Reasoning about Knowledge and Beliefs

Lecture 15

Eric Pacuit

University of Maryland, College Park

pacuit.org
epacuit@umd.edu

December 10, 2013

Let K be a belief set and φ a formula.

$K \perp \varphi$ is the **remainder set** of K.

$A \in K \perp \varphi$ iff

1. $A \subseteq K$
2. $\varphi \notin Cn(A)$
3. There is no B such that $A \subset B \subseteq K$ and $\varphi \notin Cn(B)$.

- $K \perp \alpha = \{K\}$ iff $\alpha \notin Cn(K)$
- $K \perp \alpha = \emptyset$ iff $\alpha \in Cn(\emptyset)$
- If $K' \subseteq K$ and $\alpha \notin Cn(K')$ then there is some T such that $K' \subseteq T \in K \perp \alpha$.

A **selection function** γ for K is a function on $K \perp \alpha$ such that:

- If $K \perp \alpha \neq \emptyset$, then $\gamma(K \perp \alpha) \subseteq K \perp \alpha$ and $\gamma(K \perp \alpha) \neq \emptyset$
- If $K \perp \alpha = \emptyset$, then $\gamma(K \perp \alpha) = \{K\}$

Let K be a set of formulae. A function $-$ on $\mathcal{L}$ is a **partial meet contraction** for K if there is a selection function γ for K such that for all formula α:

$$K - \alpha = \bigcap \gamma(K \perp \alpha)$$

Let K be a set of formulae. A function $-$ on $\mathcal{L}$ is a **partial meet contraction** for K if there is a selection function γ for K such that for all formula α:

$$K - \alpha = \bigcap \gamma(K \perp \alpha)$$

Then $K * \alpha = \bigcap \gamma(K \perp \neg\alpha) \cup \{\alpha\}$

Let K be a set of formulae. A function $-$ on $\mathcal{L}$ is a **partial meet contraction** for K if there is a selection function γ for K such that for all formula α:

$$K - \alpha = \bigcap \gamma(K \perp \alpha)$$

Then $K * \alpha = \bigcap \gamma(K \perp \neg\alpha) \cup \{\alpha\}$

- γ selects exactly one element of $K \perp \alpha$ (maxichoice contraction)
- γ selects the entire set $K \perp \alpha$ (full meet contraction)

AGM Postulates

AGM 1: $K * \varphi$ is deductively closed

AGM 2: $\varphi \in K * \varphi$

AGM 3: $K * \varphi \subseteq Cn(K \cup \{\varphi\})$

AGM 4: If $\neg\varphi \notin K$ then $K * \varphi = Cn(K \cup \{\varphi\})$

AGM 5: $K * \varphi$ is inconsistent only if φ is inconsistent

AGM 6: If φ and ψ are logically equivalent then $K * \varphi = K * \psi$

AGM 7: $K * (\varphi \wedge \psi) \subseteq Cn(K * \varphi \cup \{\psi\})$

AGM 8: if $\neg\psi \notin K * \varphi$ then $Cn(K * \varphi \cup \{\psi\}) \subseteq K * (\varphi \wedge \psi)$

Theorem (AGM 1985). Let K be a belief set and let $*$ be a function on $\mathcal{L}$. Then

- The function $*$ is a partial meet revision for K if and only if it satisfies the postulates $AGM1$ - $AGM6$
- The function $*$ is a transitively relational partial meet revision for K if and only if it satisfies $AGM1$ - $AGM8$.

N. Tennant. *On the Degeneracy of the Full AGM-Theory of Theory-Revision*. Journal of Symbolic Logic, 71:2, pgs. 661 - 676, 2006.

D. Osherson. *Note on an observation by Neil Tennant*. 2005.

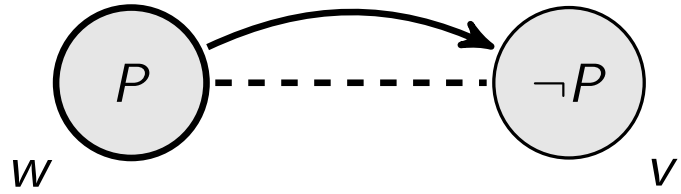
Theorem (Tennant). Let K be a belief set and φ any formula such that $K \models \varphi$. Let T be any satisfiable theory that implies $\neg\varphi$. Then for some $\Gamma \subseteq K\perp\varphi$, T is logically equivalent to $\bigcap \Gamma \cup \{\neg\varphi\}$.

Let R be a transitive relation over $\wp(K)$. and $\Gamma_{R,\varphi}$ defined as follows:
For all φ, $\Gamma_{R,\varphi} = \{\gamma \in K\bot\varphi \mid \forall\tau \in K\bot\varphi, \tau R\gamma\}$

Given K and T, φ is a K, T-disagreement just in case $K \models \varphi$ and $T \models \neg\varphi$.

Theorem (Tennant). Let K be a belief set and T a satisfiable theory. Then there is a transitive relation R over $\wp(K)$ such that for all K, T-disagreements φ, T is logically equivalent to $\bigcap \Gamma_{R,\varphi} \cup \{\neg\varphi\}$.

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 $\mathcal{M}, v \models \varphi$

Hard and Soft Updates

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

$\Downarrow$

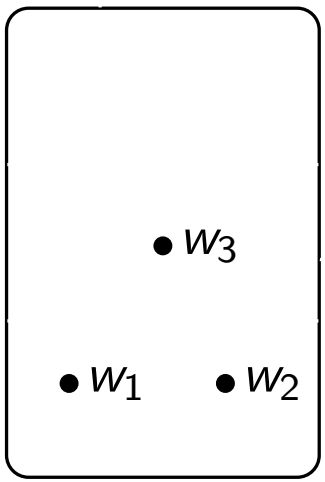
Find out that φ

$\Downarrow$

$$\mathcal{M} = \langle W', \{\sim'_i\}_{i\in\mathcal{A}}, \{\preceq'_i\}_{i\in\mathcal{A}}, V|_{W'}\rangle$$

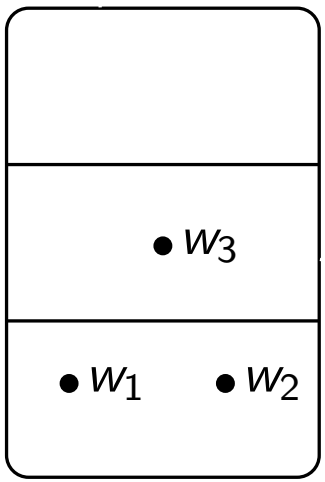
Belief Revision via Plausibility

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



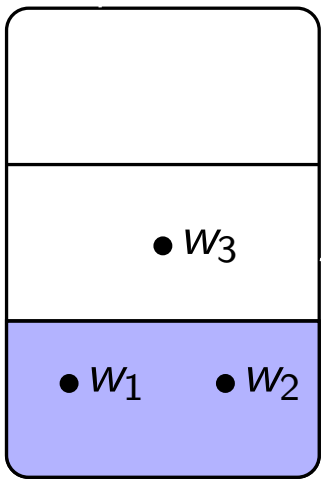
Belief Revision via Plausibility

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

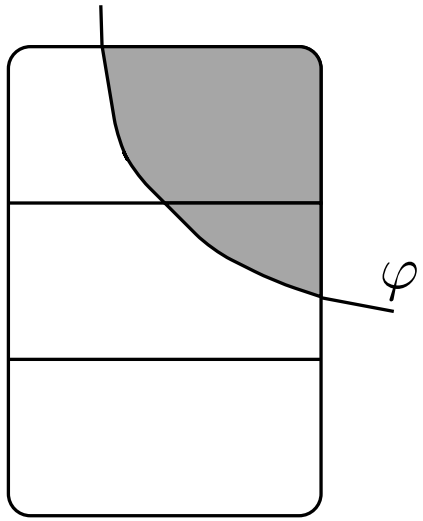


Belief Revision via Plausibility

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

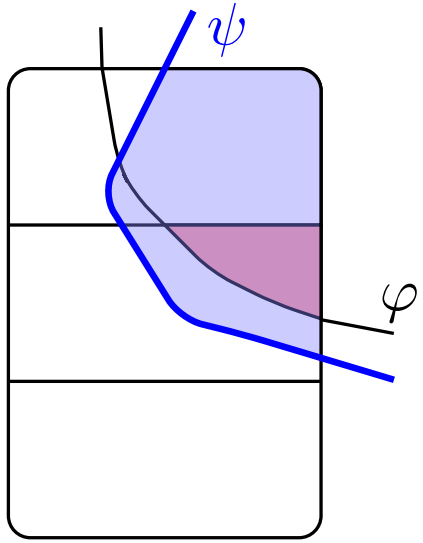


Belief Revision via Plausibility



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

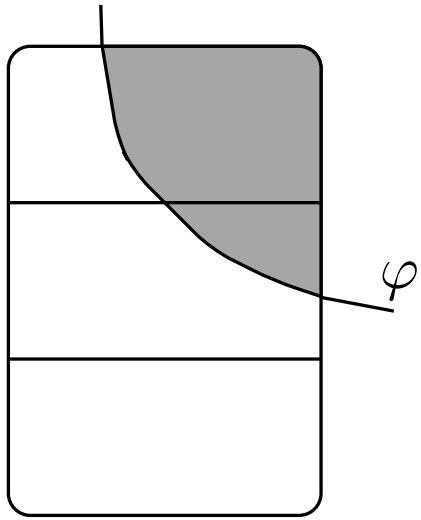
Belief Revision via Plausibility



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

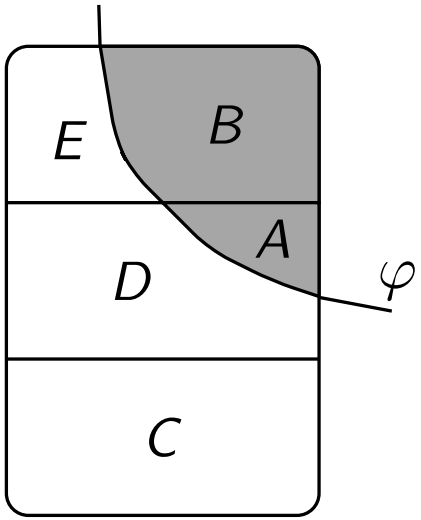
$$Min_{\preceq}([\![\varphi]\!]_{\mathcal{M}}) \subseteq [\![\psi]\!]_{\mathcal{M}}$$

Belief Revision via Plausibility



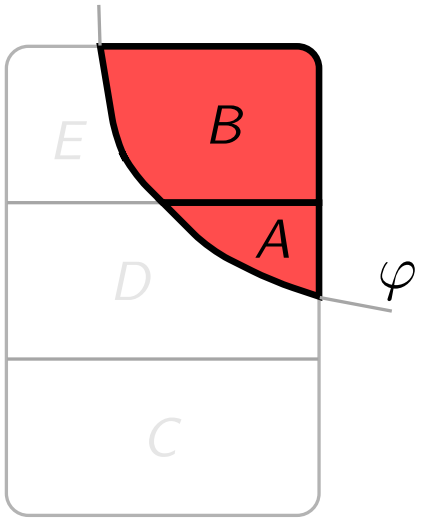
Incorporate the new information φ

Belief Revision via Plausibility



Incorporate the new information φ

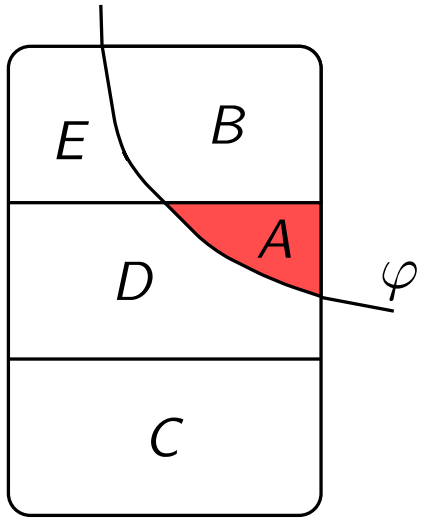
Belief Revision via Plausibility



Public Announcement: Information from an infallible source

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

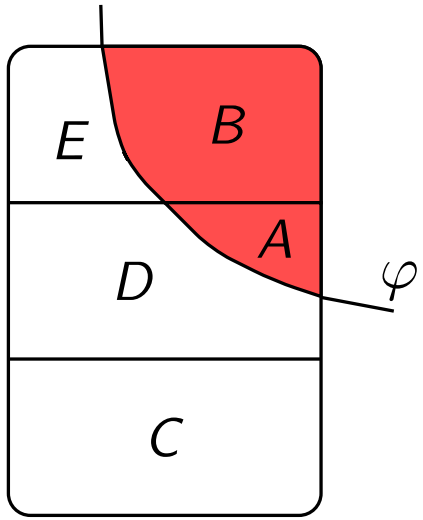
Belief Revision via Plausibility



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

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

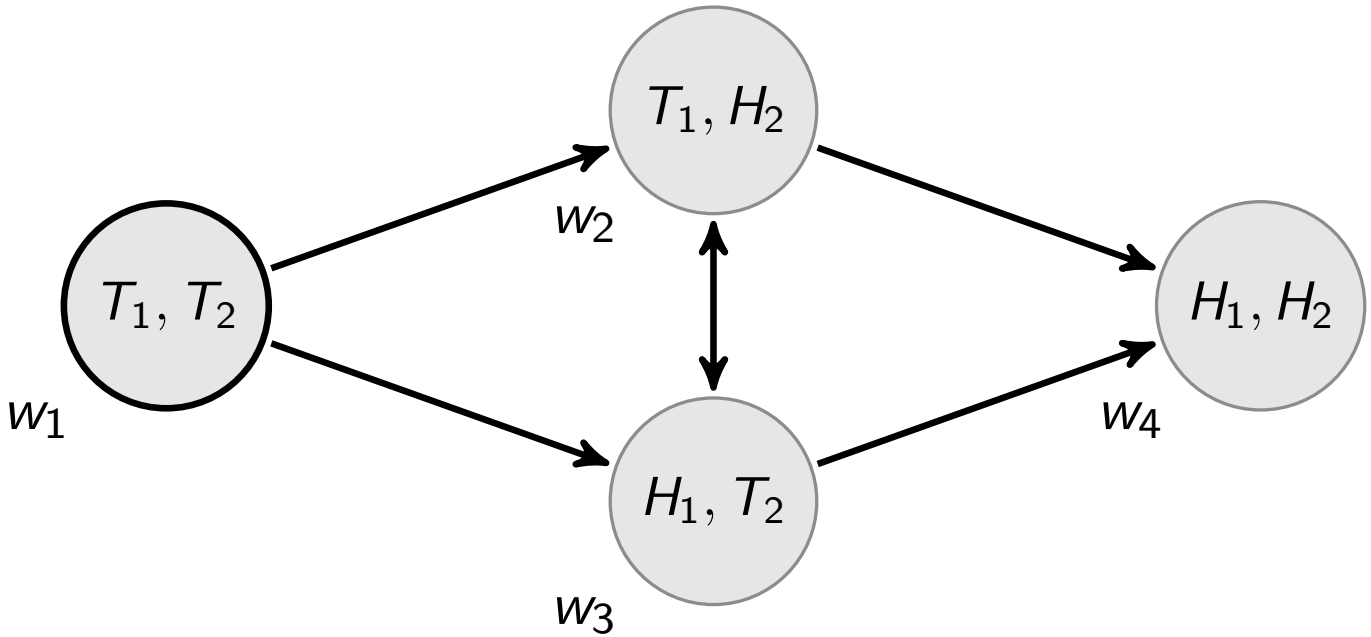
Belief Revision via Plausibility



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

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

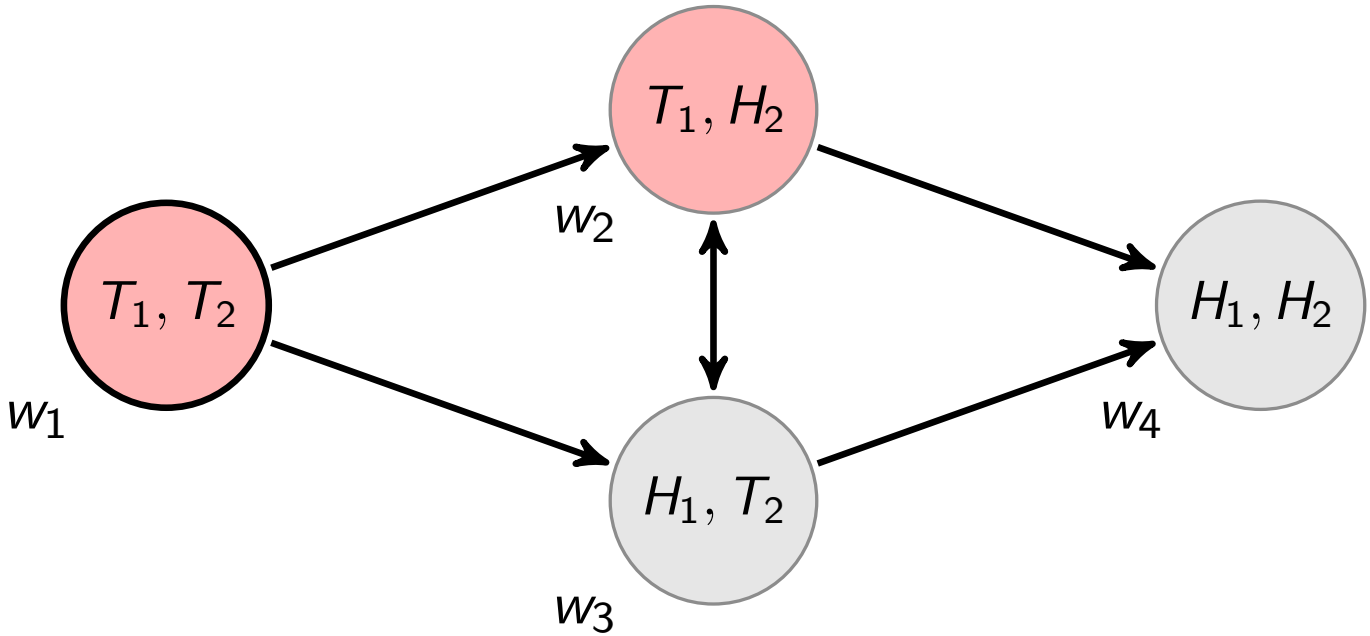
Radical Upgrade: Information from a strongly trusted source $(\Uparrow\varphi)$: $A \prec_i B \prec_i C \prec_i D \prec_i E$



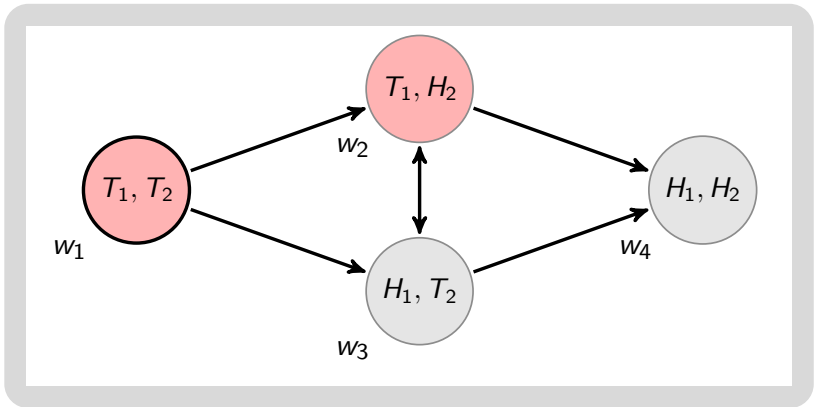
$Min_{\preceq}([w_1]) = \{w_4\}$, so $w_1 \models B(H_1 \wedge H_2)$

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

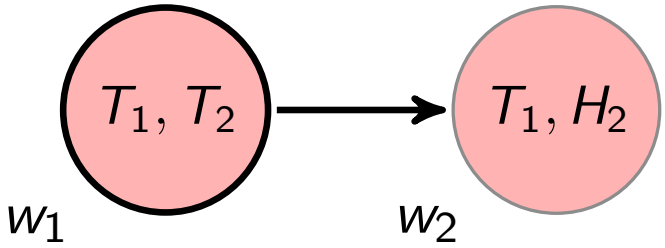
$Min_{\preceq}([w_1] \cap [\![T_1]\!]_{\mathcal{M}}) = \{w_3\}$, so $w_1 \models B^{T_2} H_1$



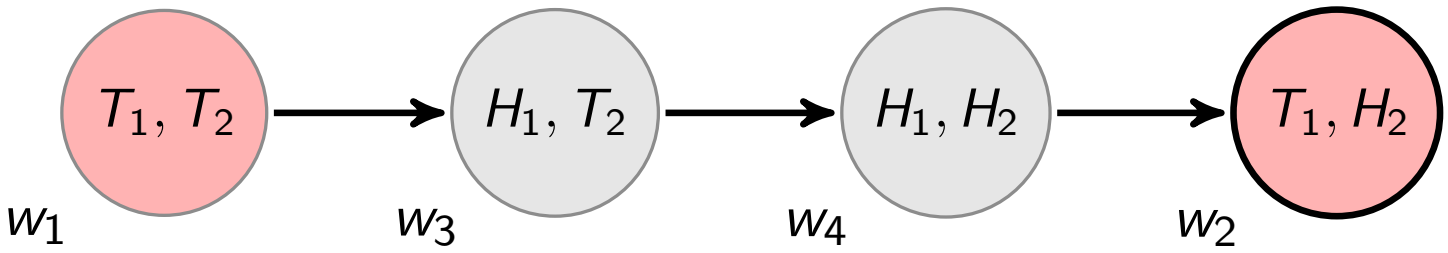
Suppose the agent *finds out that* T_1 *is true*.



$!(T_1)$

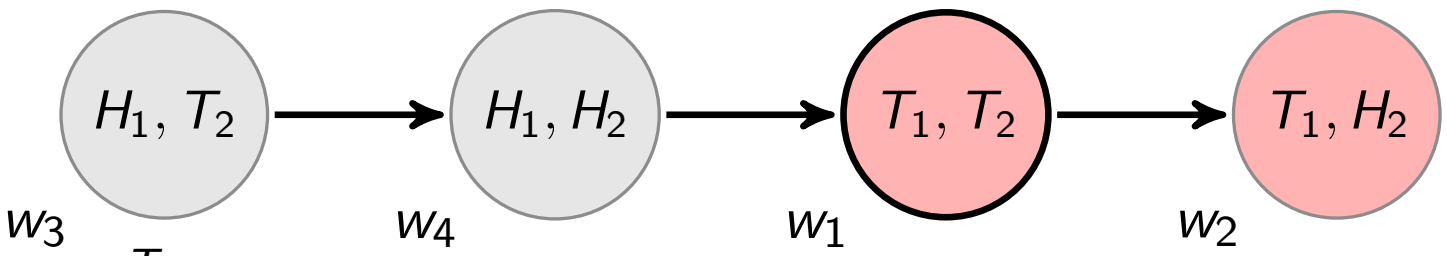


$\uparrow(T_1)$



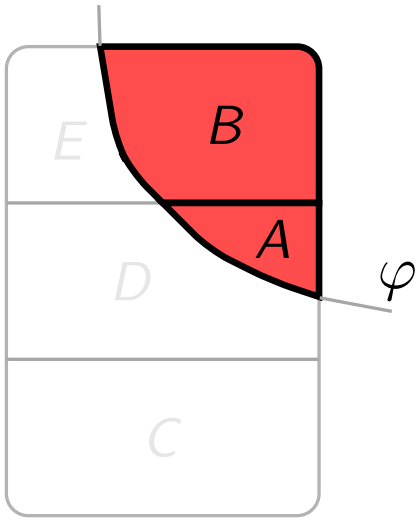
$B^{T_2} H_1$

$\Uparrow(T_1)$



$B^{T_2} T_1$

Informative Actions



Public Announcement: Information from an infallible source

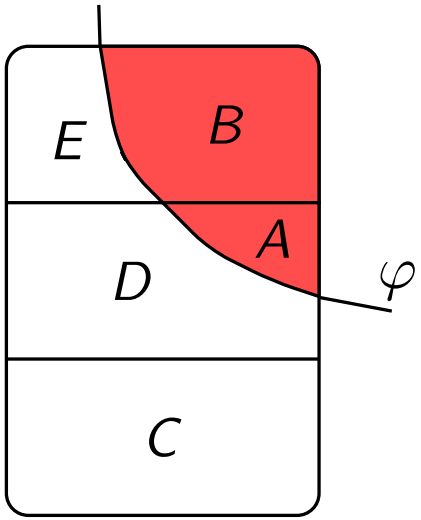
$(!\varphi)$: $A \prec_i B$ $\quad\quad$ $\mathcal{M}^{!\varphi} = \langle W^{!\varphi}, \{\sim_i^{!\varphi}\}_{i\in\mathcal{A}}, V^{!\varphi}\rangle$

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

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

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

Informative Actions

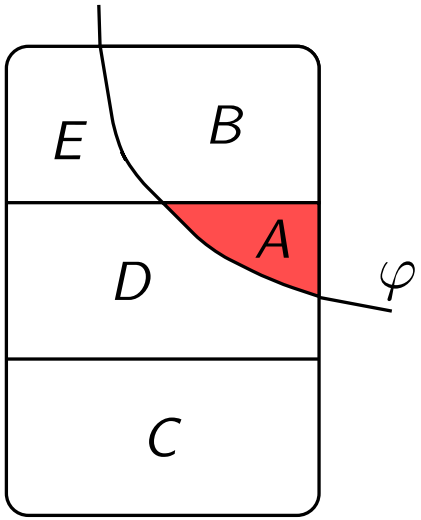


Radical Upgrade: $(\Uparrow\varphi)$: $A \prec_i B \prec_i C \prec_i D \prec_i E$,
$\mathcal{M}^{\Uparrow\varphi} = \langle W, \{\sim_i\}_{i\in\mathcal{A}}, \{\preceq_i^{\Uparrow\varphi}\}_{i\in\mathcal{A}}, V\rangle$

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

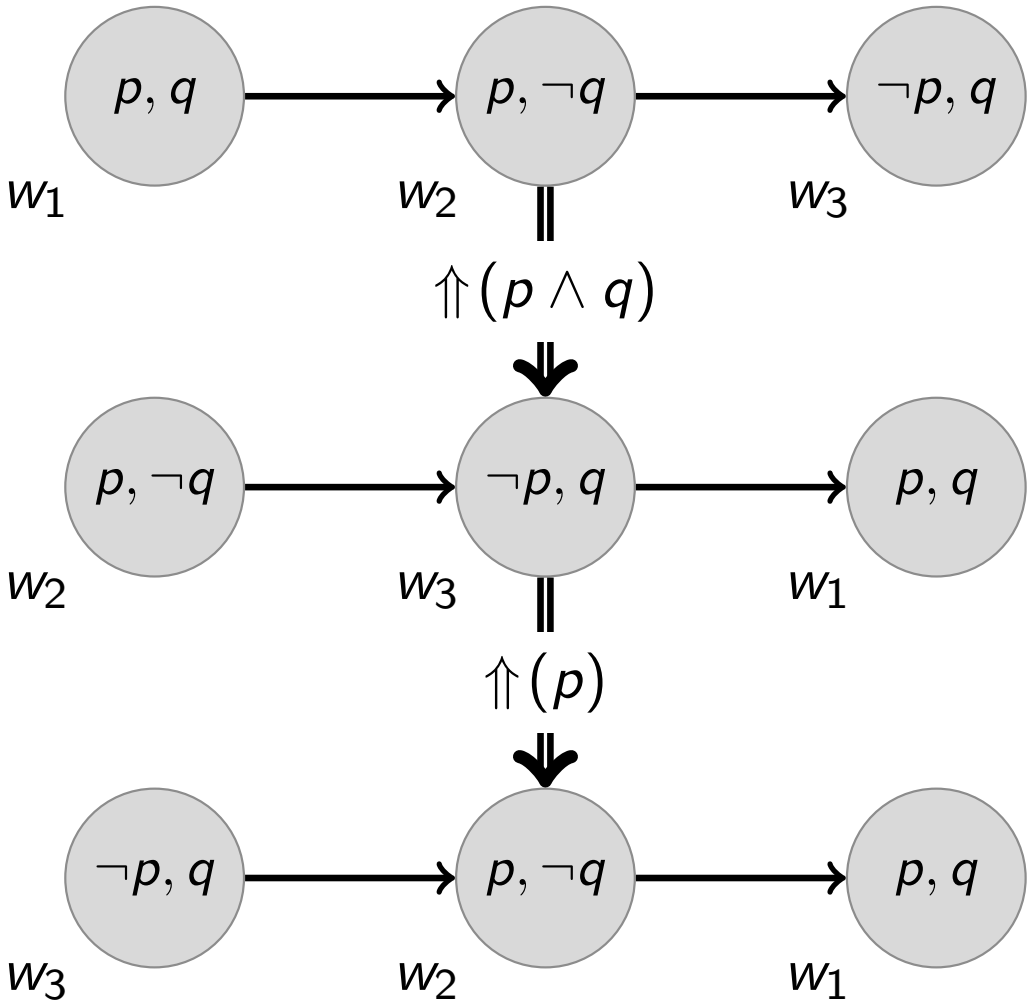
$$[\Uparrow\varphi]B^{\psi}\chi \leftrightarrow (L(\varphi \wedge [\Uparrow\varphi]\psi) \wedge B^{\varphi\wedge[\Uparrow\varphi]\psi}[\Uparrow\varphi]\chi)\vee$$
$$(\neg L(\varphi \wedge [\Uparrow\varphi]\psi) \wedge B^{[\Uparrow\varphi]\psi}[\Uparrow\varphi]\chi)$$

Recursion Axioms

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

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

Composition



Iterated Updates

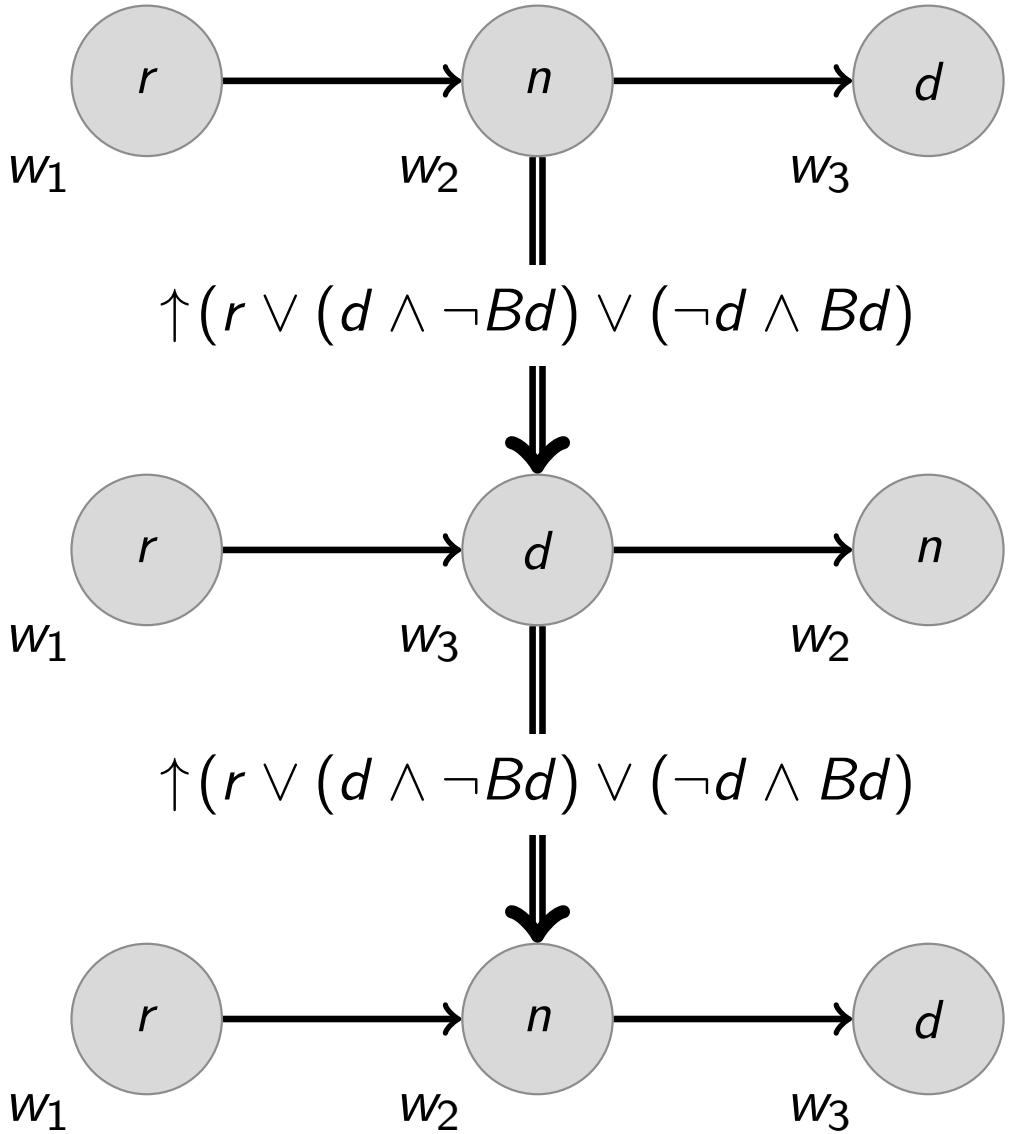
$!\varphi_1, !\varphi_2, !\varphi_3, \ldots, !\varphi_n$
always reaches a fixed-point

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

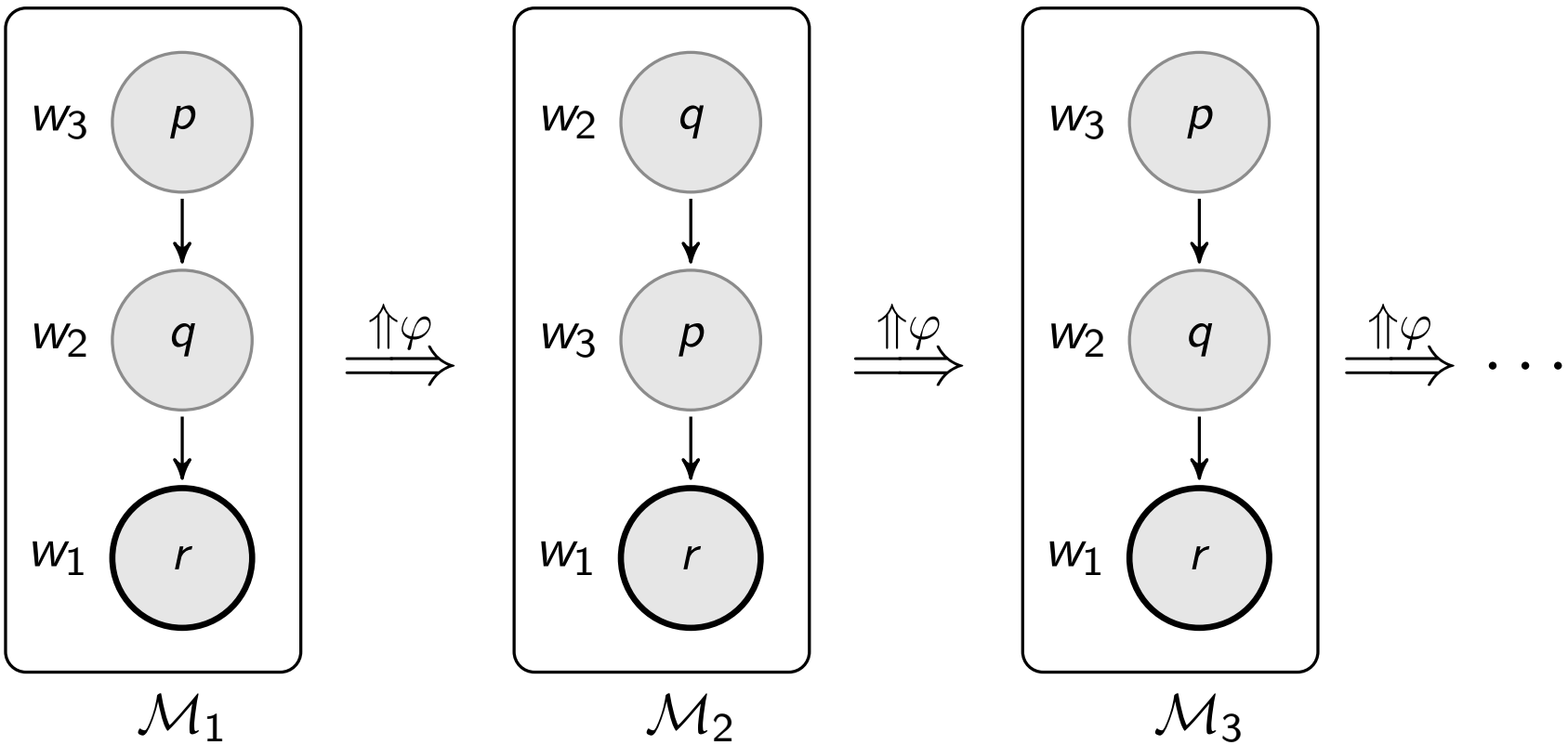
$\uparrow\varphi, \uparrow\varphi, \ldots$
Simple beliefs may never stabilize

$\Uparrow\varphi, \Uparrow\varphi, \ldots$
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 $(r \vee (B^{\neg r} q \wedge p) \vee (B^{\neg r} p \wedge q))$



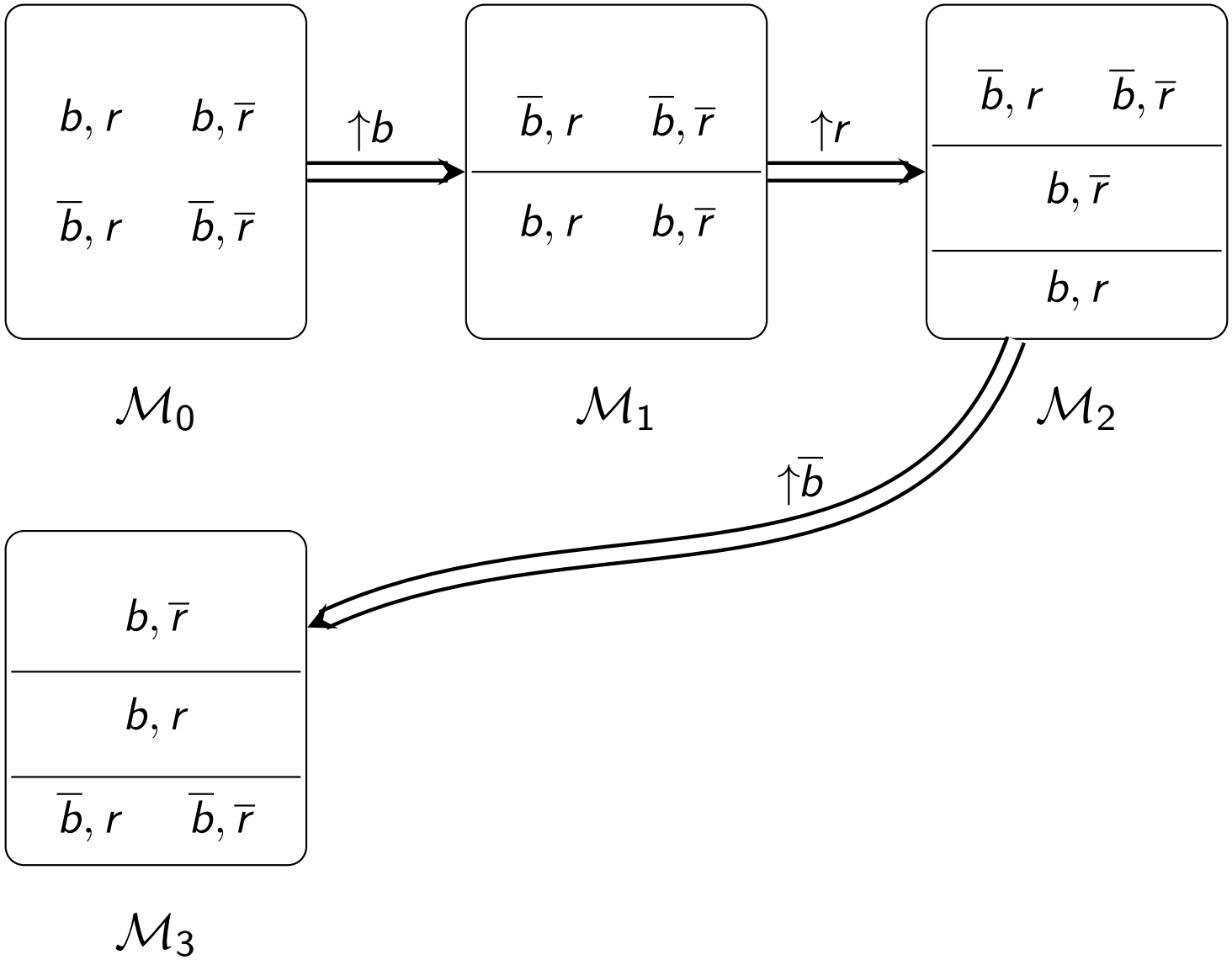
Suppose that you are in the forest and happen to a see strange-looking animal.

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.

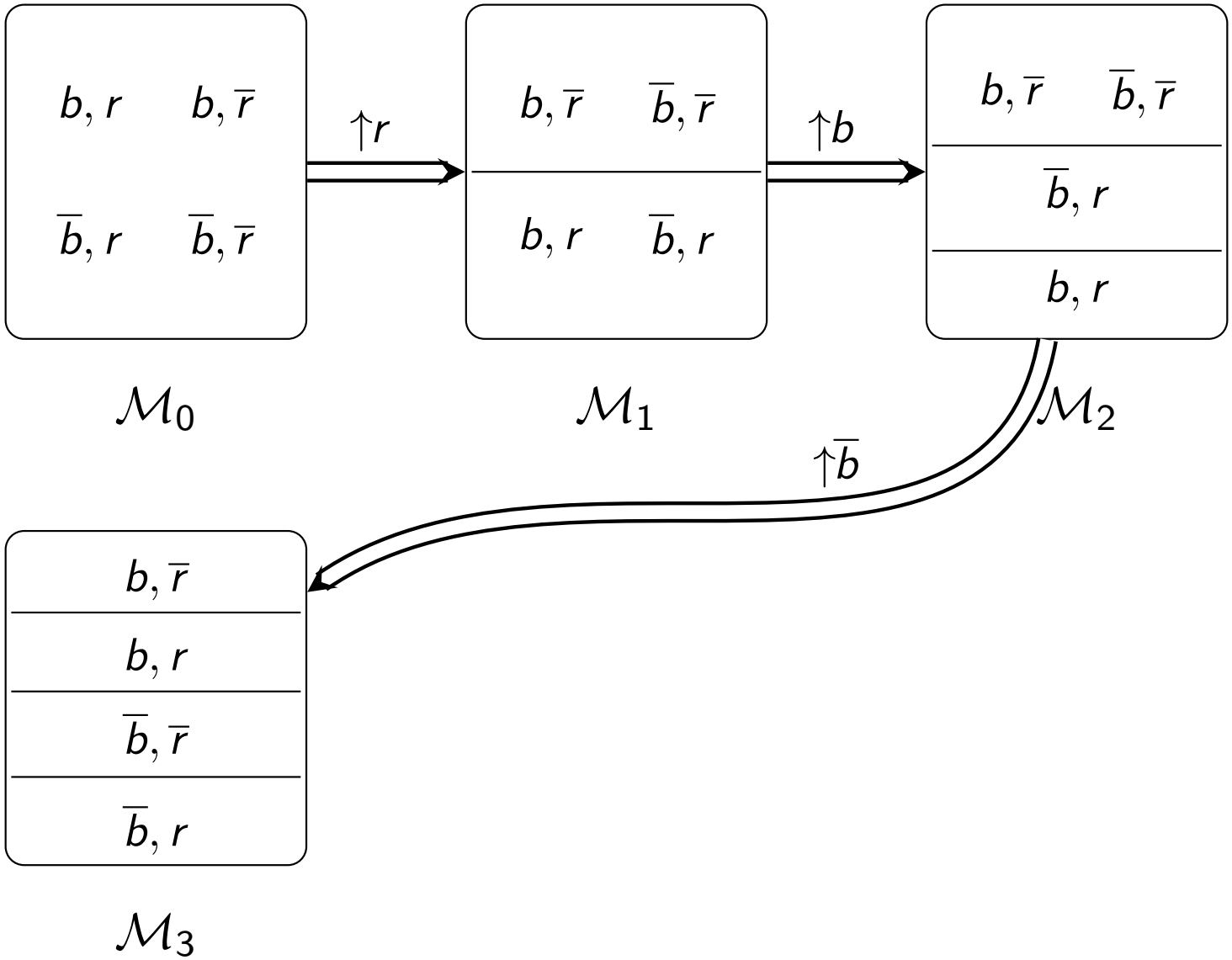
Suppose that you are in the forest and happen to a see strange-looking animal. You consult your animal guidebook and find a picture that seems to match the animal you see. The guidebook says that the animal is a type of bird, so that is what you conclude: The animal before you is a bird. After looking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

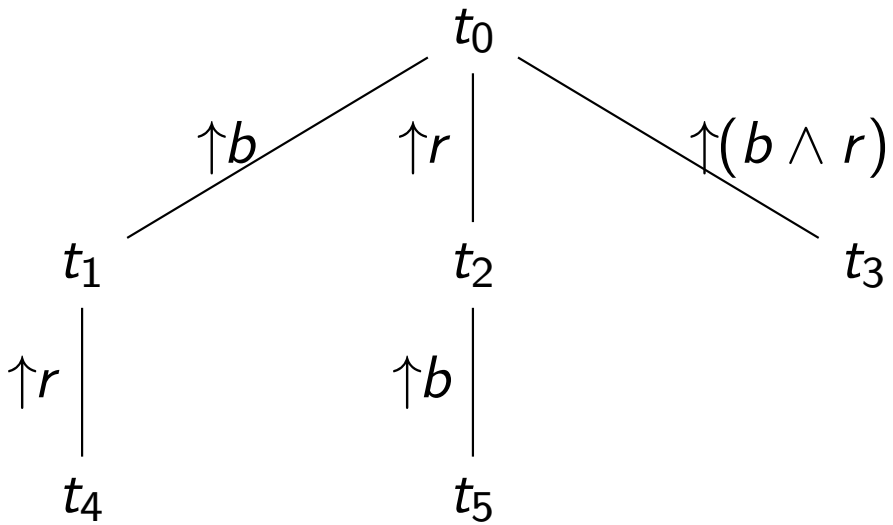

b, r
b, r
b, r
b, r
↑b
b, r
b, r
b, r
b, r
↑r
b, r
b, r
b, r
b, r
b, r
↑b
b, r
b, r
b, r
b, r
M0
M1
M2
M3

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.

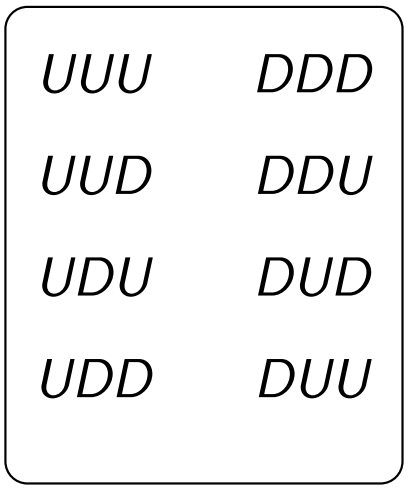
Two Postulates of Iterated Revision

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

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

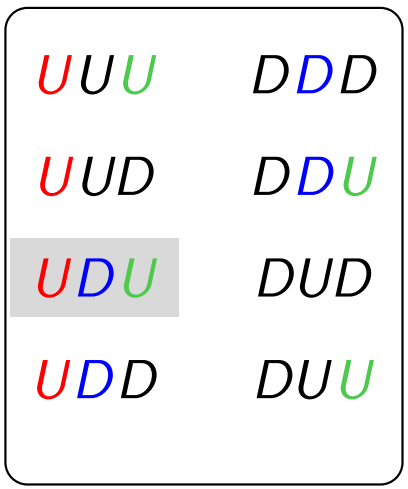
- Postulate I1 demands if $\varphi \rightarrow \psi$ is a theorem (with respect to the background theory), then first learning ψ followed by the more specific information φ is equivalent to directly learning the more specific information φ.
- Postulate I2 demands that first learning φ followed by learning 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



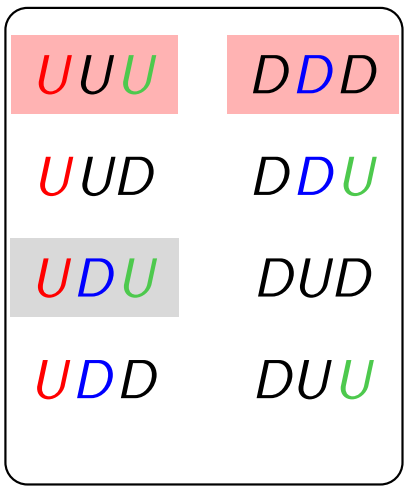
- Three switches wired such that a light is on iff all three switches are up or all three are down.

Stalnaker's Counterexample to I1



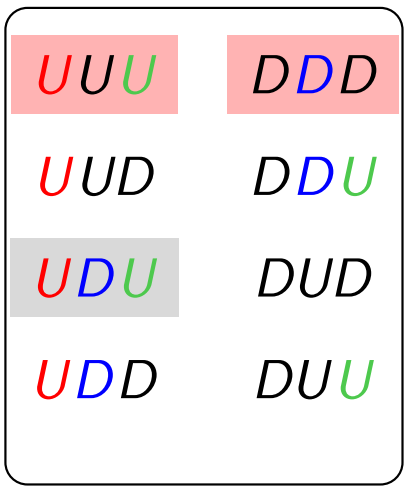
- Three switches wired such that a light is on iff all three switches are up or all three are down.
- Three independent (reliable) observers report on the switches: Alice says switch 1 is U, Bob says switch 2 is D and Carla says switch 3 is U.

Stalnaker's Counterexample to I1



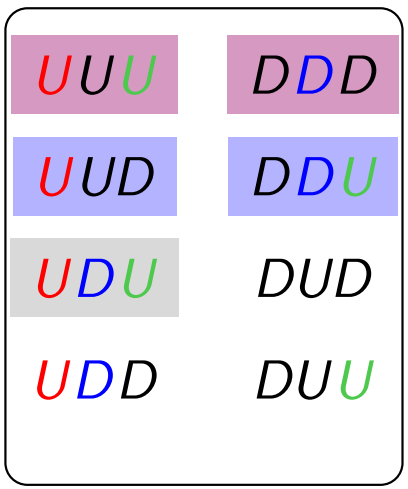
- Three switches wired such that a light is on iff all three switches are up or all three are down.
- Three independent (reliable) observers report on the switches: Alice says switch 1 is U, Bob says switch 2 is D and Carla says switch 3 is U.
- I receive the information that the light is on. What should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

Stalnaker's Counterexample to I1



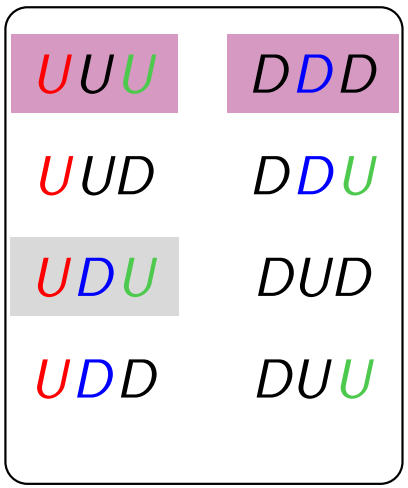
- Three switches wired such that a light is on iff all three switches are up or all three are down.
- Three independent (reliable) observers report on the switches: Alice says switch 1 is U, Bob says switch 2 is D and Carla says switch 3 is U.
- I receive the information that the light is on. What should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

Stalnaker's Counterexample to I1



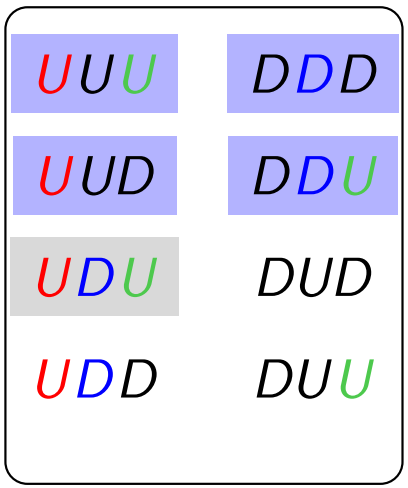
- Suppose there are two switches: L_1 is the main switch and L_2 is a secondary switch controlled by the first two lights. (So $L_1 \rightarrow L_2$, but not the converse)

Stalnaker's Counterexample to I1



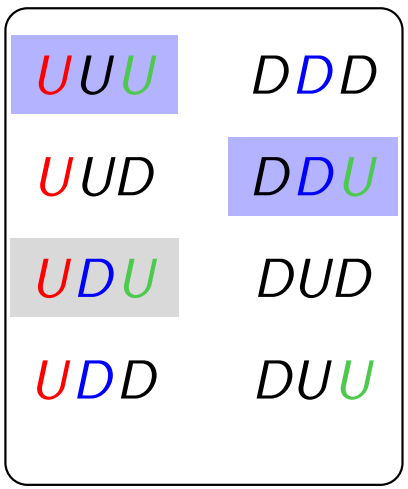
- Suppose there are two switches: L_1 is the main switch and L_2 is a secondary switch controlled by the first two lights. (So $L_1 \rightarrow L_2$, but not the converse)
- Suppose I receive $L_1 \wedge L_2$, this does not change the story.

Stalnaker's Counterexample to I1



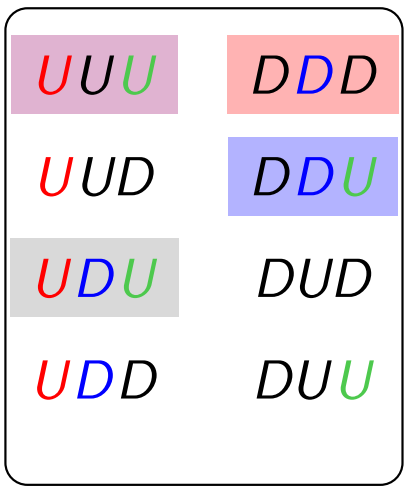
- Suppose there are two switches: L_1 is the main switch and L_2 is a secondary switch controlled by the first two lights. (So $L_1 \rightarrow L_2$, but not the converse)
- Suppose I receive $L_1 \wedge L_2$, this does not change the story.
- Suppose I learn that L_2. This is irrelevant to Carla's report,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

Stalnaker's Counterexample to I1



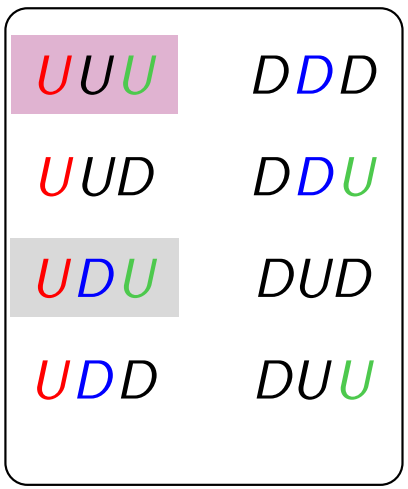
- Suppose there are two switches: L_1 is the main switch and L_2 is a secondary switch controlled by the first two lights. (So $L_1 \rightarrow L_2$, but not the converse)
- Suppose I receive $L_1 \wedge L_2$, this does not change the story.
- Suppose I learn that L_2. This is irrelevant to Carla's report,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

Stalnaker's Counterexample to I1



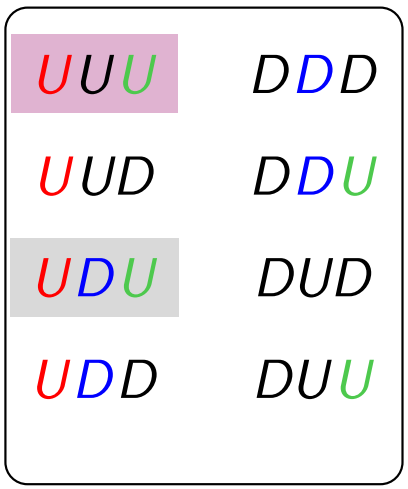
- Suppose there are two switches: L_1 is the main switch and L_2 is a secondary switch controlled by the first two lights. (So $L_1 \to L_2$, but not the converse)
- Suppose I receive $L_1 \wedge L_2$, this does not change the story.
- Suppose I learn that L_2. This is irrelevant to Carla's report, but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



- 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



- 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.

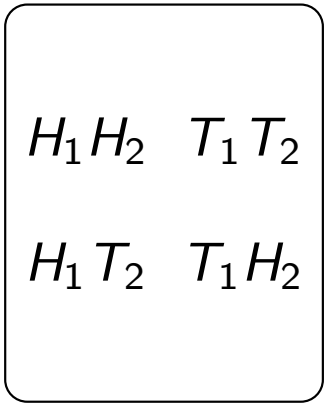
- The first revision results in the belief set $K' = K * (H_1 \wedge H_2)$, where K is the agents original set of beliefs.
- After receiving the reports, the belief set is $K' * (T_1 \wedge T_2) * H_1$.
- Since Elmers report is irrelevant to the status of the coin in box 2, it seems natural to assume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

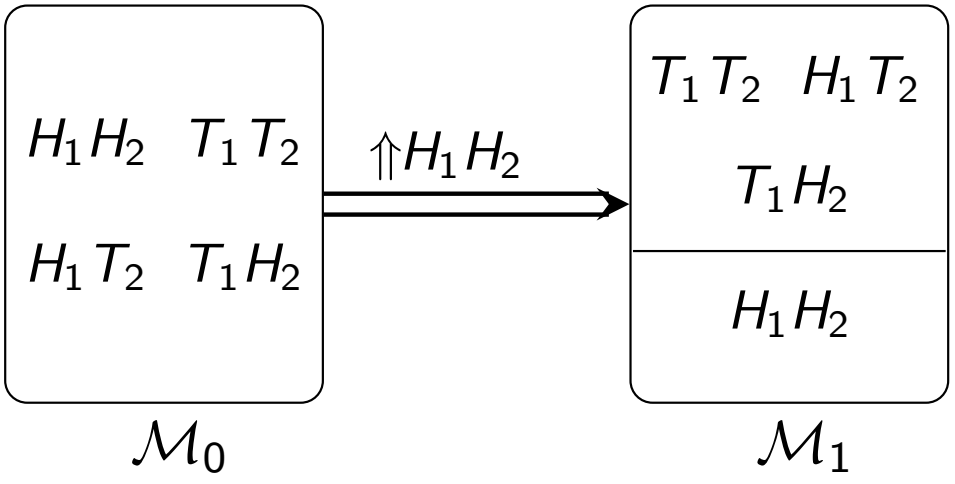
(Stalnaker, pg. 207–208)

Heuristic Diagnosis of Stalnaker's Example

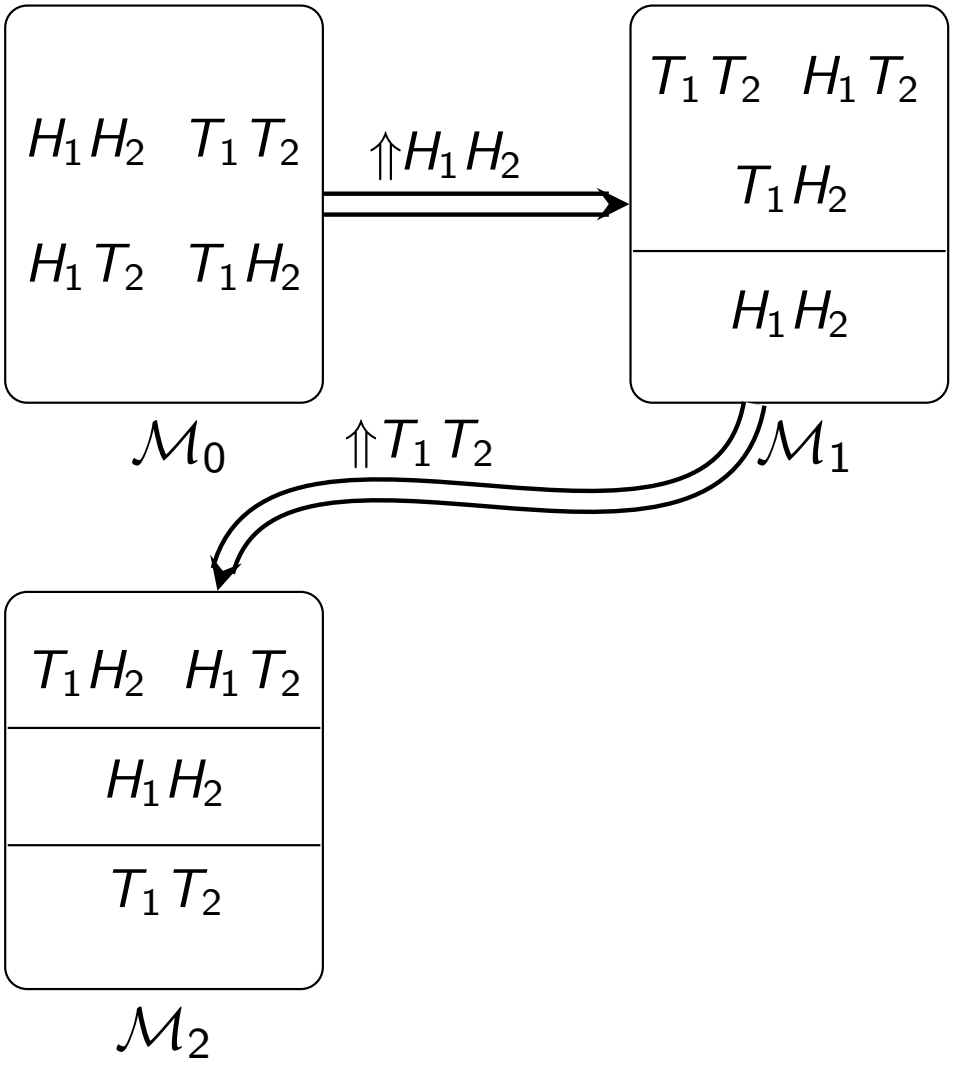


$\mathcal{M}_0$

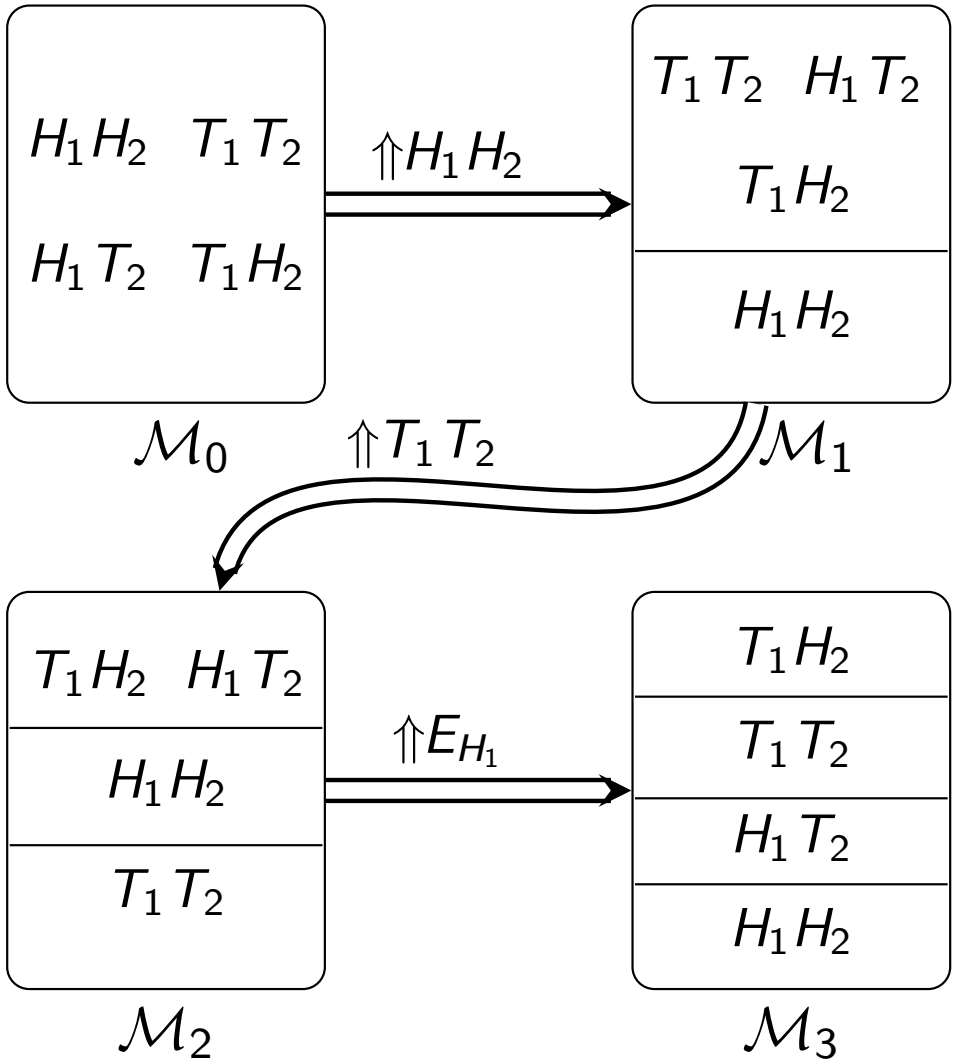
Heuristic Diagnosis of Stalnaker's Example



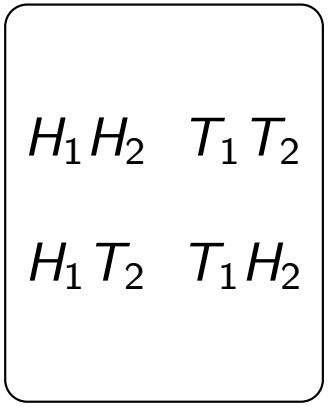
Heuristic Diagnosis of Stalnaker's Example



Heuristic Diagnosis of Stalnaker's Example

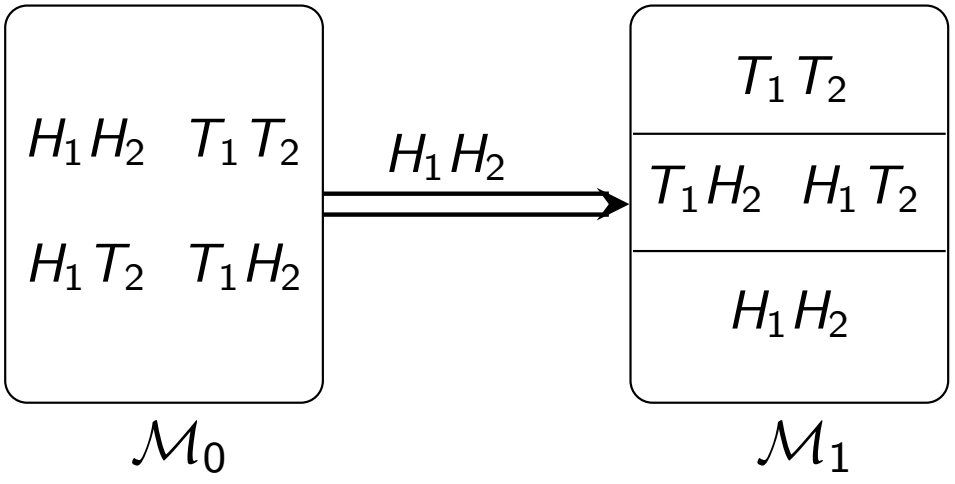


Heuristic Diagnosis of Stalnaker's Example

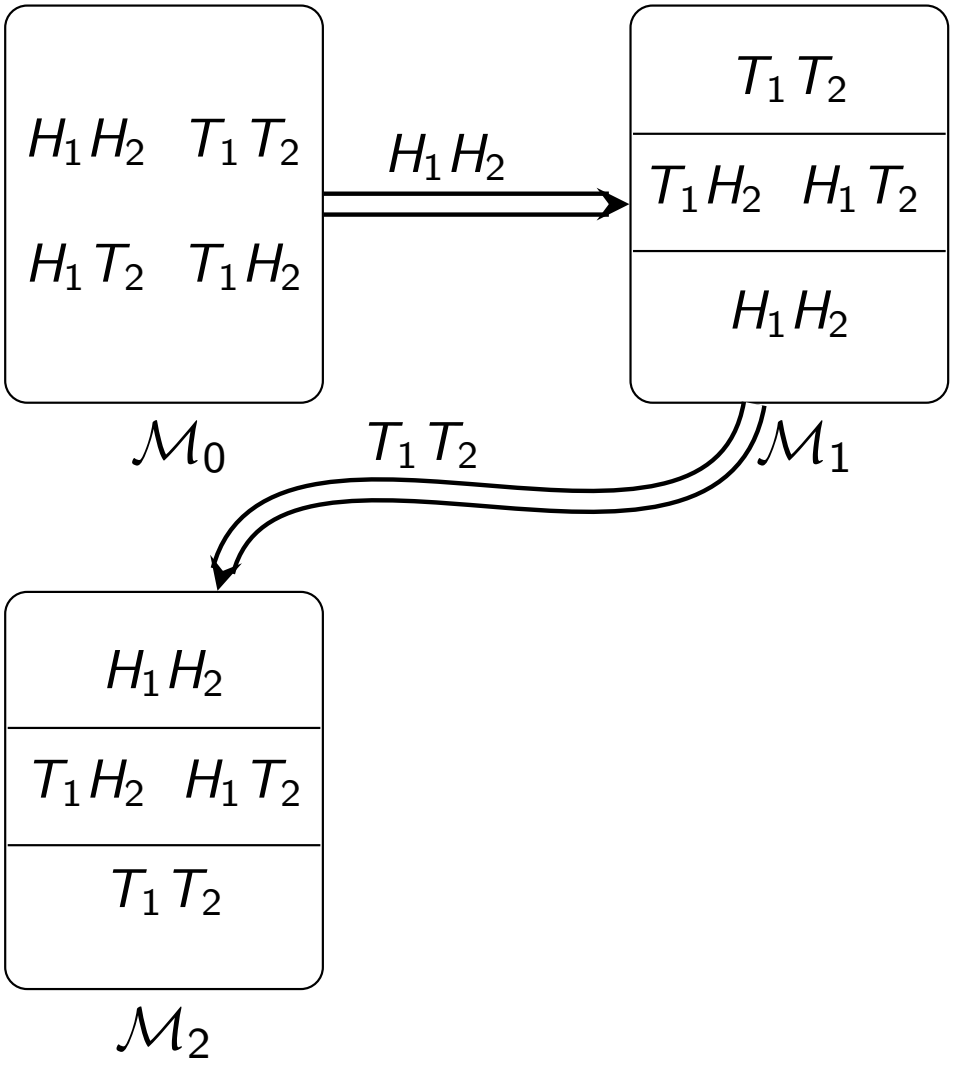


$\mathcal{M}_0$

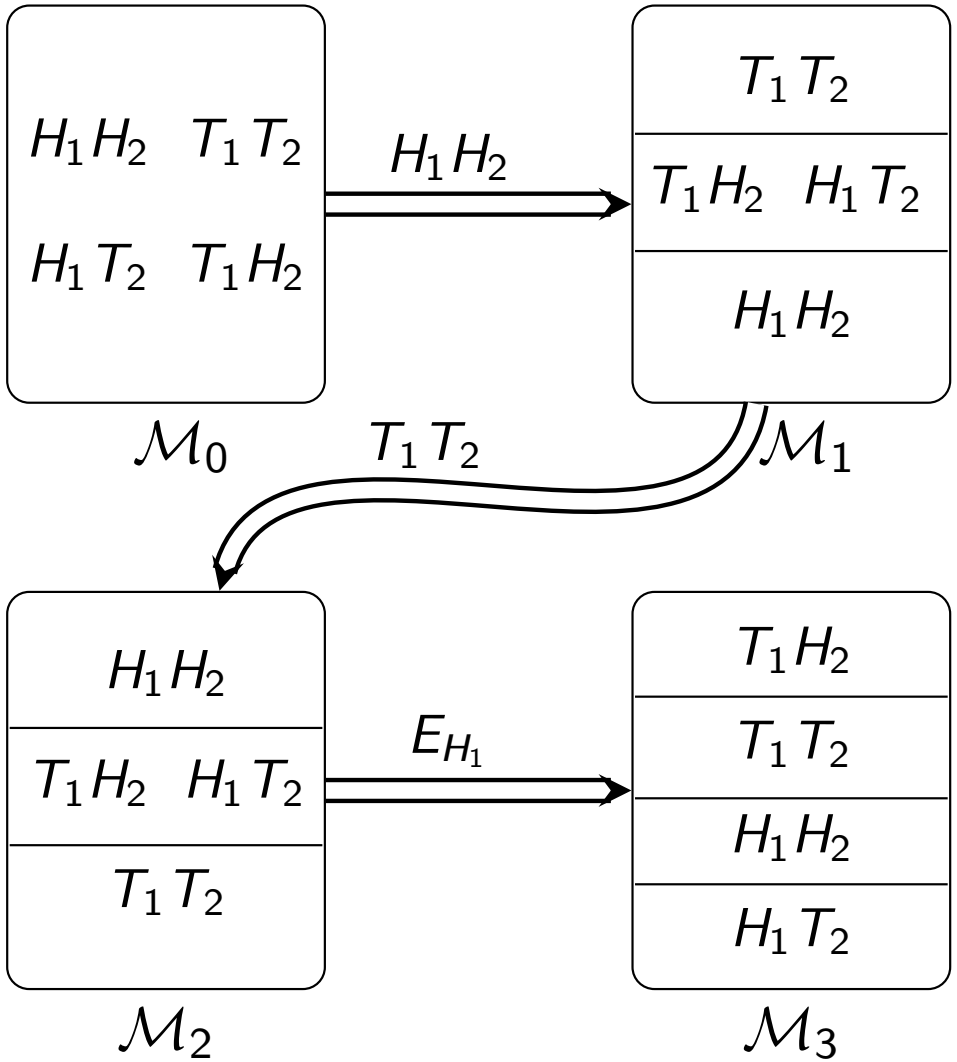
Heuristic Diagnosis of Stalnaker's Example



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.