Reasoning about Knowledge and Beliefs

Lecture 14

Eric Pacuit

University of Maryland, College Park

pacuit.org
epacuit@umd.edu

October 30, 2013

Updating vs. Revising

Revision vs. Update

Suppose φ is some incoming information that should be incorporated into the agents beliefs (represented by a theory T).

Revision vs. Update

Suppose φ is some incoming information that should be incorporated into the agents beliefs (represented by a theory T).

An important distinction:

- If φ describes facts about the current state of affairs
- If φ describes facts that have possible become true only after the original beliefs were formed.

Revision vs. Update

Suppose φ is some incoming information that should be incorporated into the agents beliefs (represented by a theory T).

An important distinction:

- If φ describes facts about the current state of affairs
- If φ describes facts that have possible become true only after the original beliefs were formed.

Complete vs. incomplete belief sets:
$K = Cn(\{p \vee q\})$ vs. $K = Cn(\{p \vee q, p, q\})$

Revision vs. Update

Suppose φ is some incoming information that should be incorporated into the agents beliefs (represented by a theory T).

An important distinction:

- If φ describes facts about the current state of affairs
- If φ describes facts that have possible become true only after the original beliefs were formed.

Complete vs. incomplete belief sets:
$K = Cn(\{p \vee q\})$ vs. $K = Cn(\{p \vee q, p, q\})$

Revising by $\neg p$ ($K * \neg p$) vs. Updating by $\neg p$ ($K \diamond \neg p$)

H. Katsuno and A. O. Mendelzon. *Propositional knowledge base revision and minimal change*. Artificial Intelligence, 52, pp. 263 - 294 (1991).

KM Postulates

KM 1: $K \diamond \varphi = Cn(K \diamond \varphi)$

KM 2: $\varphi \in K \diamond \varphi$

KM 3: If $\varphi \in K$ then $K \diamond \varphi = K$

KM 4: $K \diamond \varphi$ is inconsistent iff φ is inconsistent

KM 5: If φ and ψ are logically equivalent then $K \diamond \varphi = K \diamond \psi$

KM 6: $K \diamond (\varphi \wedge \psi) \subseteq Cn(K \diamond \varphi \cup \{\psi\})$

KM 7: If $\psi \in K \diamond \varphi$ and $\varphi \in K \diamond \psi$ then $K \diamond \varphi = K \diamond \psi$

KM 8: If K is complete then $K \diamond (\varphi \wedge \psi) \subseteq K \diamond \varphi \cap K \diamond \psi$

KM 9: $K \diamond \varphi = \bigcap_{M \in Comp(K)} M \diamond \varphi$, where $Comp(K)$ is the class of all complete theories containing K.

Updating and Revising

$$K \diamond \varphi = \bigcap_{M \in Comp(K)} M * \varphi$$

H. Katsuno and A. O. Mendelzon. *On the difference between updating a knowledge base and revising it. Belief Revision*, P. Gärdenfors (ed.), pp 182 - 203 (1992).

P. Gärdernfors. *Belief revisions and the Ramsey test for conditionals*. Philosophical Review, 95:1, pgs. 81 - 93.

When should a rational agent believe/accept/assert and *conditional*? (If A, then B: $A > B$)

Note: $A > B$ is NOT the same as $A \Rightarrow B$

Ramsey Test

If two people are arguing 'If p, then q?' and are both in doubt as to p, they are adding p hypothetically to their stock of knowledge and arguing on that basis about q; so that in a sense 'If p, q' and 'If p, $\neg q$' are contradictories. We can say that they are fixing their degree of belief in q given p. If p turns out false, these degrees of belief are rendered void. If either party believes not p for certain, the question ceases to mean anything to him except as a question about what follows from certain laws or hypotheses.

F. P. Ramsey. *General Propositions and Causality*. 1929.

AGM and the Ramsey Test

Recall that $\mathcal{L}$ is a propositional language. Let $\mathcal{L} \subseteq \mathcal{L}_{<}$ extends $\mathcal{L}$ with a conditional operator '$>$': $A > B$ means 'If A, B'

AGM and the Ramsey Test

Recall that $\mathcal{L}$ is a propositional language. Let $\mathcal{L} \subseteq \mathcal{L}_<$ extends $\mathcal{L}$ with a conditional operator '$>$': $A > B$ means 'If A, B'

$$(R) \qquad A > B \in K \text{ iff } B \in K * A$$

AGM and the Ramsey Test

Recall that $\mathcal{L}$ is a propositional language. Let $\mathcal{L} \subseteq \mathcal{L}_{<}$ extends $\mathcal{L}$ with a conditional operator '$>$': $A > B$ means 'If A, B'

$$(R) \qquad A > B \in K \text{ iff } B \in K * A$$

Belief revision model: $\langle \mathbf{K}, * \rangle$ where $* : \mathbf{K} \times \mathcal{L} \to \mathbf{K}$.

Gärdenfors Impossibility Result, I

For each $A \in \mathcal{L}$, write K_A for $K * A$.

Gärdenfors Impossibility Result, I

For each $A \in \mathcal{L}$, write K_A for $K * A$.

Assumptions

Gärdenfors Impossibility Result, I

For each $A \in \mathcal{L}$, write K_A for $K * A$.

Assumptions

(1) $A \in K_A$

Gärdenfors Impossibility Result, I

For each $A \in \mathcal{L}$, write K_A for $K * A$.

Assumptions

(1) $A \in K_A$

(2) If $K \neq K_\perp$ and $K_A = K_\perp$ then $\vdash \neg A$

Gärdenfors Impossibility Result, I

For each $A \in \mathcal{L}$, write K_A for $K * A$.

Assumptions

(1) $A \in K_A$

(2) If $K \neq K_\perp$ and $K_A = K_\perp$ then $\vdash \neg A$

(R) $A > B \in K$ iff $B \in K * A$

Gärdenfors Impossibility Result, II

From (R) we can conclude

(M) For all belief sets K, K' and propositions A, $K \subseteq K'$ implies $K_A \subseteq K'_A$

Gärdenfors Impossibility Result, II

From (R) we can conclude

(M) For all belief sets K, K' and propositions A, $K \subseteq K'$ implies $K_A \subseteq K'_A$

Preservation Criterion

(P) If $\neg A \notin K$ and $B \in K$ then $B \in K_A$

Gärdenfors Impossibility Result, III

Let $K/A = Cn(K \cup \{A\})$, We have

Gärdenfors Impossibility Result, III

Let $K/A = Cn(K \cup \{A\})$, We have

(3) $(K/A)/B = K/(A \wedge B)$
(4) $K/(A \vee B) \subseteq K/A$

Gärdenfors Impossibility Result, III

Let $K/A = Cn(K \cup \{A\})$, We have

(3) $(K/A)/B = K/(A \wedge B)$
(4) $K/(A \vee B) \subseteq K/A$

From (1) and (P)

(I) If $\neg A \notin K$ then $K/A \subseteq K_A$

Gärdenfors Impossibility Result, IV

Two propositions A and B are *disjoint* provided $\vdash \neg(A \wedge B)$

A belief revision model $\langle \mathbf{K}, \mathbf{F} \rangle$ is **nontrivial** provided there is a $K \in \mathbf{K}$ and *pairwise disjoint* propositions A, B and C such that $\neg A \notin K$, $\neg B \notin K$ and $\neg C \notin K$.

Theorem There is no nontrivial belief revision model that satisfies all the conditions (1), (2), (M) and (P).

P. Gärdenfors. *Belief Revision and the Ramsey Test*. Philosophical Review, Vol. 95, pp. 81 - 93, 1986.

Gärdenfors Impossibility Result: Proof

Nontrivial There are A, B, and C such that $\vdash \neg(A \wedge B)$, $\vdash \neg(B \wedge C)$, $\vdash \neg(A \wedge C)$ and belief set K such that $\neg A \notin K$, $\neg B \notin K$ and $\neg C \notin K$.

(1) $A \in K_A$

(2) If $K \neq K_\perp$ and $K_A = K_\perp$ then $\vdash \neg A$

(R) $A > B \in K$ iff $B \in K * A$

(M) For all K, K' and A, $K \subseteq K'$ implies $K_A \subseteq K'_A$

(P) If $\neg A \notin K$ and $B \in K$ then $B \in K_A$

(3) $(K/A)/B = K/(A \wedge B)$

(4) $K/(A \vee B) \subseteq K/A$

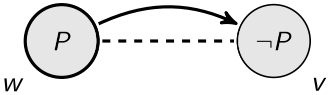
(I) If $\neg A \notin K$ then $K/A \subseteq K_A$

H. Leitgeb. *On the Ramsey Test without Triviality*. Notre Dame Journal of Formal Logic, 51:1, 2010, pp. 21 - 54.

H. Leitgeb. *Beliefs in Conditionals Vs. Conditional Beliefs*. Topoi, 26 (1), 2007, pp. 115 - 132.

I. Levi. *Iteration of Conditionals and the Ramsey Test*. Synthese, 76, 1988, pp. 49 - 81.

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

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

$$\Big\Downarrow$$

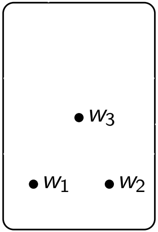
Find out that φ

$$\Big\Downarrow$$

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

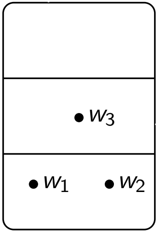
Belief Revision via Plausibility

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



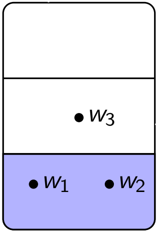
Belief Revision via Plausibility

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

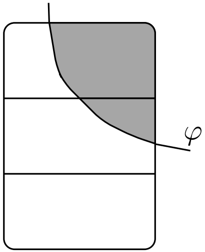


Belief Revision via Plausibility

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

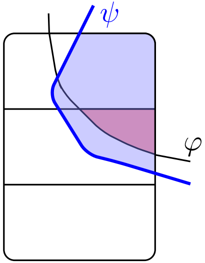


Belief Revision via Plausibility



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

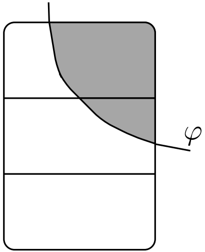
Belief Revision via Plausibility



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

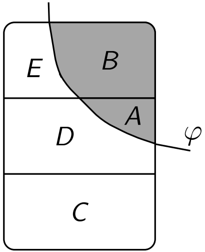
$$Min_{\preceq}(\llbracket\varphi\rrbracket_{\mathcal{M}}) \subseteq \llbracket\psi\rrbracket_{\mathcal{M}}$$

Belief Revision via Plausibility



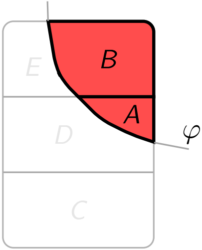
Incorporate the new information φ

Belief Revision via Plausibility



Incorporate the new information φ

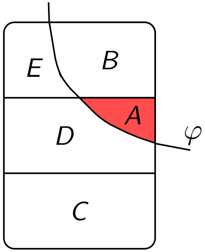
Belief Revision via Plausibility



Public Announcement: Information from an infallible source

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

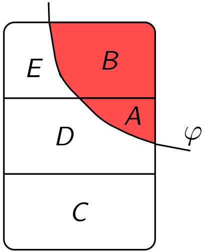
Belief Revision via Plausibility



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

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

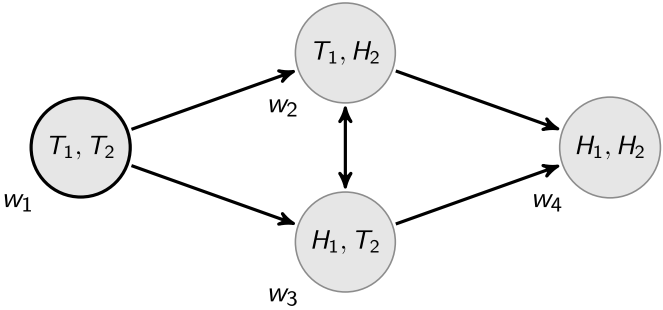
Belief Revision via Plausibility



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

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

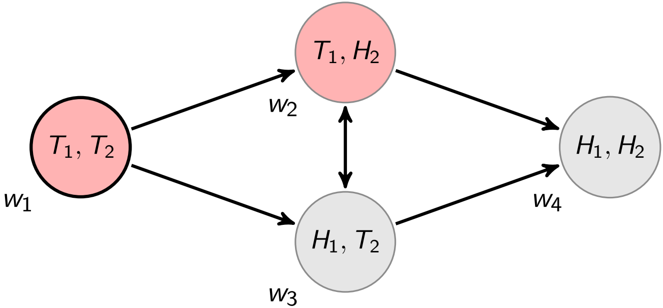
Radical Upgrade: Information from a strongly trusted source ($\Uparrow\varphi$): $A \prec_i B \prec_i C \prec_i D \prec_i E$



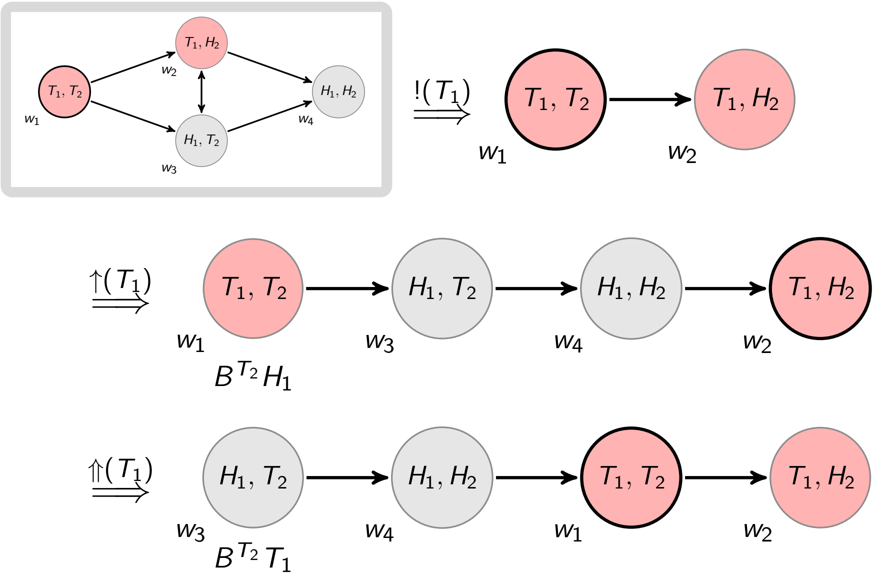
$Min_{\preceq}([w_1]) = \{w_4\}$, so $w_1 \models B(H_1 \wedge H_2)$

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

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

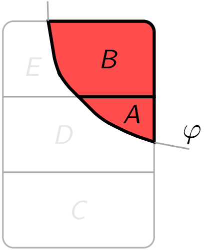


Suppose the agent *finds out that* T_1 *is true*.



T_1, H_2
w_2
T_1, T_2
w_1
H_1, H_2
w_4
H_1, T_2
w_3
$\xRightarrow{!(T_1)}$
T_1, T_2
w_1
T_1, H_2
w_2
$\xRightarrow{\uparrow(T_1)}$
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
$\xRightarrow{\Uparrow(T_1)}$
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

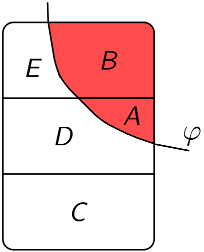


Public Announcement: Information from an infallible source

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

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

Informative Actions

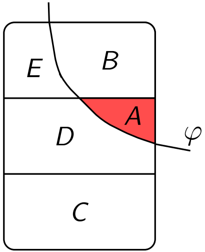


Radical Upgrade: $(\Uparrow\varphi)$: $A \prec_i B \prec_i C \prec_i D \prec_i E$,
$\mathcal{M}^{\Uparrow\varphi} = \langle W, \{\sim_i\}_{i\in\mathcal{A}}, \{\preceq_i^{\Uparrow\varphi}\}_{i\in\mathcal{A}}, V\rangle$

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

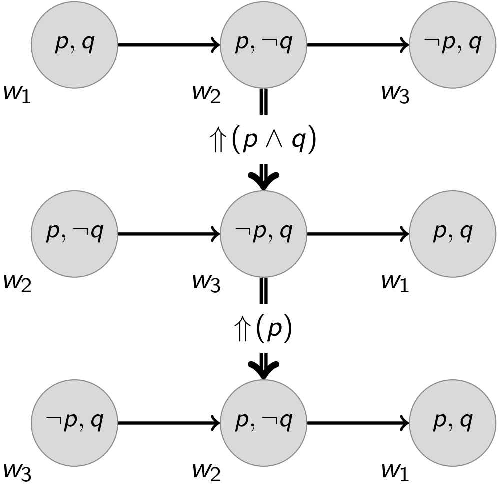
$$[\Uparrow\varphi]B^{\psi}\chi \leftrightarrow (L(\varphi \wedge [\Uparrow\varphi]\psi) \wedge B^{\varphi\wedge[\Uparrow\varphi]\psi}[\Uparrow\varphi]\chi)\vee$$
$$(\neg L(\varphi \wedge [\Uparrow\varphi]\psi) \wedge B^{[\Uparrow\varphi]\psi}[\Uparrow\varphi]\chi)$$

Recursion Axioms

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

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

Composition



Iterated Updates

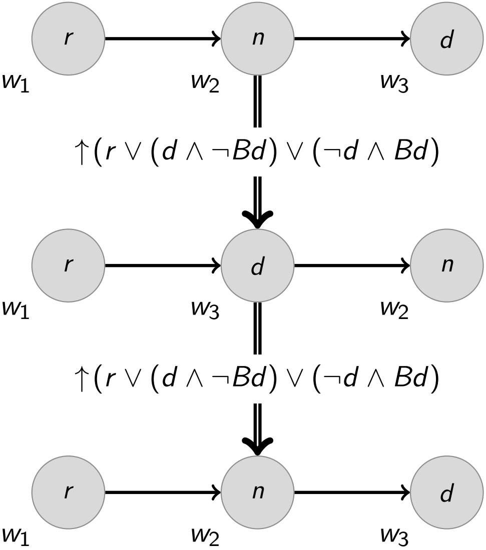
$!\varphi_1, !\varphi_2, !\varphi_3, \ldots, !\varphi_n$
always reaches a fixed-point

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

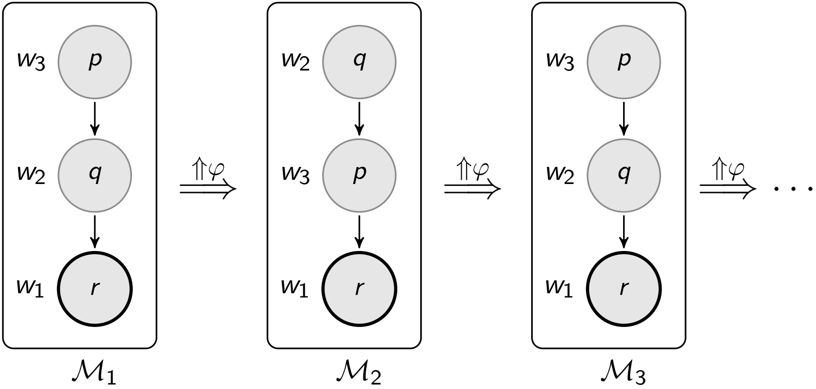
$\uparrow\varphi, \uparrow\varphi, \ldots$
Simple beliefs may never stabilize

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

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



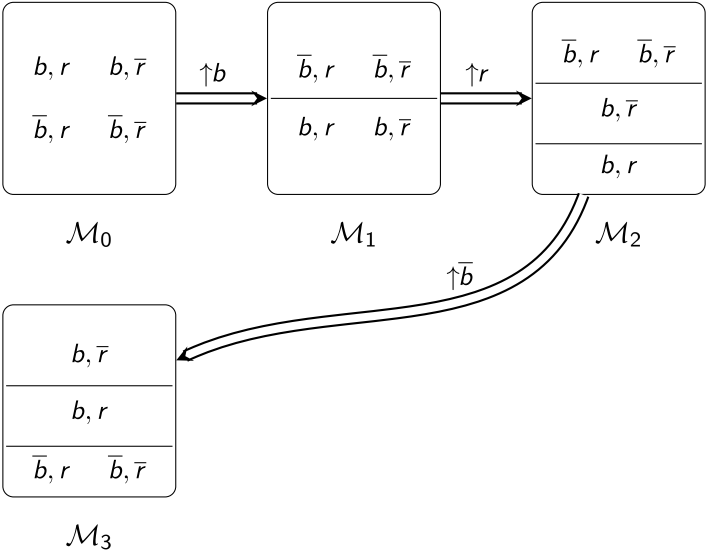
Suppose that you are in the forest and happen to a see strange-looking animal.

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.

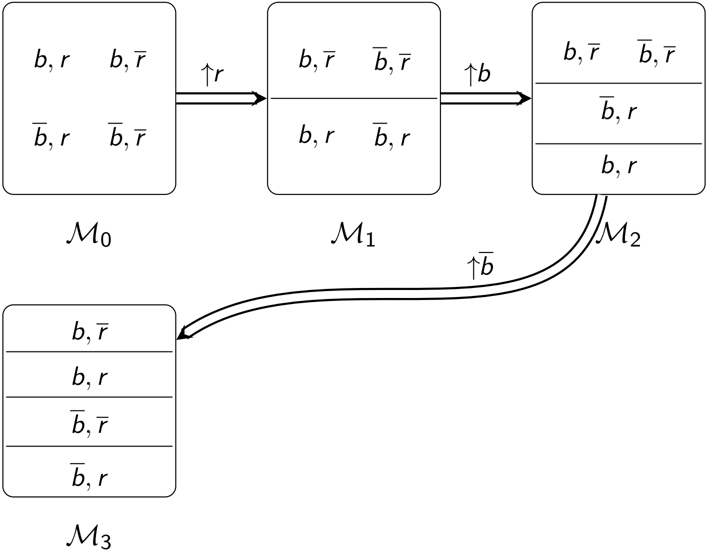
Suppose that you are in the forest and happen to a see strange-looking animal. You consult your animal guidebook and find a picture that seems to match the animal you see. The guidebook says that the animal is a type of bird, so that is what you conclude: The animal before you is a bird. After looking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

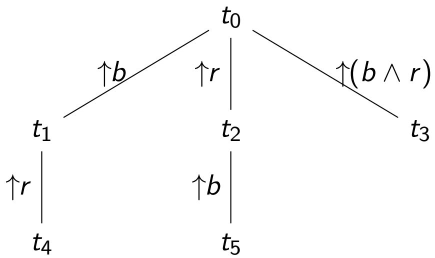

b, r $b, \overline{r}$
$\overline{b}, r$ $\overline{b}, \overline{r}$
$\mathcal{M}_0$
$\uparrow b$
$\overline{b}, r$ $\overline{b}, \overline{r}$
b, r $b, \overline{r}$
$\mathcal{M}_1$
$\uparrow r$
$\overline{b}, r$ $\overline{b}, \overline{r}$
$b, \overline{r}$
b, r
$\mathcal{M}_2$
$\uparrow \overline{b}$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

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 $L_1 \to L_2$, 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.

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 $(T_1 \wedge T_2) \to \neg H_1$ 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.*

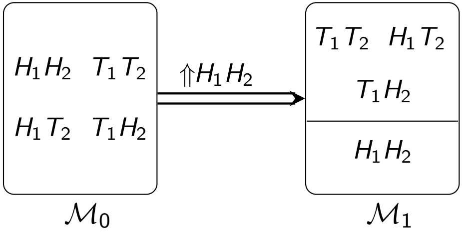
(Stalnaker, pg. 207–208)

Heuristic Diagnosis of Stalnaker's Example

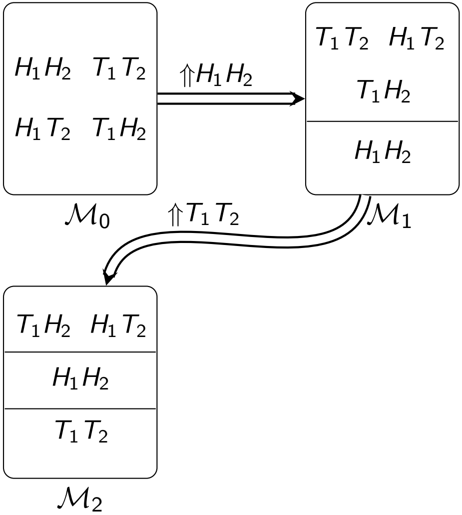
$$\boxed{\begin{array}{cc} H_1 H_2 & T_1 T_2 \\ H_1 T_2 & T_1 H_2 \end{array}}$$

$\mathcal{M}_0$

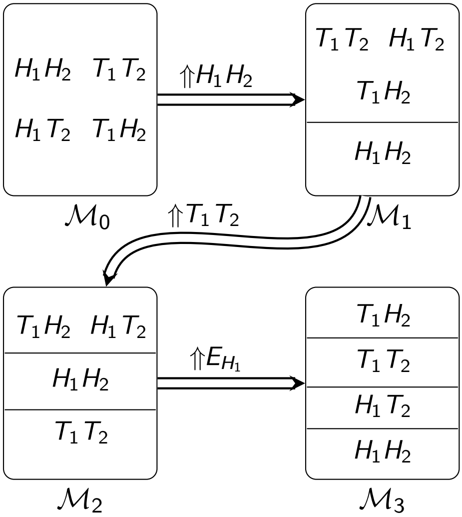
Heuristic Diagnosis of Stalnaker's Example



Heuristic Diagnosis of Stalnaker's Example



Heuristic Diagnosis of Stalnaker's Example

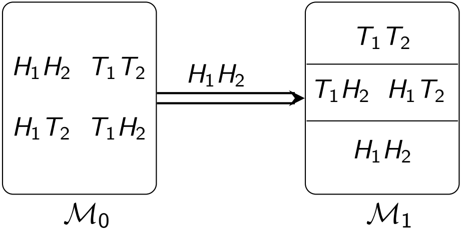


Heuristic Diagnosis of Stalnaker's Example

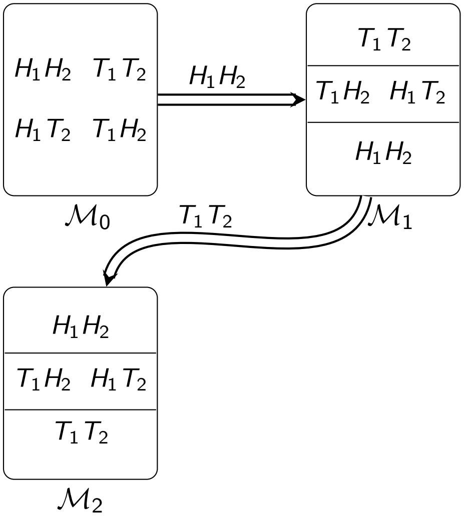
$$\boxed{\begin{array}{cc} H_1 H_2 & T_1 T_2 \\ H_1 T_2 & T_1 H_2 \end{array}}$$

$\mathcal{M}_0$

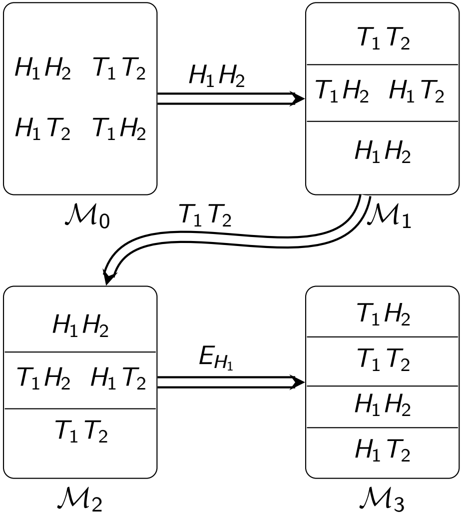
Heuristic Diagnosis of Stalnaker's Example



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.