“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

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

Tuesday: Doxastic logic and the dynamics of belief

Plan

Doxastic logic: introduction

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

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

Doxastic logic: introduction

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

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

- relevant for:
  - formal epistemology
    - what is knowledge?
    - is knowledge possible at all?
    - are all truths knowable?

Doxastic logic: introduction

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

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

- relevant for:
  - formal epistemology
    - what is knowledge?
    - is knowledge possible at all?
    - are all truths knowable?
  - distributed processes [FHMV95]
    - 'muddy children' and other puzzles
    - cryptographic protocols [Abadi et al., Ditmarsch03]

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
  - 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
  - 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" should be is consistent

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

Doxastic logic: language

- BNF:

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

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

- $\mathrm{B}_i\,\varphi$ = “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:

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

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

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

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

  - 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

1. Multiagent doxastic logic $KD45_n$
   - Doxastic logic: introduction and language
   - Doxastic logic: semantics
   - Doxastic logic: axiomatics
2. Discussions
3. Dynamics of belief

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]:
  $\mathrm{B}_i\,\varphi$ = "agent i believes that φ"
  = "*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.)

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

- $\mathcal{B}_i$ 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.)

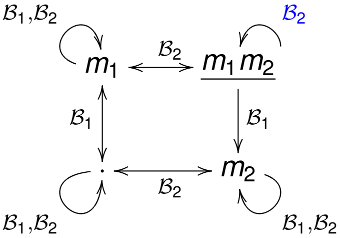
$$
\begin{aligned}
\mathcal{B}_i(w) &= \{w' : \langle w, w' \rangle \in \mathcal{B}_i\} \\
&= i\text{'s alternatives to } w \\
&= \text{worlds } i \text{ cannot distinguish from } w \text{ on basis of his beliefs} \\
&= \text{set of worlds compatible with } i\text{'s beliefs} \\
&= \textit{belief state}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

- $\mathcal{B}_i$ 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:
  - $M, w \Vdash \mathrm{B}_i \, \varphi$ iff $M, w' \Vdash \varphi$ for every $w' \in \mathcal{B}_i(w)$

Doxastic logic: semantics (ctd.)

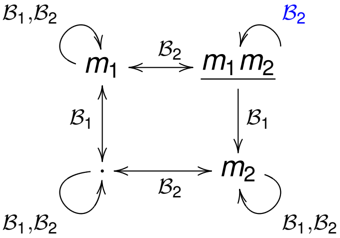
- variant of the muddy children puzzle: child 1 wrongly believes it is not muddy



$$\mathcal{B}_1(m_1 m_2) = \{(m_2)\}$$

Doxastic logic: semantics (ctd.)

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



$\mathcal{B}_1(m_1 m_2) = \{(m_2)\}$

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

Plan

Doxastic logic: axiomatics

- standard multiagent logic of belief = multimodal $KD45_n$
  - principles of multimodal K:
    - principles of classical propositional logic
    - $(B_i \varphi \wedge B_i \psi) \rightarrow B_i (\varphi \wedge \psi)$
    - from $\varphi \rightarrow \psi$ infer $B_i \varphi \rightarrow B_i \psi$
  - consistency of belief:
    - $\neg(B_i \varphi \wedge B_i \neg\varphi)$ axiom $D(B_i)$
  - positive introspection:
    - $B_i \varphi \rightarrow B_i B_i \varphi$ axiom $4(B_i)$
  - negative introspection:
    - $\neg B_i \varphi \rightarrow B_i \neg B_i \varphi$ 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 $\text{card}(Agts) = 1$
  - PSPACE-complete if $\text{card}(Agts) > 1$
- normal form if $n = 1$: modal depth ≤ 1

Plan

Omniscience problem

belief set of i at w	=	set of formulas believed by i at w
	=	$\{\varphi : M, w \Vdash \mathrm{B}_i\, \varphi\}$
	=	set of formulas true at every world of $\mathcal{B}_i(w)$

- in $KD45_n$, i's belief set is...
  - closed under theorems:
    - $\frac{\varphi}{\mathrm{B}_i\, \varphi}$ rule RN(B_i)
  - closed under logical implication:
    - $\frac{\varphi \to \psi}{\mathrm{B}_i\, \varphi \to \mathrm{B}_i\, \psi}$ rule RM(B_i)
  - closed under material implication:
    - $(\mathrm{B}_i\, \varphi \land \mathrm{B}_i\, (\varphi \to \psi)) \to \mathrm{B}_i\, \psi$ axiom K(B_i)

$\Rightarrow$ *omniscience problem*

- $KD45_n$'s belief is an idealization: rational agent, perfect reasoner
  - inadequate for human agents
  - widely accepted in AI

Plan

1. Multiagent doxastic logic $KD45_n$

2. Discussions
   - Discussion: omniscience
   - Discussion: graded belief
   - Discussion: relation between belief and knowledge
   - Discussion: belief vs. acceptance

3. Dynamics of belief

Graded belief

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

Graded belief

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

Graded belief

- language: $\mathrm{B}_i{}^{\geq d}\varphi$ = "i believes φ with degree at least d" ($d \in [0,1]$)
- semantics: $M = \langle W, \mathcal{B}, V \rangle$ 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:
  - $KD45(\mathrm{B}_i{}^{\geq d})$, for every i and d
  - $\mathrm{B}_i{}^{\geq d}\varphi \rightarrow \mathrm{B}_i{}^{\geq d'}\varphi$ if $d \geq d'$

Plan

Can knowledge be defined from belief?

[Plato, Theaetetus]

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

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’
- $\mathsf{K}_i\,\varphi \stackrel{\text{def}}{=} \mathsf{B}_i\,\varphi \wedge \varphi \wedge \mathit{hasJustif}(i, \varphi)$

Can knowledge be defined from belief?

[Plato, Theaetetus]

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

Relation between knowledge and belief?

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

Relation between knowledge and belief?

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

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

Relation between knowledge and belief?

- suppose a logic of knowledge and belief defined as:
  - $KD45(\mathsf{B}_i)$
  - $S5(\mathsf{K}_i)$
  - $\mathsf{K}_i\,\varphi \to \mathsf{B}_i\,\varphi$ ('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: $\neg\mathsf{B}_i\,\neg\mathsf{K}_i\,\varphi \to \neg\mathsf{K}_i\,\neg\mathsf{B}_i\,\varphi$

Relation between knowledge and belief?

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

Plan

Belief vs. acceptance

- A lawyer might *accept* that his client is innocent, while privately believing that his client is guilty.
- distinguishing features [Eng98, Hak06]:
  - Beliefs are not subject to the agent's will, whereas acceptances are voluntary.
  - Beliefs aim at truth, acceptances aim at utility (they depend on goals).
  - Beliefs are shaped by evidence, whereas acceptances need not be.
  - Beliefs come in degrees, while acceptances are binary.
  - Beliefs are context-independent whereas acceptances depend on context.
- logic of acceptance: see Thursday lecture on collective attitudes

Dynamics of belief

Plan

1. Multiagent doxastic logic $KD45_n$

2. Discussions

3. Dynamics of belief
   - Dynamics of belief: introduction and motivation
   - Dynamics of belief: the AGM theory

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: $\vdash_{S5_n - PAL} \mathrm{K}_i\, p \rightarrow [\neg p!]\bot$
    - ★ proof:
      $\vdash_{S5_n} \mathrm{K}_i\, p \rightarrow p$
      $\vdash_{S5_n - PAL}\, p \leftrightarrow [\neg p!]\bot$ (reduction axiom)
  - in doxastic logic:
    - ★ $\mathrm{B}_i\, p \wedge \neg p$ is $KD45_n$ satisfiable
    - ★ $\vdash_{KD45_n - PAL}\, p \leftrightarrow [\neg p!]\bot$ (reduction axiom)
    - ★ $\mathrm{B}_i\, p \wedge \neg[\neg p!]\bot$ should be $KD45_n{-}PAL$ satisfiable!

The logic of belief: getting dynamic (ctd.)

- exercise: prove $\vdash_{KD45_n-PAL} (\neg p \wedge \mathsf{B}_i\, p) \rightarrow \langle \neg p! \rangle \mathsf{B}_i \bot$

The logic of belief: getting dynamic (ctd.)

- exercise: prove $\vdash_{KD45_n - PAL} (\neg p \land \mathrm{B}_i\, p) \to \langle \neg p ! \rangle \mathrm{B}_i \bot$
  1. $\neg p \to \langle \neg p ! \rangle \top$ (red.ax.)

The logic of belief: getting dynamic (ctd.)

- exercise: prove $\vdash_{KD45_n - PAL} (\neg p \wedge B_i\, p) \rightarrow \langle \neg p! \rangle B_i \bot$
  1. $\neg p \rightarrow \langle \neg p! \rangle \top$ (red.ax.)
  2. $[\neg p!] B_i \neg p$
     - ⋆ reduction:

$$
\begin{aligned}
[\neg p!] B_i \neg p &\leftrightarrow \neg p \rightarrow B_i [\neg p!] \neg p \\
&\leftrightarrow \neg p \rightarrow B_i (\neg p \rightarrow \neg p) \\
&\leftrightarrow \neg p \rightarrow B_i \top \\
&\leftrightarrow \neg p \rightarrow \top \\
&\leftrightarrow \top
\end{aligned}
$$

The logic of belief: getting dynamic (ctd.)

- exercise: prove $\vdash_{KD45_n-PAL} (\neg p \land \mathrm{B}_i\, p) \rightarrow \langle \neg p! \rangle \mathrm{B}_i \bot$
  1. $\neg p \rightarrow \langle \neg p! \rangle \top$ (red.ax.)
  2. $[\neg p!]\mathrm{B}_i\, \neg p$
     - ★ reduction:
       $$\begin{array}{rcl} [\neg p!]\mathrm{B}_i\, \neg p & \leftrightarrow & \neg p \rightarrow \mathrm{B}_i\, [\neg p!]\neg p \\ & \leftrightarrow & \neg p \rightarrow \mathrm{B}_i\, (\neg p \rightarrow \neg p) \\ & \leftrightarrow & \neg p \rightarrow \mathrm{B}_i\, \top \\ & \leftrightarrow & \neg p \rightarrow \top \\ & \leftrightarrow & \top \end{array}$$
  3. $\mathrm{B}_i\, p \rightarrow [\neg p!]\mathrm{B}_i\, p$
     - ★ reduction:
       $$\begin{array}{rcl} [\neg p!]\mathrm{B}_i\, p & \leftrightarrow & \neg p \rightarrow \mathrm{B}_i\, [\neg p!]p \\ & \leftrightarrow & \neg p \rightarrow \mathrm{B}_i\, (\neg p \rightarrow p) \\ & \leftrightarrow & \neg p \rightarrow \mathrm{B}_i\, p \end{array}$$

The logic of belief: getting dynamic (ctd.)

- exercise: prove $\vdash_{KD45_n - PAL} (\neg p \wedge \mathrm{B}_i\, p) \rightarrow \langle \neg p! \rangle \mathrm{B}_i \bot$
  1. $\neg p \rightarrow \langle \neg p! \rangle \top$ (red.ax.)
  2. $[\neg p!]\mathrm{B}_i \neg p$
     - ★ reduction:
       $$\begin{aligned} [\neg p!]\mathrm{B}_i \neg p &\leftrightarrow \neg p \rightarrow \mathrm{B}_i [\neg p!]\neg p \\ &\leftrightarrow \neg p \rightarrow \mathrm{B}_i (\neg p \rightarrow \neg p) \\ &\leftrightarrow \neg p \rightarrow \mathrm{B}_i \top \\ &\leftrightarrow \neg p \rightarrow \top \\ &\leftrightarrow \top \end{aligned}$$
  3. $\mathrm{B}_i\, p \rightarrow [\neg p!]\mathrm{B}_i\, p$
     - ★ reduction:
       $$\begin{aligned} [\neg p!]\mathrm{B}_i\, p &\leftrightarrow \neg p \rightarrow \mathrm{B}_i [\neg p!]p \\ &\leftrightarrow \neg p \rightarrow \mathrm{B}_i (\neg p \rightarrow p) \\ &\leftrightarrow \neg p \rightarrow \mathrm{B}_i\, p \end{aligned}$$
  4. $(\neg p \wedge \mathrm{B}_i\, p) \rightarrow \langle \neg p! \rangle (\mathrm{B}_i\, p \wedge \mathrm{B}_i \neg p)$ (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

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

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
     $(M, w \Vdash \hat{B}_i\,\varphi$ and $M^{\varphi!}, w \Vdash \psi)$, or
     $(M, w \Vdash B_i\,\neg\varphi$ and $M, w \Vdash \psi)$

     - ★ reduction axiom:
       $[\varphi!]B_i\,\psi \quad \leftrightarrow \quad \neg\varphi \vee (\hat{B}_i\,\varphi \wedge B_i\,[\varphi!]\psi) \vee (B_i\,\neg\varphi \wedge B_i\,\psi)$
     - ★ 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
     ($M, w \Vdash \hat{\mathrm{B}}_i\,\varphi$ and $M^{\varphi!}, w \Vdash \psi$), or
     ($M, w \Vdash \mathrm{B}_i\,\neg\varphi$ and $M, w \Vdash \psi$)

     - ★ reduction axiom:

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

Plan

AGM theory: the internal perspective

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

- $\varphi \in S$ = "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 ("from φ infer $\mathrm{B}_i\, \varphi$")

AGM theory: coherentism vs. foundationalism

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

- 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:
  - $op : 2^{\mathcal{L}_{CPL}} \times \mathcal{L}_{CPL} \longrightarrow 2^{\mathcal{L}_{CPL}}$ [AGM85]
- agent's beliefs = formula:
  - $op : \mathcal{L}_{CPL} \times \mathcal{L}_{CPL} \longrightarrow \mathcal{L}_{CPL}$ [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 $\varphi \not\vdash \neg\psi \ldots$")

AGM theory: semantics

- model = sphere system: set of centered *spheres* surrounding $||\varphi||$
  - [Grove], 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 $\|\varphi\|$
  - [Grove], 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 $\star : \mathcal{L}_{CPL} \times \mathcal{L}_{CPL} \longrightarrow \mathcal{L}_{CPL}$ 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 $\star : \mathcal{L}_{CPL} \times \mathcal{L}_{CPL} \longrightarrow \mathcal{L}_{CPL}$ 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 [Baltag and Smets 07, 08]
  - soft beliefs: can be revised
  - hard beliefs: cannot
  - ESSLLI course "Dynamic Logics for Interactive Belief Revision" (Baltag and Smets, 2nd week)

What we saw in this lecture

- standard logic of belief: $KD45_n$
  - criticisms: omniscience
  - static
- dynamics of belief
  - belief revision

Next lecture

- logic of choice, goals and intentions

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.

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.

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

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

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

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).

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 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.