Logic and Artificial Intelligence

Lecture 13

Eric Pacuit

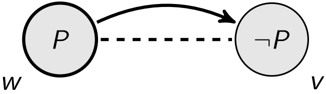
Currently Visiting the Center for Formal Epistemology, CMU

Center for Logic and Philosophy of Science
Tilburg University

ai.stanford.edu/∼epacuit
e.j.pacuit@uvt.nl

October 13, 2011

Epistemic Plausibility Models



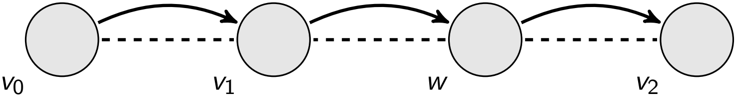
Epistemic-Plausibility Model: $\mathcal{M} = \langle W, \{\sim_i\}_{i \in \mathcal{A}}, \{\preceq_i\}_{i \in \mathcal{A}}, V \rangle$

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

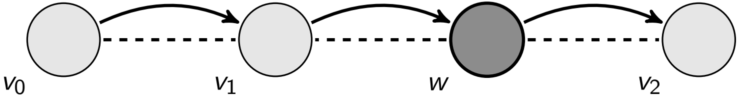
Truth:

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

Grades of Doxastic Strength

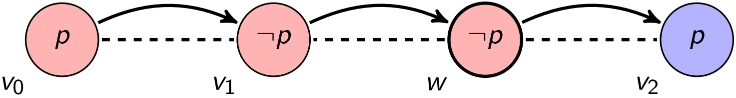


Grades of Doxastic Strength



Suppose that w is the current state.

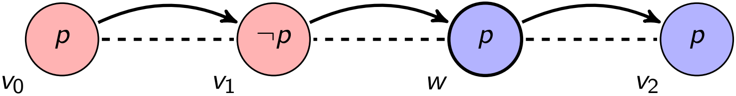
Grades of Doxastic Strength



Suppose that w is the current state.

- **Belief** (Bp)

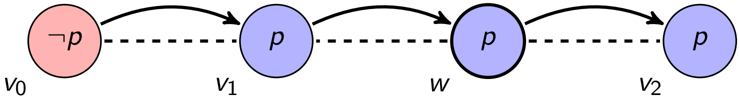
Grades of Doxastic Strength



Suppose that w is the current state.

- **Belief** (Bp)
- **Robust Belief** ($[\preceq]p$)

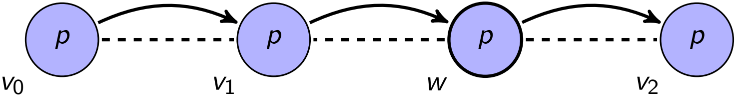
Grades of Doxastic Strength



Suppose that w is the current state.

- **Belief** (Bp)
- **Robust Belief** ($[\preceq]p$)
- **Strong Belief** ($B^s p$)

Grades of Doxastic Strength



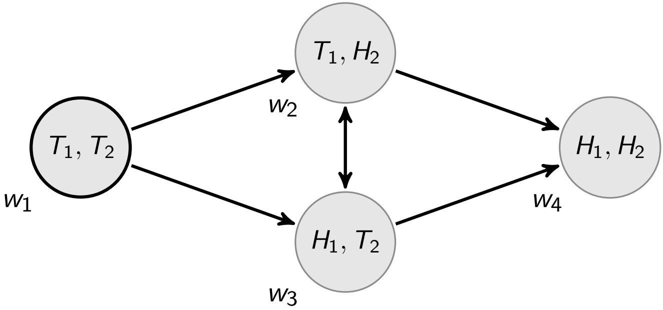
Suppose that w is the current state.

- **Belief** (Bp)
- **Robust Belief** ($[\preceq]p$)
- **Strong Belief** ($B^s p$)
- **Knowledge** (Kp)

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

Hard and Soft Updates

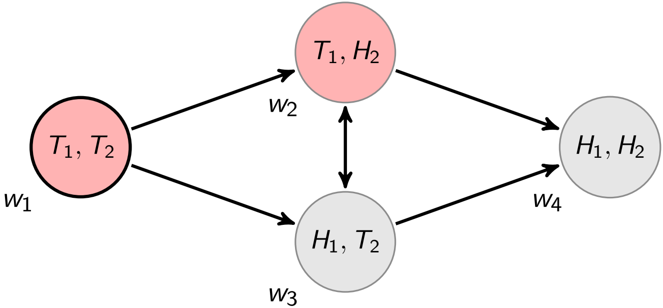
$$\begin{tikzcd}
\mathcal{M} = \langle W, \{\sim_i\}_{i \in \mathcal{A}}, \{\preceq_i\}_{i \in \mathcal{A}}, V \rangle \arrow[d, Rightarrow, "\textbf{Find out that } \varphi"] \\
\mathcal{M} = \langle W', \{\sim'_i\}_{i \in \mathcal{A}}, \{\preceq'_i\}_{i \in \mathcal{A}}, V|_{W'} \rangle
\end{tikzcd}$$



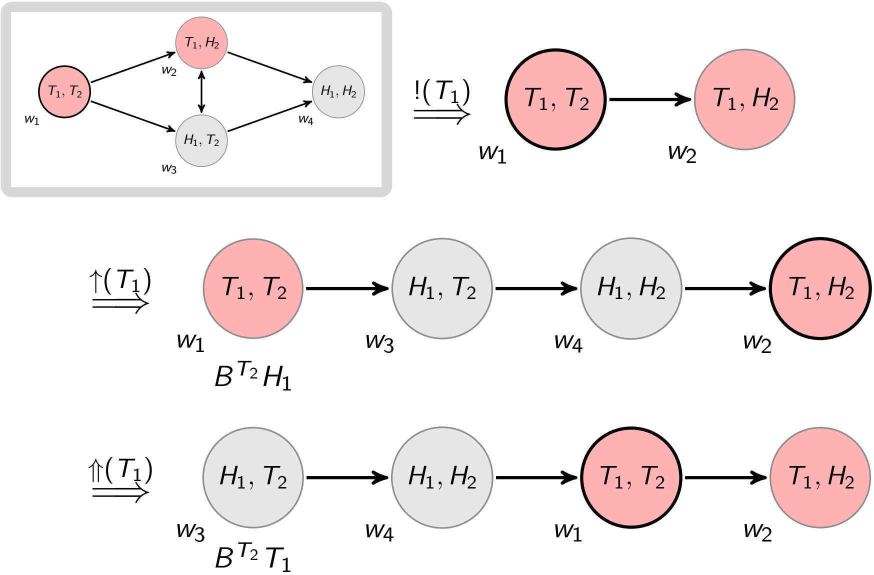
$Min_{\preceq}([w_1]) = \{w_4\}$, so $w_1 \models B(H_1 \wedge H_2)$

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

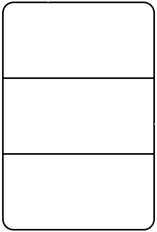
$Min_{\preceq}([w_1] \cap [\![T_1]\!]_{\mathcal{M}}) = \{w_3\}$, so $w_1 \models B^{T_2} H_1$



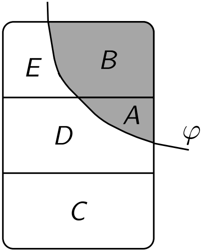
Suppose the agent *finds out that* T_1 *is/may be true*.


T_1, T_2
w_1
T_1, H_2
w_2
H_1, T_2
w_3
H_1, H_2
w_4
$\stackrel{!(T_1)}{\Longrightarrow}$
T_1, T_2
w_1
T_1, H_2
w_2
$\stackrel{\uparrow(T_1)}{\Longrightarrow}$
T_1, T_2
w_1
$B^{T_2} H_1$
H_1, T_2
w_3
H_1, H_2
w_4
T_1, H_2
w_2
$\stackrel{\Uparrow(T_1)}{\Longrightarrow}$
H_1, T_2
w_3
$B^{T_2} T_1$
H_1, H_2
w_4
T_1, T_2
w_1
T_1, H_2
w_2

Informative Actions

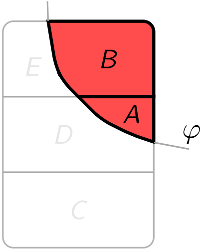


Informative Actions



Incorporate the new information φ

Informative Actions

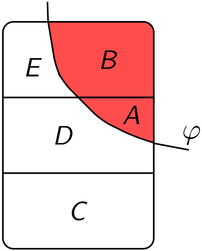


Public Announcement: Information from an infallible source

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

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

Informative Actions

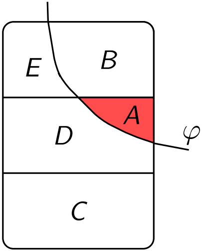


Radical Upgrade: $(\Uparrow\varphi)$: $A \prec_i B \prec_i C \prec_i D \prec_i E$,
$\mathcal{M}^{\Uparrow\varphi} = \langle W, \{\sim_i\}_{i\in\mathcal{A}}, \{\preceq_i^{\Uparrow\varphi}\}_{i\in\mathcal{A}}, V\rangle$

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 $x \preceq_i y$.

Reduction Axioms

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

Reduction Axioms

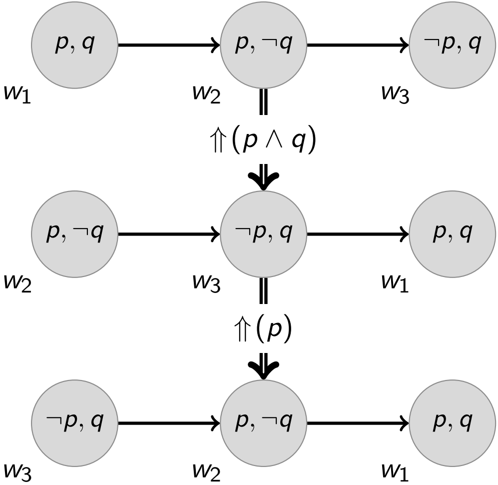
$$[\Uparrow\varphi]B^{\psi}\chi \leftrightarrow (L(\varphi \wedge [\Uparrow\varphi]\psi) \wedge B^{\varphi\wedge[\Uparrow\varphi]\psi}[\Uparrow\varphi]\chi)\vee$$
$$(\neg L(\varphi \wedge [\Uparrow\varphi]\psi) \wedge B^{[\Uparrow\varphi]\psi}[\Uparrow\varphi]\chi)$$

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

Composition

$$[!\varphi][!\psi]\chi \leftrightarrow [!(\varphi \wedge [!\varphi]\psi)]\chi$$

Composition



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

$$\begin{tikzcd}
\mathcal{M}_0 \arrow[r, Rightarrow, "!\varphi_1"] & \mathcal{M}_1 \arrow[r, Rightarrow, "!\varphi_2"] & \mathcal{M}_2 \arrow[r, Rightarrow, "!\varphi_3"] & \cdots \arrow[r, Rightarrow, "!\varphi_n"] & \mathcal{M}_f \\
\text{initial model} & & & & \text{fixed-point}
\end{tikzcd}$$

$$\begin{tikzcd}
\mathcal{M}_0 \arrow[r, Rightarrow, "\Uparrow\varphi_1"] & \mathcal{M}_1 \arrow[r, Rightarrow, "\Uparrow\varphi_2"] & \mathcal{M}_2 \arrow[r, Rightarrow, "\Uparrow\varphi_3"] & \cdots \arrow[r, Rightarrow, "\Uparrow\varphi_n"] & \mathcal{M}_f \\
\text{initial model} & & & & \text{fixed-point}
\end{tikzcd}$$

$$\begin{tikzcd}
\mathcal{M}_0 \arrow[r, Rightarrow, "!\varphi_1"] & \mathcal{M}_1 \arrow[r, Rightarrow, "\Uparrow\varphi_2"] & \mathcal{M}_2 \arrow[r, Rightarrow, "\uparrow\varphi_3"] & \cdots \arrow[r, Rightarrow, "\Uparrow\varphi_n"] & \mathcal{M}_f \\
\text{initial model} & & & & \text{fixed-point}
\end{tikzcd}$$

$$\begin{tikzcd}
\mathcal{M}_0 \arrow[r, Rightarrow, "\tau(\varphi_1)"] & \mathcal{M}_1 \arrow[r, Rightarrow, "\tau(\varphi_2)"] & \mathcal{M}_2 \arrow[r, Rightarrow, "\tau(\varphi_3)"] & \cdots \arrow[r, Rightarrow, "\tau(\varphi_n)"] & \mathcal{M}_f \\
\text{initial model} & & & & \text{fixed-point}
\end{tikzcd}$$

Where do the φ_k come from?

Iterated Updates

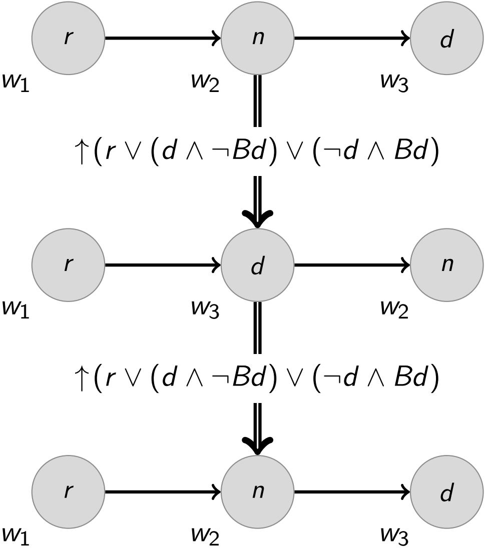
$!\varphi_1, !\varphi_2, !\varphi_3, \ldots, !\varphi_n$
always reaches a fixed-point

$\Uparrow p \Uparrow \neg p \Uparrow p \cdots$
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

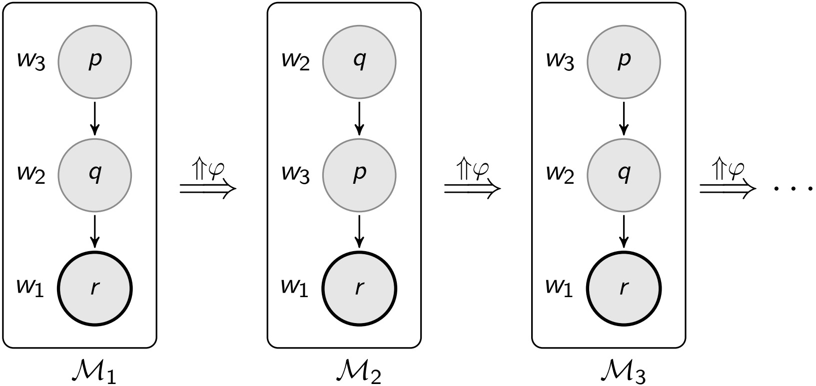
A. Baltag and S. Smets. *Group Belief Dynamics under Iterated Revision: Fixed Points and Cycles of Joint Upgrades*. TARK, 2009.



$$\Uparrow(r \vee (d \wedge \neg Bd) \vee (\neg d \wedge Bd))$$

$$\Uparrow(r \vee (d \wedge \neg Bd) \vee (\neg d \wedge Bd))$$

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



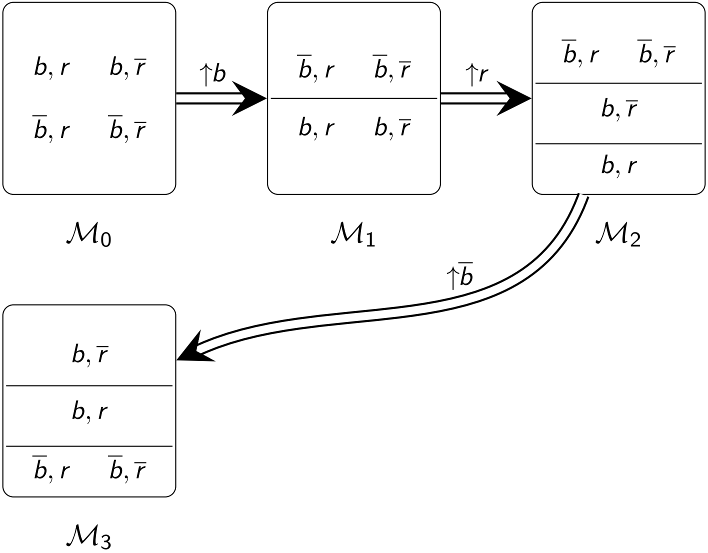
Suppose that you are in the forest and happen to a see strange-looking animal.

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.

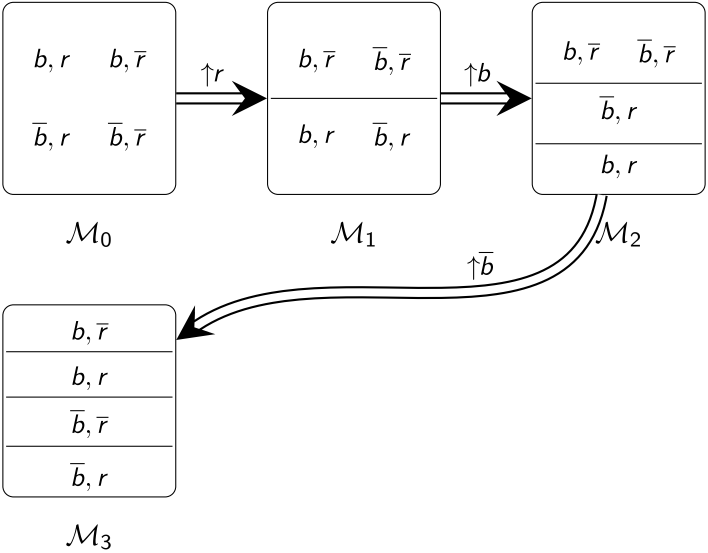
Suppose that you are in the forest and happen to a see strange-looking animal. You consult your animal guidebook and find a picture that seems to match the animal you see. The guidebook says that the animal is a type of bird, so that is what you conclude: The animal before you is a bird. After looking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

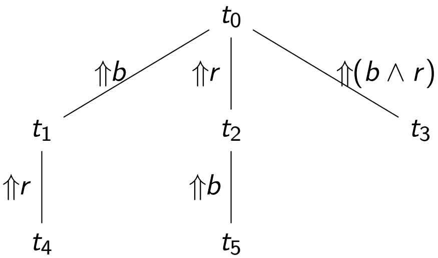

b, r
$b, \overline{r}$
$\overline{b}, r$
$\overline{b}, \overline{r}$
$\mathcal{M}_0$
$\uparrow b$
$\overline{b}, r$
$\overline{b}, \overline{r}$
b, r
$b, \overline{r}$
$\mathcal{M}_1$
$\uparrow r$
$\overline{b}, r$
$\overline{b}, \overline{r}$
$b, \overline{r}$
b, r
$\mathcal{M}_2$
$\uparrow \overline{b}$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

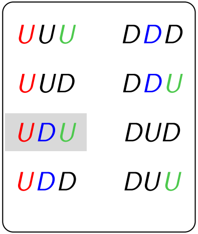
Note that in the last model, $\mathcal{M}_3$, the agent does not believe that the bird is red. The problem is that there does not seem to be any justification for why the agent drops her belief that the bird is red. This seems to result from the accidental fact that the agent started by updating with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




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

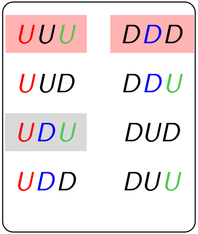
UUU	DDD
UUD	DDU
UDU	DUD
UDD	DUU

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



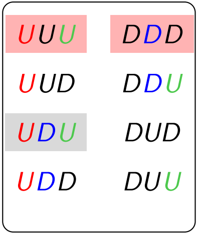
UUU	DDD
UUD	DDU
UDU	DUD
UDD	DUU

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



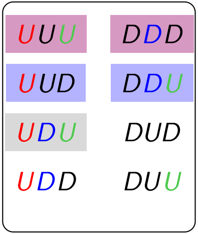
UUU	DDD
UUD	DDU
UDU	DUD
UDD	DUU

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



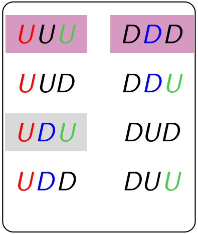
UUU	DDD
UUD	DDU
UDU	DUD
UDD	DUU

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



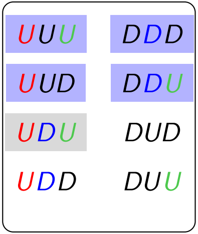
UUU	DDD
UUD	DDU
UDU	DUD
UDD	DUU

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



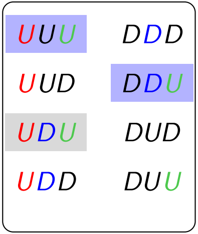
UUU	DDD
UUD	DDU
UDU	DUD
UDD	DUU

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



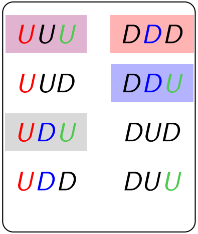
UUU	DDD
UUD	DDU
UDU	DUD
UDD	DUU

- Suppose there are two switches: L_1 is the main switch and L_2 is a secondary switch controlled by the first two lights. (So $L_1 \to L_2$, 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.



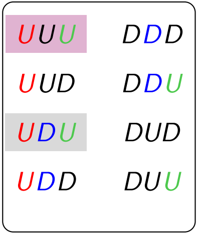
UUU	DDD
UUD	DDU
UDU	DUD
UDD	DUU

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



UUU	DDD
UUD	DDU
UDU	DUD
UDD	DUU

- Suppose there are two switches: L_1 is the main switch and L_2 is a secondary switch controlled by the first two lights. (So $L_1 \to L_2$, 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.



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.

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

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

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

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

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

Discussion

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

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

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