Epistemic logics

Andreas Herzig
CNRS, IRIT, University of Toulouse, France
www.irit.fr/~Andreas.Herzig

Institut d'Automne en Intelligence Artificielle – IA^2
Lyon, 1–2 nov. 2017

The importance of reasoning about knowledge and belief

- S. Baron Cohen's False-belief-tasks (Sally-Ann Test, ...) [BCLF85]

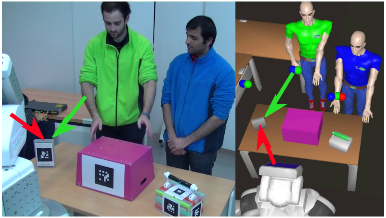
  `https://www.youtube.com/watch?v=jbL34F81Rz0`

- typically fail the test:
  - children under 3
  - autistic children
- hypothesis: specific human capacity of reasoning about other agents' beliefs ('mind reading', *'theory of mind'*)

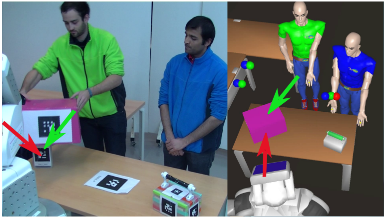
Challenge: robots with theory of mind [Milliez et al. 2014]

- at step 3, agent Green's beliefs become false
  - colored arrows = beliefs about white book position (red = robot)
  - colored spheres = reachability of an object for an agent

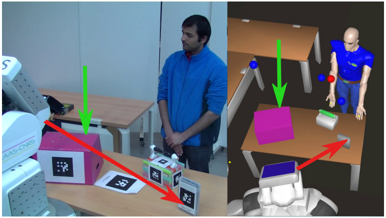
1.



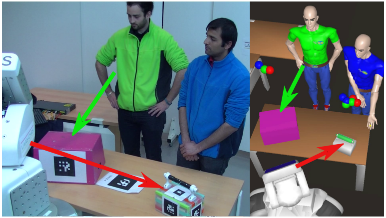
2.



3.



4.



The importance of reasoning about knowledge and belief

- concept of *mental state* of an agent
  - philosophy (philosophy of mind, epistemology)
  - psychology
  - economics
  - computer science (AI, MAS, distributed systems)
- many kinds of mental *attitudes* of an individual i:
  - i is angry; i is sad; i loves individual j; ...
  - most important: beliefs and goals
- how represented in the agent's mind?
  - language of thought [Fodor]
- which logical principles?
  - omniscience problem
- which dynamics?

Course overview

- introduction to the logics of the informational attitudes
- epistemic logics (large sense):
  1. ‘the’ logic of knowledge $S5$ (= epistemic logic in the narrow sense)
  2. ‘the’ logic of belief $KD45$ (= doxastic logic)
- brief introduction to the dynamics of knowledge and belief
  1. update of knowledge (dynamic epistemic logic)
  2. revision of belief

Epistemic logics: resources

- introductory books:
  - [Hin62] “Knowledge and Belief: An Introduction to the Logic of the Two Notions” (Hintikka, first on the topic)
  - [FHMV95] “Reasoning about Knowledge” (Fagin, Halpern, Moses & Vardi)
  - [vDHvdHK15] “Handbook of epistemic logic” (van Ditmarsch, Halpern, van der Hoek& Kooi)
- internet:
  - The Stanford Encyclopedia of Philosophy
    - “Epistemic Logic” [HS15]
    - “Dynamic Epistemic Logic”

Plan

1. The logic of knowledge $S5_n$
   - Introduction
   - Language
   - Semantics
   - Axiomatics
   - Discussions
2. Public announcement logic PAL
3. Dynamic epistemic logic DEL
4. The logic of belief $KD45_n$
5. Dynamics of belief

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 $P(x, y)$
  - there is a constructive proof ($\sim$20 pages, $\sim$1950)

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!"*

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

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')

$\Rightarrow$ which 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:

$$Knows(i, \varphi, w) = \text{"at } w, i \text{ knows that } \varphi\text{"}$$

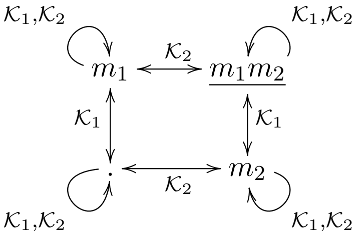
$$\stackrel{\text{def}}{=} \forall w' \, (\mathcal{K}(i, w, w') \to \varphi[w'])$$

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

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

- muddy children:
  - $Knows(1, m_2, w) \ = \ \forall w' \ (\mathcal{K}(1, w, w') \rightarrow m_2(w'))$
  - $\neg Knows(1, m_1, w) \ = \ \exists w' \ (\mathcal{K}(1, w, w') \wedge \neg m_1(w'))$

- 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:
  - $T_{Knows} \vdash_{FOL} \forall w\ \big(((S0) \land (S1) \land (S2)) \to (S3)\big)[w]$
  - consequence problem in first-order logic (FOL): semi-decidable ...

$Knows$: from first-order to modal logic

idea [Hin62, FHMV95]:

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

- K_i : modal operator
- $\mathsf{K}_i\,\varphi$ = "i knows that φ"
  - propositional language; no $\forall$, $\exists$
  - φ might contain modal operator K_j
    - precise definition requires recursive definition of language
  - will be decidable!

Epistemic language: examples

- knowing-whether:
  - $K_1\, m_2 \vee K_1\, \neg m_2$ "child 1 knows whether m_2"
- ignorance:
  - $\neg K_2\, m_2 \wedge \neg K_2\, \neg m_2$ "child 2 does not know whether m_2"
- nesting of modal operators ('higher-order knowledge'):
  - $K_1\, K_2\, (m_1 \vee m_2)$
  - $K_1\, K_2\, K_1\, (m_1 \vee m_2)$
  - ...
  - $K_2\, (K_1\, m_2 \vee K_1\, \neg m_2)$
  - $K_2\, (\neg K_1\, m_1 \wedge (K_1\, m_2 \vee K_1\, \neg m_2))$

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 non-empty 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
    - all these models are bisimular

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

Plan

1. The logic of knowledge $S5_n$
   - Introduction
   - Language
   - Semantics
   - Axiomatics
   - Discussions

2. Public announcement logic PAL

3. Dynamic epistemic logic DEL

4. The logic of belief $KD45_n$

5. Dynamics of belief

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

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:

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

Language (ctd.)

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

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

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

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 $\mathsf{K}_i\,\psi$?

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

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

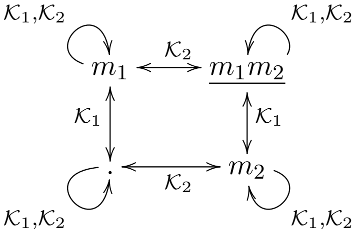
  - ... for φ contingent and non-epistemic
  - why are situations $\varphi \wedge \mathsf{K}_i\,\neg\varphi$ and $\neg\varphi \wedge \mathsf{K}_i\,\varphi$ missing?

Plan

1. The logic of knowledge $S5_n$
   - Introduction
   - Language
   - Semantics
   - Axiomatics
   - Discussions
2. Public announcement logic PAL
3. Dynamic epistemic logic DEL
4. The logic of belief $KD45_n$
5. Dynamics of belief

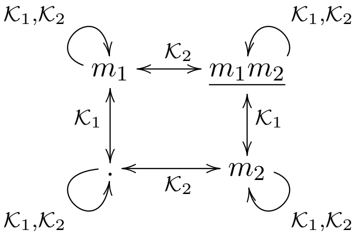
Semantics of $S5_n$: Kripke models

- 'Saul Kripke' [Kri59]
- $Agts = \{1, \ldots, n\}$ set of agents
- **$S5_n$-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 $S5_n$: truth conditions

- truth at world w of model M:
  - $M, w \Vdash p$ iff $w \in V(p)$
  - $M, w \Vdash \neg\varphi$ iff $M, w \nVdash \varphi$
  - $M, w \Vdash \varphi \wedge \psi$ 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:



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

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

Axiomatics of $S5_n$

- axiom schemas for $S5_n$:
  - every theorem schema of *classical propositional logic* (CPL)
  - $(\mathsf{K}_i\, \varphi \wedge \mathsf{K}_i\, \psi) \to \mathsf{K}_i\, (\varphi \wedge \psi)$ conjunction C(K_i)
  - $\mathsf{K}_i\, \top$ necessity N(K_i)
  - $\mathsf{K}_i\, \varphi \to \varphi$ truth T(K_i)
  - $\mathsf{K}_i\, \varphi \to \mathsf{K}_i\, \mathsf{K}_i\, \varphi$ pos. introspection 4(K_i)
  - $\neg\mathsf{K}_i\, \varphi \to \mathsf{K}_i\, \neg\mathsf{K}_i\, \varphi$ neg. introspection 5(K_i)

- inference rules for $S5_n$:
  - $\frac{\varphi, \quad \varphi \to \psi}{\psi}$ modus ponens (MP)
  - $\frac{\varphi \to \psi}{\mathsf{K}_i\, \varphi \to \mathsf{K}_i\, \psi}$ rule of monotony RM(K_i)

Axiomatics of $S5_n$: examples of theorems

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

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

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

- $\vdash_{S5_n} \mathsf{K}_i\,(\varphi \wedge \psi) \leftrightarrow (\mathsf{K}_i\,\varphi \wedge \mathsf{K}_i\,\psi)$
  - proof: ...

Axiomatics of $S5_n$: some useful theorems

- **Rule of Necessitation** RN(K$_i$): $\frac{\varphi}{\mathsf{K}_i\,\varphi}$
  ("for all φ, if $\vdash_{S5_n} \varphi$ then $\vdash_{S5_n} \mathsf{K}_i\,\varphi$")
  - proof:
    1. φ — by hyp.
    2. $\varphi \to (\top \to \varphi)$ — (CPL)
    3. $\top \to \varphi$ — from 1. and 2. by (MP)
    4. $\mathsf{K}_i\,\top \to \mathsf{K}_i\,\varphi$ — from 3. by RM(K$_i$)
    5. $\mathsf{K}_i\,\top$ — N(K$_i$)
    6. $\mathsf{K}_i\,\varphi$ — 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. $\mathsf{K}_i\,\top \to \mathsf{K}_i\,\varphi$ — from 2. by RM(K$_i$)
    4. $\mathsf{K}_i\,\top$ — N(K$_i$)
    5. $\mathsf{K}_i\,\varphi$ — 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. $\varphi = \varphi_1 \land \varphi_2$: ...
    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. $(\mathsf{K}_i \varphi \wedge \mathsf{K}_i (\varphi \to \psi)) \to \mathsf{K}_i (\varphi \wedge (\varphi \to \psi))$ — C(K_i)
    2. $(\varphi \wedge (\varphi \to \psi)) \to \psi$ — (CPL)
    3. $\mathsf{K}_i (\varphi \wedge (\varphi \to \psi)) \to \mathsf{K}_i \psi$ — from 2. by RM(K_i)
    4. $(\mathsf{K}_i \varphi \wedge \mathsf{K}_i (\varphi \to \psi)) \to \mathsf{K}_i \psi$ — 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)

- $\vdash_{S5_n} (\mathsf{K}_i \varphi \wedge \hat{\mathsf{K}}_i \psi) \to \hat{\mathsf{K}}_i (\varphi \wedge \psi)$
  - 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) :
  $(\mathsf{K}_i \varphi \wedge \mathsf{K}_i \psi) \to \mathsf{K}_i (\varphi \wedge \psi)$ 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$_i$): $K_i(\varphi \to \psi) \to (K_i\varphi \to K_i\psi)$
  - has derived rules (MP) and RN(K$_i$): $\frac{\varphi}{K_i\varphi}$
- that CPL+K(K$_i$)+RN(K$_i$)
  -
  - has theorems C(K$_i$) and N(K$_i$)
  - has derived rules (MP) and RM(K$_i$)

Plan

Knowledge: omniscience

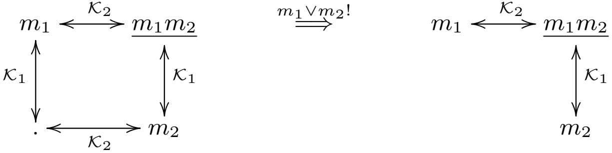
knowledge set of agent i = set of formulas known by i

- i's knowledge set is...
  - closed under theorems:
    - $\frac{\varphi}{\mathsf{K}_i\,\varphi}$ rule RN(K_i)
  - closed under logical implication:
    - $\frac{\varphi\to\psi}{\mathsf{K}_i\,\varphi\to\mathsf{K}_i\,\psi}$ rule RM(K_i)
  - closed under material implication:
    - $(\mathsf{K}_i\,\varphi \wedge \mathsf{K}_i\,(\varphi \to \psi)) \to \mathsf{K}_i\,\psi$ axiom K(K_i)
- *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
  - however widely accepted in AI
    - negative introspection criticized [Len78]

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):
  $\{\mathsf{K}_1, \ldots, \mathsf{K}_n\} \cup \{[\varphi!] : \varphi \text{ is a formula }\}$
  - 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): $\{\mathsf{K}_1, \ldots, \mathsf{K}_n\} \cup \{[\varphi!] \ : \ \varphi \text{ is a formula }\}$
  - 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"

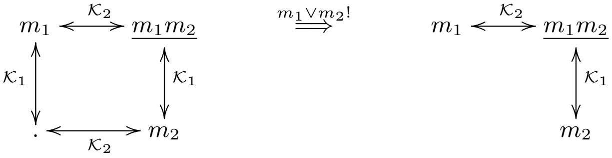
$\langle\varphi!\rangle\psi$ = $\neg[\varphi!]\neg\psi$

Public announcement logic PAL: models

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

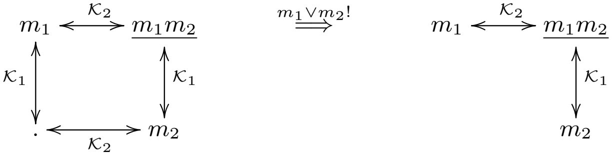
$M, w \Vdash p$	iff	$w \in V(p)$
$M, w \Vdash \neg\varphi$	iff	$\ldots$
$M, w \Vdash \varphi \wedge \psi$	iff	$\ldots$
$M, w \Vdash \mathsf{K}_i\, \varphi$	iff	$M, w' \Vdash \varphi$ for all $w' \in \mathcal{K}_i(w)$
$M, w \Vdash [\varphi!]\psi$	**iff**	$M, w \nVdash \varphi$ **or** $M^{\varphi!}, w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

- $M^{\varphi!}$ = "update of M by φ"



(reflexive arrows omitted)

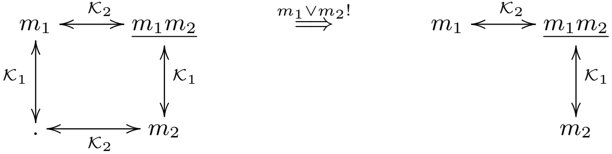
Public announcement logic PAL: models (ctd.)



(reflexive arrows omitted)

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

Public announcement logic PAL: models (ctd.)

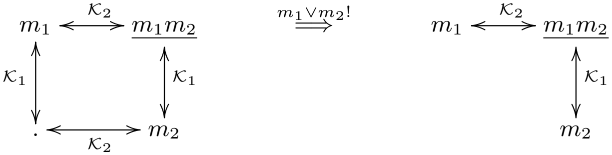


(reflexive arrows omitted)

- $M^{\varphi!} = \langle W^{\varphi!}, \mathcal{K}^{\varphi!}, V^{\varphi!} \rangle$, 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 ($\models_{\text{PAL}} \varphi$), PAL-satisfiability: defined as usual

Public announcements: non-validities!

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

Reducing PAL to $S5_n$, ctd.

Reduction Theorem.
for every PAL-formula φ:

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

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:
  - $\vdash_{\mathsf{PAL}} [p!]\mathsf{K}_1\, p$ (v.s.; p formula!)
  - $\nvdash_{\mathsf{PAL}} [\varphi!]\mathsf{K}_i\, \varphi$ (v.s.; φ schema!)

Muddy children reloaded

- *positive formula* π:

$$\pi \ ::= \ \beta \mid \varphi \land \varphi \mid \varphi \lor \varphi \mid \mathsf{K}_i \, \varphi$$

  where β ranges over Boolean formulas
- prove that $\vdash_{\mathsf{PAL}} \pi \to [\varphi!]\pi$ 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

Muddy children reloaded

- *positive formula* π:
$$\pi ::= \beta \mid \varphi \land \varphi \mid \varphi \lor \varphi \mid \mathsf{K}_i\, \varphi$$
where β ranges over Boolean formulas
- prove that $\vdash_{\mathsf{PAL}} \pi \to [\varphi!]\pi$ 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 $\vdash_{\mathsf{PAL}} [\pi!]\pi$ if π is a positive formula
  - $\vdash_{\mathsf{PAL}} \pi \to [\pi!]\pi$ because ...
  - $\vdash_{\mathsf{PAL}} \neg\pi \to [\pi!]\pi$ 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 $\vdash_{\mathsf{PAL}} \pi \to [\varphi!]\pi$ 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 $\vdash_{\mathsf{PAL}} [\pi!]\pi$ if π is a positive formula
  - $\vdash_{\mathsf{PAL}} \pi \to [\pi!]\pi$ because ...
  - $\vdash_{\mathsf{PAL}} \neg\pi \to [\pi!]\pi$ because ...
- show:
  - $\vdash_{\mathsf{PAL}} [(m_1 \vee m_2)!]\mathsf{K}_1\, \mathsf{K}_2\, (m_1 \vee m_2)$
  - $\vdash_{\mathsf{PAL}} [\neg\mathsf{K}_2\, m_2!]\mathsf{K}_1\, \neg\mathsf{K}_2\, m_2$
  - $\vdash_{S5_n} (\mathsf{K}_1\, \mathsf{K}_2\, (m_2 \vee m_1) \wedge \mathsf{K}_1\, \neg\mathsf{K}_2\, m_2 \to \mathsf{K}_1\, \neg\mathsf{K}_2\, \neg m_1$
  - $\vdash_{S5_n} (\mathsf{K}_1\, \neg\mathsf{K}_2\, \neg m_1 \wedge \mathsf{K}_1\, (\mathsf{K}_2\, \neg m_1 \vee \mathsf{K}_2\, m_1)) \to \mathsf{K}_1\, \mathsf{K}_2\, m_1$
- conclude that
$$\vdash_{\mathsf{PAL}} \mathsf{K}_1\, (\mathsf{K}_2\, \neg m_1 \vee \mathsf{K}_2\, m_1) \to [(m_1 \vee m_2)!][\neg\mathsf{K}_2\, m_2!]\mathsf{K}_1\, m_1$$

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: "either I or Bill hold 012" (and Bill: "I or 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 [Fit63]

- 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)
  - $\models^{?}_{\mathsf{PAL}} Init \to \Diamond Goal$
  - example: $\models \Diamond(\mathsf{K}_i\, p \vee \mathsf{K}_i\, \neg p)$
- Fitch's paradox of knowability:
  - verificationism: $\varphi \to \Diamond\mathsf{K}_i\, \varphi$ should be valid for every φ
  - however: $\not\models (p \wedge \neg\mathsf{K}_i\, p) \to \Diamond\mathsf{K}_i\, (p \wedge \neg\mathsf{K}_i\, p)$

Dynamic epistemic logic DEL

Dynamic epistemic logic DEL

- PAL: announcements are perceived by every agent:
  - $[p!](\mathsf{K}_1\, p \land \mathsf{K}_2\, p \land \mathsf{K}_1\, \mathsf{K}_2\, p \land \ldots)$
  - how can we model other kinds of perception?
- idea: model uncertainty about current event by *possible events*

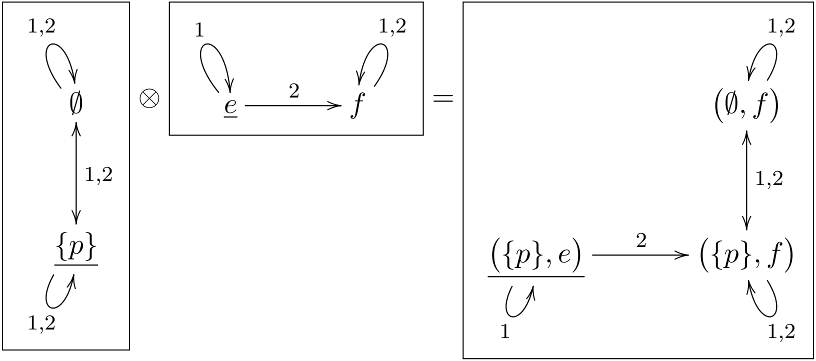
static uncertainty	dynamic uncertainty
possible worlds	possible events
indistinguishability of worlds	indistinguishability of events

- example: suppose $p \land \neg\mathsf{K}_1\, p \land \neg\mathsf{K}_1\, \neg p \land \neg\mathsf{K}_2\, p \land \neg\mathsf{K}_2\, \neg p$
  - agent 2 learns that p
  - various possible perceptions of agent 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 model $M^s = \langle W^s, \mathcal{K}^s, V^s \rangle$
- 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 $\mathcal{K}^d_i$ is an equivalence relation
    - $e\mathcal{K}_i e'$ = "i perceives occurrence of e as occurrence of e'"
  - $V^d : W^d \longrightarrow Fmls$
    - what is announced at event w^d ('precondition')
- exercise: find dynamic models for the above examples

DEL: private announcement of p to agent 1



- static model: neither 1 nor 2 knows whether p
- event model: private announcement of p to 1:

  $V^d(e) = p!$ and $V^d(f) = \top!$
- product model: update static model by event model

DEL: product construction

- given:
  - a static model $M^s = \langle W^s, \mathcal{K}^s, V^s \rangle$
  - a dynamic model $M^d = \langle W^d, \mathcal{K}^d, V^d \rangle$
- product update: $M^s \otimes M^d = \langle W, \mathcal{K}, V \rangle$ 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 $M^d = \langle W^d, \mathcal{K}^d, V^d \rangle$
- product update: $M^s \otimes M^d = \langle W, \mathcal{K}, V \rangle$ 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
  - analysis of games with imperfect information: Cluedo,...
  - epistemic planning [AB13, BJS15]
  - cryptographic protocols
  - ...

Next half of course

- logic of belief
- dynamics of belief
- group knowledge and group belief

Plan

Doxastic logic: introduction

- when is knowledge the appropriate informational attitude?
- remember: "knowledge implies truth" principle in epistemic logic:

$$\models_{S5_n} \mathsf{K}_i\, \varphi \to \varphi$$

Doxastic logic: introduction

- when is knowledge the appropriate informational attitude?
- remember: “knowledge implies truth” principle in epistemic logic:

$$\models_{S5_n} \mathsf{K}_i\, \varphi \to \varphi$$

- relevant for:
  - formal epistemology
    - what is knowledge?
    - is knowledge possible at all?
    - are all truths knowable?
  - distributed processes [FHMV95]

Doxastic logic: introduction (ctd.)

- relation of to truth less in focus in:
  - philosophy of mind
    - focus on i's mental state
  - philosophy of language
    - effects of speech acts on the participants' mental states: lies, bullshitting
  - implementation of artificial agents

Doxastic logic: introduction (ctd.)

- relation of to truth less in focus in:
  - philosophy of mind
    - focus on i's mental state
  - philosophy of language
    - effects of speech acts on the participants' mental states: lies, bullshitting
  - implementation of artificial agents
- informational mental attitude not implying truth: *belief*
  - "he knows that φ, but he is wrong": inconsistent
  - "he believes that φ, but he is wrong": consistent

  however: 'belief aims at truth' [Eng98, Hak06]
- doxastic logic [Hin62, Len78, Len95]
  - *doxa* = $\delta o \xi \alpha$ = 'believe' (Greek)

Doxastic logic: language

- BNF:

$$\varphi \quad ::= \quad p \mid \bot \mid \neg\varphi \mid (\varphi \wedge \varphi) \mid \mathsf{B}_i\,\varphi$$

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

- $\mathsf{B}_i\,\varphi$ = "agent i believes that φ"

Doxastic logic: language

- BNF:

$$\varphi \quad ::= \quad p \mid \bot \mid \neg\varphi \mid (\varphi \wedge \varphi) \mid \mathsf{B}_i\, \varphi$$

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

- $\mathsf{B}_i\, \varphi$ = "agent i believes that φ"
- examples of formulas:
  - $m_1 \wedge \mathsf{B}_1 \neg m_1$
  - $\mathsf{B}_1 \neg m_1 \wedge \mathsf{B}_2\, \mathsf{B}_1\, m_1$
  - $\mathsf{B}_1\, (\mathsf{B}_2\, m_1 \vee \mathsf{B}_2 \neg m_1)$

Doxastic logic: language

- BNF:

  $$\varphi \quad ::= \quad p \mid \bot \mid \neg\varphi \mid (\varphi \wedge \varphi) \mid \mathsf{B}_i\, \varphi$$

  where p ranges over $Atms$ and i ranges over $Agts$
- $\mathsf{B}_i\, \varphi$ = "agent i believes that φ"
- examples of formulas:
  - $m_1 \wedge \mathsf{B}_1 \neg m_1$
  - $\mathsf{B}_1 \neg m_1 \wedge \mathsf{B}_2\, \mathsf{B}_1\, m_1$
  - $\mathsf{B}_1 (\mathsf{B}_2\, m_1 \vee \mathsf{B}_2 \neg m_1)$
- abbreviation:
  - $\hat{\mathsf{B}}_i\, \varphi \stackrel{\text{def}}{=} \neg \mathsf{B}_i \neg\varphi$ "it is possible for i that φ"

Doxastic logic: language (ctd.)

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

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

  - for φ contingent and non-doxastic

Doxastic logic: language (ctd.)

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

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

  - for φ contingent and non-doxastic

- 6 possible *doxastic situations* w.r.t. a formula φ:

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

  - for φ contingent and non-doxastic

Plan

Doxastic logic: semantics

- belief explained in terms of possible worlds [Hin62, FHMV95]:

  $\mathsf{B}_i\,\varphi$ = "agent i believes that φ"

  = "*φ true in every world that is compatible with i's beliefs*"

Doxastic logic: semantics

- belief explained in terms of possible worlds [Hin62, FHMV95]:

  $\mathsf{B}_i\,\varphi \quad = \quad$ "agent i believes that φ"

  $\quad\quad\quad = \quad$ "*φ true in every world that is compatible with i's beliefs*"

- $KD45_n$**-model** $M = \langle W, \mathcal{B}, V \rangle$ where:
  - W nonempty set
  - $V : Atms \longrightarrow 2^W$ 'valuation'
  - $\mathcal{B} : Agts \longrightarrow 2^{W \times W}$ such that for every $i \in Agts$:
    - for every w there is some w' such that $\langle w, w' \rangle \in \mathcal{B}_i$ (*serial*)
    - if $\langle w, w' \rangle \in \mathcal{B}_i$ and $\langle w', w'' \rangle \in \mathcal{B}_i$ then $\langle w, w'' \rangle \in \mathcal{B}_i$ (*transitive*)
    - if $\langle w, w' \rangle \in \mathcal{B}_i$ and $\langle w, w'' \rangle \in \mathcal{B}_i$ then $\langle w', w'' \rangle \in \mathcal{B}_i$ (*Euclidian*)

Doxastic logic: semantics (ctd.)

$\mathcal{B}_i(w)$	=	$\{w' : \langle w, w' \rangle \in \mathcal{B}_i\}$
	=	i's alternatives to w
	=	worlds i cannot distinguish from w on basis of his beliefs
	=	set of worlds compatible with i's beliefs
	=	*belief state* of agent i at w

- $\mathcal{B}_i$ serial $\Leftrightarrow$ $\mathcal{B}_i(w) \neq \emptyset$
- $\mathcal{B}_i$ transitive + Euclidian $\Leftrightarrow$ if $w' \in \mathcal{B}_i(w)$ then $\mathcal{B}_i(w) = \mathcal{B}_i(w')$

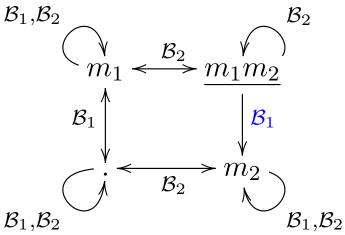
Doxastic logic: semantics (ctd.)

$\mathcal{B}_i(w)$	=	$\{w' : \langle w, w' \rangle \in \mathcal{B}_i\}$
	=	i's alternatives to w
	=	worlds i cannot distinguish from w on basis of his beliefs
	=	set of worlds compatible with i's beliefs
	=	*belief state* of agent i at w

- $\mathcal{B}_i$ serial $\Leftrightarrow$ $\mathcal{B}_i(w) \neq \emptyset$
- $\mathcal{B}_i$ transitive + Euclidian $\Leftrightarrow$ if $w' \in \mathcal{B}_i(w)$ then $\mathcal{B}_i(w) = \mathcal{B}_i(w')$
- truth condition:
  - $M, w \Vdash \mathsf{B}_i\, \varphi$ iff $M, w' \Vdash \varphi$ for every $w' \in \mathcal{B}_i(w)$

Doxastic logic: semantics (ctd.)

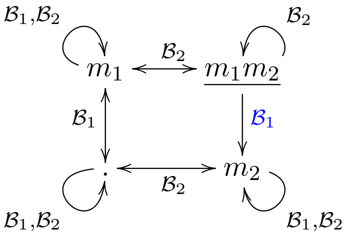
- variant of the muddy children puzzle: child 1 wrongly believes it is not muddy



$\mathcal{B}_1(m_1m_2) = \{(m_2)\}$

Doxastic logic: semantics (ctd.)

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



$\mathcal{B}_1(m_1m_2) = \{(m_2)\}$

$M, (m_1m_2) \Vdash m_1 \wedge \mathsf{B}_1 \neg m_1$

Plan

Doxastic logic: axiomatics

- standard multiagent logic of belief = multimodal $KD45_n$
  - principles of multimodal K:
    - ★ principles of classical propositional logic
    - ★ $(\mathsf{B}_i\,\varphi \land \mathsf{B}_i\,\psi) \to \mathsf{B}_i\,(\varphi \land \psi)$
    - ★ from $\varphi \to \psi$ infer $\mathsf{B}_i\,\varphi \to \mathsf{B}_i\,\psi$
  - consistency of belief:
    - ★ $\neg(\mathsf{B}_i\,\varphi \land \mathsf{B}_i\,\neg\varphi)$ — axiom D(B_i)
  - positive introspection:
    - ★ $\mathsf{B}_i\,\varphi \to \mathsf{B}_i\,\mathsf{B}_i\,\varphi$ — axiom 4(B_i)
  - negative introspection:
    - ★ $\neg\mathsf{B}_i\,\varphi \to \mathsf{B}_i\,\neg\mathsf{B}_i\,\varphi$ — axiom 5(B_i)

Doxastic logic: properties

- sound and complete: $\vdash_{KD45_n} \varphi$ iff $\models_{KD45_n} \varphi$
- decidable
- complexity of $KD45_n$-satisfiability:
  - NP-complete if $n = 1$
  - PSPACE-complete if $n > 1$
- for $n = 1$ there exists a *normal form*: modal depth ≤ 1

Plan

Discussion: omniscience problem

...

(cf. logic of knowledge)

Discussion: belief and probability

- $KD45_n$'s belief is a strong form of belief ('conviction')
- weaker form of belief:

$$\mathsf{B}_i\,\varphi = \text{“}Proba_i(\varphi) > Proba_i(\neg\varphi)\text{”}$$

- semantics:

Discussion: belief and probability

- $KD45_n$'s belief is a strong form of belief ('conviction')
- weaker form of belief:

$$\mathsf{B}_i\,\varphi = \text{“}Proba_i(\varphi) > Proba_i(\neg\varphi)\text{”}$$

- semantics: $M = \langle W, \mathcal{B}, V \rangle$ where
  - $\mathcal{B} : (Agts \longrightarrow (W \times W))$
- $M, w \models \mathsf{B}_i\,\varphi$ iff among the i-accessible worlds there are more φ worlds than $\neg\varphi$ worlds"
  - $(\mathsf{B}_i\,\varphi \wedge \mathsf{B}_i\,\psi) \to \mathsf{B}_i\,(\varphi \wedge \psi)$ not valid!
  - weakening of Kripke semantics: neighbourhood semantics [Bur69, Len78]

Discussion: graded belief

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

Discussion: graded belief

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

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

Discussion: graded belief

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

  $w\mathcal{B}_i^{\geq d}v$ = "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
  - $\mathsf{B}_i^{\geq d}\varphi \to \mathsf{B}_i^{\geq d'}\varphi$ if $d \geq d'$

Discussion: can knowledge be defined from belief?

[Plato, Theaetetus]

- $\mathsf{K}_i\,\varphi \overset{\text{def}}{=} \mathsf{B}_i\,\varphi \land \varphi$

Discussion: can knowledge be defined from belief?

[Plato, Theaetetus]

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

Discussion: can knowledge be defined from belief?

[Plato, Theaetetus]

- $\mathsf{K}_i\,\varphi \overset{\text{def}}{=} \mathsf{B}_i\,\varphi \wedge \varphi$
  - problem: ‘knowledge by accident’
- $\mathsf{K}_i\,\varphi \overset{\text{def}}{=} \mathsf{B}_i\,\varphi \wedge \varphi \wedge \mathsf{hasJustif}(i, \varphi)$
  - problem: what is a justification?
    - ★ justification logic [Artemov]
  - Gettier Problem [1963]:
    - ★ suppose a logic of belief and justification such that

      $$\frac{\varphi \to \psi}{\mathsf{hasJustif}(i,\varphi) \to \mathsf{hasJustif}(i,\psi)}$$

    - ★ suppose i wrongly believes p, but has some justification for that:
      $\neg p \wedge \mathsf{B}_i\,p \wedge \mathsf{hasJustif}(i, p)$ (‘epistemic luck’)
    - ★ …hence i believes that $p \vee q$ and i believes that $p \vee \neg q$
      (by axiom M(B_i))
    - ★ …and $\mathsf{hasJustif}(i, (p \vee q))$ and $\mathsf{hasJustif}(i, (p \vee \neg q))$
      (use inference rule for hasJustif)
    - ★ …**and either i knows that $p \vee q$, or i knows that $p \vee \neg q$, for any q**:
      $\models \mathsf{B}_i\,p \wedge \mathsf{hasJustif}(i, p) \to (\mathsf{K}_i\,(p \vee q) \vee \mathsf{K}_i\,(p \vee \neg q))$

Discussion: relation between knowledge and belief?

- suppose a logic of knowledge and belief defined as:
  - $KD45(\mathsf{B}_i)$
  - $S5(\mathsf{K}_i)$
  - $\mathsf{K}_i \varphi \to \mathsf{B}_i \varphi$ ('knowledge implies belief'; $\neq$ natural language use)
  - $\mathsf{B}_i \varphi \to \mathsf{B}_i \mathsf{K}_i \varphi$

Discussion: relation between knowledge and belief?

- suppose a logic of knowledge and belief defined as:
  - $KD45(\mathsf{B}_i)$
  - $S5(\mathsf{K}_i)$
  - $\mathsf{K}_i\,\varphi \to \mathsf{B}_i\,\varphi$ ('knowledge implies belief'; $\neq$ natural language use)
  - $\mathsf{B}_i\,\varphi \to \mathsf{B}_i\,\mathsf{K}_i\,\varphi$

- ...but implies that $\mathsf{B}_i\,\varphi \leftrightarrow \mathsf{K}_i\,\varphi$!

Discussion: relation between knowledge and belief?

- suppose a logic of knowledge and belief defined as:
  - $KD45(\mathsf{B}_i)$
  - $S5(\mathsf{K}_i)$
  - $\mathsf{K}_i\,\varphi \to \mathsf{B}_i\,\varphi$ ('knowledge implies belief'; $\neq$ natural language use)
  - $\mathsf{B}_i\,\varphi \to \mathsf{B}_i\,\mathsf{K}_i\,\varphi$
- ...but implies that $\mathsf{B}_i\,\varphi \leftrightarrow \mathsf{K}_i\,\varphi$!
  - intermediate step: $\neg\mathsf{B}_i\,\neg\mathsf{K}_i\,\varphi \to \neg\mathsf{K}_i\,\neg\mathsf{B}_i\,\varphi$

Discussion: relation between knowledge and belief?

- suppose a logic of knowledge and belief defined as:
  - $KD45(\mathsf{B}_i)$
  - $S5(\mathsf{K}_i)$
  - $\mathsf{K}_i\, \varphi \to \mathsf{B}_i\, \varphi$ (‘knowledge implies belief’; $\neq$ natural language use)
  - $\mathsf{B}_i\, \varphi \to \mathsf{B}_i\, \mathsf{K}_i\, \varphi$
- ...but implies that $\mathsf{B}_i\, \varphi \leftrightarrow \mathsf{K}_i\, \varphi$!
  - intermediate step: $\neg\mathsf{B}_i\, \neg\mathsf{K}_i\, \varphi \to \neg\mathsf{K}_i\, \neg\mathsf{B}_i\, \varphi$
- culprit: negative introspection for knowledge [Len78, Len95]

Dynamics of belief

Plan

The logic of belief: getting dynamic

- how do i's beliefs evolve when i learns that φ is true?
- extend $KD45_n$ by public announcement operator $[\varphi!]$
  - what if agent i wrongly believes that p, and $\neg p$ is announced?
  - can't be the case in epistemic logic: $\vdash_{S5_n\text{-}PAL} \mathsf{K}_i\, p \rightarrow [\neg p!]\bot$
    - ★ proof:
      $\vdash_{S5_n} \mathsf{K}_i\, p \rightarrow p$
      $\vdash_{\mathsf{S5}_n\text{-}\mathsf{PAL}} p \leftrightarrow [\neg p!]\bot$ (reduction axiom)
  - in doxastic logic:
    - ★ $\mathsf{B}_i\, p \wedge \neg p$ is $KD45_n$ satisfiable
    - ★ $\vdash_{\mathsf{KD45}_n\text{-}\mathsf{PAL}} p \leftrightarrow [\neg p!]\bot$ (reduction axiom)
    - ★ $\mathsf{B}_i\, p \wedge \neg[\neg p!]\bot$ should be $KD45_n\text{-}PAL$ satisfiable!

The logic of belief: getting dynamic (ctd.)

- exercise: prove $\vdash_{\mathsf{KD45}_n\text{-}\mathsf{PAL}} (\neg p \land \mathsf{B}_i\, p) \to \langle \neg p! \rangle \mathsf{B}_i \bot$

The logic of belief: getting dynamic (ctd.)

- exercise: prove $\vdash_{\mathsf{KD45}_n\text{-}\mathsf{PAL}} (\neg p \land \mathsf{B}_i\, p) \to \langle \neg p! \rangle \mathsf{B}_i \bot$
  1. $\neg p \to \langle \neg p! \rangle \top$ (red.ax.)

The logic of belief: getting dynamic (ctd.)

- exercise: prove $\vdash_{\mathsf{KD45}_n\text{-PAL}} (\neg p \land \mathsf{B}_i\, p) \to \langle \neg p! \rangle \mathsf{B}_i \perp$
  1. $\neg p \to \langle \neg p! \rangle \top$ (red.ax.)
  2. $[\neg p!]\mathsf{B}_i \neg p$
     - ★ reduction:

$$\begin{aligned} [\neg p!]\mathsf{B}_i \neg p \quad &\leftrightarrow \quad \neg p \to \mathsf{B}_i\, [\neg p!]\neg p \\ &\leftrightarrow \quad \neg p \to \mathsf{B}_i\, (\neg p \to \neg p) \\ &\leftrightarrow \quad \neg p \to \mathsf{B}_i\, \top \\ &\leftrightarrow \quad \neg p \to \top \\ &\leftrightarrow \quad \top \end{aligned}$$

The logic of belief: getting dynamic (ctd.)

- exercise: prove $\vdash_{\mathsf{KD45}_n\text{-PAL}} (\neg p \wedge \mathsf{B}_i\, p) \to \langle \neg p! \rangle \mathsf{B}_i \bot$
  1. $\neg p \to \langle \neg p! \rangle \top$ (red.ax.)
  2. $[\neg p!]\mathsf{B}_i \neg p$
     - ★ reduction:

$$\begin{aligned} [\neg p!]\mathsf{B}_i \neg p &\leftrightarrow \neg p \to \mathsf{B}_i [\neg p!]\neg p \\ &\leftrightarrow \neg p \to \mathsf{B}_i (\neg p \to \neg p) \\ &\leftrightarrow \neg p \to \mathsf{B}_i \top \\ &\leftrightarrow \neg p \to \top \\ &\leftrightarrow \top \end{aligned}$$

  3. $\mathsf{B}_i\, p \to [\neg p!]\mathsf{B}_i\, p$
     - ★ reduction:

$$\begin{aligned} [\neg p!]\mathsf{B}_i\, p &\leftrightarrow \neg p \to \mathsf{B}_i [\neg p!]p \\ &\leftrightarrow \neg p \to \mathsf{B}_i (\neg p \to p) \\ &\leftrightarrow \neg p \to \mathsf{B}_i\, p \end{aligned}$$

The logic of belief: getting dynamic (ctd.)

- exercise: prove $\vdash_{\mathsf{KD45}_n\text{-}\mathsf{PAL}} (\neg p \land \mathsf{B}_i\, p) \to \langle \neg p! \rangle \mathsf{B}_i \bot$
  1. $\neg p \to \langle \neg p! \rangle \top$ (red.ax.)
  2. $[\neg p!]\mathsf{B}_i\, \neg p$
     - ★ reduction:
       $$\begin{array}{rcl} [\neg p!]\mathsf{B}_i\, \neg p & \leftrightarrow & \neg p \to \mathsf{B}_i\, [\neg p!]\neg p \\ & \leftrightarrow & \neg p \to \mathsf{B}_i\, (\neg p \to \neg p) \\ & \leftrightarrow & \neg p \to \mathsf{B}_i\, \top \\ & \leftrightarrow & \neg p \to \top \\ & \leftrightarrow & \top \end{array}$$
  3. $\mathsf{B}_i\, p \to [\neg p!]\mathsf{B}_i\, p$
     - ★ reduction:
       $$\begin{array}{rcl} [\neg p!]\mathsf{B}_i\, p & \leftrightarrow & \neg p \to \mathsf{B}_i\, [\neg p!]p \\ & \leftrightarrow & \neg p \to \mathsf{B}_i\, (\neg p \to p) \\ & \leftrightarrow & \neg p \to \mathsf{B}_i\, p \end{array}$$
  4. $(\neg p \land \mathsf{B}_i\, p) \to \langle \neg p! \rangle (\mathsf{B}_i\, p \land \mathsf{B}_i\, \neg p)$ (from 1,2,3)

The logic of belief: getting dynamic (ctd.)

- ways out:
  1. drop seriality: beliefs might get inconsistent
  2. modify truth condition for announcements

$M, w \Vdash [\varphi!]\psi$ iff $M, w \nVdash \varphi$ or
($M, w \Vdash \hat{\mathsf{B}}_i\, \varphi$ and $M^{\varphi!}, w \Vdash \psi$), or
($M, w \Vdash \mathsf{B}_i\, \neg\varphi$ and $M, w \Vdash \psi$)

The logic of belief: getting dynamic (ctd.)

- ways out:
  1. drop seriality: beliefs might get inconsistent
  2. modify truth condition for announcements

     $M, w \Vdash [\varphi!]\psi$ iff $M, w \nVdash \varphi$ or
     ($M, w \Vdash \hat{\mathsf{B}}_i\, \varphi$ and $M^{\varphi!}, w \Vdash \psi$), or
     ($M, w \Vdash \mathsf{B}_i\, \neg\varphi$ and $M, w \Vdash \psi$)

     - ⋆ reduction axiom:

       $[\varphi!]\mathsf{B}_i\, \psi \quad \leftrightarrow \quad \neg\varphi \vee (\hat{\mathsf{B}}_i\, \varphi \wedge \mathsf{B}_i\, [\varphi!]\psi) \vee (\mathsf{B}_i\, \neg\varphi \wedge \mathsf{B}_i\, \psi)$
     - ⋆ believe-contravening input is rejected

The logic of belief: getting dynamic (ctd.)

- ways out:
  1. drop seriality: beliefs might get inconsistent
  2. modify truth condition for announcements

     $M, w \Vdash [\varphi!]\psi$ iff $M, w \nVdash \varphi$ or
     ($M, w \Vdash \hat{\mathsf{B}}_i\, \varphi$ and $M^{\varphi!}, w \Vdash \psi$), or
     ($M, w \Vdash \mathsf{B}_i\, \neg\varphi$ and $M, w \Vdash \psi$)

     - ★ reduction axiom:

       $[\varphi!]\mathsf{B}_i\, \psi \quad \leftrightarrow \quad \neg\varphi \vee (\hat{\mathsf{B}}_i\, \varphi \wedge \mathsf{B}_i\, [\varphi!]\psi) \vee (\mathsf{B}_i\, \neg\varphi \wedge \mathsf{B}_i\, \psi)$
     - ★ believe-contravening input is rejected
  3. **integrate belief revision mechanisms**

Plan

AGM theory: the internal perspective

beliefs of an agent = set of Boolean formulas $S \subseteq \mathcal{L}_{\mathsf{CPL}}$

- $\varphi \in S$ = "φ believed by the agent"
- internal perspective (S is 'in the agent's head')
- $\neq$ external perspective:
  - φ = "φ is (objectively) true"
  - taken in doxastic logic
- internal version of doxastic logic [Auc08]
  - distinguished agent Y ("you")
  - φ = "Y believes that φ"
  - wanted: $\vdash \varphi \leftrightarrow \mathrm{B}_Y\, \varphi$
  - abandon inference rule of necessitation
    - $\models \mathrm{B}_Y\, \varphi \to \varphi$, but $\not\models \mathrm{B}_i\, (\mathrm{B}_Y\, \varphi \to \varphi)$

AGM theory: coherentism vs. foundationalism

beliefs of an agent = set of Boolean formulas $S \subseteq \mathcal{L}_{\mathsf{CPL}}$

- foundational view: some beliefs are more basic than others
  - *belief base* (typically finite)
- coherentist view: all beliefs support each other
  - S closed under logical consequence: *belief set*
    - ⋆ omniscience problem (v.s.)
  - can be represented by a formula [KM92]
    - ⋆ logically equivalent formulas should be revised in the same way

AGM theory: belief change operations

- agent's beliefs = set of formulas:
  - $op : 2^{\mathcal{L}_{\mathsf{CPL}}} \times \mathcal{L}_{\mathsf{CPL}} \longrightarrow 2^{\mathcal{L}_{\mathsf{CPL}}}$ [AGM85]
- agent's beliefs = formula:
  - $op : \mathcal{L}_{\mathsf{CPL}} \times \mathcal{L}_{\mathsf{CPL}} \longrightarrow \mathcal{L}_{\mathsf{CPL}}$ [KM92]
  - require that when $\vdash \varphi_1 \leftrightarrow \varphi_2$ then $\vdash op(\varphi_1, \psi) \leftrightarrow op(\varphi_2, \psi)$
    - ‘simulates’ coherentist approach
- 3 kinds of operations op:
  - $\varphi + \psi$: expansion
  - $\varphi - \psi$: contraction
  - $\varphi \star \psi$: revision

AGM theory: belief change operations (ctd.)

- expand φ by ψ:
  $\varphi + \psi$ = "add ψ without worrying about consistency"
  - desiderata:
    - $\varphi + \psi \stackrel{\text{def}}{=} \varphi \wedge \psi$

- contract φ by ψ:
  $\varphi - \psi$ = "weaken φ such that ψ no longer follows"
  - desiderata:
    - $\varphi - \psi \not\vdash \psi$
    - $\varphi \vdash \varphi - \psi$

- revise φ by ψ:
  $\varphi \star \psi$ = "weaken φ such that $\neg\psi$ no longer follows, and add ψ"
  - desiderata:
    - $\varphi \star \psi = (\varphi - \neg\psi) + \psi$ (Levi Identity)
    - $\varphi \star \psi \vdash \psi$
    - ...

The basic AGM postulates for belief revision

(R1) $\varphi \star \psi \vdash \psi$

(R2) if $\varphi \not\vdash \neg\psi$ then $\vdash \varphi \star \psi \leftrightarrow \varphi \wedge \psi$

(R3) if $\varphi \star \psi \vdash \bot$ then $\psi \vdash \bot$

(R4) if $\vdash \varphi \leftrightarrow \varphi'$ and $\vdash \psi \leftrightarrow \psi'$ then $\vdash \varphi \star \psi \leftrightarrow \varphi' \star \psi'$

(R56) if $\varphi \star \psi_1 \not\vdash \neg\psi_2$ then $\vdash \varphi \star (\psi_1 \wedge \psi_2) \leftrightarrow (\varphi \star \psi_1) \wedge \psi_2$

generalizes **(R2)**

N.B.: *postulate* $\neq$ axiom: may use metalanguage ("if $\varphi \not\vdash \neg\psi$ ...")

AGM theory: semantics

- model = sphere system: set of centered *spheres* surrounding $\|\varphi\|$
  - [Gro88], inspired from conditional logics [Lew73]
  - $\|\varphi\| = \{w \ : \ w \Vdash \varphi\}$ = extension of φ (w = interpretation of CPL)
  - total preorder $\leq_\varphi$, for every formula φ
    - ★ $w_1 \approx_\varphi w_2$ iff $w_1 <_\varphi w_2$ and $w_2 <_\varphi w_1$
  - $\leq_\varphi$ centered around $\|\varphi\|$:
    - ★ if $w_1 \Vdash \varphi$ and $w_2 \Vdash \varphi$ then $w_1 \approx_\varphi w_2$
    - ★ if $w_1 \Vdash \varphi$ and $w_2 \nVdash \varphi$ then $w_1 <_\varphi w_2$
  - insensitive to syntax:
    - ★ if $\vdash \varphi \leftrightarrow \varphi'$ then $\leq_\varphi = \leq_{\varphi'}$

AGM theory: semantics

- model = sphere system: set of centered *spheres* surrounding $\|\varphi\|$
  - [Gro88], inspired from conditional logics [Lew73]
  - $\|\varphi\| = \{w \ : \ w \Vdash \varphi\}$ = extension of φ (w = interpretation of CPL)
  - total preorder $\leq_\varphi$, for every formula φ
    - $w_1 \approx_\varphi w_2$ iff $w_1 <_\varphi w_2$ and $w_2 <_\varphi w_1$
  - $\leq_\varphi$ centered around $\|\varphi\|$:
    - if $w_1 \Vdash \varphi$ and $w_2 \Vdash \varphi$ then $w_1 \ \approx_\varphi \ w_2$
    - if $w_1 \Vdash \varphi$ and $w_2 \nVdash \varphi$ then $w_1 \ <_\varphi \ w_2$
  - insensitive to syntax:
    - if $\vdash \varphi \leftrightarrow \varphi'$ then $\leq_\varphi = \leq_{\varphi'}$
- $\leq$ defines a revision operation:
  - $\|\varphi \star_\leq \psi\| = \min_{\leq_\varphi} \|\psi\|$

AGM theory: representation theorem

- representation theorem:
  let $\star : \mathcal{L}_{\mathsf{CPL}} \times \mathcal{L}_{\mathsf{CPL}} \longrightarrow \mathcal{L}_{\mathsf{CPL}}$ be any mapping;

  $\star$ satisfies the (extended) AGM postulates iff
  there is a family of total preorders $\leq_\varphi$, one for every φ, centered around $\|\varphi\|$ and insensitive to syntax, s.th. $\|\varphi \star_\leq \psi\| = \min_{\leq_\varphi} \|\psi\|$

AGM theory: representation theorem

- representation theorem:
  let $\star : \mathcal{L}_{\text{CPL}} \times \mathcal{L}_{\text{CPL}} \longrightarrow \mathcal{L}_{\text{CPL}}$ be any mapping;

  $\star$ satisfies the (extended) AGM postulates iff
  there is a family of total preorders $\leq_\varphi$, one for every φ, centered around $||\varphi||$ and insensitive to syntax, s.th. $||\varphi \star_\leq \psi|| = \min_{\leq_\varphi} ||\psi||$

- other semantics:
  - partial meet contraction [AGM85]
    - $S \bot \psi = \{S' \subseteq S \ : \ S \not\vdash \psi\}$
    - $S \star \psi = \gamma(S \bot \neg\psi) + \psi$
  - epistemic entrenchment orderings $\leq$ on *formulas* [Gär88]
    - constraints on ordering: . . .
    - relation with possibility theory [Zadeh, Dubois and Prade]
  - . . .
  - Spohn's ordinal conditional functions [Spo88]
    - numerical version of sphere systems

AGM theory: integrations with doxastic logic

- “Two traditions in the logic of belief: bringing them together” [Seg95, Seg99]
  - modal operators B_i , $[+\psi]$, $[-\psi]$, $[\star\psi]$
  - $[\star\psi]\varphi$ = “φ is true after revision by ψ”
- internal version of doxastic logic [Auc08]
  - straightforward transfer of AGM representation theorems to multiagent case
- distinguish several versions of belief [BS07, BS08]
  - soft beliefs: can be revised
  - hard beliefs: cannot

What we have seen up to now

- 'the' logic of knowledge
  - $S5_n$ = standard epistemic logic (narrow sense)
  - dynamics of knowledge:
    - PAL = Public Announcement Logic
    - DEL = Dynamic Epistemic Logic
- 'the' logic of belief
  - $KD45_n$ = standard doxastic logic
  - dynamics of belief:
    - AGM belief revision
- ... this is all about single-agent knowledge and belief: what about groups?

Group knowledge and group belief

Shared knowledge and the gossip problem

- shared knowledge ('everybody knows'):
  - $\mathsf{EK}_{\{i_1,\ldots,i_n\}}\,\varphi \stackrel{\text{def}}{=} \mathsf{K}_{i_1}\,\varphi \land \ldots \land \mathsf{K}_{i_n}\,\varphi$
- properties:
  - $\models (\mathsf{EK}_{J_1}\,\varphi \land \mathsf{EK}_{J_2}\,\varphi) \leftrightarrow \mathsf{EK}_{J_1 \cup J_2}\,\varphi$
  - $\not\models \mathsf{EK}_J\,\varphi \to \mathsf{EK}_J\,\mathsf{EK}_J\,\varphi$
  - remember: $Agts$ finite (else 2^{Agts} uncountable)
- gossip problem:
  - each of n friends has a secret s_i only known to him
  - the agents can only communicate by one-to-one phone calls
  - shared knowledge of depth 1 can be achieved by $2(n-2)$ calls
  - shared knowledge of depth d can be achieved by $(d+1)(n-2)$ calls

Shared knowledge and the gossip problem

- shared knowledge ('everybody knows'):
  - $\mathsf{EK}_{\{i_1,\ldots,i_n\}}\,\varphi \stackrel{\text{def}}{=} \mathsf{K}_{i_1}\,\varphi \land \ldots \land \mathsf{K}_{i_n}\,\varphi$
- properties:
  - $\models (\mathsf{EK}_{J_1}\,\varphi \land \mathsf{EK}_{J_2}\,\varphi) \leftrightarrow \mathsf{EK}_{J_1 \cup J_2}\,\varphi$
  - $\not\models \mathsf{EK}_J\,\varphi \to \mathsf{EK}_J\,\mathsf{EK}_J\,\varphi$
  - remember: $Agts$ finite (else 2^{Agts} uncountable)
- gossip problem:
  - each of n friends has a secret s_i only known to him
  - the agents can only communicate by one-to-one phone calls
  - shared knowledge of depth 1 can be achieved by $2(n-2)$ calls
  - shared knowledge of depth d can be achieved by $(d+1)(n-2)$ calls
  - common knowledge cannot be achieved

Common knowledge: language, motivation, semantics

- $\mathsf{CK}_{i,j}\,\varphi$ = "it is common knowledge of i and j that φ"
- informal definition:
  - $\mathsf{CK}_{i,j}\,\varphi = \mathsf{EK}_{i,j}\,\varphi \wedge \mathsf{EK}_{i,j}\,\mathsf{EK}_{i,j}\,\varphi \wedge \mathsf{EK}_{i,j}\,\mathsf{EK}_{i,j}\,\mathsf{EK}_{i,j}\,\varphi \wedge \ldots$
  - cannot be defined as an abbreviation $\Rightarrow$ new modal operator
- fundamental for coordination
  - conventions in societies ('drive on the right') [Lew69]
  - common ground in conversation ('what we agree on') [CS89]
  - coordinated attack problem ('Byzantine Generals') [FHMV95]
- truth condition:

  $M, w \Vdash \mathsf{CK}_{i,j}\,\varphi$ iff $M, w \Vdash \mathsf{EK}_{i,j}\,\varphi$ and $M, w \Vdash \mathsf{EK}_{i,j}\,\mathsf{EK}_{i,j}\,\varphi$ and ...
- in terms of accessibility relations:
  - $\mathcal{K}_{\mathsf{CK}_{i,j}}(w) = \mathcal{K}_{\mathsf{K}_i}(w) \cup \mathcal{K}_{\mathsf{K}_j}(w) \cup (\mathcal{K}_{\mathsf{K}_i} \circ \mathcal{K}_{\mathsf{K}_j})(w) \cup \ldots$

  so:
  - $\mathcal{K}_{\mathsf{CK}_J}(w) \overset{\text{def}}{=} (\bigcup_{i \in J} \mathcal{K}_{\mathsf{K}_i})^*(w)$

Common knowledge: axiomatization

- axiomatization of $KT5(\mathsf{K}_i)$ with common knowledge:
  - axiomatics $KT5(i)$
  - fixpoint axiom:
    - $\mathsf{CK}_J\,\varphi \leftrightarrow (\varphi \wedge \mathsf{EK}_J\,\mathsf{CK}_J\,\varphi)$
    - N.B.: right-to-left direction already a theorem:
      $\vdash_{KT5(\mathsf{K}_i)} \mathsf{EK}_J\,\mathsf{CK}_J\,\varphi \rightarrow \mathsf{K}_i\,\mathsf{CK}_J\,\varphi$, and
      $\vdash_{KT5(\mathsf{K}_i)} \mathsf{EK}_i\,\mathsf{CK}_J\,\varphi \rightarrow \mathsf{CK}_J\,\varphi$
  - greatest fixpoint axiom (alias induction axiom):
    - $(\varphi \wedge \mathsf{CK}_J\,(\varphi \rightarrow \mathsf{EK}_J\,\varphi)) \rightarrow \mathsf{CK}_J\,\varphi$

Common knowledge: axiomatization

- axiomatization of $KT5(\mathsf{K}_i)$ with common knowledge:
  - axiomatics $KT5(i)$
  - fixpoint axiom:
    - $\mathsf{CK}_J\,\varphi \leftrightarrow (\varphi \wedge \mathsf{EK}_J\,\mathsf{CK}_J\,\varphi)$
    - N.B.: right-to-left direction already a theorem:
      $\vdash_{KT5(\mathsf{K}_i)} \mathsf{EK}_J\,\mathsf{CK}_J\,\varphi \to \mathsf{K}_i\,\mathsf{CK}_J\,\varphi$, and
      $\vdash_{KT5(\mathsf{K}_i)} \mathsf{EK}_i\,\mathsf{CK}_J\,\varphi \to \mathsf{CK}_J\,\varphi$
  - greatest fixpoint axiom (alias induction axiom):
    - $(\varphi \wedge \mathsf{CK}_J\,(\varphi \to \mathsf{EK}_J\,\varphi)) \to \mathsf{CK}_J\,\varphi$
- sound, complete and decidable

Common knowledge: axiomatization

- axiomatization of $KT5(\mathsf{K}_i)$ with common knowledge:
  - axiomatics $KT5(i)$
  - fixpoint axiom:
    - $\mathsf{CK}_J\,\varphi \leftrightarrow (\varphi \land \mathsf{EK}_J\,\mathsf{CK}_J\,\varphi)$
    - N.B.: right-to-left direction already a theorem:
      $\vdash_{KT5(\mathsf{K}_i)} \mathsf{EK}_J\,\mathsf{CK}_J\,\varphi \to \mathsf{K}_i\,\mathsf{CK}_J\,\varphi$, and
      $\vdash_{KT5(\mathsf{K}_i)} \mathsf{EK}_i\,\mathsf{CK}_J\,\varphi \to \mathsf{CK}_J\,\varphi$
  - greatest fixpoint axiom (alias induction axiom):
    - $(\varphi \land \mathsf{CK}_J\,(\varphi \to \mathsf{EK}_J\,\varphi)) \to \mathsf{CK}_J\,\varphi$

- sound, complete and decidable
  - only weakly complete, but not strongly:
    - $\{\mathsf{EK}_J{}^n\varphi \;:\; n \geq 0\} \models \mathsf{CK}_J\,\varphi$, but
      $\{\mathsf{EK}_J{}^n\varphi \;:\; n \geq 0\} \not\vdash \mathsf{CK}_J\,\varphi$
  - 'S5$_n$ with common knowledge not compact'
  - same for LTL

Common knowledge: axiomatization

- axiomatization of $KT5(\mathsf{K}_i)$ with common knowledge:
  - axiomatics $KT5(i)$
  - fixpoint axiom:
    - $\mathsf{CK}_J\, \varphi \leftrightarrow (\varphi \land \mathsf{EK}_J\, \mathsf{CK}_J\, \varphi)$
    - N.B.: right-to-left direction already a theorem: $\vdash_{KT5(\mathsf{K}_i)} \mathsf{EK}_J\, \mathsf{CK}_J\, \varphi \to \mathsf{K}_i\, \mathsf{CK}_J\, \varphi$, and $\vdash_{KT5(\mathsf{K}_i)} \mathsf{EK}_i\, \mathsf{CK}_J\, \varphi \to \mathsf{CK}_J\, \varphi$
  - greatest fixpoint axiom (alias induction axiom):
    - $(\varphi \land \mathsf{CK}_J\, (\varphi \to \mathsf{EK}_J\, \varphi)) \to \mathsf{CK}_J\, \varphi$
- sound, complete and decidable
  - only weakly complete, but not strongly:
    - $\{\mathsf{EK}_J{}^n \varphi \ : \ n \geq 0\} \models \mathsf{CK}_J\, \varphi$, but $\{\mathsf{EK}_J{}^n \varphi \ : \ n \geq 0\} \not\vdash \mathsf{CK}_J\, \varphi$
  - 'S5$_n$ with common knowledge not compact'
  - same for LTL
- complexity of satisfiability: EXPTIME complete

Exercises

- muddy children with n children

Exercises

- muddy children with n children
  - solution requires n rounds
  - specification requires common knowledge

Exercises

- muddy children with n children
  - solution requires n rounds
  - specification requires common knowledge
- consecutive numbers: let n_i, n_j be integers;
  $\not\models \mathsf{CK}_{i,j}\,(|n_i - n_j| = 1) \to \mathsf{CK}_{i,j}\,(n_i \leq 100)$

Exercises

- muddy children with n children
  - solution requires n rounds
  - specification requires common knowledge

- consecutive numbers: let n_i, n_j be integers;
  $\not\models \mathsf{CK}_{i,j} (|n_i - n_j| = 1) \to \mathsf{CK}_{i,j} (n_i \leq 100)$

- prove that the logic of common knowledge has all principles of S5
  - prove that the reflexive and transitive union of equivalence relations is an equivalence relation
    - ⋆ $(\bigcup_{i \in J} \mathcal{K}_{\mathsf{K}_i})^*$ is reflexive
    - ⋆ if some $\mathcal{K}_{\mathsf{K}_i}$ is reflexive then $(\bigcup_{i \in J} \mathcal{K}_{\mathsf{K}_i})^+$ is reflexive
    - ⋆ if every $\mathcal{K}_{\mathsf{K}_i}$ is symmetric then $(\bigcup_{i \in J} \mathcal{K}_{\mathsf{K}_i})^+$ is symmetric

Common belief: semantics and axiomatics

- $\mathsf{EB}_J\,\varphi \stackrel{\mathsf{def}}{=} \bigwedge_{i\in J}\mathsf{B}_i\,\varphi$ 'everybody believes'
- $\mathsf{CB}_J\,\varphi = \mathsf{EB}_J\,\varphi \land \mathsf{EB}_J\,\mathsf{EB}_J\,\varphi \land \ldots$
- $\mathcal{K}_{\mathsf{CB}_J} \stackrel{\mathsf{def}}{=} (\bigcup_{i\in J}\mathcal{K}_{\mathsf{B}_i})^+$

Common belief: semantics and axiomatics

- $\mathsf{EB}_J\,\varphi \stackrel{\text{def}}{=} \bigwedge_{i\in J} \mathsf{B}_i\,\varphi$ 'everybody believes'
- $\mathsf{CB}_J\,\varphi = \mathsf{EB}_J\,\varphi \wedge \mathsf{EB}_J\,\mathsf{EB}_J\,\varphi \wedge \ldots$
- $\mathcal{K}_{\mathsf{CB}_J} \stackrel{\text{def}}{=} (\bigcup_{i\in J} \mathcal{K}_{\mathsf{B}_i})^+$

- axiomatization of $KD45(\mathsf{B}_i)$ with common belief:
  - axiomatics $KD45(\mathsf{B}_i)$
  - fixpoint axiom:
    - $\mathsf{CB}_J\,\varphi \leftrightarrow (\mathsf{EB}_J\,\varphi \wedge \mathsf{EB}_J\,\mathsf{CB}_J\,\varphi)$
  - least fixpoint inference rule (alias induction rule):
    - $\frac{\varphi\rightarrow\mathsf{EB}_J\,\varphi}{\mathsf{EB}_J\,\varphi\rightarrow\mathsf{CB}_J\,\varphi}$

    equivalent to least fixpoint axiom
    - $(\mathsf{EB}_J\,\varphi \wedge \mathsf{CB}_J\,(\varphi \rightarrow \mathsf{EB}_J\,\varphi)) \rightarrow \mathsf{CB}_J\,\varphi$
- sound, complete and decidable
- EXPTIME complete

Exercises

- prove that if $\mathcal{K}_{\mathsf{B}_i}$ is serial then $(\bigcup_{i \in J} \mathcal{K}_{\mathsf{B}_i})^+$ is serial
- prove that $(\bigcup_{i \in J} \mathcal{K}_{\mathsf{B}_i})^+$ is transitive
- prove that $(\bigcup_{i \in J} \mathcal{K}_{\mathsf{B}_i})^+$ is not necessarily Euclidean

Exercises

- prove that if $\mathcal{K}_{\mathsf{B}_i}$ is serial then $(\bigcup_{i \in J} \mathcal{K}_{\mathsf{B}_i})^+$ is serial
- prove that $(\bigcup_{i \in J} \mathcal{K}_{\mathsf{B}_i})^+$ is transitive
- prove that $(\bigcup_{i \in J} \mathcal{K}_{\mathsf{B}_i})^+$ is not necessarily Euclidean
  - $\not\models \neg \mathsf{CB}_{i,j}\, \varphi \to \mathsf{CB}_{i,j} \neg \mathsf{CB}_{i,j}\, \varphi$

    (no negative introspection!)
  - logic of common belief weaker than $KD45$!

What we have seen in this course

- ‘the’ logic of knowledge
  - $S5_n$ = standard epistemic logic (narrow sense)
  - dynamics of knowledge:
    - PAL = Public Announcement Logic
    - DEL = Dynamic Epistemic Logic
- ‘the’ logic of belief
  - $KD45_n$ = standard doxastic logic
  - dynamics of belief:
    - AGM belief revision
- shared knowledge, shared belief
- common knowledge, common belief

Common ground and the compatriots puzzle

- . . . [HL14]

Guillaume Aucher and Thomas Bolander.
Undecidability in epistemic planning.
In Francesca Rossi, editor, *IJCAI 2013, Proceedings of the 23rd International Joint Conference on Artificial Intelligence, Beijing, China, August 3-9, 2013*, pages 27–33. IJCAI/AAAI, 2013.

Carlos Alchourrón, Peter Gärdenfors, and David Makinson.
On the logic of theory change: Partial meet contraction and revision functions.
J. of Symbolic Logic, 50:510–530, 1985.

Guillaume Aucher.
Perspectives on belief and change.
PhD thesis, Université de Toulouse, July 2008.

Simon Baron-Cohen, Alan M. Leslie, and Uta Frith.
Does the autistic child have a 'theory of mind'?
Cognition, 21(1):37–46, 1985.

Thomas Bolander, Martin Holm Jensen, and François Schwarzentruber.
Complexity results in epistemic planning.
In Qiang Yang and Michael Wooldridge, editors, *Proceedings of the Twenty-Fourth International Joint Conference on Artificial Intelligence, IJCAI*

2015, Buenos Aires, Argentina, July 25-31, 2015, pages 2791–2797. AAAI Press, 2015.

Alexandru Baltag and Sonja Smets.
From conditional probability to the logic of doxastic actions.
In Dov Samet, editor, *Proc. 11th Conference on Theoretical Aspects of Rationality and Knowledge (TARK-2007)*, pages 52–61, 2007.

Alexandru Baltag and Sonja Smets.
Probabilistic dynamic belief revision.
Synthese, 165(2):179–202, 2008.

John P. Burgess.
Probability logic.
J. of Symbolic Logic, 34:264–274, 1969.

Herbert H. Clark and Edward F. Schaefer.
Contributing to discourse.
Cognitive Science, 13:259–294, 1989.

Pascal Engel.
Believing, holding true, and accepting.
Philosophical Explorations, 1(2):140–151, 1998.

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

F. Fitch.
A logical analysis of some value concepts.
The Journal of Symbolic Logic, 28:135–142, 1963.
reprinted in Salerno (ed.) 2009, 21-28.

Peter Gärdenfors.
Knowledge in Flux: Modeling the Dynamics of Epistemic States.
MIT Press, 1988.

Adam Grove.
Two modellings for theory change.
J. of Philosophical Logic, 17:157–170, 1988.

Paul Hakli.
Group beliefs and the distinction between belief and acceptance.
Cognitive Systems Research, 7:286–297, 2006.

Jaakko Hintikka.
Knowledge and belief.

Cornell University Press, Ithaca, N.Y., 1962.

Andreas Herzig and Emiliano Lorini.
A modal logic of perceptual belief.
In Franck Lihoreau and Manuel Rebuschi, editors, *Epistemology, Context, and Formalism*, volume 369 of *Synthese Library*, pages 197–211. Springer, 2014.

Vincent Hendricks and John Symons.
Epistemic logic.
In Edward N. Zalta, editor, *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University, fall 2015 edition, 2015.

Hirofumi Katsuno and Alberto O. Mendelzon.
On the difference between updating a knowledge base and revising it.
In Peter Gärdenfors, editor, *Belief revision*, pages 183–203. Cambridge University Press, 1992.
(preliminary version in Allen, J.A., Fikes, R., and Sandewall, E., eds., Principles of Knowledge Representation and Reasoning: Proc. 2nd Int. Conf., pages 387–394. Morgan Kaufmann Publishers, 1991).

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

Wolfgang Lenzen.
Recent work in epistemic logic.
North Holland Publishing Company, Amsterdam, 1978.

Wolfgang Lenzen.
On the semantics and pragmatics of epistemic attitudes.
In Armin Laux and Heinrich Wansing, editors, *Knowledge and belief in philosophy and AI*, pages 181–197. Akademie Verlag, Berlin, 1995.

David K. Lewis.
Convention: a philosophical study.
Harvard University Press, Cambridge, 1969.

David K. Lewis.
Counterfactuals.
Basil Blackwell, Oxford, 1973.

Krister Segerberg.
Belief revision from the point of view of doxastic logic.
Bulletin of the IGPL, 3:534–553, 1995.

Krister Segerberg.
Two traditions in the logic of belief: bringing them together.

In Hans Jürgen Ohlbach and Uwe Reyle, editors, *Logic, Language and Reasoning: essays in honour of Dov Gabbay*, volume 5 of *Trends in Logic*, pages 135–147. Kluwer Academic Publishers, Dordrecht, 1999.

Wolfgang Spohn.
Ordinal conditional functions: a dynamic theory of epistemic states.
In William L. Harper and Brian Skyrms, editors, *Causation in decision, belief change and statistics*, volume 2, pages 105–134. D. Reidel, Dordrecht, 1988.

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

Hans van Ditmarsch, Joseph Y Halpern, Wiebe van der Hoek, and Barteld Pieter Kooi.
Handbook of epistemic logic.
College Publications, 2015.