    - ★ semi-decidable ...

$Knows$: from first-order to modal logic

idea [Hin62]:

don't use first-order language, but add **modal operators of knowledge** to the language of classical propositional logic CPL

- K_i : modal operator
- $\mathrm{K}_i\, \varphi$ = "i knows that φ"
- epistemic logic
  - *episteme* = $\epsilon\pi\iota\sigma\tau\eta\mu\eta$ = 'know' (Greek)
- N.B.:
  - propositional language; no $\forall$, $\exists$
  - φ might contain modal operator K_j
    - ★ precise definition requires recursive definition of language

Epistemic language: examples

- knowing-whether:
  - $\mathsf{K}_1\, m_2 \vee \mathsf{K}_1\, \neg m_2$ "child 1 knows whether m_2"
- ignorance:
  - $\neg\mathsf{K}_2\, m_2 \wedge \neg\mathsf{K}_2\, \neg m_2$ "child 2 does not know whether m_2"
- nesting of modal operators ('higher-order knowledge'):
  - $\mathsf{K}_1\, \mathsf{K}_2\, (m_1 \vee m_2)$
  - $\mathsf{K}_1\, \mathsf{K}_2\, \mathsf{K}_1\, (m_1 \vee m_2)$
  - ...
  - $\mathsf{K}_2\, (\mathsf{K}_1\, m_2 \vee \mathsf{K}_1\, \neg m_2)$
  - $\mathsf{K}_2\, (\neg\mathsf{K}_1\, m_1 \wedge (\mathsf{K}_1\, m_2 \vee \mathsf{K}_1\, \neg m_2))$

The propositional logic of knowledge

- extend CPL by axiom schemas and inference rules for the modal operator K_i
  - $\vdash \mathsf{K}_i\,\varphi \to \varphi$
  - if $\vdash \varphi$ then $\vdash \mathsf{K}_i\,\varphi$
  - ...
- reasoning about knowledge:
  - $\vdash \mathsf{K}_2\,\mathsf{K}_1\,m_2 \to \mathsf{K}_2\,m_2$
  - $\vdash ((S0) \land (S1) \land (S2)) \to (S3)$
  - ...
  - reasoning problem: given φ, do we have $\vdash \varphi$?
    - ★ decidable!
    - ★ more details later ...

Reasoning in epistemic logic

- semantics: models? truth conditions?
  - resort to first-order semantics in terms of possible worlds
  - $M = \langle W, \mathcal{K}, V \rangle$ where
    - W some set ('possible worlds')
    - $\mathcal{K} : Agts \times W \times W$
    - V valuation
  - truth conditions:
    - $M, w \Vdash \mathsf{K}_i\, \varphi$ iff $M, w' \Vdash \varphi$ for all w' such that $\mathcal{K}(i, w, w')$
  - N.B.: language of epistemic logic less expressive than that of FOL
    - $\exists$ different models that give same truth value to all formulas
    - cannot be distinguished by means of a formula
    - bisimulation ...

Recap of basic logic notions

- **logic** Λ = language $\mathcal{L}_\Lambda$ + *particular subset* of $\mathcal{L}_\Lambda$ (called theorems or validities)
- *particular subset* of $\mathcal{L}_\Lambda$ can be characterized in two ways:
  - semantically: using models $\Rightarrow$ validities
  - syntactically: using axioms and inference rules $\Rightarrow$ theorems

Recap of basic logic notions: axiomatics

- requires:
  1. **axiom schemas** = basic theorems of the logic
     - ★ in an axiom schema, we can perform *uniform substitutions*: $\mathsf{K}_i\, \varphi \rightarrow \varphi$ instantiates to: $\mathsf{K}_1\,(m_2 \vee m_1) \rightarrow (m_2 \vee m_1)$
     - ★ N.B.: the φ are *meta-variables* over the language
  2. **inference rules** = generate new theorems from existing theorems
     - ★ notation: $\{\varphi_1, \ldots, \varphi_m\} \,/\, \varphi$, or: $\frac{\varphi_1, \ldots, \varphi_m}{\varphi}$
- a **proof** of φ in Λ is a sequence of formulas $\langle \varphi_1, \ldots, \varphi_n \rangle$ such that $\varphi_n = \varphi$, and for every $i \leq n$:
  - φ_i is an (instance of) some axiom schema for Λ, or
  - there are formulas φ_{i_1},..., φ_{i_m}, such that $i_j < i$, and $\frac{\varphi_{i_1}, \ldots, \varphi_{i_m}}{\varphi_i}$ is (an instance of) some inference rule for Λ
- φ is a **theorem** of Λ iff φ is provable in Λ
  - notation: $\vdash_\Lambda \varphi$
- φ is **consistent** in Λ iff $\not\vdash_\Lambda \neg\varphi$
- **deductions** $\Gamma \vdash_\Lambda \varphi$ iff ... (several options in modal logic)

Recap of basic logic notions: semantics

- requires:
  1. a **class of models** M for Λ
  2. **truth conditions**: when is φ true in M?
     - ★ notation in general: $M \Vdash \varphi$
     - ★ in modal logic: $M, w \Vdash \varphi$ ‘φ is true in $\langle M, w \rangle$’
- φ is **valid** in Λ iff $M, w \Vdash \varphi$, for every model M for Λ and world w in M
  - notation: $\models_\Lambda \varphi$
- φ is **satisfiable** in Λ iff $\not\models_\Lambda \neg\varphi$
- **logical consequence** $\Gamma \models_\Lambda \varphi$ iff ... (several options in modal logic)

Recap of basic logic notions: soundness and completeness

syntactic and semantic characterizations should coincide!

- **soundness**: for every formula φ, if $\vdash_\Lambda \varphi$ then $\models_\Lambda \varphi$
  - proof by induction on the length of the proof of φ
- **completeness**: for every formula φ, if $\models_\Lambda \varphi$ then $\vdash_\Lambda \varphi$
  - actually proved: 'if φ is consistent in Λ then φ is satisfiable in Λ'
  - non-constructive proofs: canonical models [Henkin]
  - constructive proofs: via tableau method

Plan

1. Multiagent epistemic logic $S5_n$
   - Introduction
   - Language
   - Semantics
   - Axiomatics
2. Discussions
3. Public announcement logic PAL
4. Dynamic epistemic logic DEL

Language

- primitive symbols:
  - countable set of propositional atoms $Atms$
  - finite set of agent symbols $Agts$
- BNF:

$$\varphi \ ::= \ p \mid \neg\varphi \mid \varphi \wedge \varphi \mid \mathsf{K}_i \, \varphi$$

where p ranges over $Atms$ and i over $Agts$

Language

- primitive symbols:
  - countable set of propositional atoms $Atms$
  - finite set of agent symbols $Agts$
- BNF:

$$\varphi ::= p \mid \neg\varphi \mid \varphi \wedge \varphi \mid \mathsf{K}_i \, \varphi$$

where p ranges over $Atms$ and i over $Agts$

- abbreviations:
  - $\varphi \vee \psi \stackrel{\text{def}}{=} \neg(\neg\varphi \wedge \neg\psi)$
  - $\varphi \rightarrow \psi \stackrel{\text{def}}{=} \ldots$
  - $\varphi \leftrightarrow \psi \stackrel{\text{def}}{=} \ldots$

Language

- primitive symbols:
  - countable set of propositional atoms $Atms$
  - finite set of agent symbols $Agts$
- BNF:

$$\varphi ::= p \mid \neg\varphi \mid \varphi \wedge \varphi \mid \mathsf{K}_i\,\varphi$$

where p ranges over $Atms$ and i over $Agts$

- abbreviations:
  - $\varphi \vee \psi \stackrel{\text{def}}{=} \neg(\neg\varphi \wedge \neg\psi)$
  - $\varphi \rightarrow \psi \stackrel{\text{def}}{=} \ldots$
  - $\varphi \leftrightarrow \psi \stackrel{\text{def}}{=} \ldots$
  - $\hat{\mathsf{K}}_i\,\varphi \stackrel{\text{def}}{=} \neg\mathsf{K}_i\,\neg\varphi$ = "φ is possible for i"

Language (ctd.)

- 3 possible *epistemic attitudes* w.r.t. a formula φ:

$\mathsf{K}_i\,\varphi$	$\hat{\mathsf{K}}_i\,\varphi \wedge \hat{\mathsf{K}}_i\,\neg\varphi$	$\mathsf{K}_i\,\neg\varphi$

Language (ctd.)

- 3 possible *epistemic attitudes* w.r.t. a formula φ:

$$\mathsf{K}_i\,\varphi \qquad \hat{\mathsf{K}}_i\,\varphi \wedge \hat{\mathsf{K}}_i\,\neg\varphi \qquad \mathsf{K}_i\,\neg\varphi$$

  - φ should be *contingent*: neither theorem nor inconsistent

Language (ctd.)

- 3 possible *epistemic attitudes* w.r.t. a formula φ:

$\mathsf{K}_i\, \varphi$	$\hat{\mathsf{K}}_i\, \varphi \land \hat{\mathsf{K}}_i\, \neg\varphi$	$\mathsf{K}_i\, \neg\varphi$

  - φ should be *contingent*: neither theorem nor inconsistent
  - what if φ of the form $\mathsf{K}_i\, \psi$?

Language (ctd.)

- 3 possible *epistemic attitudes* w.r.t. a formula φ:

$\mathsf{K}_i\, \varphi$	$\hat{\mathsf{K}}_i\, \varphi \wedge \hat{\mathsf{K}}_i\, \neg\varphi$	$\mathsf{K}_i\, \neg\varphi$

  - φ should be *contingent*: neither theorem nor inconsistent
  - what if φ of the form $\mathsf{K}_i\, \psi$?
  - what if φ of the form $\hat{\mathsf{K}}_i\, \psi$?

Language (ctd.)

- 3 possible *epistemic attitudes* w.r.t. a formula φ:

$\mathsf{K}_i\,\varphi$	$\hat{\mathsf{K}}_i\,\varphi \wedge \hat{\mathsf{K}}_i\,\neg\varphi$	$\mathsf{K}_i\,\neg\varphi$

  - φ should be *contingent*: neither theorem nor inconsistent
  - what if φ of the form $\mathsf{K}_i\,\psi$?
  - what if φ of the form $\hat{\mathsf{K}}_i\,\psi$?

- 4 possible *epistemic situations* w.r.t. a formula φ:

$\varphi \wedge \mathsf{K}_i\,\varphi$	$\varphi \wedge \hat{\mathsf{K}}_i\,\varphi \wedge \hat{\mathsf{K}}_i\,\neg\varphi$	
	$\neg\varphi \wedge \hat{\mathsf{K}}_i\,\varphi \wedge \hat{\mathsf{K}}_i\,\neg\varphi$	$\neg\varphi \wedge \mathsf{K}_i\,\neg\varphi$

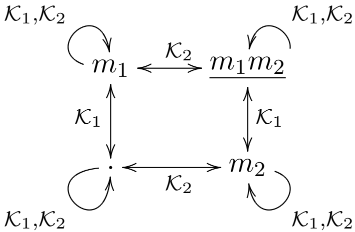
  - ... for φ contingent and non-epistemic
  - why are situations $\varphi \wedge \mathsf{K}_i\,\neg\varphi$ and $\neg\varphi \wedge \mathsf{K}_i\,\varphi$ missing?

Plan

1. Multiagent epistemic logic $S5_n$
   - Introduction
   - Language
   - Semantics
   - Axiomatics
2. Discussions
3. Public announcement logic PAL
4. Dynamic epistemic logic DEL

Semantics of $S5_n$: Kripke models

- 'Saul **K**ripke' [Kri59]
- $S5_n$**-model** = labeled graph $\langle W, \mathcal{K}, V \rangle$ where:
  - W nonempty set — 'possible worlds', 'states'
  - $\mathcal{K} : Agts \longrightarrow 2^{W \times W}$ such that every $\mathcal{K}_i$ is an *equivalence relation*
    - equivalence relation = reflexive, transitive, and symmetric relation
    - write $\mathcal{K}_i$ instead of $\mathcal{K}(i)$ — 'accessibility relation for i'
  - $V : Atms \longrightarrow 2^W$ — 'valuation'
    - $V(p) \subseteq W$
- muddy children:

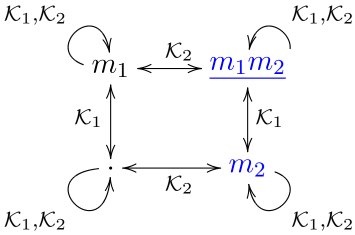


Semantics of $S5_n$: models

- equivalence relation = indistinguishability

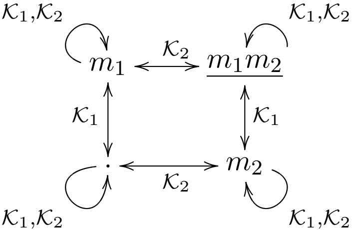
$$
\begin{aligned}
\mathcal{K}_i(m_1 m_2) &= \{w : \langle m_1 m_2, w\rangle \in \mathcal{K}_i\} \\
&= \text{“set of worlds } i \text{ cannot distinguish from } m_1 m_2\text{”} \\
&= \text{“set of worlds compatible with } i\text{'s knowledge”} \\
&= \text{“\textit{knowledge state} of agent } i \text{ at } m_1 m_2\text{”}
\end{aligned}
$$

- muddy children:



Semantics of $S5_n$: truth conditions

- truth in a pointed model:
  - $M, w \Vdash p$ iff $w \in V(p)$
  - $M, w \Vdash \neg\varphi$ iff $M, w \nVdash \varphi$
  - $M, w \Vdash \varphi \land \psi$ iff $M, w \Vdash \varphi$ and $M, w \Vdash \psi$
  - $M, w \Vdash \mathsf{K}_i\, \varphi$ **iff** $M, w' \Vdash \varphi$ **for *every*** $w' \in \mathcal{K}_i(w)$
    - hence: $M, w \Vdash \hat{\mathsf{K}}_i\, \varphi$ iff $M, w' \Vdash \varphi$ for *some* $w' \in \mathcal{K}_i(w)$
- muddy children:



$$M, (m_1 m_2) \Vdash m_1 \land m_2 \land \mathsf{K}_1\, m_2 \land \hat{\mathsf{K}}_1\, m_1 \land \hat{\mathsf{K}}_1\, \neg m_1$$

Semantics of $S5_n$: satisfiability and validity

- φ is $S5_n$**-satisfiable** iff $M, w \Vdash \varphi$ for *some* $S5_n$-model $M = \langle W, \mathcal{K}, V \rangle$ and *some* possible world $w \in W$

- φ is $S5_n$**-valid** ($\models_{S5_n} \varphi$) iff $M, w \Vdash \varphi$ for *every* $S5_n$-model $M = \langle W, \mathcal{K}, V \rangle$ and *every* possible world $w \in W$

Plan

1. Multiagent epistemic logic $S5_n$
   - Introduction
   - Language
   - Semantics
   - Axiomatics
2. Discussions
3. Public announcement logic PAL
4. Dynamic epistemic logic DEL

Axiomatics of $S5_n$

- axiom schemas for $S5_n$:
  - every theorem schema of *classical propositional logic* (CPL)
  - $(\mathsf{K}_i\,\varphi \wedge \mathsf{K}_i\,\psi) \rightarrow \mathsf{K}_i\,(\varphi \wedge \psi)$ conjunction C(K_i)
  - $\mathsf{K}_i\,\top$ necessity N(K_i)
  - $\mathsf{K}_i\,\varphi \rightarrow \varphi$ truth T(K_i)
  - $\mathsf{K}_i\,\varphi \rightarrow \mathsf{K}_i\,\mathsf{K}_i\,\varphi$ pos. introspection 4(K_i)
  - $\neg\mathsf{K}_i\,\varphi \rightarrow \mathsf{K}_i\,\neg\mathsf{K}_i\,\varphi$ neg. introspection 5(K_i)
- inference rules for $S5_n$:
  - $\frac{\varphi,\quad \varphi\rightarrow\psi}{\psi}$ modus ponens (MP)
  - $\frac{\varphi\rightarrow\psi}{\mathsf{K}_i\,\varphi\rightarrow\mathsf{K}_i\,\psi}$ rule of monotony RM(K_i)
- N.B.: in axiom schemas and rules, φ, ψ and i are meta-variables
- $S5_n$-proof, $S5_n$-theorem: as usual
- we say:
  - "CPL+C(K_i)+N(K_i)+RM(K_i)+T(K_i)+4(K_i)+5(K_i) *axiomatizes* $S5_n$"

Axiomatics of $S5_n$: examples of theorems

- $\vdash_{S5_n} \mathsf{K}_i \varphi \rightarrow \mathsf{K}_i \varphi$
  - proof:
    1. $\mathsf{K}_i \varphi \rightarrow \mathsf{K}_i \varphi$ (CPL)
- $\vdash_{S5_n} \mathsf{K}_i (\varphi \wedge \psi) \rightarrow \mathsf{K}_i \varphi$
  - proof:
    1. $(\varphi \wedge \psi) \rightarrow \varphi$ (CPL)
    2. $\mathsf{K}_i (\varphi \wedge \psi) \rightarrow \mathsf{K}_i \varphi$ from 1. by RM(K_i)
- $\vdash_{S5_n} \mathsf{K}_i (\varphi \wedge \psi) \rightarrow \mathsf{K}_i \psi$
  - proof: ...

Axiomatics of $S5_n$: examples of theorems, ctd.

- $\vdash_{S5_n} \mathsf{K}_i (\varphi \land \psi) \to (\mathsf{K}_i \varphi \land \mathsf{K}_i \psi)$
  - proof:
    1. $\mathsf{K}_i (\varphi \land \psi) \to \mathsf{K}_i \varphi$ — v.s.
    2. $\mathsf{K}_i (\varphi \land \psi) \to \mathsf{K}_i \psi$ — v.s.
    3. $1 \to (2 \to (\mathsf{K}_i (\varphi \land \psi) \to (\mathsf{K}_i \varphi \land \mathsf{K}_i \psi)))$ — (CPL)
    4. $2 \to (\mathsf{K}_i (\varphi \land \psi) \to (\mathsf{K}_i \varphi \land \mathsf{K}_i \psi))$ — from 1. and 3. by (MP)
    5. $\mathsf{K}_i (\varphi \land \psi) \to (\mathsf{K}_i \varphi \land \mathsf{K}_i \psi)$ — from 2. and 4. by (MP)
- $\vdash_{S5_n} \mathsf{K}_i (\varphi \land \psi) \leftrightarrow (\mathsf{K}_i \varphi \land \mathsf{K}_i \psi)$
  - proof: ...

Axiomatics of $S5_n$: some useful theorems

- **Rule of Necessitation** RN(K_i): $\frac{\varphi}{K_i \varphi}$
  ("for all φ, if $\vdash_{S5_n} \varphi$ then $\vdash_{S5_n} K_i \varphi$")
  - proof:
    1. φ — by hyp.
    2. $\varphi \to (\top \to \varphi)$ — (CPL)
    3. $\top \to \varphi$ — from 1. and 2. by (MP)
    4. $K_i \top \to K_i \varphi$ — from 3. by RM(K_i)
    5. $K_i \top$ — N(K_i)
    6. $K_i \varphi$ — from 4. and 5. by (MP)
  - N.B.: shorter proof using *derived* CPL *inference rules*:
    1. φ — by hyp.
    2. $\top \to \varphi$ — from 1. **by (CPL)**
    3. $K_i \top \to K_i \varphi$ — from 2. by RM(K_i)
    4. $K_i \top$ — N(K_i)
    5. $K_i \varphi$ — from 3. and 4. **by (CPL)**

Axiomatics of $S5_n$: some useful theorems

- **Rule of Equivalence** RE(K_i): $\frac{\varphi \leftrightarrow \psi}{\mathsf{K}_i \varphi \leftrightarrow \mathsf{K}_i \psi}$
  ("for all φ, if $\vdash_{S5_n} \varphi \leftrightarrow \psi$ then $\vdash_{S5_n} \mathsf{K}_i \varphi \leftrightarrow \mathsf{K}_i \psi$")
  - proof:
    1. $\varphi \leftrightarrow \psi$ — by hyp.
    2. $\varphi \rightarrow \psi$ — from 1. by (CPL)
    3. $\mathsf{K}_i \varphi \rightarrow \mathsf{K}_i \psi$ — from 2. by RM(K_i)
    4. $\psi \rightarrow \varphi$ — from 1. by (CPL)
    5. $\mathsf{K}_i \psi \rightarrow \mathsf{K}_i \varphi$ — from 4. by RM(K_i)
    6. $\mathsf{K}_i \varphi \leftrightarrow \mathsf{K}_i \psi$ — from 3. and 5. by (CPL)

Axiomatics of $S5_n$: some useful theorems, ctd.

- **Rule of Replacement of Proved Equivalents** (REq):

$$\frac{\psi \leftrightarrow \psi'}{\varphi[p/\psi] \leftrightarrow \varphi[p/\psi']}$$

(where $\varphi[p/\psi]$ obtained from φ by replacing every occurrence of p by ψ, etc.)

  - proof by induction on the *structure* of φ:
    1. φ atomic: then $\psi = \varphi$, and $\varphi' = \psi'$
    2. $\varphi = \neg\varphi_1$: if $\psi = \varphi$ then $\varphi' = \psi'$; else $\psi \in sf(\varphi_1)$; ...
    3. $\varphi = \varphi_1 \wedge \varphi_2$: ...
    4. $\varphi = \mathsf{K}_i\, \varphi_1$: ...

Axiomatics of $S5_n$: some useful theorems, ctd.

- **Kripke's axiom** K(K$_i$): $\vdash_{S5_n} \mathsf{K}_i (\varphi \to \psi) \to (\mathsf{K}_i \varphi \to \mathsf{K}_i \psi)$
  - proof:
    1. $(\mathsf{K}_i \varphi \land \mathsf{K}_i (\varphi \to \psi)) \to \mathsf{K}_i (\varphi \land (\varphi \to \psi))$ — C(K$_i$)
    2. $(\varphi \land (\varphi \to \psi)) \to \psi$ — (CPL)
    3. $\mathsf{K}_i (\varphi \land (\varphi \to \psi)) \to \mathsf{K}_i \psi$ — from 2. by RM(K$_i$)
    4. $(\mathsf{K}_i \varphi \land \mathsf{K}_i (\varphi \to \psi)) \to \mathsf{K}_i \psi$ — from 1. and 3. by (CPL)
    5. $\mathsf{K}_i (\varphi \to \psi) \to (\mathsf{K}_i \varphi \to \mathsf{K}_i \psi)$ — from 4. by (CPL)

- $\vdash_{S5_n} (\mathsf{K}_i \varphi \land \hat{\mathsf{K}}_i \psi) \to \hat{\mathsf{K}}_i (\varphi \land \psi)$
  - proof: . . . — hint: use (REq) and K(K$_i$)

Axiomatics of $S5_n$: soundness and completeness

Soundness Theorem.
If $\vdash_{S5_n} \varphi$ then $\models_{S5_n} \varphi$.

Proof.
We prove: if there is a $S5_n$-proof $\langle \varphi_1, \ldots, \varphi_n \rangle$ of φ then $\models_{S5_n} \varphi$.
We proceed by induction on n.

Base case: If $n = 1$ then φ is an instance of an axiom schema. We prove that every such instance is valid.
Let M be any $S5_n$-model, and w any world in M.

- Axiom N(K$_i$) is $S5_n$-valid:
  $M, w \Vdash \mathsf{K}_i \top$ because $M, w' \Vdash \top$ for every w'.
- Every instance of axiom schema C(K$_i$) :
  $(\mathsf{K}_i \varphi \wedge \mathsf{K}_i \psi) \rightarrow \mathsf{K}_i (\varphi \wedge \psi)$ is $S5_n$-valid:
  suppose $M, w \Vdash \mathsf{K}_i \varphi \wedge \mathsf{K}_i \psi$;
  then both φ and ψ are true in every world $w' \in \mathcal{K}_i(w)$;
  therefore $\varphi \wedge \psi$ is true in every $w' \in \mathcal{K}_i(w)$.
- ...

Axiomatics of $S5_n$: soundness and completeness, ctd.

(*Proof of Soundness Theorem, ctd.*)

Induction hypothesis (I.H.): For all $m < n$, if $\langle \varphi_1, \ldots, \varphi_m \rangle$ is a $S5_n$-proof of φ then $\models_{S5_n} \varphi$.

Induction step: Let $\langle \varphi_1, \ldots, \varphi_n \rangle$ be a $S5_n$-proof of φ. We do a case analysis, checking the possible ways φ_n is obtained:

- φ_n is an instance of an axiom schema.
  Then we already know that $\models_{S5_n} \varphi$.
- φ_n is obtained from some φ_k, $k < n$, via RM(K_i).
  Then $\varphi_k = \psi \to \chi$ and $\varphi_n = \mathsf{K}_i (\psi \to \chi)$, and
  $\langle \varphi_1, \ldots, \varphi_k \rangle$ is a $S5_n$-proof of φ_k.
  By I.H., $\models_{S5_n} \psi \to \chi$, i.e. $M, w \Vdash \psi \to \chi$ for every $S5_n$-model M and every world w in M. Therefore we must have
  $\models_{S5_n} \mathsf{K}_i (\psi \to \chi)$. "RM($\mathsf{K}_i$) preserves validity"
- φ_n is obtained from some φ_k and $\varphi_l = \varphi_k \to \varphi_n$ via (MP).
  … "(MP) preserves validity"

Axiomatics of $S5_n$: soundness and completeness, ctd.

Weak Completeness Theorem.
If $\models_{S5_n} \varphi$ then $\vdash_{S5_n} \varphi$.

Proof.
follows from more general result: Sahlqvist's completeness theorem

Decidability and complexity Theorem.
The problem of $S5_n$-satisfiability of a formula φ can be decided in polynomial space (PSPACE).

Proof.
using the tableau procedure

- $n > 1$: requires indeed polynomial space in the worst case
  - $S5_n$ is PSPACE-complete for $n > 1$
- $n = 1$: decidable in nondeterministic polynomial time (NP)
  - $S5_1$ is NP-complete (because CPL already NP-hard)

Axiomatics of $S5_n$: an equivalent axiomatization

Theorem.
The logic $S5_n$ is also axiomatized by CPL+K(K$_i$)+RN(K$_i$).

Proof.
We have to show:

- φ can be proved from CPL+C(K$_i$)+N(K$_i$)+RM(K$_i$) iff φ can be proved from CPL+K(K$_i$)+RN(K$_i$).

For that, it will suffice to prove:

- that CPL+C(K$_i$)+N(K$_i$)+RM(K$_i$)
  - has theorem K(K$_i$): $\mathsf{K}_i(\varphi \to \psi) \to (\mathsf{K}_i\varphi \to \mathsf{K}_i\psi)$
  - has derived rules (MP) and RN(K$_i$): $\frac{\varphi}{\mathsf{K}_i\varphi}$
- that CPL+K(K$_i$)+RN(K$_i$)
  - has theorems C(K$_i$) and N(K$_i$)
  - has derived rules (MP) and RM(K$_i$)

Epistemic logic: discussions

Knowledge: omniscience

knowledge set of agent i = set of formulas known by i

- i's knowledge set is...
  - closed under theorems:
    - $\frac{\varphi}{\mathsf{K}_i\,\varphi}$ rule RN(K)
  - closed under logical implication:
    - $\frac{\varphi\to\psi}{\mathsf{K}_i\,\varphi\to\mathsf{K}_i\,\psi}$ rule RM(K)
  - closed under material implication:
    - $(\mathsf{K}_i\,\varphi \wedge \mathsf{K}_i\,(\varphi \to \psi)) \to \mathsf{K}_i\,\psi$ axiom K(K)
- *omniscience problem*
  - if I know the axioms and inference rules of Peano Arithmetic then I know whether every even integer greater than 2 can be written as the sum of two prime numbers
    - Goldbach's conjecture; still unproved!
  - $S5_n$ is an idealization: rational agent, perfect reasoner
  - inadequate for human agents
  - widely accepted in AI
  - negative introspection criticized (see tomorrow)

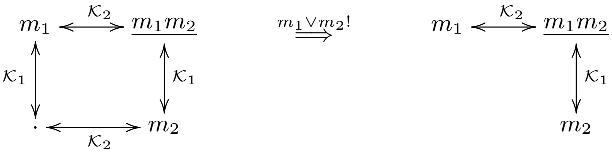
The logic of knowledge: properties

- sound and complete: $\vdash_{S5_n} \varphi$ iff $\models_{S5_n} \varphi$
- decidable
- complexity of $S5_n$-satisfiability is
  - NP-complete if $n = 1$
  - PSPACE-complete if $n > 1$
- there exists a simple normal form for the monoagent case $n = 1$
  - modal depth ≤ 1

Public announcement logic PAL

Epistemic logic: getting dynamic

- observe: after the children have heard father's announcement that $m_1 \vee m_2$, they eliminate all those worlds where $m_1 \vee m_2$ is false
- idea: public announcements transform the model ('update')
- example of muddy children puzzle: father says "$m_1 \vee m_2$!"



(reflexive arrows omitted)

Public announcement logic PAL: language

- $\varphi!$ = announcement of truth of φ
- modal operators of public announcement logic (roughly): $\{\mathsf{K}_{i_1}, \ldots, \mathsf{K}_{i_{\mathsf{card}(Agts)}}\} \cup \{[\varphi!] \ : \ \varphi$ is a formula $\}$
  - either circular definition of formulas
  - or would not allow complex announcements
    - $[([p!]q)!]\mathsf{K}_i\, q$

Public announcement logic PAL: language

- $\varphi!$ = announcement of truth of φ
- modal operators of public announcement logic (roughly):
  $\{\mathsf{K}_{i_1}, \ldots, \mathsf{K}_{i_{\mathsf{card}(Agts)}}\} \cup \{[\varphi!] \; : \; \varphi \text{ is a formula } \}$
  - either circular definition of formulas
  - or would not allow complex announcements
    - ★ $[([p!]q)!]\mathsf{K}_i\, q$
- BNF:

$$\varphi \;::=\; p \mid \neg\varphi \mid \varphi \wedge \varphi \mid \mathsf{K}_i\, \varphi \mid [\varphi!]\varphi$$

where p ranges over $Atms$ and i over $Agts$

- reading:

$[\varphi!]\psi$ = "ψ is true after every possible execution of the announcement of φ"

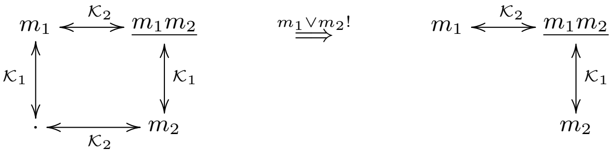
$\langle\varphi!\rangle\psi$ = $\neg[\varphi!]\neg\psi$

Public announcement logic PAL: models

- PAL-model = $S5_n$-model
- truth conditions:

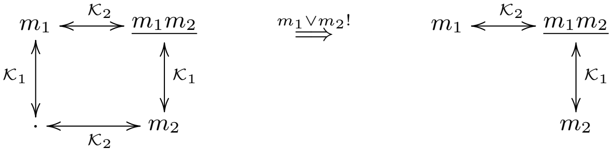
$M, w \Vdash p$	iff	$w \in V(p)$
$M, w \Vdash \neg\varphi$	iff	$\dots$
$M, w \Vdash \varphi \wedge \psi$	iff	$\dots$
$M, w \Vdash \mathsf{K}_i\, \varphi$	iff	$M, w' \Vdash \varphi$ for all $w' \in \mathcal{K}_i(w)$
$M, w \Vdash [\varphi!]\psi$	**iff**	$M, w \nVdash \varphi$ **or** $M^{\varphi!}, w \Vdash \psi$

- $M^{\varphi!}$ = "update of M by φ"



(reflexive arrows omitted)

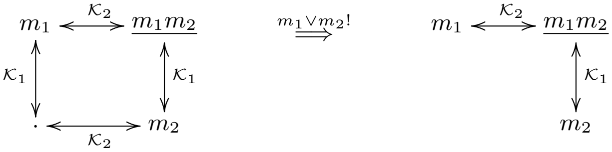
Public announcement logic PAL: models (ctd.)



(reflexive arrows omitted)

- $M^{\varphi!} = \langle W^{\varphi!}, \mathcal{K}^{\varphi!}, V^{\varphi!} \rangle$, where
$$W^{\varphi!} = \{w' \in W \;:\; M, w' \Vdash \varphi\}$$

Public announcement logic PAL: models (ctd.)

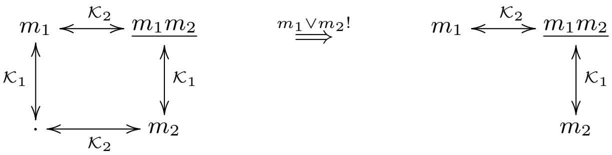


(reflexive arrows omitted)

- $M^{\varphi!}$ = $\langle W^{\varphi!}, \mathcal{K}^{\varphi!}, V^{\varphi!} \rangle$, where

$$\begin{array}{rcl} W^{\varphi!} & = & \{w' \in W \; : \; M, w' \Vdash \varphi\} \\ \mathcal{K}_i^{\varphi!} & = & \mathcal{K}_i \cap (W^{\varphi!} \times W^{\varphi!}) \\ V^{\varphi!}(p) & = & V(p) \cap W^{\varphi!} \end{array}$$

Public announcement logic PAL: models (ctd.)



(reflexive arrows omitted)

- $M^{\varphi!} = \langle W^{\varphi!}, \mathcal{K}^{\varphi!}, V^{\varphi!} \rangle$, where

$$\begin{array}{rcl} W^{\varphi!} & = & \{w' \in W \ : \ M, w' \Vdash \varphi\} \\ \mathcal{K}_i^{\varphi!} & = & \mathcal{K}_i \cap (W^{\varphi!} \times W^{\varphi!}) \\ V^{\varphi!}(p) & = & V(p) \cap W^{\varphi!} \end{array}$$

- *Remarks.*
  - announcements have to be truthful
    - else satisfaction relation $\Vdash$ would be ill-defined
  - if there is $w \in W$ such that $M, w \Vdash \varphi$ then $M^{\varphi!}$ is an $S5_n$-model
- PAL-validity ($\models_{PAL} \varphi$), PAL-satisfiability: defined as usual

Public announcements: non-validities!

- public announcements do not always preserve knowledge:
  $\not\models_{PAL} \mathsf{K}_i\, \psi \rightarrow [\varphi!]\mathsf{K}_i\, \psi$
  - consider $\psi = \neg \mathsf{K}_i\, p$ ...
- public announcements are not always successful:
  $\not\models_{PAL} [\varphi!]\mathsf{K}_i\, \varphi$
  - consider $\varphi = p \wedge \neg \mathsf{K}_i\, p$ ('Moore sentence'),
    and remember: $\mathsf{K}_i\, (p \wedge \neg \mathsf{K}_i\, p)$ is $S5_n$-unsatisfiable!

Reducing PAL to $S5_n$

- useful PAL validities:

$$
\begin{array}{lcll}
[\varphi!]\psi & \leftrightarrow & (\neg\varphi \vee \psi) & \text{if } \psi \text{ is atomic} \\
[\varphi!]\neg\psi & \leftrightarrow & (\neg\varphi \vee \neg[\varphi!]\psi) & \\
[\varphi!](\psi_1 \wedge \psi_2) & \leftrightarrow & ([\varphi!]\psi_1 \wedge [\varphi!]\psi_2) & \\
[\varphi!]\mathsf{K}_i\,\psi & \leftrightarrow & (\neg\varphi \vee \mathsf{K}_i\,[\varphi!]\psi) &
\end{array}
$$

Reducing PAL to $S5_n$

- useful PAL validities:

$$
\begin{array}{lcll}
[\varphi!]\psi & \leftrightarrow & (\neg\varphi \vee \psi) & \text{if } \psi \text{ is atomic} \\
[\varphi!]\neg\psi & \leftrightarrow & (\neg\varphi \vee \neg[\varphi!]\psi) & \\
[\varphi!](\psi_1 \wedge \psi_2) & \leftrightarrow & ([\varphi!]\psi_1 \wedge [\varphi!]\psi_2) & \\
[\varphi!]\mathsf{K}_i\,\psi & \leftrightarrow & (\neg\varphi \vee \mathsf{K}_i\,[\varphi!]\psi) &
\end{array}
$$

- idea: use equivalences as reduction axioms (rewriting from left to right)
  - 'push down' announcement operators
  - eliminate when a Boolean formula is attained
  - $red(\varphi)$ = result of reduction of φ

Reducing PAL to $S5_n$

- useful PAL validities:

$$
\begin{array}{lcll}
[\varphi!]\psi & \leftrightarrow & (\neg\varphi \vee \psi) & \text{if } \psi \text{ is atomic} \\
[\varphi!]\neg\psi & \leftrightarrow & (\neg\varphi \vee \neg[\varphi!]\psi) & \\
[\varphi!](\psi_1 \wedge \psi_2) & \leftrightarrow & ([\varphi!]\psi_1 \wedge [\varphi!]\psi_2) & \\
[\varphi!]\mathsf{K}_i\,\psi & \leftrightarrow & (\neg\varphi \vee \mathsf{K}_i\,[\varphi!]\psi) &
\end{array}
$$

- idea: use equivalences as reduction axioms (rewriting from left to right)
  - 'push down' announcement operators
  - eliminate when a Boolean formula is attained
  - $red(\varphi) =$ result of reduction of φ
- exercises:
  - $red([p!]\mathsf{K}_1\,p) =$?
  - $red([p!]\mathsf{K}_1\,\mathsf{K}_2\,p) =$?
  - $red([(p \wedge \neg\mathsf{K}_1\,p)!]\mathsf{K}_1\,p) =$?
- reduction axioms also provide axiomatics (together with rule of substitution of equivalents)
  - while the other axiom schemas of K are PAL-valid, too, reduction axioms suffice to prove all valid formula instances

Reduction Theorem.
for every PAL-formula φ:

1. $red(\varphi)$ is an $S5_n$-formula
2. $\vdash_{PAL} \varphi \leftrightarrow red(\varphi)$

Sketch of proof.

- equivalences are theorems
- substitution of proved equivalents (REq) preserves PAL-theoremhood
- define a decreasing counter (sum of the number of announcements governing subformulas)
  $\Rightarrow$ rewriting terminates

PAL: properties

- satisfiability in PAL is decidable
  - apply red + decision procedure for $S5_n$
- reduction to $S5_n$ leads to suboptimal decision procedure
- N.B.: rule of uniform substitution not PAL-valid:
  - $\vdash_{PAL} [p!]\mathsf{K}_1\, p$ (v.s.; p formula!)
  - $\not\vdash_{PAL} [\varphi!]\mathsf{K}_i\, \varphi$ (v.s.; φ schema!)

Muddy children reloaded

- *positive formula* π:

$$\pi \ ::= \ \beta \mid \varphi \wedge \varphi \mid \varphi \vee \varphi \mid \mathsf{K}_i \, \varphi$$

  where β ranges over Boolean formulas
- prove that $\vdash_{PAL} \pi \rightarrow [\varphi!]\pi$ if π is a positive formula
  - induction step for $\pi = \mathsf{K}_i \, \pi_1$:
    1. $\pi_1 \rightarrow [\varphi!]\pi_1$ — by induction hyp.
    2. $\mathsf{K}_i \, \pi_1 \rightarrow \mathsf{K}_i \, [\varphi!]\pi_1$ — by rule RM(K_i)
    3. $\mathsf{K}_i \, [\varphi!]\pi_1 \rightarrow [\varphi!]\mathsf{K}_i \, \pi_1$ — no forgetting
    4. $\mathsf{K}_i \, \pi_1 \rightarrow [\varphi!]\mathsf{K}_i \, \pi_1$ — from 2. and 3. by CPL

Muddy children reloaded

- *positive formula* π:
$$\pi \ ::= \ \beta \mid \varphi \wedge \varphi \mid \varphi \vee \varphi \mid \mathsf{K}_i\, \varphi$$
where β ranges over Boolean formulas
- prove that $\vdash_{PAL} \pi \to [\varphi!]\pi$ if π is a positive formula
  - induction step for $\pi = \mathsf{K}_i\, \pi_1$:
    1. $\pi_1 \to [\varphi!]\pi_1$ — by induction hyp.
    2. $\mathsf{K}_i\, \pi_1 \to \mathsf{K}_i\, [\varphi!]\pi_1$ — by rule RM(K_i)
    3. $\mathsf{K}_i\, [\varphi!]\pi_1 \to [\varphi!]\mathsf{K}_i\, \pi_1$ — no forgetting
    4. $\mathsf{K}_i\, \pi_1 \to [\varphi!]\mathsf{K}_i\, \pi_1$ — from 2. and 3. by CPL
- prove that $\vdash_{PAL} [\pi!]\pi$ if π is a positive formula
  - $\vdash_{PAL} \pi \to [\pi!]\pi$ because …
  - $\vdash_{PAL} \neg\pi \to [\pi!]\pi$ because …

Muddy children reloaded

- *positive formula* π:
$$\pi \ ::= \ \beta \mid \varphi \wedge \varphi \mid \varphi \vee \varphi \mid \mathsf{K}_i\, \varphi$$
where β ranges over Boolean formulas
- prove that $\vdash_{PAL} \pi \to [\varphi!]\pi$ if π is a positive formula
  - induction step for $\pi = \mathsf{K}_i\, \pi_1$:
    1. $\pi_1 \to [\varphi!]\pi_1$ — by induction hyp.
    2. $\mathsf{K}_i\, \pi_1 \to \mathsf{K}_i\, [\varphi!]\pi_1$ — by rule RM(K_i)
    3. $\mathsf{K}_i\, [\varphi!]\pi_1 \to [\varphi!]\mathsf{K}_i\, \pi_1$ — no forgetting
    4. $\mathsf{K}_i\, \pi_1 \to [\varphi!]\mathsf{K}_i\, \pi_1$ — from 2. and 3. by CPL
- prove that $\vdash_{PAL} [\pi!]\pi$ if π is a positive formula
  - $\vdash_{PAL} \pi \to [\pi!]\pi$ because …
  - $\vdash_{PAL} \neg\pi \to [\pi!]\pi$ because …
- show:
  - $\vdash_{PAL} [(m_1 \vee m_2)!]\mathsf{K}_1\, \mathsf{K}_2\, (m_1 \vee m_2)$
  - $\vdash_{PAL} [\neg\mathsf{K}_2\, m_2!]\mathsf{K}_1\, \neg\mathsf{K}_2\, m_2$
  - $\vdash_{S5_n} (\mathsf{K}_1\, \mathsf{K}_2\, (m_2 \vee m_1) \wedge \mathsf{K}_1\, \neg\mathsf{K}_2\, m_2 \to \mathsf{K}_1\, \neg\mathsf{K}_2\, \neg m_1$
  - $\vdash_{S5_n} (\mathsf{K}_1\, \neg\mathsf{K}_2\, \neg m_1 \wedge \mathsf{K}_1\, (\mathsf{K}_2\, \neg m_1 \vee \mathsf{K}_2\, m_1)) \to \mathsf{K}_1\, \mathsf{K}_2\, m_1$
- conclude that
$$\vdash_{PAL} \mathsf{K}_1\, (\mathsf{K}_2\, \neg m_1 \vee \mathsf{K}_2\, m_1) \to [(m_1 \vee m_2)!][\neg\mathsf{K}_2\, m_2!]\mathsf{K}_1\, m_1$$

Excursion: the Russian Cards problem [vD03]

Moscow Mathematics Olympiad in 2000:

From a pack of seven known cards Ann and Bill each draw three cars and Cath gets the remaining card.
How can Ann and Bill publicly and truthfully inform each other about their cards, without Cath learning from any of their cards who holds it?

Excursion: the Russian Cards problem [vD03]

Moscow Mathematics Olympiad in 2000:

From a pack of seven known cards Ann and Bill each draw three cars and Cath gets the remaining card.
How can Ann and Bill publicly and truthfully inform each other about their cards, without Cath learning from any of their cards who holds it?

- cards are 0,1,...,6; Ann holds 012 and Bill holds 345
- some bad solutions:
  - Ann says: "Cath holds 6"
    - Ann can only announce what she knows!
  - Ann says: "I don't hold 6"
    - Ann should know that Cath doesn't learn anything!
  - Ann says: "I our Bill hold 012" (and Bill: "I our Ann hold 345")
    - Cath learns that Ann has 012!
  - Ann says: "either I hold 012, or I hold none of 0, 1, 2"
    - Cath doesn't learn any card,
    - Ann knows that,
    - but Cath does not know *that*!

⇒ that Cath remains ignorant should be *common knowledge*

Excursion: the Russian Cards problem [vD03]

- solutions:
  - Ann says: “My cards are among 012, 034, 056, 135 and 246”, and then Bill says: “Cath has 6”
  - . . .
- can be modeled in PAL
- does not work for any number and any distribution of cards
  - for which numbers there is a solution? (open problem)

Excursion: the Russian Cards problem [vD03]

- solutions:
  - Ann says: "My cards are among 012, 034, 056, 135 and 246", and then Bill says: "Cath has 6"
  - ...
- can be modeled in PAL
- does not work for any number and any distribution of cards
  - for which numbers there is a solution? (open problem)
- perspective: unconditionally sure cryptographic protocols (perfect reasoners, public communication)
  - RSA algorithm presupposes non-omniscience (decomposition into prime factors not feasible)

Excursion: the paradox of knowability [Fitch]

- add a new modal operator quantifying over announcements:
  - $M, w \Vdash \Diamond\varphi$ iff there is ψ such that $M, w \Vdash \langle\psi\rangle\varphi$
    - N.B.: ψ should have no occurrence of $\Diamond$ (why?)
- allows to reason about plan existence (epistemic actions only)
  - $\models^? \mathit{Init} \rightarrow \Diamond \mathit{Goal}$

  example: $\models \Diamond(\mathsf{K}_i\, p \vee \mathsf{K}_i \neg p)$
- verificationist thesis:
  - $\varphi \rightarrow \Diamond\mathsf{K}_i\,\varphi$ should be valid for every φ
- paradox of knowability:
  - $\not\models (p \wedge \neg\mathsf{K}_i\, p) \rightarrow \Diamond\mathsf{K}_i\,(p \wedge \neg\mathsf{K}_i\, p)$

Dynamic epistemic logic DEL

Dynamic epistemic logic DEL

- PAL: announcements are perceived by every agent:
  - $[p!](\mathsf{K}_1\, p \wedge \mathsf{K}_2\, p \wedge \mathsf{K}_3\, p \wedge \ldots)$
- idea: $S5_n$ models the agents' uncertainty about current state by means of *possible states*
  $\Rightarrow$ model uncertainty about current event by *possible events*

static uncertainty	dynamic uncertainty
possible worlds indistinguishability of worlds	possible events indistinguishability of events

- example: suppose $p \wedge \neg\mathsf{K}_1\, p \wedge \neg\mathsf{K}_1\, \neg p \wedge \neg\mathsf{K}_2\, p \wedge \neg\mathsf{K}_2\, \neg p$
  - agent 2 learns that p
  - various possible perceptions of 1:
    - 1 also learns that p, and 2 knows that, etc. $\Rightarrow PAL$
    - 1 sees that 2 learns whether p, but does learn it himself (and 2 knows that, etc.)
    - 1 does not sees this (and 2 knows that, etc.)
    - 1 *suspects this*
    - ...

DEL: event models

- static epistemic logic: static model $M^s = \langle W^s, \mathcal{K}^s, V^s \rangle$
- dynamic epistemic logic: dynamic model $M^d = \langle W^d, \mathcal{K}^d, V^d \rangle$, where
  - W^d is a nonempty set of events
  - $\mathcal{K}^d : Agts \longrightarrow W^d \times W^d$
    - every $\mathcal{K}_i^d$ is an equivalence relation
    - $e\mathcal{K}_i e'$ = "i perceives occurrence of e as occurrence of e'"
  - $V^d : W^d \longrightarrow Fmls$
    - precondition of event w^d
- exercise: find dynamic models for the above examples

DEL: product construction

- given:
  - a static model $M^s = \langle W^s, \mathcal{K}^s, V^s \rangle$
  - a dynamic model $M^d = \langle W^d, \mathcal{K}^d, V^d \rangle$

  what is the resulting static model?

DEL: product construction

- given:
  - a static model $M^s = \langle W^s, \mathcal{K}^s, V^s \rangle$
  - a dynamic model $M^d = \langle W^d, \mathcal{K}^d, V^d \rangle$

  what is the resulting static model?

- $M = M^s \otimes M^d = \langle W, \mathcal{K}, V \rangle$ where
  - $W = \{\langle w^s, w^d \rangle \ : \ w^s \in W^s, w^d \in W^d$, and $M, w^s \Vdash V^d(w^d)\}$
  - $\mathcal{K}_i = \{\langle \langle w^s, w^d \rangle, \langle v^s, v^d \rangle \rangle \ : \ w^s \mathcal{K}_i^s v^s$ and $w^d \mathcal{K}_i^d v^d\}$
  - $V(\langle w^s, w^d \rangle) = V^s(w^s)$

restricted product

DEL: product construction

- given:
  - a static model $M^s = \langle W^s, \mathcal{K}^s, V^s \rangle$
  - a dynamic model $M^d = \langle W^d, \mathcal{K}^d, V^d \rangle$

  what is the resulting static model?

- $M = M^s \otimes M^d = \langle W, \mathcal{K}, V \rangle$ where
  - $W = \{\langle w^s, w^d \rangle \ : \ w^s \in W^s, w^d \in W^d$, and $M, w^s \Vdash V^d(w^d)\}$
  - $\mathcal{K}_i = \{\langle \langle w^s, w^d \rangle, \langle v^s, v^d \rangle \rangle \ : \ w^s \mathcal{K}_i^s v^s$ and $w^d \mathcal{K}_i^d v^d\}$
  - $V(\langle w^s, w^d \rangle) = V^s(w^s)$

  restricted product

- exercise: build outcome models for the above examples

DEL: properties

- reduction axioms
- completeness (via reduction axioms)
- applications
  - Cluedo
  - cryptographic protocols
  - ...

What we saw in this lecture

- standard logic of knowledge: $S5_n$
  - criticisms: omniscience
  - static
- dynamics of knowledge
  - public announcement logic
  - dynamic epistemic logic

Next lecture

- logic of belief
- dynamics of belief

Ronald Fagin, Joseph Y. Halpern, Yoram Moses, and Moshe Y. Vardi.
Reasoning about knowledge.
MIT Press, 1995.

Jaakko K. K. Hintikka.
Knowledge and belief.
Cornell University Press, Ithaca, N.Y., 1962.

Saul Kripke.
A completeness theorem in modal logic.
Journal of Symbolic Logic, 24(1):1–14, 1959.

Hans P. van Ditmarsch.
The russian cards problem.
Studia Logica, 75:31–62, 2003.