Introduction to Logics of Knowledge and Belief

Eric Pacuit

University of Maryland

pacuit.org

epacuit@umd.edu

May 6, 2019

Plausibility Models

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

Truth: $\mathcal{M}, w \models \varphi$ is defined as follows:

- $\mathcal{M}, w \models p$ iff $w \in V(p)$ (with $p \in \mathrm{At}$)
- $\mathcal{M}, w \models \neg\varphi$ if $\mathcal{M}, w \not\models \varphi$
- $\mathcal{M}, w \models \varphi \wedge \psi$ if $\mathcal{M}, w \models \varphi$ and $\mathcal{M}, w \models \psi$
- $\mathcal{M}, w \models K_i\varphi$ if for each $v \in W$, if $w \sim_i v$, then $\mathcal{M}, v \models \varphi$

Plausibility Models

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

Truth: $\mathcal{M}, w \models \varphi$ is defined as follows:

- $\mathcal{M}, w \models p$ iff $w \in V(p)$ (with $p \in \mathrm{At}$)
- $\mathcal{M}, w \models \neg\varphi$ if $\mathcal{M}, w \not\models \varphi$
- $\mathcal{M}, w \models \varphi \wedge \psi$ if $\mathcal{M}, w \models \varphi$ and $\mathcal{M}, w \models \psi$
- $\mathcal{M}, w \models K_i\varphi$ if for each $v \in W$, if $w \sim_i v$, then $\mathcal{M}, v \models \varphi$

Plausibility Models

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

Plausibility Relation: $\preceq_i \subseteq W \times W$. $w \preceq_i v$ means

"w is at least as plausible as v."

Plausibility Models

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

Plausibility Relation: $\preceq_i \subseteq W \times W$. $w \preceq_i v$ means

"w is at least as plausible as v."

Properties of $\preceq_i$: reflexive, transitive, and *well-founded*.

Plausibility Models

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

Plausibility Relation: $\preceq_i \subseteq W \times W$. $w \preceq_i v$ means

"w is at least as plausible as v."

Properties of $\preceq_i$: reflexive, transitive, and *well-founded*.

Most Plausible: For $X \subseteq W$, let

$$Min_{\preceq_i}(X) = \{v \in W \mid v \preceq_i w \text{ for all } w \in X\}$$

Plausibility Models

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

Plausibility Relation: $\preceq_i \subseteq W \times W$. $w \preceq_i v$ means

"w is at least as plausible as v."

Properties of $\preceq_i$: reflexive, transitive, and *well-founded*.

Most Plausible: For $X \subseteq W$, let

$$Min_{\preceq_i}(X) = \{v \in W \mid v \preceq_i w \text{ for all } w \in X\}$$

Assumptions:

1. *plausibility implies possibility*: if $w \preceq_i v$ then $w \sim_i v$.
2. *locally-connected*: if $w \sim_i v$ then either $w \preceq_i v$ or $v \preceq_i w$.

Plausibility Models

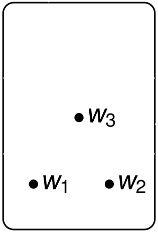
Epistemic-Plausibility Models: $\mathcal{M} = \langle W, \{\sim_i\}_{i\in\mathcal{A}}, \{\preceq_i\}_{i\in\mathcal{A}}, V\rangle$

Truth: $\mathcal{M}, w \models \varphi$ is defined as follows:

- $\mathcal{M}, w \models p$ iff $w \in V(p)$ (with $p \in \mathrm{At}$)
- $\mathcal{M}, w \models \neg\varphi$ if $\mathcal{M}, w \not\models \varphi$
- $\mathcal{M}, w \models \varphi \wedge \psi$ if $\mathcal{M}, w \models \varphi$ and $\mathcal{M}, w \models \psi$
- $\mathcal{M}, w \models K_i\varphi$ if for each $v \in W$, if $w \sim_i v$, then $\mathcal{M}, v \models \varphi$
- $\mathcal{M}, w \models B_i\varphi$ if for each $v \in Min_{\preceq_i}([w]_i)$, $\mathcal{M}, v \models \varphi$
  $[w]_i = \{v \mid w \sim_i v\}$ is the agent's **information cell.**

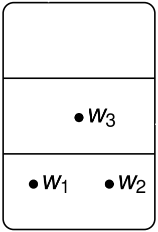
Beliefs via Plausibility

- $W = \{w_1, w_2, w_3\}$



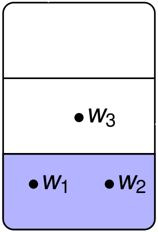
Beliefs via Plausibility

- $W = \{w_1, w_2, w_3\}$
- $w_1 \preceq w_2$ 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 $w_3 \not\preceq w_2$)

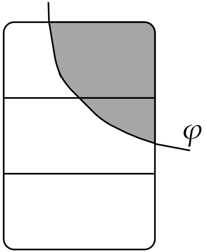


Beliefs via Plausibility

- $W = \{w_1, w_2, w_3\}$
- $w_1 \preceq w_2$ 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 $w_3 \not\preceq w_2$)
- $\{w_1, w_2\} \subseteq Min_{\preceq}([w_i])$

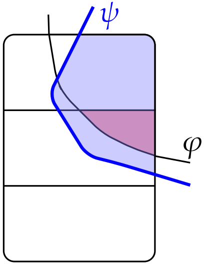


Beliefs via Plausibility



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

Beliefs via Plausibility



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

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

Finding out that φ

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

$$\Big\Downarrow$$

Find out that φ

$$\Big\Downarrow$$

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

Model Update

Given $\mathcal{M} = \langle W, \{R_a \mid a \in \text{Agt}\}, V\rangle$, the *updated model* $\mathcal{M}_{|\varphi}$ is obtained by deleting from $\mathcal{M}$ all worlds in which φ was false.

Model Update

Given $\mathcal{M} = \langle W, \{R_a \mid a \in \mathsf{Agt}\}, V\rangle$, the *updated model* $\mathcal{M}_{|\varphi}$ is obtained by deleting from $\mathcal{M}$ all worlds in which φ was false.

Formally, $\mathcal{M}_{|\varphi} = \langle W_{|\varphi}, \{R_{a_{|\varphi}} \mid a \in \mathsf{Agt}\}, V_{|\varphi}\rangle$ is the model where:

$$W_{|\varphi} = \{v \in W \mid \mathcal{M}, v \models \varphi\};$$

Model Update

Given $\mathcal{M} = \langle W, \{R_a \mid a \in \text{Agt}\}, V\rangle$, the *updated model* $\mathcal{M}_{|\varphi}$ is obtained by deleting from $\mathcal{M}$ all worlds in which φ was false.

Formally, $\mathcal{M}_{|\varphi} = \langle W_{|\varphi}, \{R_{a_{|\varphi}} \mid a \in \text{Agt}\}, V_{|\varphi}\rangle$ is the model where:

$W_{|\varphi} = \{v \in W \mid \mathcal{M}, v \models \varphi\}$;

$R_{a_{|\varphi}}$ is the restriction of R_a to $W_{|\varphi}$;

Model Update

Given $\mathcal{M} = \langle W, \{R_a \mid a \in \text{Agt}\}, V\rangle$, the *updated model* $\mathcal{M}_{|\varphi}$ is obtained by deleting from $\mathcal{M}$ all worlds in which φ was false.

Formally, $\mathcal{M}_{|\varphi} = \langle W_{|\varphi}, \{R_{a_{|\varphi}} \mid a \in \text{Agt}\}, V_{|\varphi}\rangle$ is the model where:

$W_{|\varphi} = \{v \in W \mid \mathcal{M}, v \models \varphi\}$;

$R_{a_{|\varphi}}$ is the restriction of R_a to $W_{|\varphi}$;

$V_{|\varphi}(p)$ is the intersection of $V(p)$ and $W_{|\varphi}$.

Model Update

Given $\mathcal{M} = \langle W, \{R_a \mid a \in \text{Agt}\}, V\rangle$, the *updated model* $\mathcal{M}_{|\varphi}$ is obtained by deleting from $\mathcal{M}$ all worlds in which φ was false.

Formally, $\mathcal{M}_{|\varphi} = \langle W_{|\varphi}, \{R_{a_{|\varphi}} \mid a \in \text{Agt}\}, V_{|\varphi}\rangle$ is the model where:

$W_{|\varphi} = \{v \in W \mid \mathcal{M}, v \models \varphi\}$;

$R_{a_{|\varphi}}$ is the restriction of R_a to $W_{|\varphi}$;

$V_{|\varphi}(p)$ is the intersection of $V(p)$ and $W_{|\varphi}$.

Public Announcement Logic

The language of Public Announcement Logic (PAL) is given by:

$$\varphi ::= p \mid \neg\varphi \mid (\varphi \wedge \varphi) \mid K_a\varphi \mid [!\varphi]\varphi$$

Public Announcement Logic

The language of Public Announcement Logic (PAL) is given by:

$$\varphi ::= p \mid \neg\varphi \mid (\varphi \wedge \varphi) \mid K_a\varphi \mid [!\varphi]\varphi$$

Read $[!\varphi]\psi$ as "after (every) true announcement of φ, ψ."

Public Announcement Logic

The language of Public Announcement Logic (PAL) is given by:

$$\varphi ::= p \mid \neg\varphi \mid (\varphi \wedge \varphi) \mid K_a\varphi \mid [!\varphi]\varphi$$

Read $[!\varphi]\psi$ as "after (every) true announcement of φ, ψ."

Read $\langle !\varphi\rangle\psi := \neg[!\varphi]\neg\psi$ as "after a true announcement of φ, ψ."

Public Announcement Logic

Read $[!\varphi]\psi$ as "after (every) true announcement of φ, ψ."

Read $\langle!\varphi\rangle\psi := \neg[!\varphi]\neg\psi$ as "after a true announcement of φ, ψ."

The truth clause for the dynamic operator $[!\varphi]$ is:

Public Announcement Logic

Read $[!\varphi]\psi$ as "after (every) true announcement of φ, ψ."

Read $\langle !\varphi\rangle\psi := \neg[!\varphi]\neg\psi$ as "after a true announcement of φ, ψ."

The truth clause for the dynamic operator $[!\varphi]$ is:

- $\mathcal{M}, w \models [!\varphi]\psi$ iff $\mathcal{M}, w \models \varphi$ *implies* $\mathcal{M}_{|\varphi}, w \models \psi$.

Public Announcement Logic

Read $[!\varphi]\psi$ as "after (every) true announcement of φ, ψ."

Read $\langle !\varphi\rangle\psi := \neg[!\varphi]\neg\psi$ as "after a true announcement of φ, ψ."

The truth clause for the dynamic operator $[!\varphi]$ is:

- $\mathcal{M}, w \models [!\varphi]\psi$ iff $\mathcal{M}, w \models \varphi$ *implies* $\mathcal{M}_{|\varphi}, w \models \psi$.

So if φ is false, $[!\varphi]\psi$ is vacuously true.

Public Announcement Logic

Read $[!\varphi]\psi$ as "after (every) true announcement of φ, ψ."

Read $\langle !\varphi\rangle\psi := \neg[!\varphi]\neg\psi$ as "after a true announcement of φ, ψ."

The truth clause for the dynamic operator $[!\varphi]$ is:

- $\mathcal{M}, w \models [!\varphi]\psi$ iff $\mathcal{M}, w \models \varphi$ *implies* $\mathcal{M}_{|\varphi}, w \models \psi$.

So if φ is false, $[!\varphi]\psi$ is vacuously true. Here is the $\langle !\varphi\rangle$ clause:

Public Announcement Logic

Read $[!\varphi]\psi$ as "after (every) true announcement of φ, ψ."

Read $\langle !\varphi\rangle\psi := \neg[!\varphi]\neg\psi$ as "after a true announcement of φ, ψ."

The truth clause for the dynamic operator $[!\varphi]$ is:

- $\mathcal{M}, w \models [!\varphi]\psi$ iff $\mathcal{M}, w \models \varphi$ *implies* $\mathcal{M}_{|\varphi}, w \models \psi$.

So if φ is false, $[!\varphi]\psi$ is vacuously true. Here is the $\langle !\varphi\rangle$ clause:

- $\mathcal{M}, w \models \langle !\varphi\rangle\psi$ iff $\mathcal{M}, w \models \varphi$ *and* $\mathcal{M}_{|\varphi}, w \models \psi$.

Public Announcement Logic

Read $[!\varphi]\psi$ as "after (every) true announcement of φ, ψ."

Read $\langle !\varphi\rangle\psi := \neg[!\varphi]\neg\psi$ as "after a true announcement of φ, ψ."

The truth clause for the dynamic operator $[!\varphi]$ is:

- $\mathcal{M}, w \models [!\varphi]\psi$ iff $\mathcal{M}, w \models \varphi$ *implies* $\mathcal{M}_{|\varphi}, w \models \psi$.

So if φ is false, $[!\varphi]\psi$ is vacuously true. Here is the $\langle !\varphi\rangle$ clause:

- $\mathcal{M}, w \models \langle !\varphi\rangle\psi$ iff $\mathcal{M}, w \models \varphi$ *and* $\mathcal{M}_{|\varphi}, w \models \psi$.

Key Idea: we evaluate $[!\varphi]\psi$ and $\langle !\varphi\rangle\psi$ not by looking at *other worlds in the same model*, but rather by looking at a new model.

Public Announcement Logic

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

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

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

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

Public Announcement Logic

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

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

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

Public Announcement Logic

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

Public Announcement Logic

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

Public Announcement Logic

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

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

Finding out that φ

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

$$\Big\Downarrow$$

Find out that φ

$$\Big\Downarrow$$

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

- Epistemic states: AGM, Plausibility Models, Bayesian Model (and the many variations)

- Epistemic states: AGM, Plausibility Models, Bayesian Model (and the many variations)

- "Finding out that φ"
  - Learn that φ
  - Suppose that φ
  - Accept φ
  - ...

- Epistemic states: AGM, Plausibility Models, Bayesian Model (and the many variations)

- “Finding out that φ”
  - Learn that φ
  - Suppose that φ
  - Accept φ
  - ...

- *How* did you find out that φ?
  - Directly observed φ
  - Indirectly observed φ
  - Told ‘φ’ (by an epistemic peer, by an expert, by a trusted individual)
  - ...

- Belief change over time

The Theory of Belief Revision

C. Alchourrón, P. Gärdenfors and D. Makinson. *On the logic of theory change: Partial meet contraction and revision functions*. Journal of Symbolic Logic, 50, 510 - 530, 1985.

Hans Rott. *Change, Choice and Inference: A Study of Belief Revision and Nonmonotonic Reasoning*. Oxford University Press, 2001.

A.P. Pedersen and H. Arló-Costa. *"Belief Revision."*. In Continuum Companion to Philosophical Logic. Continuum Press, 2011.

$$\mathcal{B} * \varphi$$

$\mathcal{B} * \varphi$

Initial set of beliefs

$\mathcal{B} * \varphi$

Initial set of beliefs

New evidence φ

Revision operator: $* : \mathcal{B} \times \mathcal{L} \to \mathcal{B}$

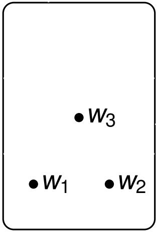
$$\mathcal{B} * \varphi$$

Initial set of beliefs

New evidence φ

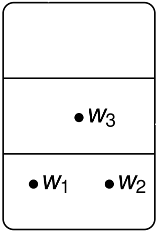
Belief Revision via Plausibility

- $W = \{w_1, w_2, w_3\}$



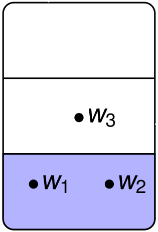
Belief Revision via Plausibility

- $W = \{w_1, w_2, w_3\}$
- $w_1 \preceq w_2$ 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 $w_3 \not\preceq w_2$)

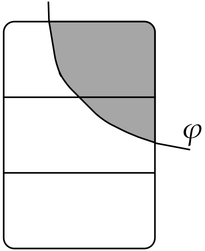


Belief Revision via Plausibility

- $W = \{w_1, w_2, w_3\}$
- $w_1 \preceq w_2$ 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 $w_3 \not\preceq w_2$)
- $\{w_1, w_2\} \subseteq Min_{\preceq}([w_i])$

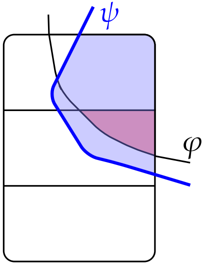


Belief Revision via Plausibility



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

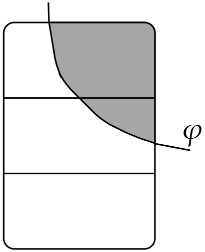
Belief Revision via Plausibility



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

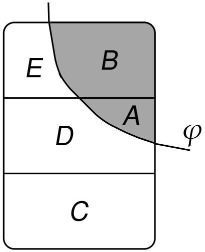
$$Min_{\leq}([\![\varphi]\!]_{\mathcal{M}}) \subseteq [\![\psi]\!]_{\mathcal{M}}$$

Belief Revision via Plausibility



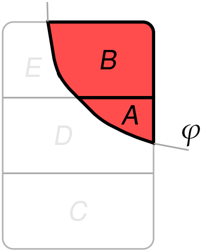
Incorporate the new information φ

Belief Revision via Plausibility



Incorporate the new information φ

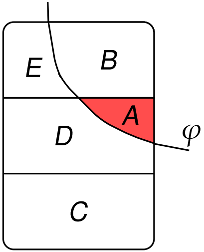
Belief Revision via Plausibility



Public Announcement: Information from an infallible source

$(!\varphi)$: $A \prec_i B$

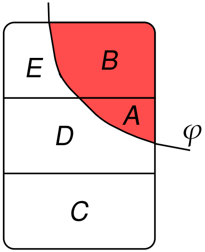
Belief Revision via Plausibility



Public Announcement: Information from an infallible source
$(!\varphi)$: $A \prec_i B$

Conservative Upgrade: Information from a trusted source
$(\uparrow\varphi)$: $A \prec_i C \prec_i D \prec_i B \cup E$

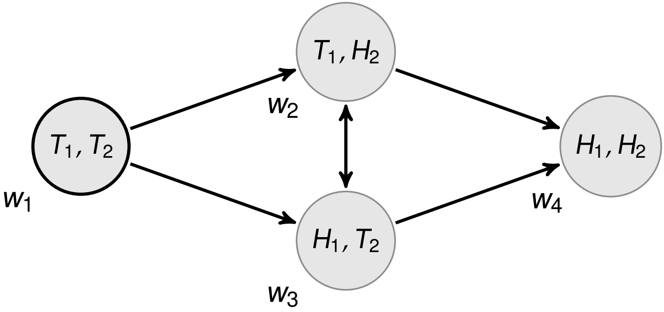
Belief Revision via Plausibility



Public Announcement: Information from an infallible source
$(!\varphi)$: $A \prec_i B$

Conservative Upgrade: Information from a trusted source
$(\uparrow\varphi)$: $A \prec_i C \prec_i D \prec_i B \cup E$

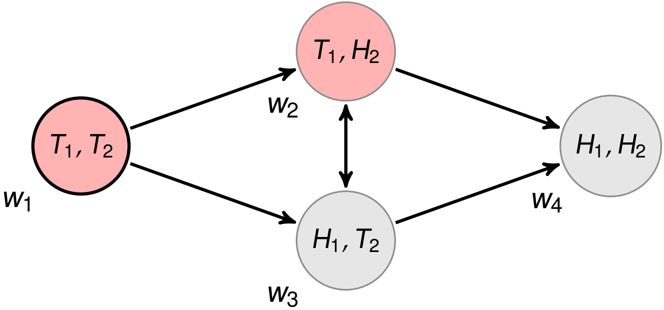
Radical Upgrade: Information from a strongly trusted source
$(\Uparrow\varphi)$: $A \prec_i B \prec_i C \prec_i D \prec_i E$



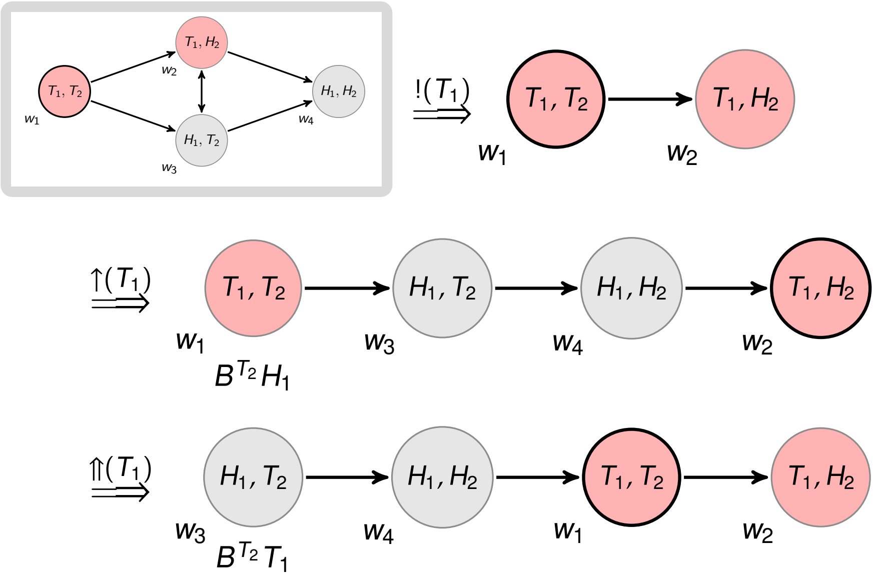
$Min_{\leq}([w_1]) = \{w_4\}$, so $w_1 \models B(H_1 \wedge H_2)$

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

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

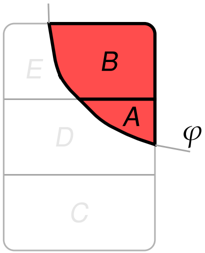


Suppose the agent *finds out that* T_1 *is true*.



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

Informative Actions



Public Announcement: Information from an infallible source

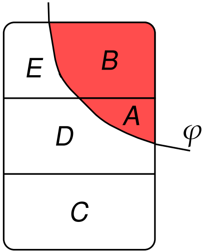
$(!\varphi)$: $A \prec_i B$ $\qquad \mathcal{M}^{!\varphi} = \langle W^{!\varphi}, \{\sim_i^{!\varphi}\}_{i \in \mathcal{A}}, V^{!\varphi} \rangle$

$W^{!\varphi} = [\![\varphi]\!]_{\mathcal{M}}$

$\sim_i^{!\varphi} = \sim_i \cap (W^{!\varphi} \times W^{!\varphi})$

$\preceq_i^{!\varphi} = \preceq_i \cap (W^{!\varphi} \times W^{!\varphi})$

Informative Actions



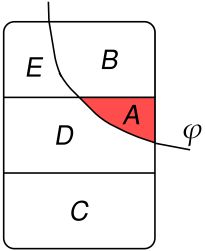
Radical Upgrade: $(\Uparrow\varphi)$: $A \prec_i B \prec_i C \prec_i D \prec_i E$,

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

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 $x \preceq_i y$.

Recursion Axioms

$$[\Uparrow\varphi]B^{\psi}\chi \leftrightarrow (L(\varphi \wedge [\Uparrow\varphi]\psi) \wedge B^{\varphi\wedge[\Uparrow\varphi]\psi}[\Uparrow\varphi]\chi)\vee$$
$$(\neg L(\varphi \wedge [\Uparrow\varphi]\psi) \wedge B^{[\Uparrow\varphi]\psi}[\Uparrow\varphi]\chi)$$

Recursion Axioms

$$[\Uparrow\varphi]B^{\psi}\chi \leftrightarrow (L(\varphi \wedge [\Uparrow\varphi]\psi) \wedge B^{\varphi\wedge[\Uparrow\varphi]\psi}[\Uparrow\varphi]\chi) \vee$$
$$(\neg L(\varphi \wedge [\Uparrow\varphi]\psi) \wedge B^{[\Uparrow\varphi]\psi}[\Uparrow\varphi]\chi)$$

$$[\uparrow\varphi]B^{\psi}\chi \leftrightarrow (B^{\varphi}\neg[\uparrow\varphi]\psi \wedge B^{[\uparrow\varphi]\psi}[\uparrow\varphi]\chi) \vee (\neg B^{\varphi}\neg[\uparrow\varphi]\psi \wedge B^{\varphi\wedge[\uparrow\varphi]\psi}[\uparrow\varphi]\chi)$$

Iterated Updates

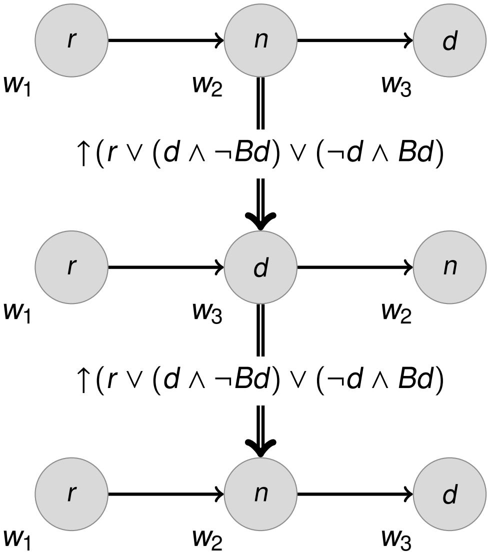
$!\varphi_1, !\varphi_2, !\varphi_3, \ldots, !\varphi_n$
always reaches a fixed-point

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

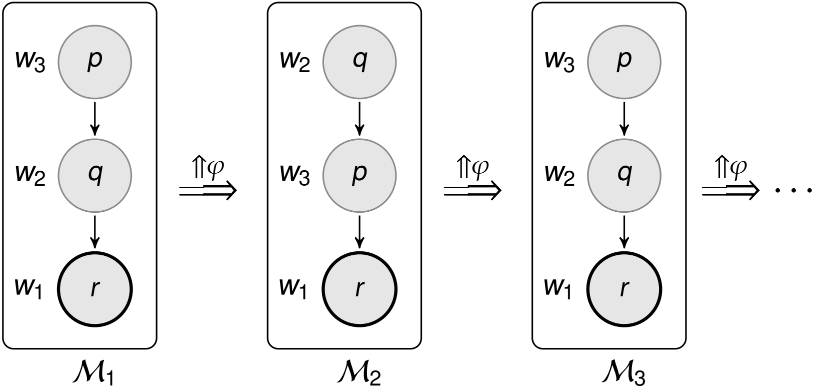
$\uparrow\varphi, \uparrow\varphi, \ldots$
Simple beliefs may never stabilize

$\Uparrow\varphi, \Uparrow\varphi, \ldots$
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
w_1
w_2
w_3
$\uparrow(r \vee (d \wedge \neg Bd) \vee (\neg d \wedge Bd))$
r
d
n
w_1
w_3
w_2
$\uparrow(r \vee (d \wedge \neg Bd) \vee (\neg d \wedge Bd))$
r
n
d
w_1
w_2
w_3

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



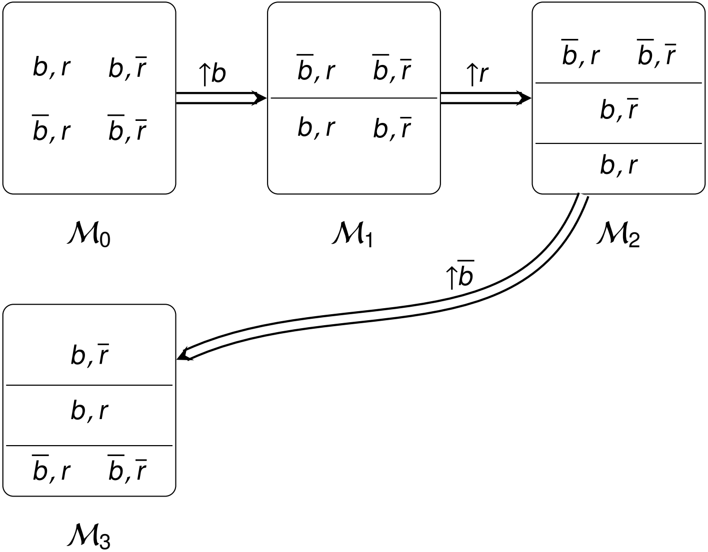
Suppose that you are in the forest and happen to a see strange-looking animal.

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.

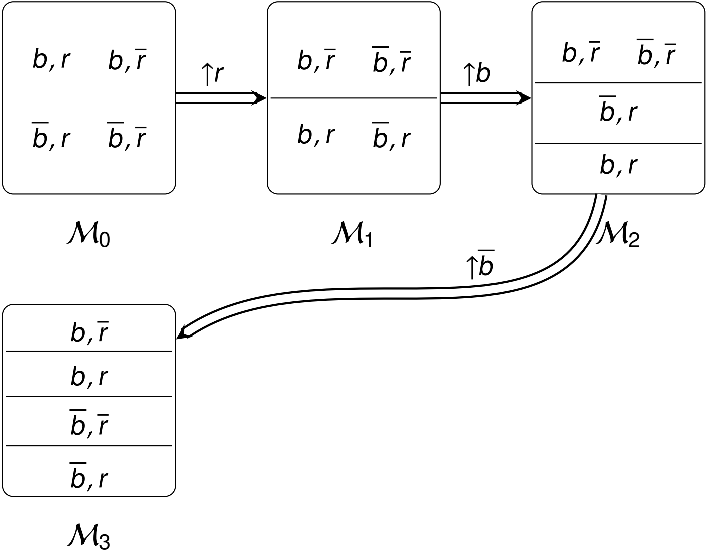
Suppose that you are in the forest and happen to a see strange-looking animal. You consult your animal guidebook and find a picture that seems to match the animal you see. The guidebook says that the animal is a type of bird, so that is what you conclude: The animal before you is a bird. After looking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

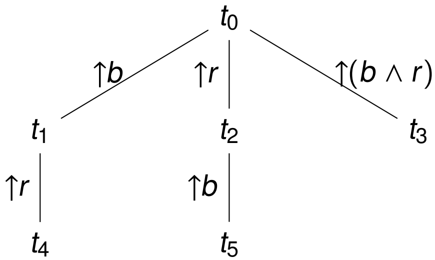

b, r $b, \overline{r}$
$\overline{b}, r$ $\overline{b}, \overline{r}$
$\mathcal{M}_0$
$\uparrow b$
$\overline{b}, r$ $\overline{b}, \overline{r}$
b, r $b, \overline{r}$
$\mathcal{M}_1$
$\uparrow r$
$\overline{b}, r$ $\overline{b}, \overline{r}$
$b, \overline{r}$
b, r
$\mathcal{M}_2$
$\uparrow \overline{b}$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


t_0
$\uparrow b$
$\uparrow r$
$\uparrow (b \wedge r)$
t_1
t_2
t_3
$\uparrow r$
$\uparrow b$
t_4
t_5

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

Two Postulates of Iterated Revision

I1 If $\psi \in Cn(\{\varphi\})$ then $(K * \psi) * \varphi = K * \varphi$.

I2 If $\neg\psi \in Cn(\{\varphi\})$ then $(K * \varphi) * \psi = K * \psi$

Two Postulates of Iterated Revision

I1 If $\psi \in Cn(\{\varphi\})$ then $(K * \psi) * \varphi = K * \varphi$.

I2 If $\neg\psi \in Cn(\{\varphi\})$ then $(K * \varphi) * \psi = K * \psi$

- Postulate I1 demands if $\varphi \rightarrow \psi$ is a theorem (with respect to the background theory), then first learning ψ followed by the more specific information φ is equivalent to directly learning the more specific information φ.

Two Postulates of Iterated Revision

I1 If $\psi \in Cn(\{\varphi\})$ then $(K * \psi) * \varphi = K * \varphi$.

I2 If $\neg\psi \in Cn(\{\varphi\})$ then $(K * \varphi) * \psi = K * \psi$

- Postulate I1 demands if $\varphi \to \psi$ is a theorem (with respect to the background theory), then first learning ψ followed by the more specific information φ is equivalent to directly learning the more specific information φ.

- Postulate I2 demands that first learning φ followed by learning a piece of information ψ incompatible with φ is the same as simply learning ψ outright. So, for example, first learning φ and then $\neg\varphi$ should result in the same belief state as directly learning $\neg\varphi$.

Stalnaker's Counterexample to I1

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.

Stalnaker's Counterexample to I1

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.

Stalnaker's Counterexample to I1

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?

Stalnaker's Counterexample to I1

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*

Stalnaker's Counterexample to I1

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)

Stalnaker's Counterexample to I1

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.

Stalnaker's Counterexample to I1

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.

Stalnaker's Counterexample to I1

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.

Stalnaker's Counterexample to I1

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.

Stalnaker's Counterexample to I1

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.
- 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.

Stalnaker's Counterexample to I1

UUU	*DDD*
UUD	*DDU*
UDU	*DUD*
UDD	*DUU*

- So, $L_2 \in Cn(\{L_1\})$ but (potentially) $(K * L_2) * L_1 \neq K * L_1$.

Stalnaker's Counterexample to I2

- Two fair coins are flipped and placed in two boxes and two independent and reliable observers deliver reports about the status (heads up or tails up) of the coins in the opaque boxes.

Stalnaker's Counterexample to I2

- Two fair coins are flipped and placed in two boxes and two independent and reliable observers deliver reports about the status (heads up or tails up) of the coins in the opaque boxes.
- Alice reports that the coin in box 1 is lying heads up, Bert reports that the coin in box 2 is lying heads up.

Stalnaker's Counterexample to I2

- Two fair coins are flipped and placed in two boxes and two independent and reliable observers deliver reports about the status (heads up or tails up) of the coins in the opaque boxes.
- Alice reports that the coin in box 1 is lying heads up, Bert reports that the coin in box 2 is lying heads up.
- Two new independent witnesses, whose reliability trumps that of Alice's and Bert's, provide additional reports on the status of the coins. Carla reports that the coin in box 1 is lying tails up, and Dora reports that the coin in box 2 is lying tails up.

Stalnaker's Counterexample to I2

- Two fair coins are flipped and placed in two boxes and two independent and reliable observers deliver reports about the status (heads up or tails up) of the coins in the opaque boxes.
- Alice reports that the coin in box 1 is lying heads up, Bert reports that the coin in box 2 is lying heads up.
- Two new independent witnesses, whose reliability trumps that of Alice's and Bert's, provide additional reports on the status of the coins. Carla reports that the coin in box 1 is lying tails up, and Dora reports that the coin in box 2 is lying tails up.
- Finally, Elmer, a third witness considered the most reliable overall, reports that the coin in box 1 is lying heads up.

H_i (T_i): the coin in box i facing heads (tails) up.

H_i (T_i): the coin in box i facing heads (tails) up.

- The first revision results in the belief set $K' = K * (H_1 \wedge H_2)$, where K is the agents original set of beliefs.

H_i (T_i): the coin in box i facing heads (tails) up.

- The first revision results in the belief set $K' = K * (H_1 \wedge H_2)$, where K is the agents original set of beliefs.
- After receiving the reports, the belief set is $K' * (T_1 \wedge T_2) * H_1$.

H_i (T_i): the coin in box i facing heads (tails) up.

- The first revision results in the belief set $K' = K * (H_1 \wedge H_2)$, where K is the agents original set of beliefs.
- After receiving the reports, the belief set is $K' * (T_1 \wedge T_2) * H_1$.
- Since Elmers report is irrelevant to the status of the coin in box 2, it seems natural to assume that $H_1 \wedge T_2 \in K' * (T_1 \wedge T_2) * H_1$.

H_i (T_i): the coin in box i facing heads (tails) up.

- The first revision results in the belief set $K' = K * (H_1 \wedge H_2)$, where K is the agents original set of beliefs.
- After receiving the reports, the belief set is $K' * (T_1 \wedge T_2) * H_1$.
- Since Elmers report is irrelevant to the status of the coin in box 2, it seems natural to assume that $H_1 \wedge T_2 \in K' * (T_1 \wedge T_2) * H_1$.
- The problem: Since $(T_1 \wedge T_2) \rightarrow \neg H_1$ is a theorem (given the background theory), by I2 it follows that $K' * (T_1 \wedge T_2) * H_1 = K' * H_1$.

H_i (T_i): the coin in box i facing heads (tails) up.

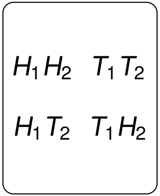
- The first revision results in the belief set $K' = K * (H_1 \wedge H_2)$, where K is the agents original set of beliefs.
- After receiving the reports, the belief set is $K' * (T_1 \wedge T_2) * H_1$.
- Since Elmers report is irrelevant to the status of the coin in box 2, it seems natural to assume that $H_1 \wedge T_2 \in K' * (T_1 \wedge T_2) * H_1$.
- The problem: Since $(T_1 \wedge T_2) \rightarrow \neg H_1$ is a theorem (given the background theory), by I2 it follows that $K' * (T_1 \wedge T_2) * H_1 = K' * H_1$.

  Yet, since $H_1 \wedge H_2 \in K'$ and H_1 is consistent with H_2, we must have $H_1 \wedge H_2 \in K' * H_1$, which yields a conflict with the assumption that $H_1 \wedge T_2 \in K' * (T_1 \wedge T_2) * H_1$.

...[Postulate I2] directs us to take back the totality of any information that is overturned. Specifically, if we first receive information α, and then receive information that conflicts with α, we should return to the belief state we were previously in, before learning α. But this directive is too strong. Even if the new information conflicts with the information just received, it need not necessarily cast doubt on all *of that information.*

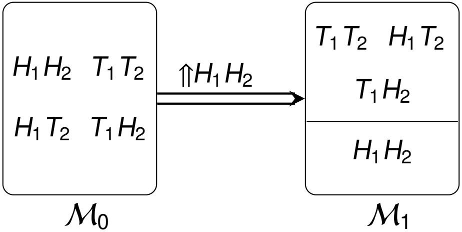
(Stalnaker, pg. 207–208)

Heuristic Diagnosis of Stalnaker's Example

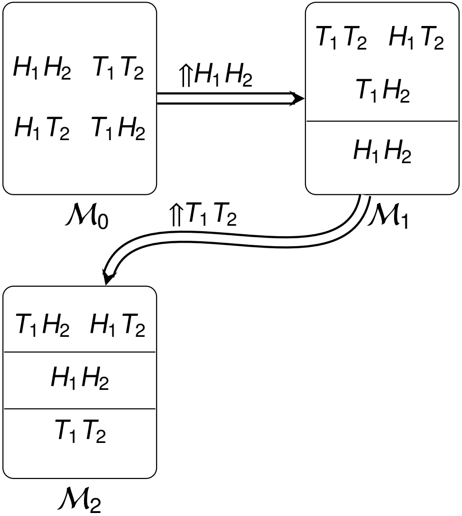


$\mathcal{M}_0$

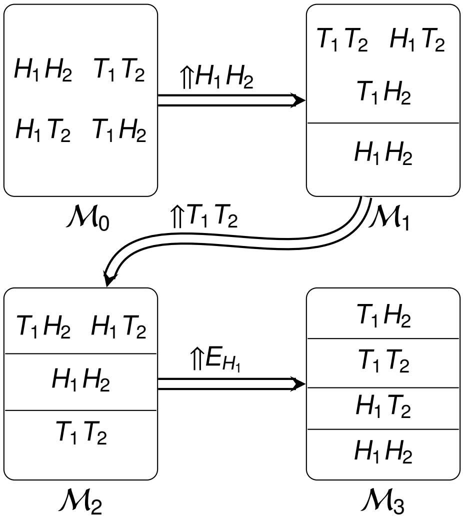
Heuristic Diagnosis of Stalnaker's Example



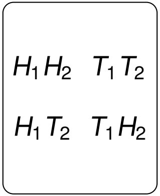
Heuristic Diagnosis of Stalnaker's Example



Heuristic Diagnosis of Stalnaker's Example

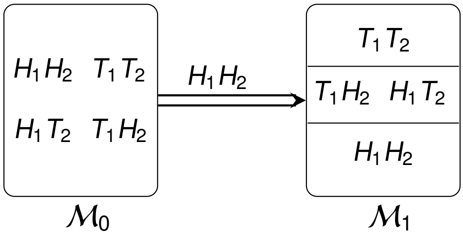


Heuristic Diagnosis of Stalnaker's Example

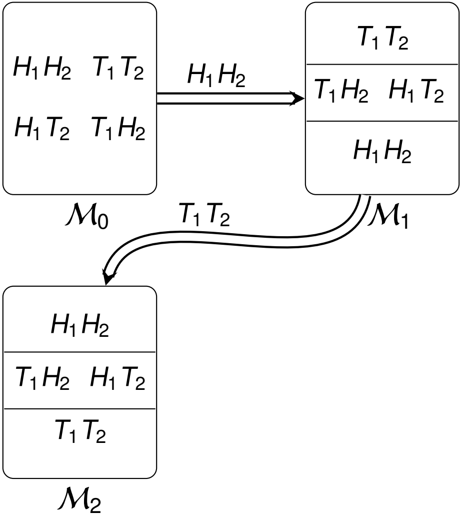


$\mathcal{M}_0$

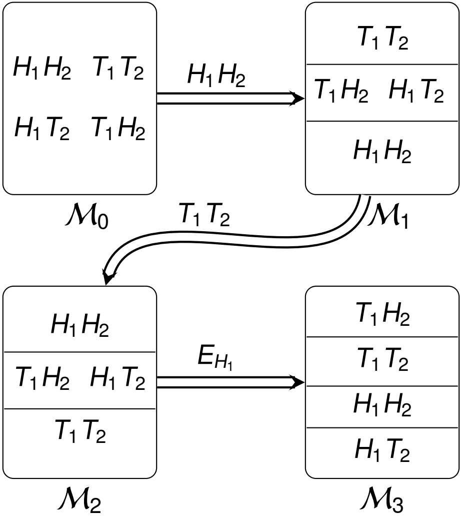
Heuristic Diagnosis of Stalnaker's Example



Heuristic Diagnosis of Stalnaker’s Example



Heuristic Diagnosis of Stalnaker's Example



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.

Discussion, I

A proper conceptualization of the event and report structure is crucial (the event space must be ‘rich enough’): A theory must be able to accommodate the conceptualization, but other than that it hardly counts in favor of a theory that the modeler gets this conceptualization right.

Discussion, II

There seems to be a trade-off between a rich set of states and event structure, and a rich theory of ‘doxastic actions’.

How should we resolve this trade-off when analyzing counterexamples to postulates of belief changes over time?

meta-information: information about how “trusted” or “reliable” the sources of the information are.

meta-information: information about how “trusted” or “reliable” the sources of the information are.

This is particularly important when analyzing how an agent’s beliefs change over an extended period of time. For example, rather than taking a stream of contradictory incoming evidence (i.e., the agent receives the information that p, then the information that q, then the information that $\neg p$, then the information that $\neg q$) at face value (and performing the suggested belief revisions), a rational agent may consider the stream itself as evidence that the source is not reliable

procedural information: information about the underlying *protocol* specifying which events (observations, messages, actions) are available (or permitted) at any given moment.

procedural information: information about the underlying *protocol* specifying which events (observations, messages, actions) are available (or permitted) at any given moment.

A *protocol* describes what the agents “can” or “cannot” do (say, observe) in a social interactive situation or rational inquiry.

meta-information: information about how “trusted” or “reliable” the sources of the information are.

procedural information: information about the underlying *protocol* specifying which events (observations, messages, actions) are available (or permitted) at any given moment.

Counterexamples

Are there any genuine counterexamples or do we want to reduce everything to misapplication? Under what conditions we can ignore the meta-information, which is often not specified in the description of an example (cf. the work of Halpern and Grünwald on *coarsening at random*).

P. Grünwald and J. Halpern. *Updating Probabilities*. Journal of Artificial Intelligence Research 19, pgs. 243 - 278, 2003.

Three Prisoner's Problem

Three prisoners A, B and C have been tried for murder and their verdicts will told to them tomorrow morning. They know only that one of them will be declared guilty and will be executed while the others will be set free. The identity of the condemned prisoner is revealed to the very reliable prison guard, but not to the prisoners themselves. Prisoner A asks the guard "Please give this letter to one of my friends — to the one who is to be released. We both know that at least one of them will be released".

Three Prisoner's Problem

An hour later, A asks the guard "Can you tell me which of my friends you gave the letter to? It should give me no clue regarding my own status because, regardless of my fate, each of my friends had an equal chance of receiving my letter." The guard told him that B received his letter.

Prisoner A then concluded that the probability that he will be released is 1/2 (since the only people without a verdict are A and C).

Three Prisoner's Problem

But, A thinks to himself:

Three Prisoner's Problem

But, A thinks to himself:

> Before I talked to the guard my chance of being executed was 1 in 3. Now that he told me B has been released, only C and I remain, so my chances of being executed have gone from 33.33% to 50%. What happened? I made certain not to ask for any information relevant to my own fate...

Three Prisoner's Problem

But, A thinks to himself:

> Before I talked to the guard my chance of being executed was 1 in 3. Now that he told me B has been released, only C and I remain, so my chances of being executed have gone from 33.33% to 50%. What happened? I made certain not to ask for any information relevant to my own fate...

Explain what is wrong with A's reasoning.

A's reasoning

Consider the following events:

G_A: "Prisoner A will be declared guilty" (we have $p(G_A) = 1/3$)

I_B: "Prisoner B will be declared innocent" (we have $p(I_B) = 2/3$)

A's reasoning

Consider the following events:

G_A: "Prisoner A will be declared guilty" (we have $p(G_A) = 1/3$)

I_B: "Prisoner B will be declared innocent" (we have $p(I_B) = 2/3$)

We have $p(I_B \mid G_A) = 1$: "If A is declared guilty then B will be declared innocent."

A's reasoning

Consider the following events:

G_A: "Prisoner A will be declared guilty" (we have $p(G_A) = 1/3$)

I_B: "Prisoner B will be declared innocent" (we have $p(I_B) = 2/3$)

We have $p(I_B \mid G_A) = 1$: "If A is declared guilty then B will be declared innocent."

Bayes Theorem:

$$p(G_A \mid I_B) =$$

A's reasoning

Consider the following events:

G_A: "Prisoner A will be declared guilty" (we have $p(G_A) = 1/3$)

I_B: "Prisoner B will be declared innocent" (we have $p(I_B) = 2/3$)

We have $p(I_B \mid G_A) = 1$: "If A is declared guilty then B will be declared innocent."

Bayes Theorem:

$$p(G_A \mid I_B) = p(I_B \mid G_A)\frac{p(G_A)}{p(I_B)} =$$

A's reasoning

Consider the following events:

G_A: "Prisoner A will be declared guilty" (we have $p(G_A) = 1/3$)

I_B: "Prisoner B will be declared innocent" (we have $p(I_B) = 2/3$)

We have $p(I_B \mid G_A) = 1$: "If A is declared guilty then B will be declared innocent."

Bayes Theorem:

$$p(G_A \mid I_B) = p(I_B \mid G_A)\frac{p(G_A)}{p(I_B)} = 1 \cdot \frac{1/3}{2/3} = 1/2$$

A's reasoning, corrected

But, A did not receive the information that B will be declared innocent, but rather that "the guard said that B will be declared innocent." So, A should have conditioned on the event:

I'_B: "The guard said that B will be declared innocent"

A's reasoning, corrected

But, A did not receive the information that B will be declared innocent, but rather that "the guard said that B will be declared innocent." So, A should have conditioned on the event:

I'_B: "The guard said that B will be declared innocent"

Given that $p(I'_B \mid G_A)$ is 1/2 (given that A is guilty, there is a 50-50 chance that the guard could have given the letter to B or C).

A's reasoning, corrected

But, A did not receive the information that B will be declared innocent, but rather that "the guard said that B will be declared innocent." So, A should have conditioned on the event:

I'_B: "The guard said that B will be declared innocent"

Given that $p(I'_B \mid G_A)$ is 1/2 (given that A is guilty, there is a 50-50 chance that the guard could have given the letter to B or C). This gives us the following correct calculation:

$$p(G_A \mid I'_B) = p(I'_B \mid G_A)\frac{p(G_A)}{p(I'_B)} = 1/2 \cdot \frac{1/3}{1/2} = 1/3$$

When does conditioning on the “naive” space give the same results as conditioning on the “sophisticated” space?

When does conditioning on the “naive” space give the same results as conditioning on the “sophisticated” space? Grünwald and Halpern’s answer: When the CAR (coarsening at random) condition is satisfied (and this only happens in trivial cases).

$$\begin{tikzcd}[column sep=huge, row sep=huge]
K_0 \arrow[r, Rightarrow, "\substack{\text{Learn that } \varphi \\ \text{Suppose that } \varphi}"] \arrow[d, leftrightarrow] & K_t = K_0 * \varphi \arrow[d, leftrightarrow] \\
p_0 \arrow[r, Rightarrow, "\text{Learning experience}"] & p_t = ???
\end{tikzcd}$$

$$
\begin{tikzcd}[column sep=huge, row sep=huge]
K_0 \arrow[r, Rightarrow, "\substack{\text{Learn that } \varphi \\ \text{Suppose that } \varphi}"] \arrow[d, leftrightarrow] & K_t = K_0 * \varphi \arrow[d, leftrightarrow] \\
p_0 \arrow[r, Rightarrow, "\text{Learning experience}"] & p_t = ???
\end{tikzcd}
$$

Bridge Principles

Probability 1: $Bel(A)$ iff $P(A) = 1$

Bridge Principles

Probability 1: $Bel(A)$ iff $P(A) = 1$

The Lockean Thesis: $Bel(A)$ iff $P(A) > r$

Bridge Principles

Probability 1: $Bel(A)$ iff $P(A) = 1$

The Lockean Thesis: $Bel(A)$ iff $P(A) > r$

Decision-theoretic accounts: $Bel(A)$ iff
$\sum_{w\in W} P(\{w\})\cdot u(bel\ A, w)$ has such-and-such property

Bridge Principles

Probability 1: $Bel(A)$ iff $P(A) = 1$

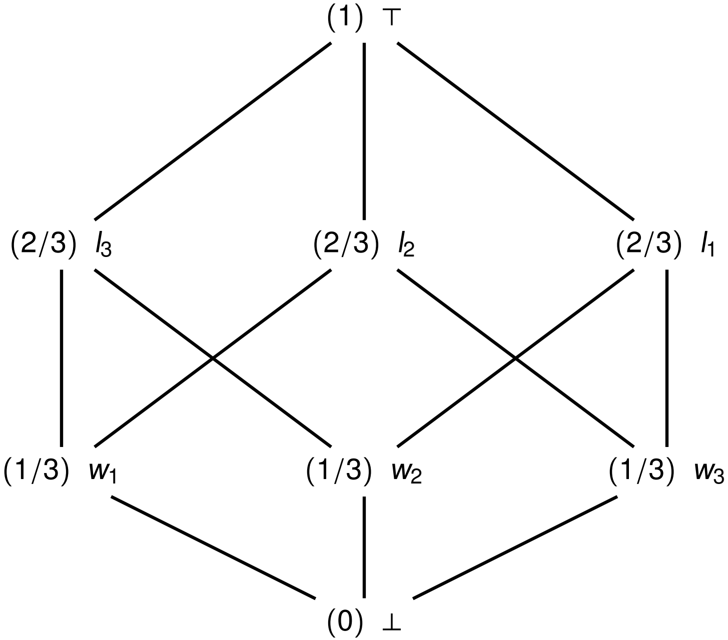
The Lockean Thesis: $Bel(A)$ iff $P(A) > r$

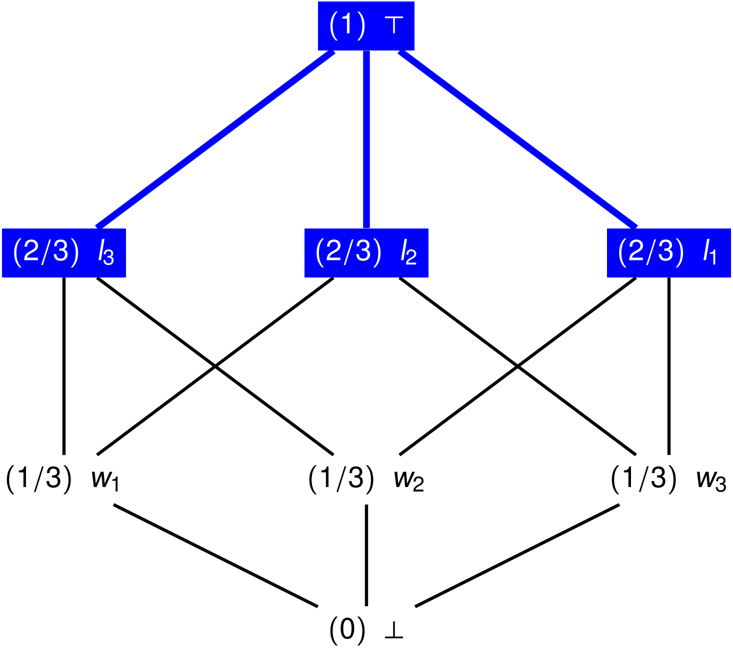
Decision-theoretic accounts: $Bel(A)$ iff
$\sum_{w \in W} P(\{w\}) \cdot u(bel\ A, w)$ has such-and-such property

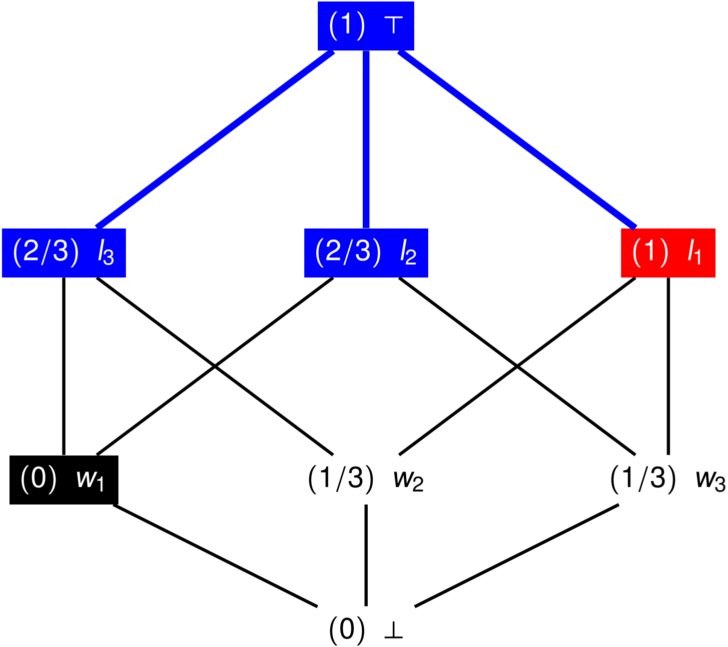
The Nihilistic proposal: "...no explication of belief is possible within the confines of the probability model."

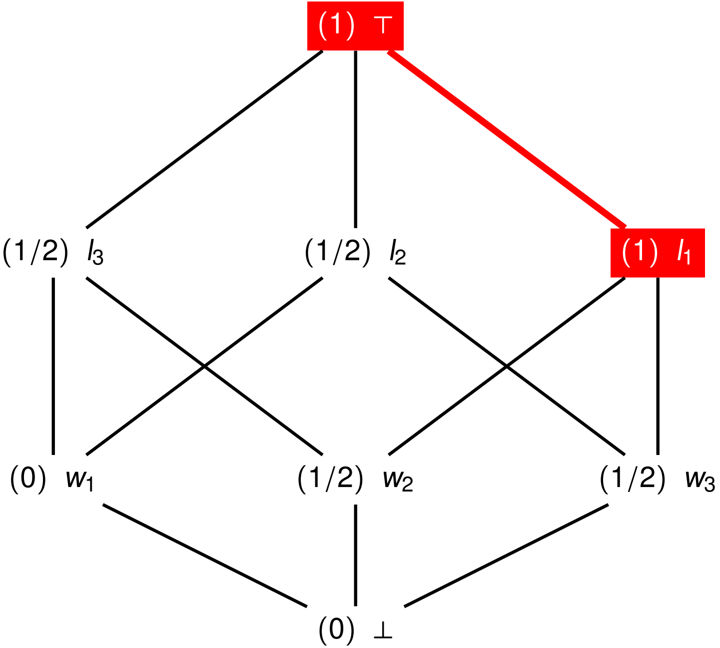
Beliefs that obey the Lockean thesis can be undermined by new evidence that is consistent with the agent's current beliefs.

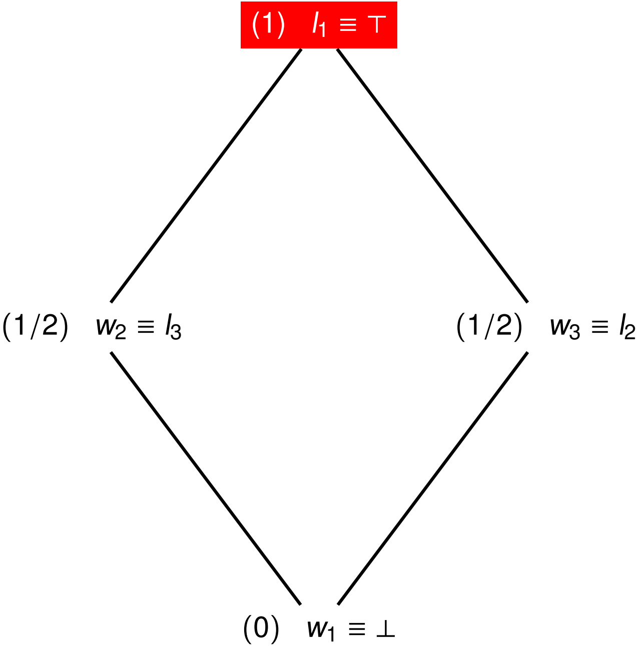
For each $i = 1, 2, 3$, let l_i be the proposition Ticket i won't win (and w_i is the proposition that "ticket i will win"). And let us set our threshold for Lockean belief at $r = 0.6$.


(1) ⊤
(2/3) l_3
(2/3) l_2
(2/3) l_1
(1/3) w_1
(1/3) w_2
(1/3) w_3
(0) ⊥


(1) ⊤
(2/3) l_3
(2/3) l_2
(2/3) l_1
(1/3) w_1
(1/3) w_2
(1/3) w_3
(0) ⊥


(1) ⊤
(2/3) l_3
(2/3) l_2
(1) l_1
(0) w_1
(1/3) w_2
(1/3) w_3
(0) ⊥


(1) ⊤
(1/2) l_3
(1/2) l_2
(1) l_1
(0) w_1
(1/2) w_2
(1/2) w_3
(0) ⊥


(1) $l_1 \equiv \top$
(1/2) $w_2 \equiv l_3$
(1/2) $w_3 \equiv l_2$
(0) $w_1 \equiv \bot$

Resiliency, Robust Belief, Stable Belief

B. Skyrms. *Resiliency, propensities, and causal necessity*. Journal of Philosophy, 74:11, pgs. 704 - 713, 1977.

A. Baltag and S. Smets. *Probabilistic Belief Revision*. Synthese, 2008.

H. Leitgeb. *Reducing belief simpliciter to degrees of belief*. Annals of Pure and Applied Logic, 16:4, pgs. 1338 - 1380, 2013.

R. Stalnaker. *Belief revision in games: forward and backward induction*. Mathematical Social Sciences, 36, pgs. 31 - 56, 1998.

Probability

Let W be a set of states and $\mathfrak{A}$ a σ-algebra: $\mathfrak{A} \subseteq \wp(W)$ such that

- $W, \emptyset \in \mathfrak{A}$
- if $X \in \mathfrak{A}$ then $W - X \in \mathfrak{A}$
- if $X, Y \in \mathfrak{A}$ then $X \cup Y \in \mathfrak{A}$
- if $X_0, X_1, \ldots \in \mathfrak{A}$ then $\bigcup_{i \in \mathbb{N}} X_i \in \mathfrak{A}$.

Probability

$P : \mathfrak{A} \to [0, 1]$ satisfying the usual constraints

- $P(W) = 1$
- (finite additivity) If $X_1, X_2 \in \mathfrak{A}$ are pairwise disjoint, then $P(X_1 \cup X_2) = P(X_1) + P(X_2)$

$P(Y|X) = \frac{P(Y \cap X)}{P(X)}$ whenever $P(X) > 0$. So, $P(Y|W)$ is $P(Y)$.

- P is countably additive (σ-additive): if $X_1, X_2, \ldots, X_n, \ldots$ are pairwise disjoint members of $\mathfrak{A}$, then $P(\bigcup_{n\in\mathbb{N}} X_n) = \sum_{n\in\mathbb{N}} P(X_n)$

P-stability^r

Definition. Let P be a probability measure on $\mathfrak{A}$ over W, let $0 \leq t < 1$. For all $X \in \mathfrak{A}$:

X is P-stable^t if and only if for all $Y \in \mathfrak{A}$ with $Y \cap X \neq \emptyset$ and $P(Y) > 0$: $P(X|Y) > t$.

P-stability^r

Definition. Let P be a probability measure on $\mathfrak{A}$ over W, let $0 \leq t < 1$. For all $X \in \mathfrak{A}$:

X is P-stable^t if and only if for all $Y \in \mathfrak{A}$ with $Y \cap X \neq \emptyset$ and $P(Y) > 0$: $P(X|Y) > t$.

- Trivially, the empty set of P-stable^t.

P-stability^r

Definition. Let P be a probability measure on $\mathfrak{A}$ over W, let $0 \leq t < 1$. For all $X \in \mathfrak{A}$:

$$X \text{ is } P\text{-stable}^t \text{ if and only if for all } Y \in \mathfrak{A} \text{ with } Y \cap X \neq \emptyset \text{ and } P(Y) > 0\text{: } P(X|Y) > t.$$

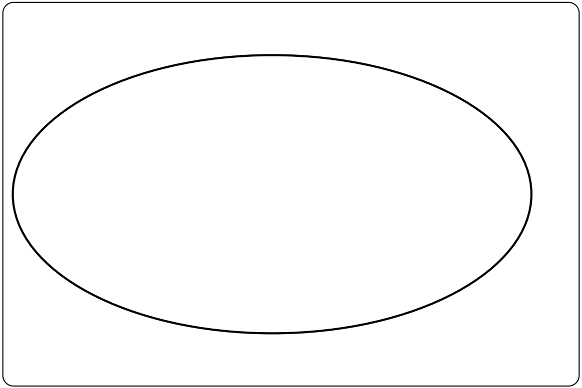
- Trivially, the empty set of P-stable^t.
- If $P(X) = 1$, then X is P-stable^t.

P-stability^r

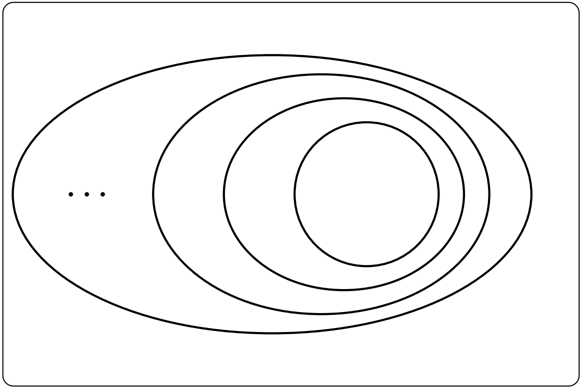
Definition. Let P be a probability measure on $\mathfrak{A}$ over W, let $0 \leq t < 1$. For all $X \in \mathfrak{A}$:

$$X \text{ is } P\text{-stable}^t \text{ if and only if for all } Y \in \mathfrak{A} \text{ with } Y \cap X \neq \emptyset \text{ and } P(Y) > 0\text{: } P(X|Y) > t.$$

- Trivially, the empty set of P-stable^t.
- If $P(X) = 1$, then X is P-stable^t.
- There are P-stable^t sets with $0 < P(X) < 1$.



- Assuming countable additivity and $t \geq \frac{1}{2}$, The class of P-stable^t propositions X in $\mathfrak{A}$ with $P(X) < 1$ is well-ordered with respect to the subset relation.
- If there is a non-empty P-stable^r $X \in \mathfrak{A}$ with $P(X) < 1$, then there is also a least such X.



- Assuming countable additivity and $t \geq \frac{1}{2}$, The class of P-stable^t propositions X in $\mathfrak{A}$ with $P(X) < 1$ is well-ordered with respect to the subset relation.
- If there is a non-empty P-stable^r $X \in \mathfrak{A}$ with $P(X) < 1$, then there is also a least such X.

$w \in SB(H)$ iff for all $E \in \mathfrak{A}(W)$ with $H \cap E \neq \emptyset$ and $P(E) \neq 0$:
$P(H \mid E) \geq t$

$w \in SB(H)$ iff for all $E \in \mathfrak{A}(W)$ with $H \cap E \neq \emptyset$ and $P(E) \neq 0$: $P(H \mid E) \geq t_C$

1. The threshold t is determined contextually (the "cautiousness level")

$w \in SB(H)$ iff for all $E \in \mathfrak{A}_H(W)$ with $H \cap E \neq \emptyset$ and $P(E) \neq 0$: $P(H \mid E) \geq t_C$

1. The threshold t is determined contextually (the "cautiousness level")
2. The evidence "relevant" to H

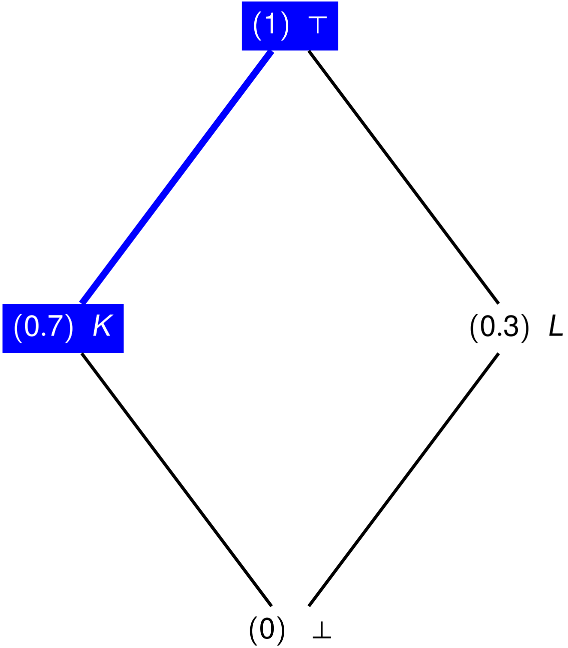
$w \in SB(H)$ iff for all $E \in \mathfrak{A}_H(W_\Pi)$ with $H \cap E \neq \emptyset$ and $P(E) \neq 0$:
$P(H \mid E) \geq t_C$

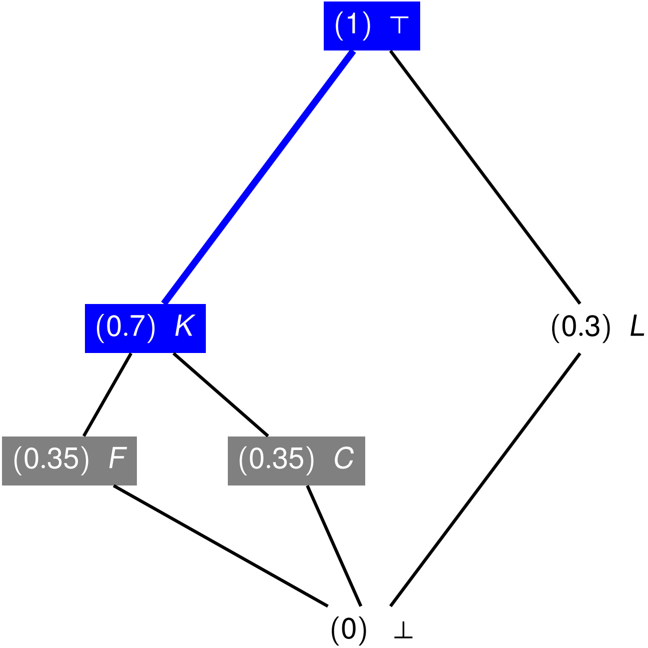
1. The threshold t is determined contextually (the "cautiousness level")
2. The evidence "relevant" to H
3. The states may be contextually determined (by a partition Π on a set W of "maximally specific worlds")

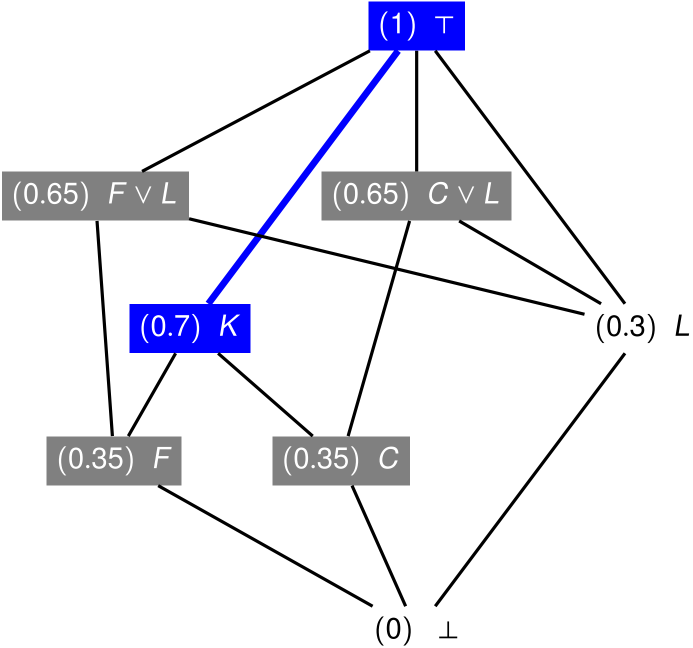
H. Leitgeb. *The Stability Theory of Belief*. The Philosophical Review 123/2, 131171, 2014.

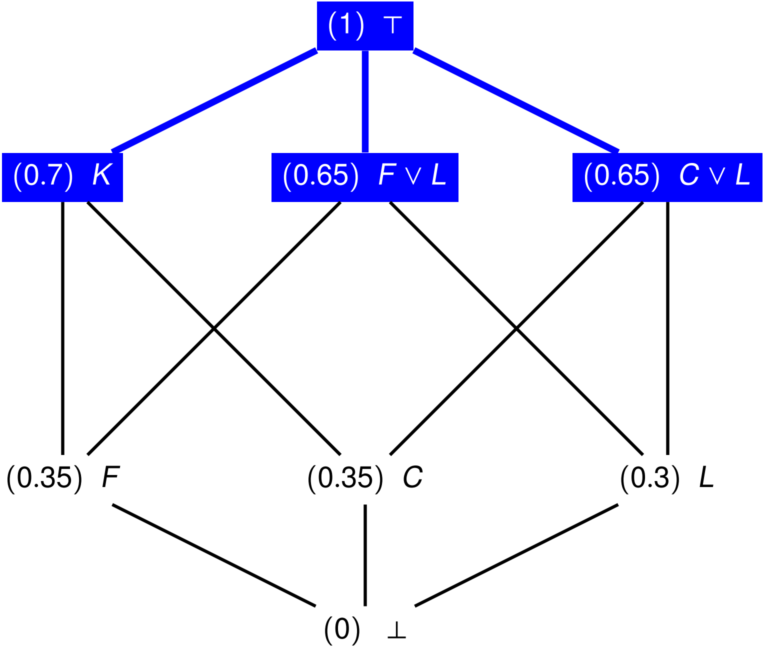
H. Leitgeb. *The Humean Thesis on Belief*. Proceedings of the Aristotelian Society of Philosophy 89(1), 143185, 2015.

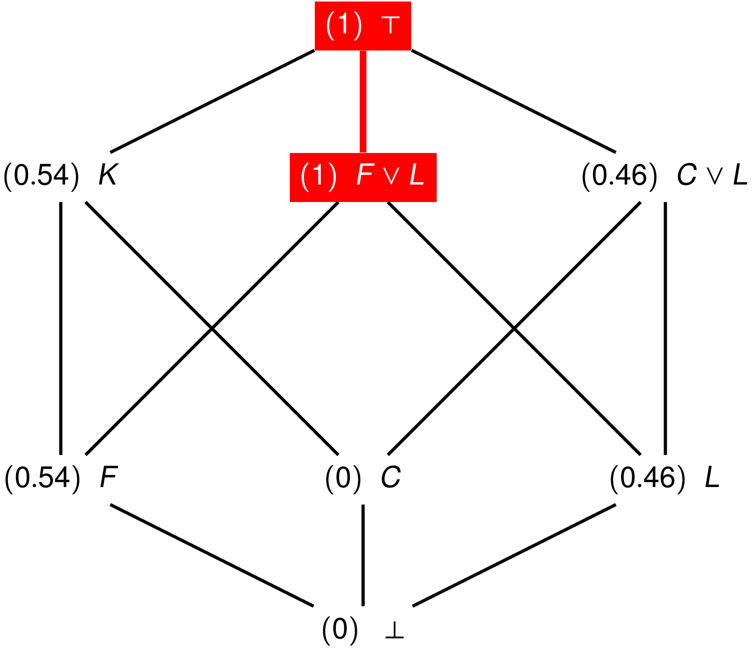
R. Pettigrew. *Pluralism about belief states*. Proceedings of the Aristotelian Society 89(1):187-204, 2015.


(1) ⊤
(0.7) K
(0.3) L
(0) ⊥


(1) ⊤
(0.7) K
(0.3) L
(0.35) F
(0.35) C
(0) ⊥


(1) $\top$
(0.65) $F \vee L$
(0.65) $C \vee L$
(0.7) K
(0.3) L
(0.35) F
(0.35) C
(0) $\bot$


(1) ⊤
(0.7) K
(0.65) $F \vee L$
(0.65) $C \vee L$
(0.35) F
(0.35) C
(0.3) L
(0) ⊥


(1) $\top$
(0.54) K
(1) $F \vee L$
(0.46) $C \vee L$
(0.54) F
(0) C
(0.46) L
(0) $\bot$

Thus, while **robust belief** is stable under acquisition of new (doxastically possible) evidence and Lockean belief is not, **robust belief** is not stable under fine-graining of possibilities while Lockean belief is.

Leitgeb's Solution to the Lottery Paradox

In a context in which the agent is interested in *whether ticket i will be drawn*; for example, for $i = 1$: Let Π be the corresponding partition:

$$\{\{w_1\}, \{w_2, \ldots, w_{1,000,000}\}\}$$

The resulting probability measure P_Π is given so that P is given by P so that:

$$P_\Pi(\{\{w_1\}\}) = \frac{1}{1,000,000} \qquad P_\Pi(\{\{w_2, \ldots, w_{1,000,000}\}\}) = \frac{999,999}{1,000,000}$$

There are two P_Π-stable sets, and one of the two possible choices for the strongest believed proposition $B_W^\Pi = \{\{w_2, \ldots, w_{1,000,000}\}\}$.

If B_W^Π is chosen as such, our perfectly rational agent believes of ticket $i = 1$ that it will not be drawn, (and of course P1 -P3 are satisfied).

For example, this might be a context in which a single ticket holder—the person holding ticket 1—would be inclined to say of his or her ticket: “I believe it wont win.”

In a context in which the agent is interested in *which ticket will be drawn*: Let Π' be the corresponding partition that consists of all singleton subsets of W. The probability measure P^{Π} is the uniform probability on W.

The only P-stable set—and hence the only choice for the strongest believed proposition $B_W^{\Pi'}$—is W itself: our perfectly rational agent believes that some ticket will be drawn, but he or she does not believe of any ticket that it will not win

For example, this might be a context in which a salesperson of tickets in a lottery would be inclined to say of each ticket: “It might win” (that is, it is not the case that I believe that it won’t win).

In either of the two contexts from before, the theory avoids the absurd conclusion of the Lottery Paradox; in each context, it preserves the closure of belief under conjunction; and in each context, it preserves the Lockean thesis for some threshold ($r = \frac{999{,}999}{1{,}000{,}000}$ in the first case, $r = 1$ in the second case)-all of this follows from P-stability and the theorem.

In the first Π-context, the intuition is preserved that, in some sense, one believes of ticket *i* that it will lose since it is so likely to lose.

In the second Π'-context, the intuition is preserved that, in a different sense, one should not believe of any ticket that it will lose since the situation is symmetric with respect to tickets, as expressed by the uniform probability measure, and of course some ticket must win.

Finally, by disregarding or mixing the contexts, it becomes apparent why one might have regarded all of the premises of the Lottery Paradox as true.

But according to the present theory, contexts should not be disregarded or mixed: partitions Π and Π' differ from each other, and different partitions may lead to different beliefs, as observed in the last section and as exemplified in the Lottery Paradox.

Accordingly, the thresholds in the Lockean thesis may have to be chosen differently in different contexts, and once again, this is what happens in the Lottery Paradox—which makes good sense: in the second Π'-context, by uniformity, the agents degrees of belief do not give him or her much of a hint of what to believe. That is why the agent ought to be supercautious about her beliefs in that

Knowledge, Questions and Issues

J. van Benthem and S. Minica. *Toward a Dynamic Logic of Questions*. Journal of Philosophical Logic, 41(4), pp. 633 - 669, 2012.

A. Baltag, R. Boddy and S. Smets. *Group Knowledge in Interrogative Epistemology*. in *Jaakko Hintikka on Knowledge and Game-Theoretical Semantics*, pp. 131-164.

I. Ciardelli. *Modalities in the realm of questions: axiomatizing inquisitive epistemic logic*. Advances in Modal Logic, 2014.

Inquisitive Semantics

Ivano Ciardelli, Jeroen Groenendijk, and Floris Roelofsen. *Inquisitive Semantics*. Oxford University Press, 2018.

Questions

Suppose that W is a set of states.

A **question** is a partition on W.

Questions

Suppose that W is a set of states.

A **question** is a partition on W.

$Quest_W = \{\approx_Q \mid \approx_Q \text{ is a partition on } W\}$

Given $P \subseteq W$, a **binary question** is the partition $\{P, W \setminus P\}$, so $s \approx^P t$ iff either $s, t \in P$ or $s, t \notin P$

Questions

Suppose that W is a set of states.

A **question** is a partition on W.

$Quest_W = \{\approx_Q \mid \approx_Q \text{ is a partition on } W\}$

Given $P \subseteq W$, a **binary question** is the partition $\{P, W \setminus P\}$, so $s \approx^P t$ iff either $s, t \in P$ or $s, t \notin P$

Every family of questions $Quest \subseteq Quest_W$ can be 'compressed' into one big 'conjunctive' question: this is the least refined partition that refines every question in $Quest$, $\approx_{Quest} = \bigcap\{\approx_Q \mid Q \in Quest\}$

For $i \in \mathcal{A}$, let $\approx_i$ represent i's, *total question*.

"van Benthem and Minica call $\approx_i$ the agent i's *issue relation*.... it essentially captures agent i's conceptual indistinguishability relation, since it specifies the finest relevant world-distinctions that agent i makes....Two worlds $s \approx_i t$ are conceptually indistinguishable for agent i (since the answers to all i's questions are the same in both worlds): one can say that s and t will correspond to the same world in agent i's own "subjective model"." (Baltag et al.)

Epistemic Issue Model

$\mathcal{M} = \langle W, \{\rightarrow_i\}_{i\in\mathcal{A}}, \{\approx_i\}_{i\in\mathcal{A}}, V\rangle$, where

- W is a non-empty set of states
- For $i \in \mathcal{A}$, $\approx_i \subseteq W \times W$ is an equivalence relation (the issue relation)
- For $i \in \mathcal{A}$, $\rightarrow_i \subseteq W \times W$ is reflexive (the epistemic alternative relation)
- $V : \mathsf{At} \rightarrow \wp(W)$ is a valuation funciton

For $s \in W$, $s(i) = \{s' \mid s \to_i s'\}$ is the set of epistemic possibilities for i at s.

Open questions: The restriction $\approx_{i|_{s(a)}} = \approx_i \cap (s(a) \times s(a))$ represents i's current open isues at world s.

For $s \in W$, $s(i) = \{s' \mid s \to_i s'\}$ is the set of epistemic possibilities for i at s.

Open questions: The restriction $\approx_{i|_{s(a)}} = \approx_i \cap (s(a) \times s(a))$ represents i's current open isues at world s.

Suppose that $P \subseteq W$ is a proposition. Then,

$K_i P = \{s \mid s \in W, s(i) \subseteq P\}$

$CP = \{s \mid \text{ for all } t, \text{ if } s(\bigcup_i \to_i)^+ t, \text{ then } t \in P\}$

$DP = \{s \mid \text{ for all } t, \text{ if } s(\bigcap_i \to_i) t, \text{ then } t \in P\}$

$Q_i P = \{s \mid \text{ for all } t, \text{ if } s \approx_i t, \text{ then } t \in P\}$

Conceptual indistinguishability implies epistemic indistinguishability: For all $i \in \mathcal{A}$, $\approx_i \subseteq \rightarrow_i$.

For all φ, $K_i\varphi \Rightarrow Q_i\varphi$

To know is to know the answer to a question: For all $i \in \mathcal{A}$, $\rightarrow_i \approx_i \subseteq \rightarrow_i$

For all φ, $K_i\varphi \Rightarrow K_iQ_i\varphi$

Selective Public Announcement

Principle of Selective Learning. When confronted with information, agents come to know only the information that is relevant for their issues.

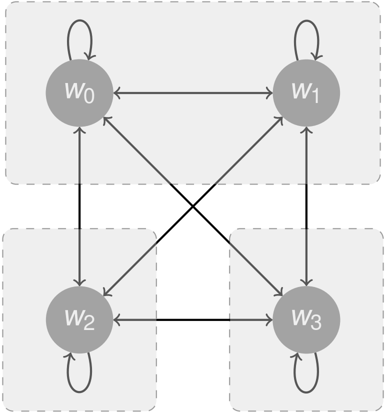
For any proposition $P \subseteq W$ and $i \in \mathcal{A}$, let P_i the *strongest i-relevant proposition entailed by P*:

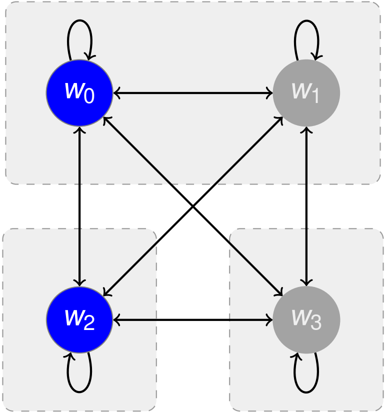
$$P_i = \{s \in W \mid s \approx_i s' \text{ for some } s' \in P\}$$

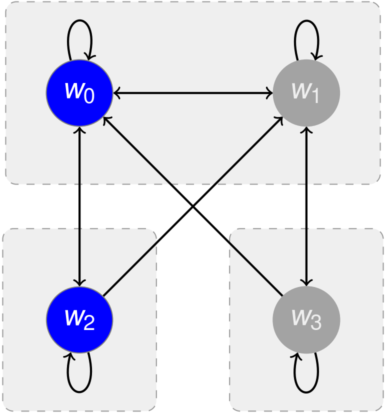
Selective Public Announcement

Suppose that $\mathcal{M} = \langle W, \{\rightarrow_i\}_{i\in\mathcal{A}}, \{\approx_i\}_{i\in\mathcal{A}}, V\rangle$ is an epistemic issue model and $P \subseteq W$ is a proposition. A **selective public announcement** $!P$ is an action that changes $\mathcal{M}$ to $\mathcal{M}^P = \langle W^P, \{\rightarrow_i^P\}_{i\in\mathcal{A}}, \{\approx_i^P\}_{i\in\mathcal{A}}, V\rangle$, where

- $W^P = W$
- $\rightarrow_i^P = \rightarrow_i \cap \approx^{P_i}$
- $\approx_i^P = \approx_i$
- For all $p \in \mathrm{At}$, $V^P(p) = V(p)$.


w_0
w_1
w_2
w_3


w_0
w_1
w_2
w_3


w_0
w_1
w_2
w_3

A. Baltag, R. Boddy and S. Smets. *Group Knowledge in Interrogative Epistemology*. in *Jaakko Hintikka on Knowledge and Game-Theoretical Semantics*, pp. 131-164.

I. Ciardelli and F. Roelofsen. *Inquisitive dynamic epistemic logic*. Synthese, 2015.

An **issue** is a non-empty, downward closed set of information states. We say that an information state t settles an issue I in case $t \in I$.

Let Π be the set of all issues.

An **issue** is a non-empty, downward closed set of information states. We say that an information state t settles an issue I in case $t \in I$.

Let Π be the set of all issues.

An **inquisitive model** is a tuple $\langle W, (\Sigma_i)_{i \in \mathcal{A}}, V\rangle$ where

- W is a non-empty set of possible worlds
- $V : W \to \wp(\mathrm{At})$ is a valuation function
- $\Sigma_i : W \to \Pi$ where $\Sigma_i(w)$ is an issue, satisfying:

  Factivity For all $w \in W$, $w \in \sigma_i(w)$
  Introspection For any $w, v \in W$ if $v \in \sigma_i(w)$, then $\Sigma_i(v) = \Sigma_i(w)$.

  where $\sigma_i(w) := \Sigma_i(w)$ represents the information state of agent i in w.

1. For all $p \in \text{At}$, $p \in \mathcal{L}_!$
2. For all $\bot \in \mathcal{L}_!$
3. If $\alpha_1, \ldots, \alpha_n \in\in \mathcal{L}_!$, then $?\{\alpha_1, \ldots, \alpha_n\}$
4. If $\varphi \in \mathcal{L}_\circ$ and $\psi \in \mathcal{L}_\circ$, then $\varphi \wedge \psi \in \mathcal{L}_\circ$
5. If $\varphi \in \mathcal{L}_\circ$ and $\psi \in \mathcal{L}_\circ$, then $\varphi \wedge \psi \in \mathcal{L}_\circ$
6. If $\alpha \in \mathcal{L}_!$ and $\psi \in \mathcal{L}_\circ$, then $\alpha \to \psi \in \mathcal{L}_\circ$
7. If $\varphi \in \mathcal{L}_\circ$, then $E_i\varphi \in \mathcal{L}_!$
8. If $\varphi \in \mathcal{L}_\circ$, then $K_i\varphi \in \mathcal{L}_!$

Interrogative: $?\{\alpha_1, \ldots, \alpha_n\}$.

$?p$ means $?\{p, \neg p\}$

$K_i\varphi$: i knows that φ is true

$E_i\varphi$: i entertains φ being true

$K_i?p$ means "i knows whether p is true

$K_i?K_j?p$ "i knows whether j knows whether p is true

The following definition specifies recursively when a sentence is **supported** by a state s. Intuitively, for declaratives being supported amounts to being established, or true everywhere in s, while for interrogatives it amounts to being resolved in s.

1. $\mathcal{M}, s \models p$ iff $p \in V(w)$ for all $w \in s$.
2. $\mathcal{M}, s \models \bot$ iff $s = \emptyset$.
3. $\mathcal{M}, s \models ?\{\alpha_1, \ldots, \alpha_n\}$ iff $s = \emptyset$.
4. $\mathcal{M}, s \models \varphi \wedge \psi$ iff $\mathcal{M}, s \models \varphi$ and $\mathcal{M}, s \models \psi$.
5. $\mathcal{M}, s \models \alpha \rightarrow \psi$ iff for any $t \subseteq s$, if $\mathcal{M}, t \models \alpha$, then $\mathcal{M}, t \models \varphi$.
6. $\mathcal{M}, s \models K_i\psi$ iff for any $w \in s$, $\mathcal{M}, \sigma_i(w) \models \varphi$.
7. $\mathcal{M}, s \models E_i\psi$ iff for any $w \in s$, for any $t \in \Sigma_i(w)$, $\mathcal{M}, t \models \varphi$.

Fact 1 (Persistency of support) If $\mathcal{M}, s \models \varphi$ and $t \subseteq s$, then $\mathcal{M}, t \models \varphi$.

Fact 2 (The empty state supports everything) For any $\mathcal{M}$ and any φ, $\mathcal{M}, \emptyset \models \varphi$

Fact 3 (Support for negation, disjunction, and polar interrogatives)

- $\mathcal{M}, s \models \neg\alpha$ iff for any non-empty $t \subseteq s$, $\mathcal{M}, t \not\models \alpha$
- $\mathcal{M}, s \models \alpha \vee \beta$ iff there are t_1, t_2 such that $s = t_1 \cup t_2$, and $\mathcal{M}, t_1 \models \alpha$ and $\mathcal{M}, t_2 \models \beta$
- $\mathcal{M}, s \models ?\alpha$ iff $\mathcal{M}, t \models \alpha$ or $\mathcal{M}, t \models \neg\alpha$

We say that a sentence φ **entails** ψ, notation $\varphi \models \psi$, just in case for all models $\mathcal{M}$ and states s, if $\mathcal{M}, s \models \varphi$ then $\mathcal{M}, s \models \psi$.

We say that a sentence φ is **valid** in case it is supported by all states in all models.

We say that two sentences φ and ψ are **equivalent**, notation $\varphi \equiv \psi$, just in case for all models $\mathcal{M}$ and states s, $\mathcal{M}, s \models \varphi$ iff $\mathcal{M}, s \models \psi$.

φ is **true** at w in $\mathcal{M}$ iff φ is supported by $\{w\}$ in $\mathcal{M}$

The **truth set** of a sentence φ in a model $\mathcal{M}$, denoted $|\varphi|_{\mathcal{M}}$, is defined as the set of worlds in $\mathcal{M}$ where φ is true:
$|\varphi|_{\mathcal{M}} := \{w \in W \mid \mathcal{M}, w \models \varphi\}$

The **proposition** $[\varphi]_{\mathcal{M}}$ expressed by a sentence φ in a model $\mathcal{M}$ is the set of all states in $\mathcal{M}$ that support φ:
$[\varphi]_{\mathcal{M}} := \{s \subseteq W \mid \mathcal{M}, s \models \varphi\}$

φ is **true** at w in $\mathcal{M}$ iff φ is supported by $\{w\}$ in $\mathcal{M}$

The **truth set** of a sentence φ in a model $\mathcal{M}$, denoted $|\varphi|_{\mathcal{M}}$, is defined as the set of worlds in $\mathcal{M}$ where φ is true:
$|\varphi|_{\mathcal{M}} := \{w \in W \mid \mathcal{M}, w \models \varphi\}$

The **proposition** $[\varphi]_{\mathcal{M}}$ expressed by a sentence φ in a model $\mathcal{M}$ is the set of all states in $\mathcal{M}$ that support φ:
$[\varphi]_{\mathcal{M}} := \{s \subseteq W \mid \mathcal{M}, s \models \varphi\}$

We have that $|?p|_{\mathcal{M}} = |?q|_{\mathcal{M}}$, but $[?p]_{\mathcal{M}} \neq [?q]_{\mathcal{M}}$

Fact: For any φ and any model $\mathcal{M}$, $|\varphi|_{\mathcal{M}} = \bigcup[\varphi]_{\mathcal{M}}$

Fact (Truth and support) For any model $\mathcal{M}$, any state s and any declarative α, the following holds:

$$\mathcal{M}, s \models \alpha \text{ iff } \mathcal{M}, w \models \alpha \text{ for all } w \in s$$

$\mathcal{M}, s \models \alpha \to \varphi$ iff $\mathcal{M}, s \cap |\alpha|_{\mathcal{M}} \models \varphi$

$\mathcal{M}, s \models \alpha \to \varphi$ iff $\mathcal{M}, s \cap |\alpha|_{\mathcal{M}} \models \varphi$

If Ann invites Bill to the party, will he go? ($p \to ?q$)

Answers:

- Yes, if Ann invites Bill, he will go. ($p \to q$)
- No, if Ann invites Bill, he will not go. ($p \to \neg q$)

Knowledge

For declaratives α, $K_i\alpha$ boils down to the usual definition of truth of a modality familiar from modal logic.

For interrogatives μ, $K_i\mu$ holds when μ is resolved in $\sigma_i(w)$, which means that $K_i\mu$ expresses the fact that i has sufficient information to resolve μ at w.

For instance, $K_i?p$ is true at w just in case that $\sigma_i(w)$ supports either p or $\neg p$. That is, when i knows whether p is true.

Entertaining

$E_i\varphi$ is true at w just in case φ is supported by any state $t \in \Sigma_i(w)$

Fact. For any φ, $K_i\varphi \models E_i\varphi$

Fact. For any declarative α, $K_i\alpha \equiv E_i\alpha$

$W_i\varphi$ means "i wonders about φ: $W_i\varphi := \neg K_i\varphi \wedge E_i\varphi$

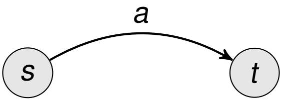
- $\mathcal{M}, w \models K_i\varphi$ iff $\bigcup \Sigma_i(w) \in [\varphi]_{\mathcal{M}}$
- $\mathcal{M}, w \models E_i\varphi$ iff $\Sigma_i(w) \subseteq [\varphi]_{\mathcal{M}}$

Actions

1. Actions as *transitions between states, or situations*:

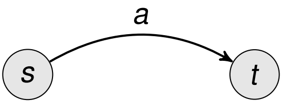
Actions

1. Actions as *transitions between states, or situations*:

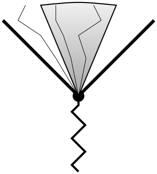


Actions

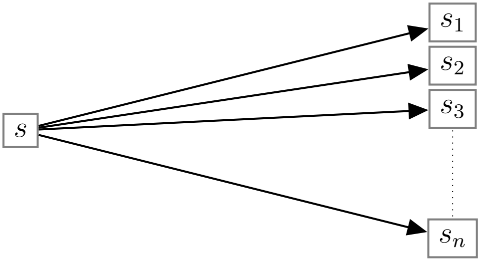
1. Actions as *transitions between states, or situations*:



2. Actions *restrict* the set of possible future histories.



J. van Benthem, H. van Ditmarsch, J. van Eijck and J. Jaspers. *Chapter 6: Propositional Dynamic Logic*. Logic in Action Online Course Project, 2011.


s
s_1
s_2
s_3
s_n

Propositional Dynamic Logic

Language: The language of propositional dynamic logic is generated by the following grammar:

$$p \mid \neg\varphi \mid \varphi \wedge \psi \mid [\alpha]\varphi$$

where $p \in \mathrm{At}$ and α is generated by the following grammar:

$$a \mid \alpha \cup \beta \mid \alpha;\beta \mid \alpha^* \mid \varphi?$$

where $a \in \mathrm{Act}$ and φ is a formula.

Propositional Dynamic Logic

Language: The language of propositional dynamic logic is generated by the following grammar:

$$p \mid \neg\varphi \mid \varphi \wedge \psi \mid [\alpha]\varphi$$

where $p \in \mathsf{At}$ and α is generated by the following grammar:

$$a \mid \alpha \cup \beta \mid \alpha;\beta \mid \alpha^* \mid \varphi?$$

where $a \in \mathsf{Act}$ and φ is a formula.

Semantics: $\mathcal{M} = \langle W, \{R_a \mid a \in \mathsf{P}\}, V\rangle$ where for each $a \in \mathsf{P}$, $R_a \subseteq W \times W$ and $V : \mathsf{At} \to \wp(W)$

Propositional Dynamic Logic

Language: The language of propositional dynamic logic is generated by the following grammar:

$$p \mid \neg\varphi \mid \varphi \wedge \psi \mid [\alpha]\varphi$$

where $p \in \mathsf{At}$ and α is generated by the following grammar:

$$a \mid \alpha \cup \beta \mid \alpha;\beta \mid \alpha^* \mid \varphi?$$

where $a \in \mathsf{Act}$ and φ is a formula.

Semantics: $\mathcal{M} = \langle W, \{R_a \mid a \in \mathsf{P}\}, V\rangle$ where for each $a \in \mathsf{P}$, $R_a \subseteq W \times W$ and $V : \mathsf{At} \rightarrow \wp(W)$

$[\alpha]\varphi$ means "after doing α, φ will be true"

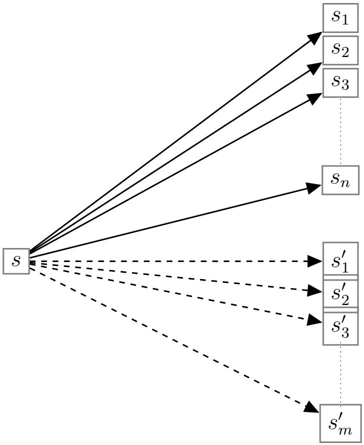
$\langle\alpha\rangle\varphi$ means "after doing α, φ may be true"

$\mathcal{M}, w \models [\alpha]\varphi$ iff for each v, if $wR_\alpha v$ then $\mathcal{M}, v \models \varphi$

$\mathcal{M}, w \models \langle\alpha\rangle\varphi$ iff there is a v such that $wR_\alpha v$ and $\mathcal{M}, v \models \varphi$

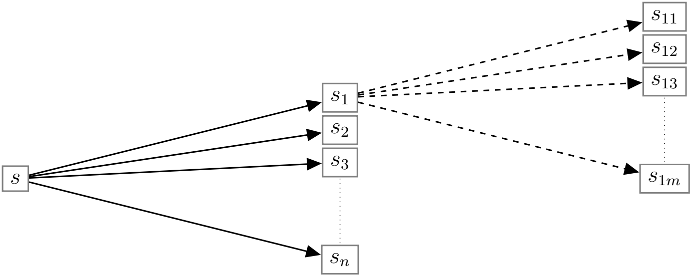
Union

$R_{\alpha\cup\beta} := R_\alpha \cup R_\beta$



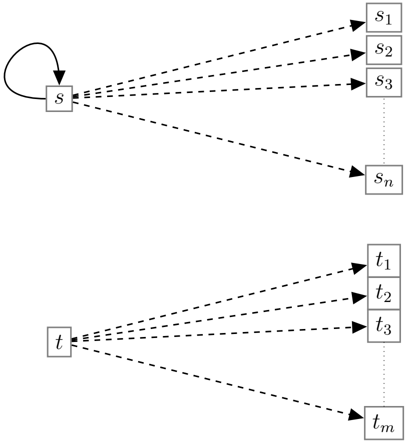
Sequence

$$R_{\alpha;\beta} := R_\alpha \circ R_\beta$$



Test

$$R_{\varphi?} = \{(w, w) \mid \mathcal{M}, w \models \varphi\}$$



Iteration

$$R_{\alpha^*} := \cup_{n \geq 0} R_\alpha^n$$

Propositional Dynamic Logic

1. Axioms of propositional logic
2. $[\alpha](\varphi \to \psi) \to ([\alpha]\varphi \to [\alpha]\psi)$
3. $[\alpha \cup \beta]\varphi \leftrightarrow [\alpha]\varphi \wedge [\beta]\varphi$
4. $[\alpha; \beta]\varphi \leftrightarrow [\alpha][\beta]\varphi$
5. $[\psi?]\varphi \leftrightarrow (\psi \to \varphi)$
6. $\varphi \wedge [\alpha][\alpha^*]\varphi \leftrightarrow [\alpha^*]\varphi$
7. $\varphi \wedge [\alpha^*](\varphi \to [\alpha]\varphi) \to [\alpha^*]\varphi$
8. Modus Ponens and Necessitation (for each program α)

Propositional Dynamic Logic

1. Axioms of propositional logic
2. $[\alpha](\varphi \to \psi) \to ([\alpha]\varphi \to [\alpha]\psi)$
3. $[\alpha \cup \beta]\varphi \leftrightarrow [\alpha]\varphi \wedge [\beta]\varphi$
4. $[\alpha; \beta]\varphi \leftrightarrow [\alpha][\beta]\varphi$
5. $[\psi?]\varphi \leftrightarrow (\psi \to \varphi)$
6. $\varphi \wedge [\alpha][\alpha^*]\varphi \leftrightarrow [\alpha^*]\varphi$ (Fixed-Point Axiom)
7. $\varphi \wedge [\alpha^*](\varphi \to [\alpha]\varphi) \to [\alpha^*]\varphi$ (Induction Axiom)
8. Modus Ponens and Necessitation (for each program α)

Actions and Ability

An early approach to interpret PDL as logic of actions was put forward by Krister Segerberg.

Segerberg adds an "agency" program to the PDL language δA where A is a formula.

K. Segerberg. *Bringing it about*. JPL, 1989.

Actions and Agency

The intended meaning of the program 'δA' is that the agent "brings it about that A': *formally*, δA is the set of all paths p such that

Actions and Agency

The intended meaning of the program 'δA' is that the agent "brings it about that A': *formally*, δA is the set of all paths p such that

1. p is the computation according to some program α, and
2. α only terminates at states in which it is true that A

Actions and Agency

The intended meaning of the program 'δA' is that the agent "brings it about that A': *formally*, δA is the set of all paths p such that

1. p is the computation according to some program α, and
2. α only terminates at states in which it is true that A

Interestingly, Segerberg also briefly considers a third condition:

3. p is optimal (in some sense: shortest, maximally efficient, most convenient, etc.) in the set of computations satisfying conditions (1) and (2).

Actions and Agency

The intended meaning of the program 'δA' is that the agent "brings it about that A": *formally*, δA is the set of all paths p such that

1. p is the computation according to some program α, and
2. α only terminates at states in which it is true that A

Interestingly, Segerberg also briefly considers a third condition:

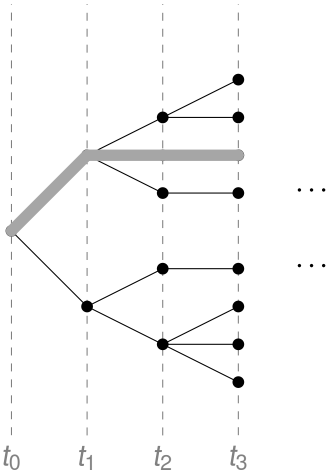
3. p is optimal (in some sense: shortest, maximally efficient, most convenient, etc.) in the set of computations satisfying conditions (1) and (2).

The axioms:

1. $[\delta A]A$
2. $[\delta A]B \to ([\delta B]C \to [\delta A]C)$

Actions and Agency in Branching Time

Alternative accounts of agency do not include explicit description of the actions:



STIT

- Each node represents a choice point for the agent.

STIT

- Each node represents a choice point for the agent.
- A **history** is a maximal branch in the above tree.

STIT

- Each node represents a choice point for the agent.
- A **history** is a maximal branch in the above tree.
- Formulas are interpreted at history moment pairs.

STIT

- Each node represents a choice point for the agent.
- A **history** is a maximal branch in the above tree.
- Formulas are interpreted at history moment pairs.
- At each moment there is a choice available to the agent (partition of the histories through that moment)

STIT

- Each node represents a choice point for the agent.
- A **history** is a maximal branch in the above tree.
- Formulas are interpreted at history moment pairs.
- At each moment there is a choice available to the agent (partition of the histories through that moment)
- The key modality is $[i\ \mathit{stit}]\varphi$ which is intended to mean that the agent *i* can "see to it that φ is true".
  - $[i\ \mathit{stit}]\varphi$ is true at a history moment pair provided the agent can choose a (set of) branch(es) such that every future history-moment pair satisfies φ

STIT

We use the modality '$\Diamond$' to mean historic possibility.

$\Diamond[i\ \mathit{stit}]\varphi$: "the agent has the ability to bring about φ".

STIT Model

A STIT models is $\mathcal{M} = \langle T, <, Choice, V\rangle$ where

STIT Model

A STIT models is $\mathcal{M} = \langle T, <, Choice, V\rangle$ where

- $\langle T, <\rangle$: T a set of moments, $<$ a tree-like ordering on T (irreflexive, transitive, linear-past)

STIT Model

A STIT models is $\mathcal{M} = \langle T, <, Choice, V \rangle$ where

- $\langle T, < \rangle$: T a set of moments, $<$ a tree-like ordering on T (irreflexive, transitive, linear-past)
- Let *Hist* be the set of all histories, and $H_t = \{h \in Hist \mid t \in h\}$ the histories through t.

STIT Model

A STIT models is $\mathcal{M} = \langle T, <, Choice, V\rangle$ where

- $\langle T, <\rangle$: T a set of moments, $<$ a tree-like ordering on T (irreflexive, transitive, linear-past)
- Let $Hist$ be the set of all histories, and $H_t = \{h \in Hist \mid t \in h\}$ the histories through t.
- $Choice : \mathcal{A} \times T \to \wp(\wp(H))$ is a function mapping each agent to a partition of H_t
  - $Choice_i^t \neq \emptyset$
  - $K \neq \emptyset$ for each $K \in Choice_i^t$
  - For all t and mappings $s_t : \mathcal{A} \to \wp(H_t)$ such that $s_t(i) \in Choice_i^t$, we have $\bigcap_{i \in \mathcal{A}} s_t(i) \neq \emptyset$

STIT Model

A STIT models is $\mathcal{M} = \langle T, <, Choice, V\rangle$ where

- $\langle T, <\rangle$: T a set of moments, $<$ a tree-like ordering on T (irreflexive, transitive, linear-past)
- Let *Hist* be the set of all histories, and $H_t = \{h \in Hist \mid t \in h\}$ the histories through t.
- $Choice : \mathcal{A} \times T \rightarrow \wp(\wp(H))$ is a function mapping each agent to a partition of H_t
  - $Choice_i^t \neq \emptyset$
  - $K \neq \emptyset$ for each $K \in Choice_i^t$
  - For all t and mappings $s_t : \mathcal{A} \rightarrow \wp(H_t)$ such that $s_t(i) \in Choice_i^t$, we have $\bigcap_{i\in\mathcal{A}} s_t(i) \neq \emptyset$
- V : At $\rightarrow \wp(T \times Hist)$ is a valuation function assigning to each atomic proposition

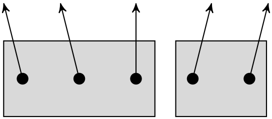
STIT Model

A STIT models is $\mathcal{M} = \langle T, <, Choice, V\rangle$ where

- $\langle T, <\rangle$: T a set of moments, $<$ a tree-like ordering on T (irreflexive, transitive, linear-past)
- Let $Hist$ be the set of all histories, and $H_t = \{h \in Hist \mid t \in h\}$ the histories through t.
- $Choice : \mathcal{A} \times T \to \wp(\wp(H))$ is a function mapping each agent to a partition of H_t
  - $Choice_i^t \neq \emptyset$
  - $K \neq \emptyset$ for each $K \in Choice_i^t$
  - For all t and mappings $s_t : \mathcal{A} \to \wp(H_t)$ such that $s_t(i) \in Choice_i^t$, we have $\bigcap_{i\in\mathcal{A}} s_t(i) \neq \emptyset$
- V : At $\to \wp(T \times Hist)$ is a valuation function assigning to each atomic proposition

Many Agents

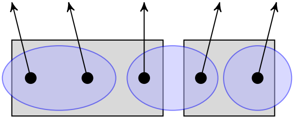
The previous model assumes there is *one* agent that "controls" the transition system.



Many Agents

The previous model assumes there is *one* agent that "controls" the transition system.

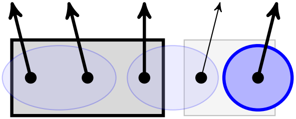
What if there is more than one agent?



Many Agents

The previous model assumes there is *one* agent that “controls” the transition system.

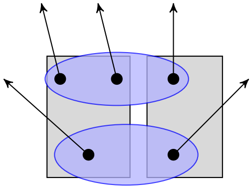
What if there is more than one agent? *Independence of agents*



Many Agents

The previous model assumes there is *one* agent that "controls" the transition system.

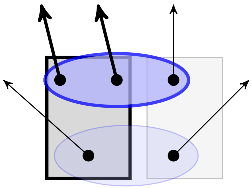
What if there is more than one agent? *Independence of agents*



Many Agents

The previous model assumes there is *one* agent that "controls" the transition system.

What if there is more than one agent? *Independence of agents*



STIT Language

$$\varphi \; = \; p \mid \neg\varphi \mid \varphi \wedge \psi \mid [i\ \mathit{stit}]\varphi \mid [i\ \mathit{dstit}:\ \varphi] \mid \Box\varphi$$

STIT Language

$$\varphi \;=\; p \mid \neg\varphi \mid \varphi \wedge \psi \mid [i\ stit]\varphi \mid [i\ dstit:\ \varphi] \mid \Box\varphi$$

- $\mathcal{M}, t/h \models p$ iff $t/h \in V(p)$

STIT Language

$$\varphi \;=\; p \mid \neg\varphi \mid \varphi \wedge \psi \mid [i\ stit]\varphi \mid [i\ dstit:\ \varphi] \mid \Box\varphi$$

- $\mathcal{M}, t/h \models p$ iff $t/h \in V(p)$
- $\mathcal{M}, t/h \models \neg\varphi$ iff $\mathcal{M}, t/h \not\models \varphi$

STIT Language

$$\varphi \ = \ p \mid \neg\varphi \mid \varphi \wedge \psi \mid [i\ stit]\varphi \mid [i\ dstit:\ \varphi] \mid \Box\varphi$$

- $\mathcal{M}, t/h \models p$ iff $t/h \in V(p)$
- $\mathcal{M}, t/h \models \neg\varphi$ iff $\mathcal{M}, t/h \not\models \varphi$
- $\mathcal{M}, t/h \models \varphi \wedge \psi$ iff $\mathcal{M}, t/h \models \varphi$ and $\mathcal{M}, t/h \models \psi$

STIT Language

$$\varphi \; = \; p \mid \neg\varphi \mid \varphi \wedge \psi \mid [i\ stit]\varphi \mid [i\ dstit:\ \varphi] \mid \Box\varphi$$

- $\mathcal{M}, t/h \models p$ iff $t/h \in V(p)$
- $\mathcal{M}, t/h \models \neg\varphi$ iff $\mathcal{M}, t/h \not\models \varphi$
- $\mathcal{M}, t/h \models \varphi \wedge \psi$ iff $\mathcal{M}, t/h \models \varphi$ and $\mathcal{M}, t/h \models \psi$
- $\mathcal{M}, t/h \models \Box\varphi$ iff $\mathcal{M}, t/h' \models \varphi$ for all $h' \in H_t$

STIT Language

$$\varphi \;=\; p \mid \neg\varphi \mid \varphi \wedge \psi \mid [i\ stit]\varphi \mid [i\ dstit:\ \varphi] \mid \Box\varphi$$

- $\mathcal{M}, t/h \models p$ iff $t/h \in V(p)$
- $\mathcal{M}, t/h \models \neg\varphi$ iff $\mathcal{M}, t/h \not\models \varphi$
- $\mathcal{M}, t/h \models \varphi \wedge \psi$ iff $\mathcal{M}, t/h \models \varphi$ and $\mathcal{M}, t/h \models \psi$
- $\mathcal{M}, t/h \models \Box\varphi$ iff $\mathcal{M}, t/h' \models \varphi$ for all $h' \in H_t$
- $\mathcal{M}, t/h \models [i\ stit]\varphi$ iff $\mathcal{M}, t/h' \models \varphi$ for all $h' \in Choice_i^t(h)$

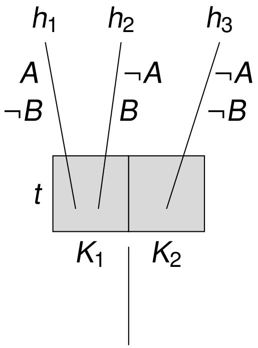
STIT Language

$$\varphi \;=\; p \mid \neg\varphi \mid \varphi \wedge \psi \mid [i\ stit]\varphi \mid [i\ dstit:\ \varphi] \mid \Box\varphi$$

- $\mathcal{M}, t/h \models p$ iff $t/h \in V(p)$
- $\mathcal{M}, t/h \models \neg\varphi$ iff $\mathcal{M}, t/h \not\models \varphi$
- $\mathcal{M}, t/h \models \varphi \wedge \psi$ iff $\mathcal{M}, t/h \models \varphi$ and $\mathcal{M}, t/h \models \psi$
- $\mathcal{M}, t/h \models \Box\varphi$ iff $\mathcal{M}, t/h' \models \varphi$ for all $h' \in H_t$
- $\mathcal{M}, t/h \models [i\ stit]\varphi$ iff $\mathcal{M}, t/h' \models \varphi$ for all $h' \in Choice_i^t(h)$
- $\mathcal{M}, t/h \models [i\ dstit]\varphi$ iff $\mathcal{M}, t/h' \models \varphi$ for all $h' \in Choice_i^t(h)$ and there is a $h'' \in H_t$ such that $\mathcal{M}, t/h \models \neg\varphi$

STIT: Example

The following are false: $A \to \Diamond[stit]A$ and
$\Diamond[stit](A \vee B) \to \Diamond[stit]A \vee \Diamond[stit]B$.



J. Horty. *Agency and Deontic Logic*. 2001.

STIT: Axiomatics

- **S5** for $\Box$: $\Box(\varphi \to \psi) \to (\Box\varphi \to \Box\psi)$, $\Box\varphi \to \varphi$, $\Box\varphi \to \Box\Box\varphi$, $\neg\Box\varphi \to \Box\neg\Box\varphi$

STIT: Axiomatics

- **S5** for $\Box$: $\Box(\varphi \rightarrow \psi) \rightarrow (\Box\varphi \rightarrow \Box\psi)$, $\Box\varphi \rightarrow \varphi$, $\Box\varphi \rightarrow \Box\Box\varphi$, $\neg\Box\varphi \rightarrow \Box\neg\Box\varphi$
- **S5** for [*i stit*]: $[i\ \mathit{stit}](\varphi \rightarrow \psi) \rightarrow ([i\ \mathit{stit}]\varphi \rightarrow [i\ \mathit{stit}]\psi)$, $[i\ \mathit{stit}]\varphi \rightarrow \varphi$, $[i\ \mathit{stit}]\varphi \rightarrow [i\ \mathit{stit}][i\ \mathit{stit}]\varphi$, $\neg[i\ \mathit{stit}]\varphi \rightarrow [i\ \mathit{stit}]\neg[i\ \mathit{stit}]\varphi$

STIT: Axiomatics

- **S5** for $\Box$: $\Box(\varphi \to \psi) \to (\Box\varphi \to \Box\psi)$, $\Box\varphi \to \varphi$, $\Box\varphi \to \Box\Box\varphi$, $\neg\Box\varphi \to \Box\neg\Box\varphi$
- **S5** for $[i\ stit]$: $[i\ stit](\varphi \to \psi) \to ([i\ stit]\varphi \to [i\ stit]\psi)$, $[i\ stit]\varphi \to \varphi$, $[i\ stit]\varphi \to [i\ stit][i\ stit]\varphi$, $\neg[i\ stit]\varphi \to [i\ stit]\neg[i\ stit]\varphi$
- $\Box\varphi \to [i\ stit]\varphi$

STIT: Axiomatics

- **S5** for $\Box$: $\Box(\varphi \to \psi) \to (\Box\varphi \to \Box\psi)$, $\Box\varphi \to \varphi$, $\Box\varphi \to \Box\Box\varphi$, $\neg\Box\varphi \to \Box\neg\Box\varphi$
- **S5** for $[i\ \mathit{stit}]$: $[i\ \mathit{stit}](\varphi \to \psi) \to ([i\ \mathit{stit}]\varphi \to [i\ \mathit{stit}]\psi)$, $[i\ \mathit{stit}]\varphi \to \varphi$, $[i\ \mathit{stit}]\varphi \to [i\ \mathit{stit}][i\ \mathit{stit}]\varphi$, $\neg[i\ \mathit{stit}]\varphi \to [i\ \mathit{stit}]\neg[i\ \mathit{stit}]\varphi$
- $\Box\varphi \to [i\ \mathit{stit}]\varphi$
- $(\bigwedge_{i\in\mathcal{A}} \Diamond[i\ \mathit{stit}]\varphi_i) \to \Diamond(\bigwedge_{i\in\mathcal{A}}[i\ \mathit{stit}]\varphi_i)$

STIT: Axiomatics

- **S5** for $\Box$: $\Box(\varphi \to \psi) \to (\Box\varphi \to \Box\psi)$, $\Box\varphi \to \varphi$, $\Box\varphi \to \Box\Box\varphi$, $\neg\Box\varphi \to \Box\neg\Box\varphi$
- **S5** for $[i\ stit]$: $[i\ stit](\varphi \to \psi) \to ([i\ stit]\varphi \to [i\ stit]\psi)$, $[i\ stit]\varphi \to \varphi$, $[i\ stit]\varphi \to [i\ stit][i\ stit]\varphi$, $\neg[i\ stit]\varphi \to [i\ stit]\neg[i\ stit]\varphi$
- $\Box\varphi \to [i\ stit]\varphi$
- $(\bigwedge_{i\in\mathcal{A}} \Diamond[i\ stit]\varphi_i) \to \Diamond(\bigwedge_{i\in\mathcal{A}}[i\ stit]\varphi_i)$
- Modus Ponens and Necessitation for $\Box$

M. Xu. *Axioms for deliberative STIT*. Journal of Philosophical Logic, Volume 27, pp. 505 - 552, 1998.

P. Balbiani, A. Herzig and N. Troquard. *Alternative axiomatics and complexity of deliberative STIT theories*. Journal of Philosophical Logic, 37:4, pp. 387 - 406, 2008.

Recap: Logics of Action and Ability

- $F\varphi$: φ is true at some moment in *the* future

Recap: Logics of Action and Ability

- $F\varphi$: φ is true at some moment in *the* future
- $\exists F\varphi$: there is a history where φ is true some moment in the future

Recap: Logics of Action and Ability

- $F\varphi$: φ is true at some moment in *the* future
- $\exists F\varphi$: there is a history where φ is true some moment in the future
- $[\alpha]\varphi$: after doing action α, φ is true

Recap: Logics of Action and Ability

- $F\varphi$: φ is true at some moment in *the* future
- $\exists F\varphi$: there is a history where φ is true some moment in the future
- $[\alpha]\varphi$: after doing action α, φ is true
- $[\delta\varphi]\psi$: after bringing about φ, ψ is true

Recap: Logics of Action and Ability

- $F\varphi$: φ is true at some moment in *the* future
- $\exists F\varphi$: there is a history where φ is true some moment in the future
- $[\alpha]\varphi$: after doing action α, φ is true
- $[\delta\varphi]\psi$: after bringing about φ, ψ is true
- $[i\ stit]\varphi$: the agent can "see to it that" φ is true

Recap: Logics of Action and Ability

- $F\varphi$: φ is true at some moment in *the* future
- $\exists F\varphi$: there is a history where φ is true some moment in the future
- $[\alpha]\varphi$: after doing action α, φ is true
- $[\delta\varphi]\psi$: after bringing about φ, ψ is true
- $[i\ \mathit{stit}]\varphi$: the agent can "see to it that" φ is true
- $\Diamond[i\ \mathit{stit}]\varphi$: the agent has the ability to bring about φ

Epistemizing logics of action and ability

Knowledge, action, abilities

A. Herzig. *Logics of knowledge and action: critical analysis and challenges*. Autonomous Agent and Multi-Agent Systems, 2014.

J. Broeresen, A. Herzig and N. Troquard. *What groups do, can do and know they can do: An analysis in normal modal logics*. Journal of Applied and Non-Classical Logics, 19:3, pgs. 261 - 289, 2009.

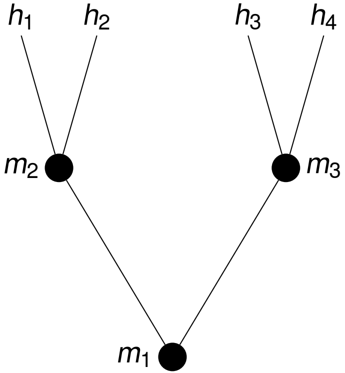
W. van der Hoek and M. Wooldridge. *Cooperation, knowledge and time: Alternating-time temporal epistemic logic and its applications*. Studia Logica, 75, pgs. 125 - 157, 2003.

Epistemic stit model

$$\langle \mathit{Tree}, <, \mathit{Agent}, \mathit{Choice}, \{\sim_\alpha\}_{\alpha \in \mathit{Agent}}, V \rangle$$

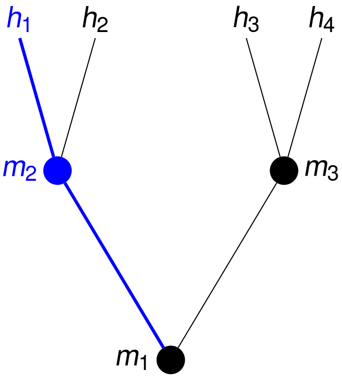
Epistemic stit model

$$\langle Tree, <, Agent, Choice, \{\sim_\alpha\}_{\alpha \in Agent}, V\rangle$$



Epistemic stit model

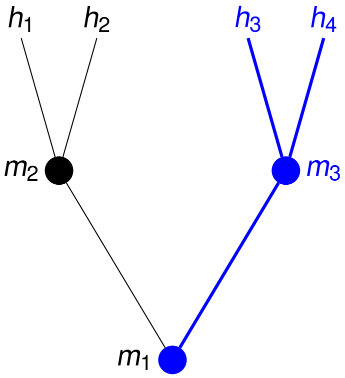
$$\langle Tree, <, Agent, Choice, \{\sim_\alpha\}_{\alpha \in Agent}, V\rangle$$



m/h denotes (m, h) with $m \in h$ is called an **index**

Epistemic stit model

$$\langle Tree, <, Agent, Choice, \{\sim_\alpha\}_{\alpha \in Agent}, V\rangle$$

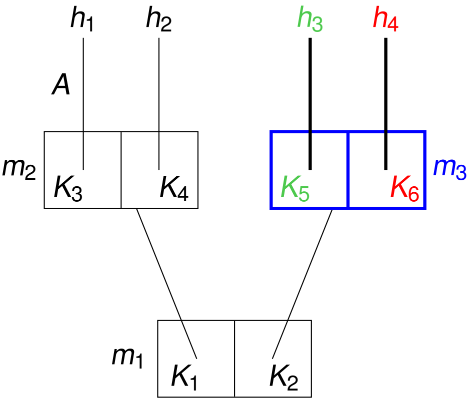


m/h denotes (m, h) with $m \in h$ is called an **index**

$H^m = \{h \mid m \in h\}$

Epistemic stit model

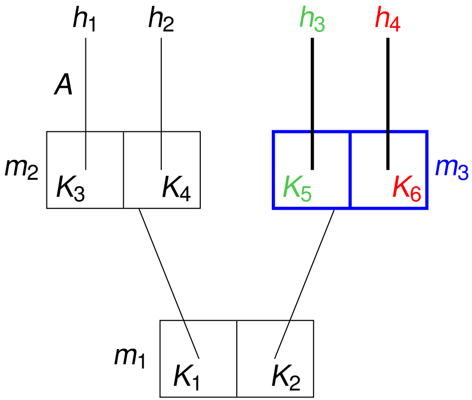
$$\langle Tree, <, Agent, Choice, \{\sim_\alpha\}_{\alpha \in Agent}, V\rangle$$



For $\alpha \in Agent$, $Choice^m_\alpha$ is a partition on H^m

Epistemic stit model

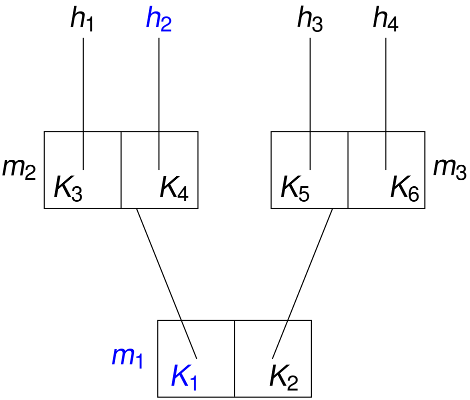
$$\langle Tree, <, Agent, Choice, \{\sim_\alpha\}_{\alpha \in Agent}, V\rangle$$



For $\alpha \in Agent$, $Choice^m_\alpha$ is a partition on H^m

Epistemic stit model

$$\langle Tree, <, Agent, Choice, \{\sim_\alpha\}_{\alpha \in Agent}, V\rangle$$

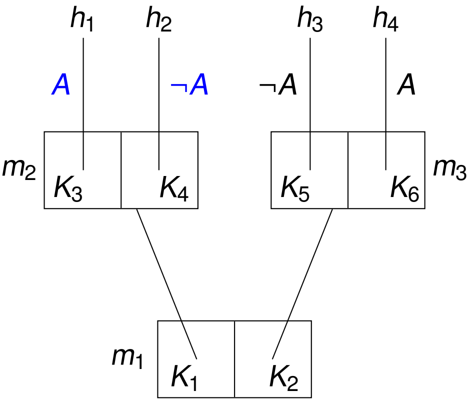


For $\alpha \in Agent$, $Choice^m_\alpha$ is a partition on H^m

$Choice^m_\alpha(h)$ is the particular action at m that contains h

Epistemic stit model

$$\langle \mathit{Tree}, <, \mathit{Agent}, \mathit{Choice}, \{\sim_\alpha\}_{\alpha \in \mathit{Agent}}, V \rangle$$

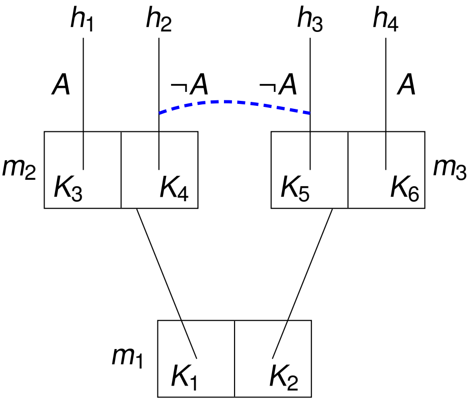


V assigns sets of indices to atomic propositions.

$m_2/h_1 \models A$ $\quad$ $m_2/h_2 \not\models A$

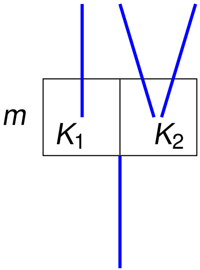
Epistemic stit model

$$\langle Tree, <, Agent, Choice, \{\sim_\alpha\}_{\alpha \in Agent}, V\rangle$$

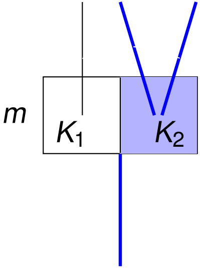


$\sim_\alpha$ is an (equivalence) relation on indices

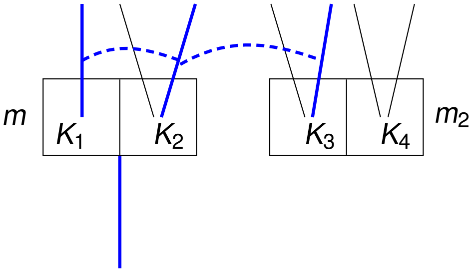
$m/h \sim_\alpha m'/h'$: everything α knows at m/h is true at m'/h', α cannot distinguish m/h and m'/h', ...



- $\mathcal{M}, m/h \models \Box A$ if and only if $\mathcal{M}, m/h' \models A$ for all $h' \in H^m$,



- $\mathcal{M}, m/h \models \Box A$ if and only if $\mathcal{M}, m/h' \models A$,
- $\mathcal{M}, m/h \models [\alpha\ stit\colon A]$ if and only if $Choice^m_\alpha(h) \subseteq |A|^m_{\mathcal{M}}$,



- $\mathcal{M}, m/h \models \Box A$ if and only if $\mathcal{M}, m/h' \models A$,
- $\mathcal{M}, m/h \models [\alpha\ stit\colon A]$ if and only if $Choice^m_\alpha(h) \subseteq |A|^m_{\mathcal{M}}$,
- $\mathcal{M}, m/h \models \mathsf{K}_\alpha A$ if and only if, for all m'/h', if $m/h \sim_\alpha m'/h'$, then $\mathcal{M}, m'/h' \models A$

Action labels

Let *Type* $= \{\tau_1, \tau_2, \ldots, \tau_n\}$ be a set of action types—general kinds of action, as opposed to the concrete action tokens.

An action type τ is interpreted as a partial function mapping each agent α and moment m into the particular action token $[\tau]^m_\alpha$ that results when τ is executed by α at m (so, $[\tau]^m_\alpha \in Choice^m_\alpha$)

Labeled stit frames

$$\langle \mathit{Tree}, <, \mathit{Agent}, \mathit{Choice}, \{\sim_\alpha\}_{\alpha \in \mathit{Agent}}, \mathit{Type}, \mathit{Label}, V\rangle,$$

Label maps each action token $K \in \mathit{Choice}^m_\alpha$ to a particular action type $\mathit{Label}(K) \in \mathit{Type}$.

1. If $K \in \mathit{Choice}^m_\alpha$, then $[\mathit{Label}(K)]^m_\alpha = K$,
2. If $\tau \in \mathit{Type}$ and $[\tau]^m_\alpha$ is defined, then $\mathit{Label}([\tau]^m_\alpha) = \tau$.

Labeled stit frames

$$\langle \mathit{Tree}, <, \mathit{Agent}, \mathit{Choice}, \{\sim_\alpha\}_{\alpha\in \mathit{Agent}}, \mathit{Type}, \mathit{Label}, V\rangle,$$

Label maps each action token $K \in \mathit{Choice}^m_\alpha$ to a particular action type $\mathit{Label}(K) \in \mathit{Type}$.

1. If $K \in \mathit{Choice}^m_\alpha$, then $[\mathit{Label}(K)]^m_\alpha = K$,
2. If $\tau \in \mathit{Type}$ and $[\tau]^m_\alpha$ is defined, then $\mathit{Label}([\tau]^m_\alpha) = \tau$.

$$\mathit{Type}^m_\alpha = \{\mathit{Label}(K) : K \in \mathit{Choice}^m_\alpha\}$$

$$\mathit{Type}^m_\alpha(h) = \mathit{Label}(\mathit{Choice}^m_\alpha(h))$$

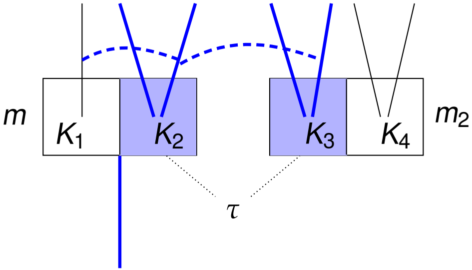
Frame properties

- If $m/h \sim_\alpha m'/h'$, then $m/h'' \sim_\alpha m'/h'''$ for each $h'' \in H^m$ and $h''' \in H^{m'}$.
- For all m/h, $Know_\alpha(m/h) \subseteq H^m$.
- If $m/h \sim_\alpha m'/h'$, then $Type_\alpha^m = Type_\alpha^{m'}$.
- If $m/h \sim_\alpha m'/h'$, then $Type_\alpha^m(h) = Type_\alpha^{m'}(h')$.

kstit

- $\mathcal{M}, m/h \models [\alpha\ kstit\colon A]$ if and only if $[Type^{m}_{\alpha}(h)]^{m'}_{\alpha} \subseteq |A|^{m'}_{\mathcal{M}}$ for all m'/h' such that $m'/h' \sim_{\alpha} m/h$.

kstit

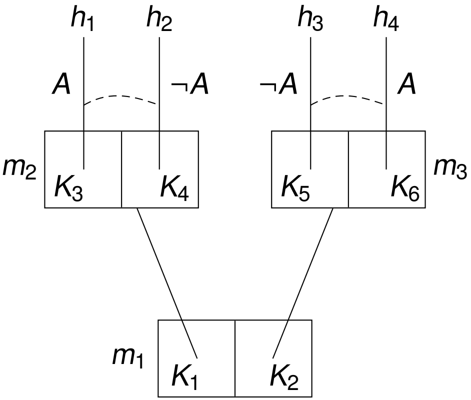


- $\mathcal{M}, m/h \models [\alpha\ kstit\colon A]$ if and only if $[Type^m_\alpha(h)]^{m'}_\alpha \subseteq |A|^{m'}_{\mathcal{M}}$ for all m'/h' such that $m'/h' \sim_\alpha m/h$.

Causal vs. epistemic ability

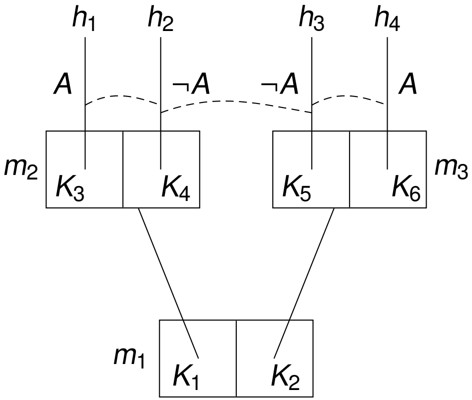
$$\Diamond[\alpha \; \textit{stit}\!: A]$$

Causal vs. epistemic ability



$\Diamond[\alpha$ *stit*: $A]$ is settled true at m_2

Causal vs. epistemic ability



$\Diamond[\alpha$ *stit*: $A]$ is settled true at m_2

Causal vs. epistemic ability

$$\Diamond[\alpha\ \mathit{stit}\colon A]$$

$$\mathsf{K}_\alpha\Diamond[\alpha\ \mathit{stit}\colon A]$$

$$\Diamond\mathsf{K}_\alpha[\alpha\ \mathit{stit}\colon A]$$

$$\Diamond[\alpha\ \mathit{kstit}\colon A]$$

Ex ante vs. *ex interim* knowledge

- $\mathcal{M}, m/h \models \mathsf{K}_\alpha A$ if and only if, for all m'/h', if $m/h \sim_\alpha m'/h'$, then $\mathcal{M}, m'/h' \models A$
- $\mathcal{M}, m/h \models \mathsf{K}^{\mathrm{act}}_\alpha A$ if and only if, for all m'/h', if $m/h \sim_\alpha m'/h'$ and $h' \in [\mathit{Type}^m_\alpha(h)]^{m'}_\alpha$, $\mathcal{M}, m'/h' \models A$

Discussion

- Language/validities

  $\Box A \supset [\alpha\ \mathit{stit}\text{: } A]$
  $\mathsf{K}_\alpha \Box A \supset [\alpha\ \mathit{kstit}\text{: } A]$
  $[\alpha\ \mathit{kstit}\text{: } A] \equiv \mathsf{K}_\alpha^{\mathsf{act}}[\alpha\ \mathit{stit}\text{: } A]$
  $\ldots$

- What do the agents know vs. What do the agents know *given what they are doing*.

- Equivalence between labeled stit models (cf. Thompson transformations specifying when two imperfect information games reduce to the same Normal form)