Models of Strategic Reasoning
Lecture 4

Eric Pacuit

University of Maryland, College Park
ai.stanford.edu/~epacuit

August 9, 2012

Game Plan

✓ Introduction, Motivation and Background

✓ The Dynamics of Rational Deliberation

✓ Reasoning to a Solution: Common Modes of Reasoning in Games

Lecture 4: Reasoning to a Model: Iterated Belief Change as Deliberation

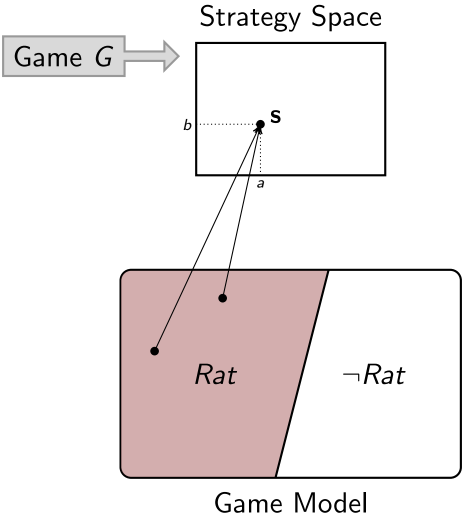
Lecture 5: Reasoning in Specific Games: Experimental Results

Informational Context of a Game

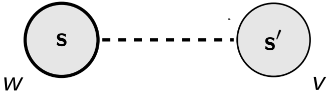
"In any particular structure, certain beliefs, beliefs about belief, ..., will be present and others won't be. So, there is an important implicit assumption behind the choice of a structure. This is that it is "transparent" to the players that the beliefs in the type structure — and only those beliefs — are possible ....The idea is that there is a "context" to the strategic situation (eg., history, conventions, etc.) and this "context" causes the players to rule out certain beliefs." (pg. 810)

Adam Brandenburger and Amanda Friedenberg. *Self-Admissible Sets*. Journal of Economic Theory, 145, 785 - 811, 2010.

Informational Context of a Game



Informational Context of a Game



Epistemic Model: $\mathcal{M} = \langle W, \{\sim_i\}_{i \in \mathcal{A}}, V \rangle$

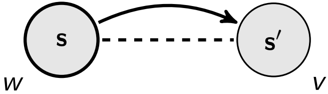
- $w \sim_i v$ means i cannot rule out v according to her information.

Language: $\varphi := p \mid \neg\varphi \mid \varphi \wedge \psi \mid K_i\varphi$

Truth:

- $\mathcal{M}, w \models p$ iff $w \in V(p)$ (p an atomic proposition)
- Boolean connectives as usual
- $\mathcal{M}, w \models K_i\varphi$ iff for all $v \in W$, if $w \sim_i v$ then $\mathcal{M}, v \models \varphi$

Informational Context of a Game



Epistemic-Plausibility Model: $\mathcal{M} = \langle W, \{\sim_i\}_{i \in \mathcal{A}}, \{\preceq_i\}_{i \in \mathcal{A}}, V \rangle$

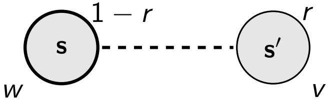
- $w \preceq_i v$ means v is at least as plausibility as w for agent i.

Language: $\varphi := p \mid \neg\varphi \mid \varphi \wedge \psi \mid K_i\varphi \mid B^\varphi\psi \mid [\preceq_i]\varphi$

Truth:

- $[\![\varphi]\!]_\mathcal{M} = \{w \mid \mathcal{M}, w \models \varphi\}$
- $\mathcal{M}, w \models B_i^\varphi\psi$ iff for all $v \in Min_{\preceq_i}([\![\varphi]\!]_\mathcal{M} \cap [w]_i)$, $\mathcal{M}, v \models \psi$
- $\mathcal{M}, w \models [\preceq_i]\varphi$ iff for all $v \in W$, if $v \preceq_i w$ then $\mathcal{M}, v \models \varphi$

Informational Context of a Game



Epistemic-Plausibility Model: $\mathcal{M} = \langle W, \{\sim_i\}_{i\in\mathcal{A}}, \{\pi_i\}_{i\in\mathcal{A}}, V\rangle$

- $\pi_i : W \to [0,1]$ is a probability measure

Language: $\varphi := p \mid \neg\varphi \mid \varphi\wedge\psi \mid K_i\varphi \mid B^p\psi$

Truth:

- $[\![\varphi]\!]_{\mathcal{M}} = \{w \mid \mathcal{M}, w \models \varphi\}$
- $\mathcal{M}, w \models B^p\varphi$ iff $\pi_i([\![\varphi]\!]_{\mathcal{M}} \mid [w]_i) = \frac{\pi_i([\![\varphi]\!]_{\mathcal{M}}\cap[w]_i)}{\pi_i([w]_i)} \geq p$, $\mathcal{M}, v \models \psi$
- $\mathcal{M}, w \models K_i\varphi$ iff for all $v \in W$, if $w \sim_i v$ then $\mathcal{M}, v \models \varphi$

Reasoning *to* a context

"It is important to understand that we have two forms of irrationality in this paper...For us, a player is rational if he optimizes and also rules nothing out. So irrationality might mean not optimizing. But it can also mean optimizing while not considering everything possible."

(pg. 314)

A. Brandenburger, A. Friedenberg and H. J. Keisler. *Admissibility in Games*. Econometrica, 76:2, 2008, pgs. 307 - 352.

Reasoning *to* a context

"It is important to understand that we have two forms of irrationality in this paper...For us, a player is rational if he optimizes and also rules nothing out. So irrationality might mean not optimizing. But it can also mean optimizing while not considering everything possible."

(pg. 314)

A. Brandenburger, A. Friedenberg and H. J. Keisler. *Admissibility in Games*. Econometrica, 76:2, 2008, pgs. 307 - 352.

A player can be rationally criticized for

Reasoning *to* a context

"It is important to understand that we have two forms of irrationality in this paper...For us, a player is rational if he optimizes and also rules nothing out. So irrationality might mean not optimizing. But it can also mean optimizing while not considering everything possible."

(pg. 314)

A. Brandenburger, A. Friedenberg and H. J. Keisler. *Admissibility in Games*. Econometrica, 76:2, 2008, pgs. 307 - 352.

A player can be rationally criticized for

1. not choosing what is *best* or what is *rationally permissible*, *given one's information*.

Reasoning *to* a context

"It is important to understand that we have two forms of irrationality in this paper...For us, a player is rational if he optimizes and also rules nothing out. So irrationality might mean not optimizing. But it can also mean optimizing while not considering everything possible."

(pg. 314)

A. Brandenburger, A. Friedenberg and H. J. Keisler. *Admissibility in Games*. Econometrica, 76:2, 2008, pgs. 307 - 352.

A player can be rationally criticized for

1. not choosing what is *best* or what is *rationally permissible*, *given one's information*.
2. not reasoning to a "proper" informational context.

Key Idea

Informational contexts of a game arise as fixed points of iterated "rationality announcements".

Key Idea

Informational contexts of a game arise as fixed points of iterated "rationality announcements".

J. van Benthem. *Rational dynamics and epistemic logic in games*. International Game Theory Review 9, 1 (2007), 13-45.

A. Baltag, S. Smets, and J. Zvesper. *Keep hoping for rationality: a solution to the backwards induction paradox*. Synthese 169 (2009), 301-333.

K. Apt and J. Zvesper. *Public announcements in strategic games with arbitrary strategy sets*. Proceedings of LOFT 2010 (2010).

J. van Benthem, and A. Gheerbrant. *Game solution, epistemic dynamics and fixed-point logics*. Fund. Inform. 100 (2010), 1-23.

Dynamic Epistemic/Doxastic Logic

J. van Benthem. *Logical Dynamics of Information and Interaction*. Cambridge University Press, 2011.

EP. *Dynamic Epistemic Logic Part I: Modeling Knowledge and Belief*. Philosophy Compass, forthcoming.

EP. *Dynamic Epistemic Logic Part II: Logics of Information Change*. Philosophy Compass, forthcoming.

Modeling Information Change: Two Methodologies

1. "Change-based modeling": describe the effect a *learning experience* has on a model

2. "Explicit-temporal modeling": explicitly describe different moments *in the model*

Modeling Information Change: Two Methodologies

1. "Change-based modeling": describe the effect a *learning experience* has on a model
2. "Explicit-temporal modeling": explicitly describe different moments *in the model*

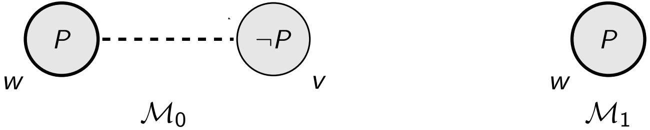
$[\psi]\varphi$: after everyone *finds out* that ψ is true, φ is true

$[\psi]\varphi$: after everyone *finds out* that ψ is true, φ is true

How did you find out that ψ?

- direct observation of ψ
- public announcement of ψ
- ...

Finding out that p is true



Public Announcement Logic

J. Plaza. *Logics of Public Communications*. 1989.

J. Gerbrandy. *Bisimulations on Planet Kripke*. 1999.

J. van Benthem. *One is a lonely number*. 2002.

Public Announcement Logic

$$p \mid \neg\varphi \mid \varphi \wedge \varphi \mid K_i\varphi \mid C\varphi \mid [\psi]\varphi$$

where $p \in \mathsf{At}$ and $i \in \mathcal{A}$.

Public Announcement Logic

$$p \mid \neg\varphi \mid \varphi \wedge \varphi \mid K_i\varphi \mid C\varphi \mid [\psi]\varphi$$

where $p \in \mathrm{At}$ and $i \in \mathcal{A}$.

- $[\psi]\varphi$ is intended to mean "After ψ is publicly announced, φ is true".

Public Announcement Logic

$$p \mid \neg\varphi \mid \varphi \wedge \varphi \mid K_i\varphi \mid C\varphi \mid [\psi]\varphi$$

where $p \in \mathsf{At}$ and $i \in \mathcal{A}$.

- $[p]K_ip$: *after publicly announcing P, agent i knows P*

Public Announcement Logic

$$p \mid \neg\varphi \mid \varphi \wedge \varphi \mid K_i\varphi \mid C\varphi \mid [\psi]\varphi$$

where $p \in \mathsf{At}$ and $i \in \mathcal{A}$.

- $[\neg K_i p]Cp$: *after announcing that agent i does not know p, then p is common knowledge*

Public Announcement Logic

$$p \mid \neg\varphi \mid \varphi \wedge \varphi \mid K_i\varphi \mid C\varphi \mid [\psi]\varphi$$

where $p \in \mathrm{At}$ and $i \in \mathcal{A}$.

- $[\neg K_i p]K_i p$: *after announcing i does not know p, then i knows p*

Public Announcement Logic

Suppose $\mathcal{M} = \langle W, \{\sim_i\}_{i\in\mathcal{A}}, V\rangle$ is a multi-agent epistemic models

$$\mathcal{M}, w \models [\psi]\varphi \text{ iff } \mathcal{M}, w \models \psi \text{ implies } \mathcal{M}^{\psi}, w \models \varphi$$

where $\mathcal{M}^{\psi} = \langle W', \{\sim_i'\}_{i\in\mathcal{A}}, V'\rangle$ with

- $W' = W \cap \{w \mid \mathcal{M}, w \models \psi\}$
- For each i, $\sim_i' = \sim_i \cap (W' \times W')$
- for all $p \in \mathrm{At}$, $V'(p) = V(p) \cap W'$

Public Announcement Logic

$$[\psi]p \quad \leftrightarrow \quad (\psi \rightarrow p)$$

Public Announcement Logic

$$
\begin{aligned}
[\psi]p &\leftrightarrow (\psi \rightarrow p) \\
[\psi]\neg\varphi &\leftrightarrow (\psi \rightarrow \neg[\psi]\varphi)
\end{aligned}
$$

Public Announcement Logic

$$
\begin{array}{rcl}
[\psi]p & \leftrightarrow & (\psi \rightarrow p) \\
[\psi]\neg\varphi & \leftrightarrow & (\psi \rightarrow \neg[\psi]\varphi) \\
[\psi](\varphi \wedge \chi) & \leftrightarrow & ([\psi]\varphi \wedge [\psi]\chi)
\end{array}
$$

Public Announcement Logic

$$
\begin{array}{rcl}
[\psi]p & \leftrightarrow & (\psi \rightarrow p) \\
[\psi]\neg\varphi & \leftrightarrow & (\psi \rightarrow \neg[\psi]\varphi) \\
[\psi](\varphi \wedge \chi) & \leftrightarrow & ([\psi]\varphi \wedge [\psi]\chi) \\
[\psi][\varphi]\chi & \leftrightarrow & [\psi \wedge [\psi]\varphi]\chi
\end{array}
$$

Public Announcement Logic

$$
\begin{aligned}
[\psi]p &\leftrightarrow (\psi \to p) \\
[\psi]\neg\varphi &\leftrightarrow (\psi \to \neg[\psi]\varphi) \\
[\psi](\varphi \wedge \chi) &\leftrightarrow ([\psi]\varphi \wedge [\psi]\chi) \\
[\psi][\varphi]\chi &\leftrightarrow [\psi \wedge [\psi]\varphi]\chi \\
[\psi]K_i\varphi &\leftrightarrow (\psi \to K_i(\psi \to [\psi]\varphi))
\end{aligned}
$$

Public Announcement Logic

$$
\begin{array}{rcl}
[\psi]p & \leftrightarrow & (\psi \rightarrow p) \\
[\psi]\neg\varphi & \leftrightarrow & (\psi \rightarrow \neg[\psi]\varphi) \\
[\psi](\varphi \wedge \chi) & \leftrightarrow & ([\psi]\varphi \wedge [\psi]\chi) \\
[\psi][\varphi]\chi & \leftrightarrow & [\psi \wedge [\psi]\varphi]\chi \\
[\psi]K_i\varphi & \leftrightarrow & (\psi \rightarrow K_i(\psi \rightarrow [\psi]\varphi))
\end{array}
$$

Theorem Every formula of Public Announcement Logic is equivalent to a formula of Epistemic Logic.

Public Announcement Logic

$$
\begin{array}{rcl}
[\psi]p & \leftrightarrow & (\psi \rightarrow p) \\
[\psi]\neg\varphi & \leftrightarrow & (\psi \rightarrow \neg[\psi]\varphi) \\
[\psi](\varphi \wedge \chi) & \leftrightarrow & ([\psi]\varphi \wedge [\psi]\chi) \\
[\psi][\varphi]\chi & \leftrightarrow & [\psi \wedge [\psi]\varphi]\chi \\
[\psi]K_i\varphi & \leftrightarrow & (\psi \rightarrow K_i(\psi \rightarrow [\psi]\varphi))
\end{array}
$$

The situation is more complicated with common knowledge.

J. van Benthem, J. van Eijk, B. Kooi. *Logics of Communication and Change*. 2006.

Aspects of Informative Events

1. The agents' *observational* powers.

   Agents may perceive the same event differently and this can be described in terms of what agents do or do not observe. Examples range from *public announcements* where everyone witnesses the same event to private communications between two or more agents with the other agents not even being aware that an event took place.

Aspects of Informative Events

1. The agents' *observational* powers.

2. The *type* of change triggered by the event.

   Agents may differ in precisely how they incorporate new information into their epistemic states. These differences are based, in part, on the agents' perception of the *source* of the information. For example, an agent may consider a particular source of information *infallible* (not allowing for the possibility that the source is mistaken) or merely *trustworthy* (accepting the information as reliable though allowing for the possibility of a mistake).

Aspects of Informative Events

1. The agents' *observational* powers.
2. The *type* of change triggered by the event.
3. The underlying *protocol* specifying which events (observations, messages, actions) are available (or permitted) at any given moment.

   This is intended to represent the rules or conventions that govern many of our social interactions. For example, in a conversation, it is typically not polite to "blurt everything out at the beginning", as we must speak in small chunks. Other natural conversational protocol rules include "do not repeat yourself", "let others speak in turn", and "be honest". Imposing such rules *restricts* the legitimate sequences of possible statements or events.

Aspects of Informative Events

1. The agents' *observational* powers.

2. The *type* of change triggered by the event.

3. The underlying *protocol* specifying which events (observations, messages, actions) are available (or permitted) at any given moment.

Aspects of Informative Events

1. The agents' *observational* powers.

2. The *type* of change triggered by the event.

3. The underlying *protocol* specifying which events (observations, messages, actions) are available (or permitted) at any given moment.

Finding out that φ

$$\mathcal{M} = \langle W, \{\sim_i\}_{i\in\mathcal{A}}, \{\preceq_i\}_{i\in\mathcal{A}}, V\rangle$$

$$\Downarrow$$

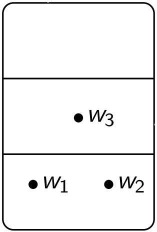
Find out that φ

$$\Downarrow$$

$$\mathcal{M}' = \langle W', \{\sim_i'\}_{i\in\mathcal{A}}, \{\preceq_i'\}_{i\in\mathcal{A}}, V|_{W'}\rangle$$

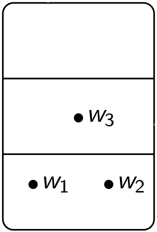
Informative Actions

- $w_1 \sim w_2 \sim w_3$



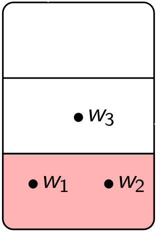
Informative Actions

- $w_1 \sim w_2 \sim w_3$
- $w_1 \preceq w_2$ and $w_2 \preceq w_1$ (w_1 and w_2 are equi-plausbile)
- $w_1 \prec w_3$ ($w_1 \preceq w_3$ and $w_3 \not\preceq w_1$)
- $w_2 \prec w_3$ ($w_2 \preceq w_3$ and $w_3 \not\preceq w_2$)

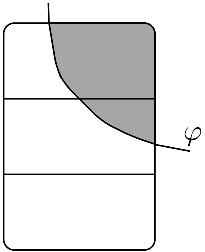


Informative Actions

- $w_1 \sim w_2 \sim w_3$
- $w_1 \preceq w_2$ and $w_2 \preceq w_1$ (w_1 and w_2 are equi-plausbile)
- $w_1 \prec w_3$ ($w_1 \preceq w_3$ and $w_3 \not\preceq w_1$)
- $w_2 \prec w_3$ ($w_2 \preceq w_3$ and $w_3 \not\preceq w_2$)
- $\{w_1, w_2\} \subseteq Min_{\preceq}([w_i])$

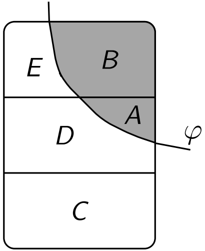


Informative Actions



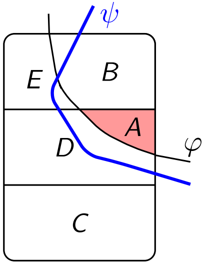
Incorporate the new information φ

Informative Actions



Incorporate the information that φ

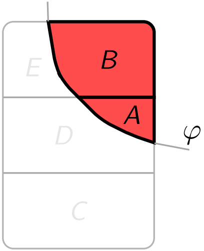
Informative Actions



Conditional Belief: $B^{\varphi}\psi$

$$Min_{\preceq}(W \cap [\![\varphi]\!]_{\mathcal{M}}) \subseteq [\![\psi]\!]_{\mathcal{M}}$$

Informative Actions

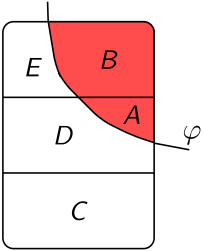


Public Announcement: Information from an infallible source

$(!\varphi)$: $A \prec_i B$ $\qquad \mathcal{M}^{!\varphi} = \langle W^{!\varphi}, \{\sim_i^{!\varphi}\}_{i\in\mathcal{A}}, V^{!\varphi}\rangle$

$W^{!\varphi} = [\![\varphi]\!]_{\mathcal{M}}$
$\sim_i^{!\varphi} = \sim_i \cap (W^{!\varphi} \times W^{!\varphi})$
$\preceq_i^{!\varphi} = \preceq_i \cap (W^{!\varphi} \times W^{!\varphi})$

Informative Actions

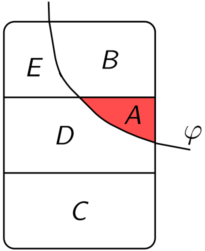


Radical Upgrade: $(\Uparrow\varphi)$: $A \prec_i B \prec_i C \prec_i D \prec_i E$,
$\mathcal{M}^{\Uparrow\varphi} = \langle W, \{\sim_i\}_{i\in\mathcal{A}}, \{\preceq_i^{\Uparrow\varphi}\}_{i\in\mathcal{A}}, V\rangle$

Let $[\![\varphi]\!]_i^w = \{x \mid \mathcal{M}, x \models \varphi\} \cap [w]_i$

- for all $x \in [\![\varphi]\!]_i^w$ and $y \in [\![\neg\varphi]\!]_i^w$, set $x \prec_i^{\Uparrow\varphi} y$,
- for all $x, y \in [\![\varphi]\!]_i^w$, set $x \preceq_i^{\Uparrow\varphi} y$ iff $x \preceq_i y$, and
- for all $x, y \in [\![\neg\varphi]\!]_i^w$, set $x \preceq_i^{\Uparrow\varphi} y$ iff $x \preceq_i y$.

Informative Actions



Conservative Upgrade: $(\uparrow\varphi)$: $A \prec_i C \prec_i D \prec_i B \cup E$

Conservative upgrade is radical upgrade with the formula

$$best_i(\varphi, w) := Min_{\preceq_i}([w]_i \cap \{x \mid \mathcal{M}, x \models \varphi\})$$

1. If $v \in best_i(\varphi, w)$ then $v \prec_i^{\uparrow\varphi} x$ for all $x \in [w]_i$, and
2. for all $x, y \in [w]_i - best_i(\varphi, w)$, $x \preceq_i^{\uparrow\varphi} y$ iff $x \preceq_i y$.

- $[q]Kq$

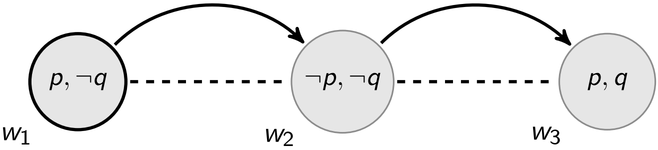
- $[q]Kq$
- $Kp \to [q]Kp$

- $[q]Kq$
- $Kp \rightarrow [q]Kp$
- $B\varphi \rightarrow [\psi]B\varphi$

- $[q]Kq$

- $Kp \to [q]Kp$

- $B\varphi \to [\psi]B\varphi$



- $[\varphi]\varphi$

Public Announcement vs. Conditional Belief

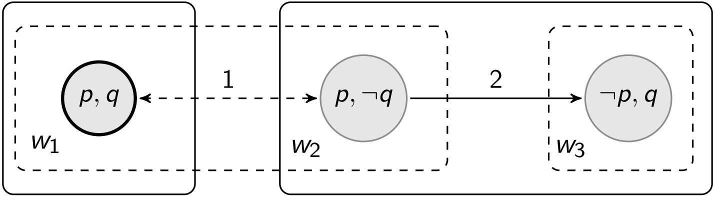
Are $[\varphi]B\psi$ and $B^{\varphi}\psi$ different?

Public Announcement vs. Conditional Belief

Are $[\varphi]B\psi$ and $B^{\varphi}\psi$ different? Yes!

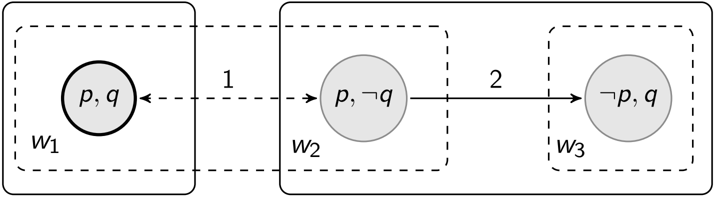
Public Announcement vs. Conditional Belief

Are $[\varphi]B\psi$ and $B^{\varphi}\psi$ different? Yes!



Public Announcement vs. Conditional Belief

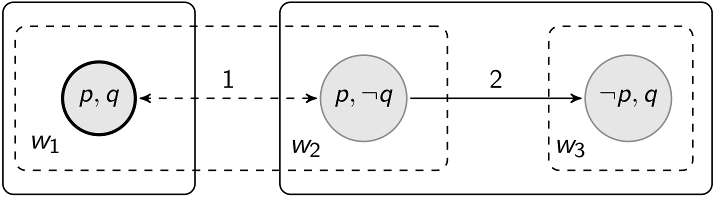
Are $[\varphi]B\psi$ and $B^{\varphi}\psi$ different? Yes!



- $w_1 \models B_1 B_2 q$

Public Announcement vs. Conditional Belief

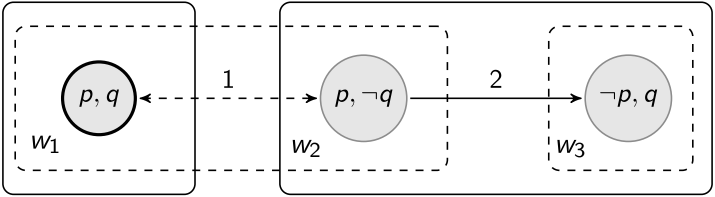
Are $[\varphi]B\psi$ and $B^{\varphi}\psi$ different? Yes!



- $w_1 \models B_1 B_2 q$
- $w_1 \models B_1^p B_2 q$

Public Announcement vs. Conditional Belief

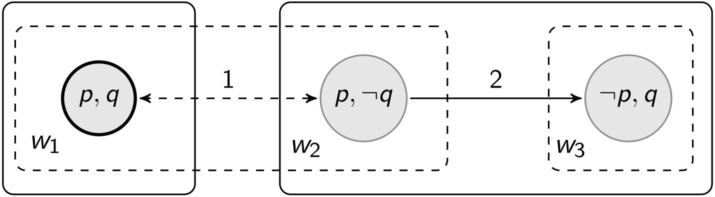
Are $[\varphi]B\psi$ and $B^{\varphi}\psi$ different? Yes!



- $w_1 \models B_1 B_2 q$
- $w_1 \models B_1^p B_2 q$
- $w_1 \models [p]\neg B_1 B_2 q$

Public Announcement vs. Conditional Belief

Are $[\varphi]B\psi$ and $B^{\varphi}\psi$ different? Yes!



- $w_1 \models B_1 B_2 q$
- $w_1 \models B_1^p B_2 q$
- $w_1 \models [p]\neg B_1 B_2 q$
- More generally, $B_i^p(p \wedge \neg K_i p)$ is satisfiable but $[p]B_i(p \wedge \neg K_i p)$ is not.

Hard and Soft Updates

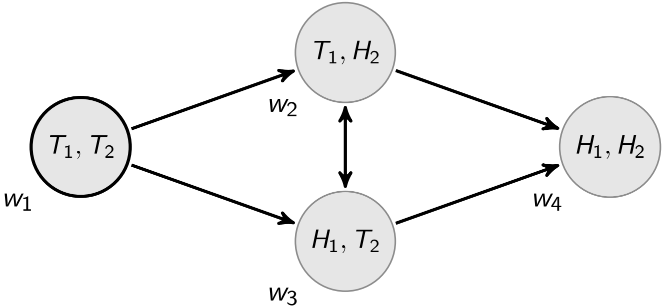
$$\mathcal{M} = \langle W, \{\sim_i\}_{i\in\mathcal{A}}, \{\preceq_i\}_{i\in\mathcal{A}}, V\rangle$$

$$\Big\Downarrow$$

Find out that φ

$$\Big\Downarrow$$

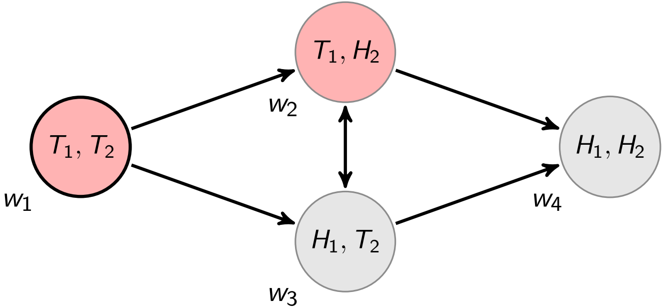
$$\mathcal{M} = \langle W', \{\sim_i'\}_{i\in\mathcal{A}}, \{\preceq_i'\}_{i\in\mathcal{A}}, V|_{W'}\rangle$$



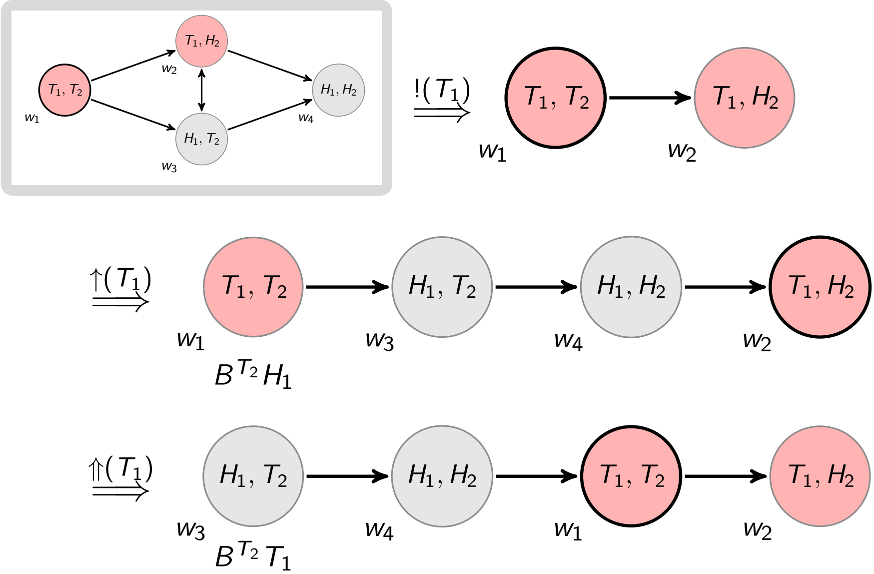
$Min_{\preceq}([w_1]) = \{w_4\}$, so $w_1 \models B(H_1 \wedge H_2)$

$Min_{\preceq}([w_1] \cap [\![T_1]\!]_{\mathcal{M}}) = \{w_2\}$, so $w_1 \models B^{T_1} H_2$

$Min_{\preceq}([w_1] \cap [\![T_1]\!]_{\mathcal{M}}) = \{w_3\}$, so $w_1 \models B^{T_2} H_1$



Suppose the agent *finds out that* T_1 *is/may be true*.


T1, T2
T1, H2
H1, T2
H1, H2
w1
w2
w3
w4
!(T1)
↑(T1)
⇑(T1)
B^T2 H1
B^T2 T1

What happens as beliefs change over time (iterated belief revision)?

$$\underset{\text{initial model}}{\mathcal{M}_0} \overset{!\varphi_1}{\Longrightarrow} \mathcal{M}_1 \overset{!\varphi_2}{\Longrightarrow} \mathcal{M}_2 \overset{!\varphi_3}{\Longrightarrow} \cdots \overset{!\varphi_n}{\Longrightarrow} \underset{\text{fixed-point}}{\mathcal{M}_f}$$

$$\underset{\text{initial model}}{\mathcal{M}_0} \xRightarrow{{\color{red}\Uparrow}\varphi_1} \mathcal{M}_1 \xRightarrow{{\color{red}\Uparrow}\varphi_2} \mathcal{M}_2 \xRightarrow{{\color{red}\Uparrow}\varphi_3} \cdots \xRightarrow{{\color{red}\Uparrow}\varphi_n} \underset{\text{fixed-point}}{\mathcal{M}_f}$$

$$\underset{\text{initial model}}{\mathcal{M}_0} \xRightarrow{!\varphi_1} \mathcal{M}_1 \xRightarrow{\Uparrow\varphi_2} \mathcal{M}_2 \xRightarrow{\uparrow\varphi_3} \cdots \xRightarrow{\Uparrow\varphi_n} \underset{\text{fixed-point}}{\mathcal{M}_f}$$

$$\underset{\text{initial model}}{\mathcal{M}_0} \xRightarrow{\tau(\varphi_1)} \mathcal{M}_1 \xRightarrow{\tau(\varphi_2)} \mathcal{M}_2 \xRightarrow{\tau(\varphi_3)} \cdots \xRightarrow{\tau(\varphi_n)} \underset{\text{fixed-point}}{\mathcal{M}_f}$$

Where do the φ_k come from?

Iterated Updates

$!\varphi_1, !\varphi_2, !\varphi_3, \ldots, !\varphi_n$
always reaches a fixed-point

Iterated Updates

$!\varphi_1, !\varphi_2, !\varphi_3, \ldots, !\varphi_n$
always reaches a fixed-point

$\Uparrow p \Uparrow \neg p \Uparrow p \cdots$
Contradictory beliefs leads to oscillations

Iterated Updates

$!\varphi_1, !\varphi_2, !\varphi_3, \ldots, !\varphi_n$
always reaches a fixed-point

$\Uparrow p \Uparrow\neg p \Uparrow p \cdots$
Contradictory beliefs leads to oscillations

$\uparrow\varphi, \uparrow\varphi, \ldots$
Simple beliefs may never stabilize

Iterated Updates

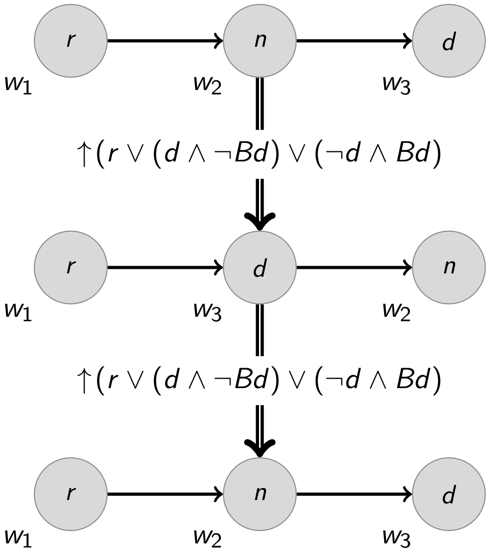
$!\varphi_1, !\varphi_2, !\varphi_3, \ldots, !\varphi_n$
always reaches a fixed-point

$\Uparrow p \Uparrow \neg p \Uparrow p \cdots$
Contradictory beliefs leads to oscillations

$\uparrow\varphi, \uparrow\varphi, \ldots$
Simple beliefs may never stabilize

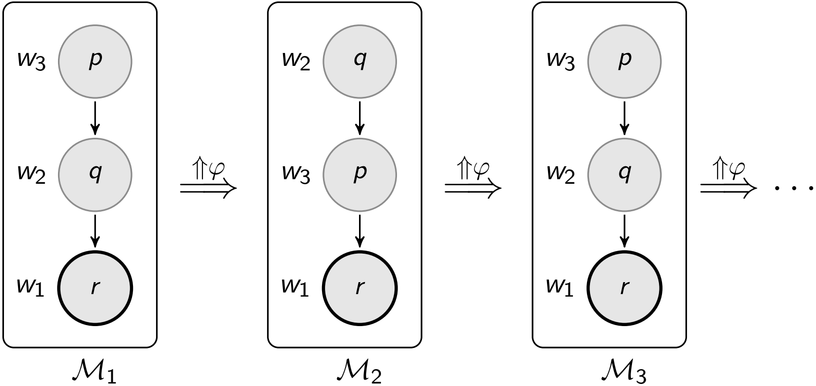
$\Uparrow\varphi, \Uparrow\varphi, \ldots$
Simple beliefs stabilize, but conditional beliefs do not

A. Baltag and S. Smets. *Group Belief Dynamics under Iterated Revision: Fixed Points and Cycles of Joint Upgrades*. TARK, 2009.



r
n
d
w1
w2
w3
↑(r ∨ (d ∧ ¬Bd) ∨ (¬d ∧ Bd)
r
d
n
w1
w3
w2
↑(r ∨ (d ∧ ¬Bd) ∨ (¬d ∧ Bd)
r
n
d
w1
w2
w3

Let φ be $(r \vee (B^{\neg r} q \wedge p) \vee (B^{\neg r} p \wedge q))$



Iterated Updates

$!\varphi_1, !\varphi_2, !\varphi_3, \ldots, !\varphi_n$
always reaches a fixed-point

$\Uparrow p \; \Uparrow \neg p \; \Uparrow p \cdots$
Contradictory beliefs leads to oscillations

$\uparrow\varphi, \uparrow\varphi, \ldots$
Simple beliefs may never stabilize

$\Uparrow\varphi, \Uparrow\varphi, \ldots$
Simple beliefs stabilize, but conditional beliefs do not

A. Baltag and S. Smets. *Group Belief Dynamics under Iterated Revision: Fixed Points and Cycles of Joint Upgrades*. TARK, 2009.

Iterated belief revision: two issues

C1: If $\alpha \to \varphi$ then $\Psi(\beta_1, \ldots, \beta_n, \varphi, \alpha) = \Psi(\beta_1, \ldots, \beta_n, \alpha)$
C2: If $\alpha \to \neg\varphi$ then $\Psi(\beta_1, \ldots, \beta_n, \varphi, \alpha) = \Psi(\beta_1, \ldots, \beta_n, \alpha)$

R. Stalnaker. *Iterated Belief Revision*. Erkentnis 70, pgs. 189 209, 2009.

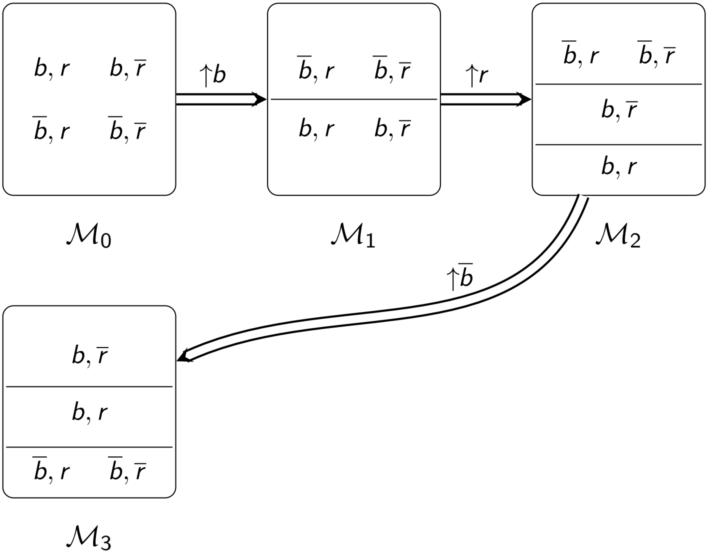
Suppose that you are in the forest and happen to a see strange-looking animal.

Suppose that you are in the forest and happen to a see strange-looking animal. You consult your animal guidebook and find a picture that seems to match the animal you see.

Suppose that you are in the forest and happen to a see strange-looking animal. You consult your animal guidebook and find a picture that seems to match the animal you see. The guidebook says that the animal is a type of bird, so that is what you conclude: The animal before you is a bird. After looking more closely, you also notice that the animal is also red.

Suppose that you are in the forest and happen to a see strange-looking animal. You consult your animal guidebook and find a picture that seems to match the animal you see. The guidebook says that the animal is a type of bird, so that is what you conclude: The animal before you is a bird. After looking more closely, you also notice that the animal is also red. So, you also update your beliefs with that fact.

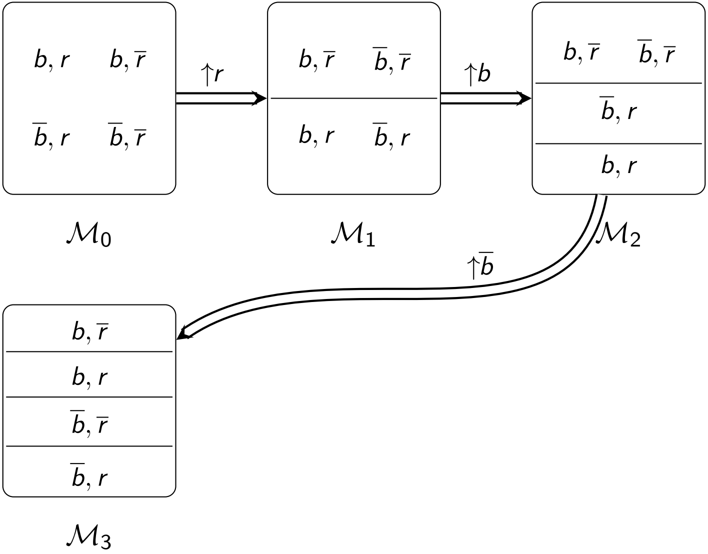
Suppose that you are in the forest and happen to a see strange-looking animal. You consult your animal guidebook and find a picture that seems to match the animal you see. The guidebook says that the animal is a type of bird, so that is what you conclude: The animal before you is a bird. After looking more closely, you also notice that the animal is also red. So, you also update your beliefs with that fact. Now, suppose that an expert (whom you trust) happens to walk by and tells you that the animal is, in fact, not a bird.


b, r $b, \overline{r}$
$\overline{b}, r$ $\overline{b}, \overline{r}$
$\mathcal{M}_0$
$\uparrow b$
$\overline{b}, r$ $\overline{b}, \overline{r}$
b, r $b, \overline{r}$
$\mathcal{M}_1$
$\uparrow r$
$\overline{b}, r$ $\overline{b}, \overline{r}$
$b, \overline{r}$
b, r
$\mathcal{M}_2$
$\uparrow \overline{b}$
$b, \overline{r}$
b, r
$\overline{b}, r$ $\overline{b}, \overline{r}$
$\mathcal{M}_3$

Note that in the last model, $\mathcal{M}_3$, the agent does not believe that the bird is red.

Note that in the last model, $\mathcal{M}_3$, the agent does not believe that the bird is red. The problem is that there does not seem to be any justification for why the agent drops her belief that the bird is red. This seems to result from the accidental fact that the agent started by updating with the information that the animal is a bird.

Note that in the last model, $\mathcal{M}_3$, the agent does not believe that the bird is red. The problem is that there does not seem to be any justification for why the agent drops her belief that the bird is red. This seems to result from the accidental fact that the agent started by updating with the information that the animal is a bird. In particular, note that the following sequence of updates is not problematic:


b, r $b, \overline{r}$
$\overline{b}, r$ $\overline{b}, \overline{r}$
$\mathcal{M}_0$
$\uparrow r$
$b, \overline{r}$ $\overline{b}, \overline{r}$
b, r $\overline{b}, r$
$\mathcal{M}_1$
$\uparrow b$
$b, \overline{r}$ $\overline{b}, \overline{r}$
$\overline{b}, r$
b, r
$\mathcal{M}_2$
$\uparrow \overline{b}$
$b, \overline{r}$
b, r
$\overline{b}, \overline{r}$
$\overline{b}, r$
$\mathcal{M}_3$

C1: If $\alpha \rightarrow \varphi$ then $\Psi(\beta_1, \ldots, \beta_n, \varphi, \alpha) = \Psi(\beta_1, \ldots, \beta_n, \alpha)$

UUU	DDD
UUD	DDU
UDU	DUD
UDD	DUU

- Three switches wired such that a light is on iff all three switches are up or all three are down.

UUU	DDD
UUD	DDU
UDU	DUD
UDD	DUU

- Three switches wired such that a light is on iff all three switches are up or all three are down.
- Three independent (reliable) observers report on the switches: Alice says switch 1 is U, Bob says switch 2 is D and Carla says switch 3 is U.

UUU	DDD
UUD	DDU
UDU	DUD
UDD	DUU

- Three switches wired such that a light is on iff all three switches are up or all three are down.
- Three independent (reliable) observers report on the switches: Alice says switch 1 is U, Bob says switch 2 is D and Carla says switch 3 is U.
- I receive the information that the light is on. What should I believe?

UUU	DDD
UUD	DDU
UDU	DUD
UDD	DUU

- Three switches wired such that a light is on iff all three switches are up or all three are down.
- Three independent (reliable) observers report on the switches: Alice says switch 1 is U, Bob says switch 2 is D and Carla says switch 3 is U.
- I receive the information that the light is on. What should I believe?
- Cautious: UUU, DDD; Bold: UUU

UUU	DDD
UUD	DDU
UDU	DUD
UDD	DUU

- Suppose there are two switches: L_1 is the main switch and L_2 is a secondary switch controlled by the first two lights. (So $L_1 \rightarrow L_2$, but not the converse)

UUU	DDD
UUD	DDU
UDU	DUD
UDD	DUU

- Suppose there are two switches: L_1 is the main switch and L_2 is a secondary switch controlled by the first two lights. (So $L_1 \to L_2$, but not the converse)
- Suppose I receive $L_1 \wedge L_2$, this does not change the story.

UUU	DDD
UUD	DDU
UDU	DUD
UDD	DUU

- Suppose there are two switches: L_1 is the main switch and L_2 is a secondary switch controlled by the first two lights. (So $L_1 \rightarrow L_2$, but not the converse)
- Suppose I receive $L_1 \wedge L_2$, this does not change the story.
- Suppose I learn that L_2. This is irrelevant to Carla's report, but it means either Ann or Bob is wrong.

UUU	DDD
UUD	DDU
UDU	DUD
UDD	DUU

- Suppose there are two switches: L_1 is the main switch and L_2 is a secondary switch controlled by the first two lights. (So $L_1 \rightarrow L_2$, but not the converse)
- Suppose I receive $L_1 \wedge L_2$, this does not change the story.
- Suppose I learn that L_2. This is irrelevant to Carla's report, but it means either Ann or Bob is wrong.

UUU	DDD
UUD	DDU
UDU	DUD
UDD	DUU

- Suppose there are two switches: L_1 is the main switch and L_2 is a secondary switch controlled by the first two lights. (So $L_1 \rightarrow L_2$, but not the converse)
- Suppose I receive $L_1 \wedge L_2$, this does not change the story.
- Suppose I learn that L_2. This is irrelevant to Carla's report, but it means either Ann or Bob is wrong.
- Now, after learning L_1, the only rational thing to believe is that all three switches are up.

UUU	DDD
UUD	DDU
UDU	DUD
UDD	DUU

- Suppose there are two switches: L_1 is the main switch and L_2 is a secondary switch controlled by the first two lights. (So $L_1 \rightarrow L_2$, but not the converse)
- Suppose I receive $L_1 \wedge L_2$, this does not change the story.
- Suppose I learn that L_2. This is irrelevant to Carla's report, but it means either Ann or Bob is wrong.
- Now, after learning L_1, the only rational thing to believe is that all three switches are up.

C2: If $\alpha \rightarrow \neg\varphi$ then $\Psi(\beta_1, \ldots, \beta_n, \varphi, \alpha) = \Psi(\beta_1, \ldots, \beta_n, \alpha)$

1. Alice and Bert (who I initially take to be reliable) report to me, independently, about the results: Alice tells me that the coin in room A came up heads (H_A), while Bert tells me that the coin in room B came up heads (H_B).

1. Alice and Bert (who I initially take to be reliable) report to me, independently, about the results: Alice tells me that the coin in room A came up heads (H_A), while Bert tells me that the coin in room B came up heads (H_B).

2. Carla and Dora, also two independent witnesses whose reliability, in my view, trumps that of Alice and Bert, give me conflicting information: Carla tells me that the coin in room A came up tails (T_A), and Dora tells me the same about the coin in room B (T_B)

1. Alice and Bert (who I initially take to be reliable) report to me, independently, about the results: Alice tells me that the coin in room A came up heads (H_A), while Bert tells me that the coin in room B came up heads (H_B).

2. Carla and Dora, also two independent witnesses whose reliability, in my view, trumps that of Alice and Bert, give me conflicting information: Carla tells me that the coin in room A came up tails (T_A), and Dora tells me the same about the coin in room B (T_B)

3. Elmer, whose reliability trumps everyone else, tells me that that the coin in room A in fact landed heads (H_A)

1. Alice and Bert (who I initially take to be reliable) report to me, independently, about the results: Alice tells me that the coin in room A came up heads (H_A), while Bert tells me that the coin in room B came up heads (H_B).

2. Carla and Dora, also two independent witnesses whose reliability, in my view, trumps that of Alice and Bert, give me conflicting information: Carla tells me that the coin in room A came up tails (T_A), and Dora tells me the same about the coin in room B (T_B)

3. Elmer, whose reliability trumps everyone else, tells me that that the coin in room A in fact landed heads (H_A)

What should I now believe about the coin in room B?

Many of the recent developments in this area have been driven by analyzing *concrete* examples.

This raises an important methodological issue: Implicit assumptions about what the actors know and believe about the situation being modeled often guide the analyst's intuitions. In many cases, it is crucial to make these underlying assumptions explicit.

The general point is that *how* the agent(s) come to know or believe that some proposition p is true is as important (or, perhaps, more important) than the fact that the agent(s) knows or believes that p is the case

Discussion

A key aspect of any formal model of a (social) interactive situation or situation of rational inquiry is the way it accounts for the

> *...information about how I learn some of the things I learn, about the sources of my information, or about what I believe about what I believe and don't believe. If the story we tell in an example makes certain information about any of these things relevant, then it needs to be included in a proper model of the story, if it is to play the right role in the evaluation of the abstract principles of the model.* (Stalnaker, pg. 203)

R. Stalnaker. *Iterated Belief Revision*. Erkentnis 70, pgs. 189 209, 2009.

Iterated belief revision as a model of deliberation

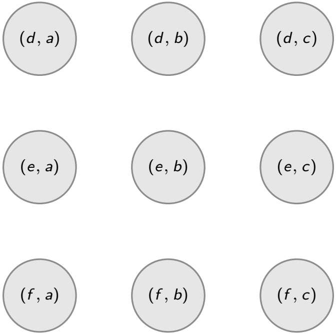
Reasoning *about* (strategic) games

Reasoning *about* (strategic) games

There is Kripke structure "built in" a strategic game.

$W = \{\sigma \mid \sigma \text{ is a strategy profile: } \sigma \in \Pi_{i \in N} S_i\}$

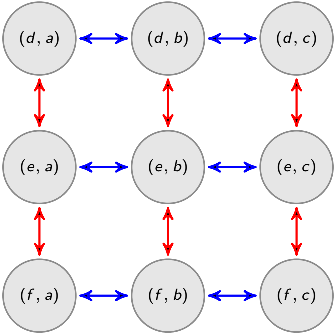
	a	b	c
d	(2,3)	(2,2)	(1,1)
e	(0,2)	(4,0)	(1,0)
f	(0,1)	(1,4)	(2,0)



Reasoning *about* (strategic) games

$\sigma \sim_i \sigma'$ iff $\sigma_i = \sigma'_i$: this epistemic relation represents player i's "view of the game" at the *ex interim* stage where i's choice is fixed but the choices of the other players' are unknown

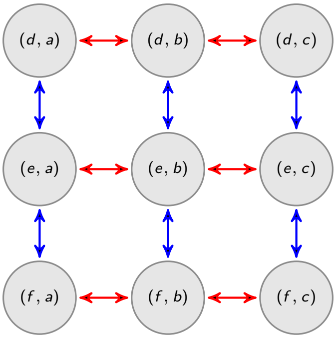
	a	b	c
d	(2,3)	(2,2)	(1,1)
e	(0,2)	(4,0)	(1,0)
f	(0,1)	(1,4)	(2,0)



Reasoning *about* (strategic) games

$\sigma \approx_i \sigma'$ iff $\sigma_{-i} = \sigma_{-i}$: this relation of "action freedom" gives the alternative choices for player i when the other players' choices are fixed.

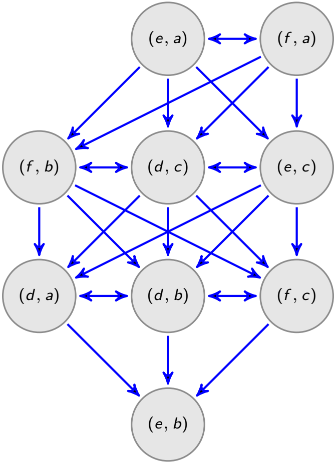
	a	b	c
d	(2,3)	(2,2)	(1,1)
e	(0,2)	(4,0)	(1,0)
f	(0,1)	(1,4)	(2,0)



Reasoning *about* (strategic) games

$\sigma \succeq_i \sigma'$ iff player i prefers the outcome σ at least as much as outcome σ'

	a	b	c
d	(2,3)	(2,2)	(1,1)
e	(0,2)	(4,0)	(1,0)
f	(0,1)	(1,4)	(2,0)



Reasoning *about* (strategic) games

$$\mathcal{M} = \langle W, \{\sim_i\}_{i\in N}, \{\approx_i\}_{i\in N}, \{\succeq_i\}_{i\in N}\rangle$$

- $\sigma \models [\sim_i]\varphi$ iff for all σ', if $\sigma \sim_i \sigma'$ then $\sigma' \models \varphi$.
- $\sigma \models [\approx_i]\varphi$ iff for all σ', if $\sigma \approx_i \sigma'$ then $\sigma' \models \varphi$.
- $\sigma \models \langle \succeq_i \rangle\varphi$ iff there exists σ' such that $\sigma' \succeq_i \sigma$ and $\sigma' \models \varphi$.
- $\sigma \models \langle \succ_i \rangle\varphi$ iff there is a σ' with $\sigma' \succeq_i \sigma$, $\sigma \not\succeq_i \sigma'$, and $\sigma' \models \varphi$

Rationality Announcements: Theorem

Weak Rationality: $w \models WR_j$ means $\bigwedge_{a \neq w(j)}$ 'j *thinks* that j's current action is at least as good for j as a.', where the a's run over the *current* model.

Theorem The following are equivalent for all states s in a full game model

1. s survives iterated removal of strongly dominated strategies
2. repeated successive **public announcements** of WR for the players stabilizes at a submodel whose domain contains s.

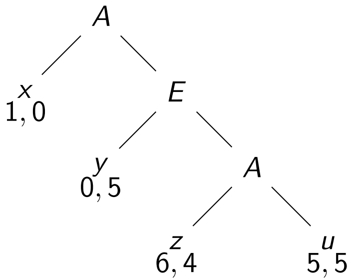
J. van Benthem. *Rational dynamics and epistemic logic in games*. International Game Theory Review 9, 1 (2007), 13-45.

Dynamics for the tree

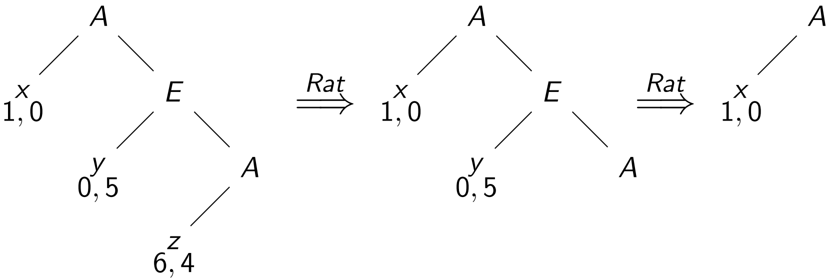
Where do the models satisfying common knowledge/belief of rationality come from?

J. van Benthem and A. Gheerbrant. *Game solution, epistemic dynamics and fixed-point logics*. Fund. Inform., 100 (2010) 1–23..

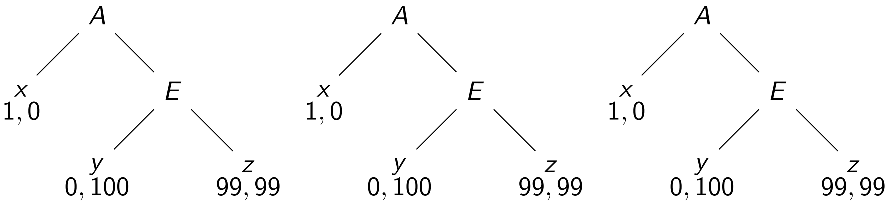
Dynamics for the tree



Dynamics for the tree



Dynamics for the tree



$$x \quad y \quad z \quad \xRightarrow{\Uparrow rat} \quad x \quad y > z \quad \xRightarrow{\Uparrow rat} \quad x > y > z$$

The Dynamics of Rational Play

A. Baltag, S. Smets and J. Zvesper. *Keep 'hoping' for rationality: a solution to the backward induction paradox*. Synthese, 169, pgs. 301 - 333, 2009.

Hard vs. Soft Information in a Game

The structure of the game and past moves are ‘hard information’: *irrevocably known*

Hard vs. Soft Information in a Game

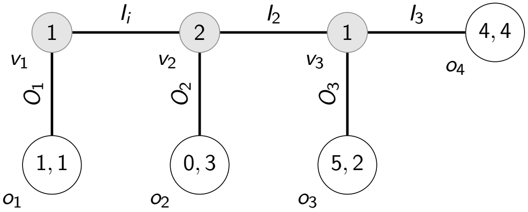
The structure of the game and past moves are ‘hard information’: *irrevocably known*

Players’ ‘knowledge’ of other players’ rationality and ‘knowledge’ of her own future moves at nodes not yet reached are not of the same degree of certainty.

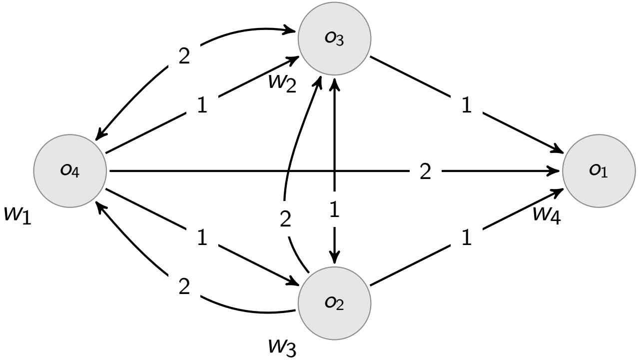
Hard vs. Soft Information in a Game

The structure of the game and past moves are ‘hard information’: *irrevocably known*

Players’ ‘knowledge’ of other players’ rationality and ‘knowledge’ of her own future moves at nodes not yet reached are not of the same degree of certainty.



1
2
1
4, 4
l_i
l_2
l_3
v_1
v_2
v_3
o_4
O_1
O_2
O_3
1, 1
0, 3
5, 2
o_1
o_2
o_3



o_3
o_4
o_1
o_2
w_1
w_2
w_3
w_4
2
1
1
2
2
1
1
2
1
2

There are atomic propositions for each possible outcome $\mathrm{o_i}$ is true only at state o_i).

There are atomic propositions for each possible outcome $\mathsf{o_i}$ is true only at state o_i).

The non-terminal nodes $v \in V$ are then identified with the set of outcomes reachable from that node:

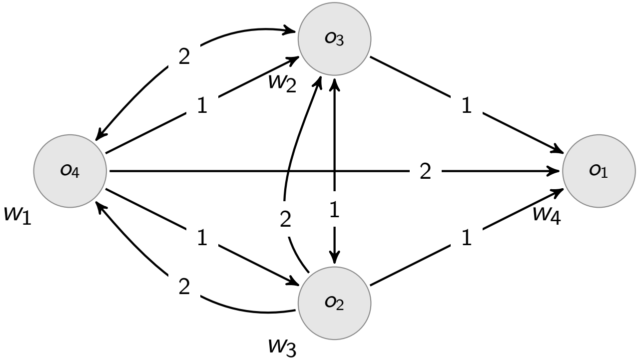
$$\mathsf{v} := \bigvee_{v \rightsquigarrow o} \mathsf{o}$$

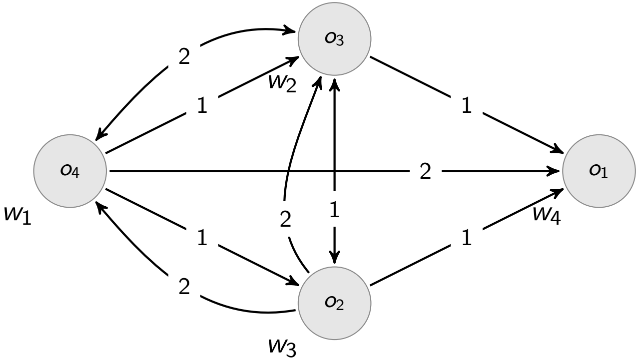
There are atomic propositions for each possible outcome $\mathsf{o_i}$ is true only at state o_i).

The non-terminal nodes $v \in V$ are then identified with the set of outcomes reachable from that node:

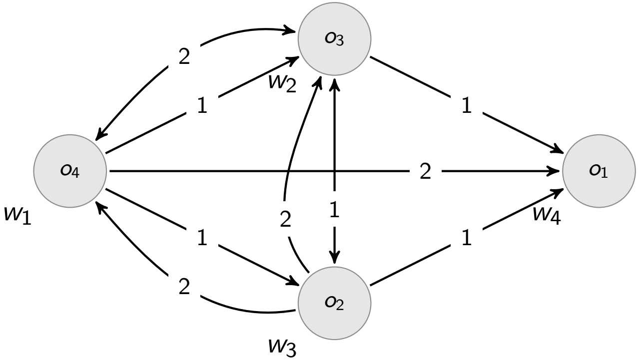
$$\mathsf{v} := \bigvee_{v \rightsquigarrow o} \mathsf{o}$$

Open future: none of the players have "hard information" that an outcome is ruled out


o_3
o_4
o_1
o_2
w_1
w_2
w_3
w_4
2
1
1
2
2
1
1
1
2
2



Player 1 is committed to the BI strategy is encoded in the conditional beliefs of the player: both $B_1^{v_1} o_1$ and $B_1^{v_3} o_3$ are true in the previous model.



Player 1 is committed to the BI strategy is encoded in the conditional beliefs of the player: both $B_1^{v_1} o_1$ and $B_1^{v_3} o_3$ are true in the previous model.

For player 2, $B_2^{v_2}(o_3 \vee o_4)$ is true in the above model, which implies player 2 plans on choosing action l_2 at node v_2.

The players' belief change as they learn (irrevocably) which of the nodes in the game are reached:

$$\mathcal{M} = \mathcal{M}^{!\mathsf{v}_1}; \mathcal{M}^{!\mathsf{v}_2}; \mathcal{M}^{!\mathsf{v}_3}; \mathcal{M}^{!\mathsf{o}_4}$$

The players' belief change as they learn (irrevocably) which of the nodes in the game are reached:

$$\mathcal{M} = \mathcal{M}^{!\mathsf{v}_1}; \mathcal{M}^{!\mathsf{v}_2}; \mathcal{M}^{!\mathsf{v}_3}; \mathcal{M}^{!\mathsf{o}_4}$$

The assumption that the players are "incurably optimistic" is represented as follows: no matter what true formula is publicly announced (i.e., no matter how the game proceeds), there is common belief that the players will make a rational choice (when it is their turn to move).

The players' belief change as they learn (irrevocably) which of the nodes in the game are reached:

$$\mathcal{M} = \mathcal{M}^{!\mathsf{v_1}}; \mathcal{M}^{!\mathsf{v_2}}; \mathcal{M}^{!\mathsf{v_3}}; \mathcal{M}^{!\mathsf{o_4}}$$

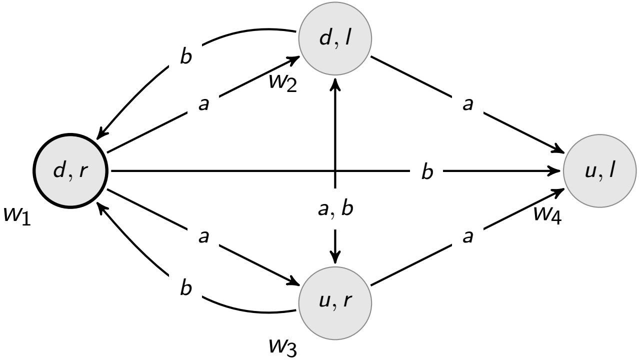
The assumption that the players are "incurably optimistic" is represented as follows: no matter what true formula is publicly announced (i.e., no matter how the game proceeds), there is common belief that the players will make a rational choice (when it is their turn to move).

$\mathcal{M}, w \models [\ !\ ]\varphi$ provided for all formulas ψ if $\mathcal{M}, w \models \psi$ then $\mathcal{M}, w \models [!\psi]\varphi$.

Theorem (Baltag, Smets and Zvesper). Common knowledge of the game structure, of open future and *common stable belief* in dynamic rationality implies common belief in the backward induction outcome.

$$Ck(Struct_G \wedge F_G \wedge [\,!\,]CbRat) \to Cb(BI_G)$$

	l	r
u	3, 3	0, 0
d	0, 0	1, 1



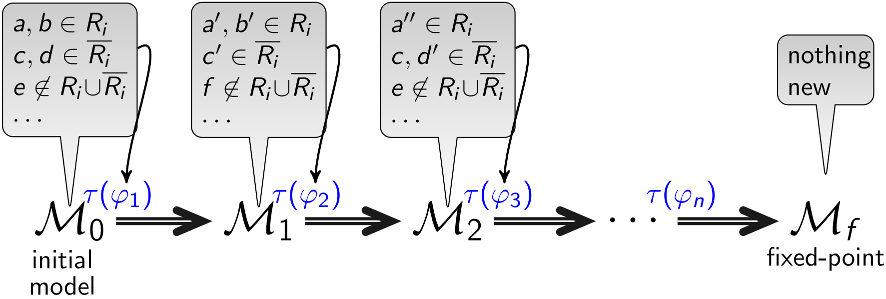
EP and O. Roy. *A Dynamic Analysis of Interactive Rationality*. Proceedings of LORI-III, 2011.

$$\mathcal{M}_0 \xRightarrow{\tau(\varphi_1)} \mathcal{M}_1 \xRightarrow{\tau(\varphi_2)} \mathcal{M}_2 \xRightarrow{\tau(\varphi_3)} \cdots \xRightarrow{\tau(\varphi_n)} \mathcal{M}_f$$

initial model

fixed-point

Where do the φ_k come from?



Where do the φ_k come from? from the players' practical reasoning (i.e., their *categorization* of their feasible moves)

Iterated Admissibility

Bob

Ann		L	R
	T	1,1	1,0
	B	1,0	0,1

Iterated Admissibility

		Bob L	Bob R
Ann	T	1,1	1,0
Ann	B	1,0	0,1

T weakly dominates B

Iterated Admissibility

		Bob L	Bob R
Ann	T	1,1	1,0
Ann	B	1,0	0,1

Then L strictly dominates R.

Iterated Admissibility

		Bob L	Bob R
Ann	T	1,1	1,0
	B	1,0	0,1

The IA set

Iterated Admissibility

Bob

Ann		L	R
	T	1,1	1,0
	B	1,0	0,1

But, now what is the reason for not playing B?

A Dynamic Analysis of Iterated Admissibility

	L	R
u	1,1	1,0
d	1,0	0,1

A Dynamic Analysis of Iterated Admissibility

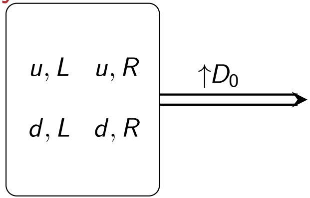
	L	R
u	1,1	1,0
d	1,0	0,1

$$\boxed{\begin{array}{cc} u,L & u,R \\ d,L & d,R \end{array}}$$

$\mathcal{M}_0$

A Dynamic Analysis of Iterated Admissibility

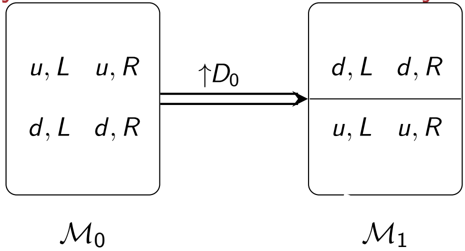
	L	R
u	1,1	1,0
d	1,0	0,1



$\mathcal{M}_0$

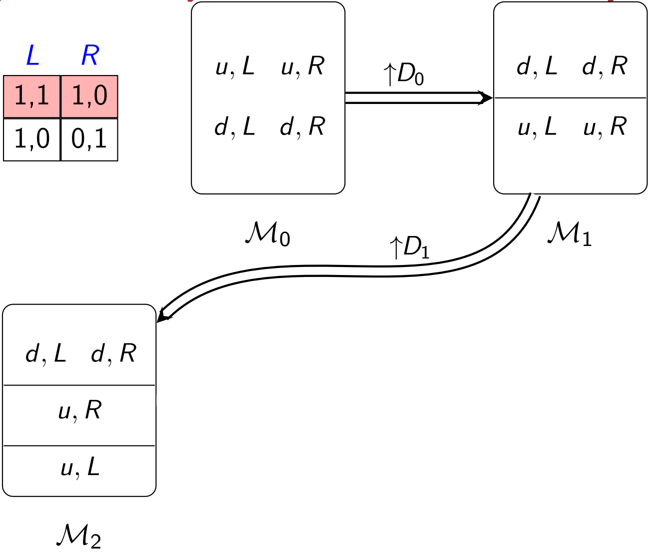
A Dynamic Analysis of Iterated Admissibility

	L	R
u	1,1	1,0
d	1,0	0,1



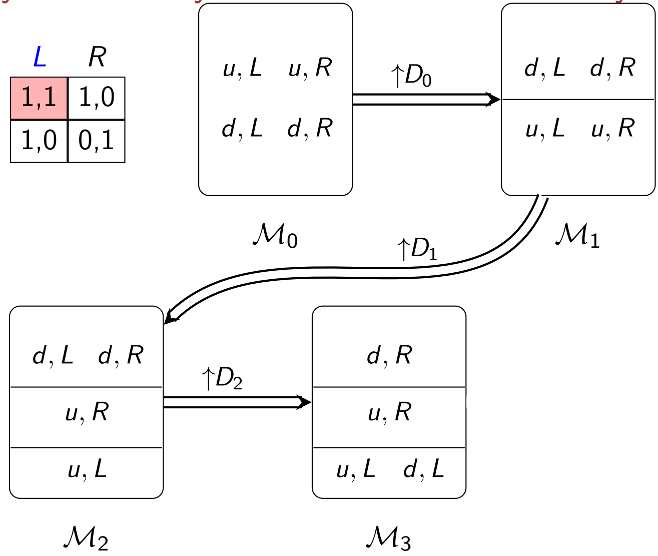
A Dynamic Analysis of Iterated Admissibility

	L	R
u	1,1	1,0
d	1,0	0,1



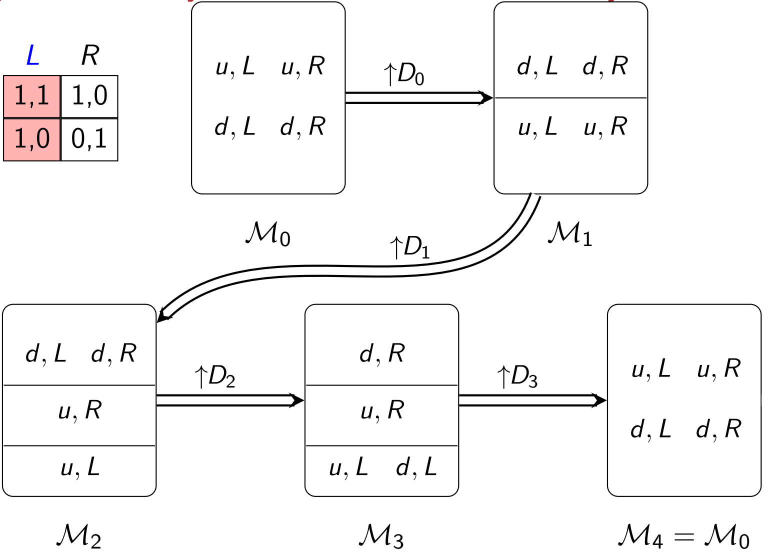
A Dynamic Analysis of Iterated Admissibility

	L	R
u	1,1	1,0
d	1,0	0,1



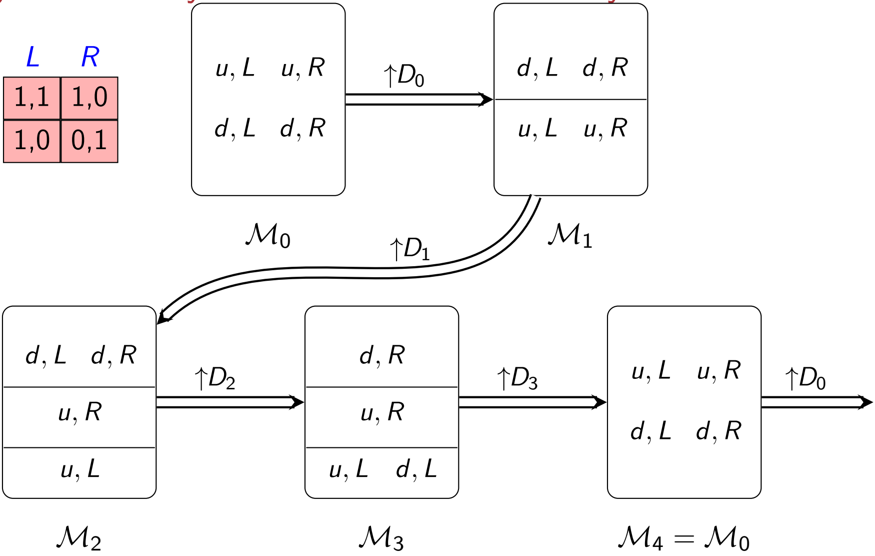
A Dynamic Analysis of Iterated Admissibility

	L	R
u	1,1	1,0
d	1,0	0,1



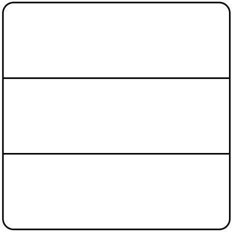
A Dynamic Analysis of Iterated Admissibility

	L	R
u	1,1	1,0
d	1,0	0,1



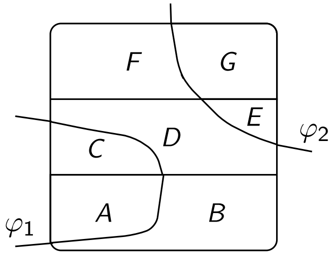
Suspending Judgement

Both φ_1 and φ_2 describe "good" options...



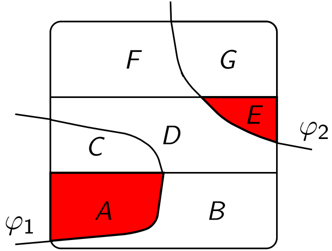
Suspending Judgement

Both φ_1 and φ_2 describe "good" options...



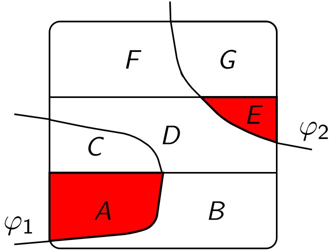
Suspending Judgement

Both φ_1 and φ_2 describe “good” options...



Suspending Judgement

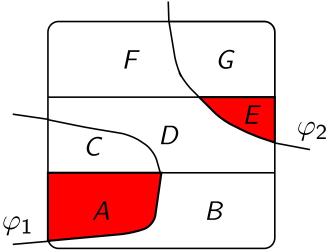
Both φ_1 and φ_2 describe "good" options...



$$\uparrow\{\varphi_1, \varphi_2\} : A \cup E \prec B \prec C \cup D \prec F \cup G$$

Suspending Judgement

Both φ_1 and φ_2 describe "good" options...



$$\uparrow\{\varphi_1, \varphi_2\} : A \cup E \prec B \prec C \cup D \prec F \cup G$$

$$\Uparrow\{\varphi_1, \varphi_2\} : A \prec E \prec B \prec C \cup D \prec F \cup G$$

Remembering Reasons

	L	R
u	1,1	1,0
d	1,0	0,1