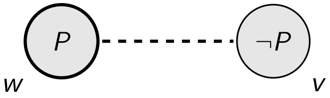
Ten Puzzles and Paradoxes about Knowledge and Belief

ESSLLI 2013, Düsseldorf

Wes Holliday **Eric Pacuit**

August 16, 2013

Taking Stock



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

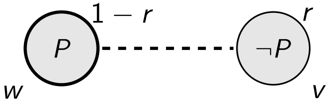
- wR_iv means v is compatible with everything i knows at w.

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

Truth:

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

Taking Stock



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

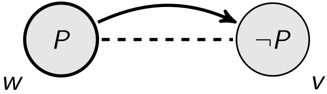
- $p_i : W \to [0,1]$ are probabilities, $\sim_i$ is an equivalence relation

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

Truth:

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

Taking Stock



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

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

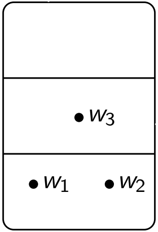
Language: $\varphi := p \mid \neg\varphi \mid \varphi\wedge\psi \mid K_i\varphi \mid B^{\varphi}\psi \mid [\preceq_i]\varphi$

Truth:

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

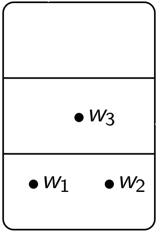
More on Plausibility Structures

- $w_1 \sim w_2 \sim w_3$



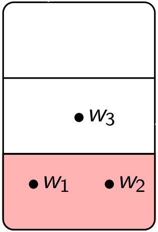
More on Plausibility Structures

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

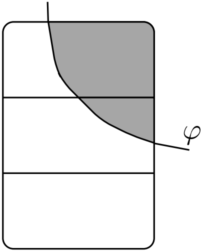


More on Plausibility Structures

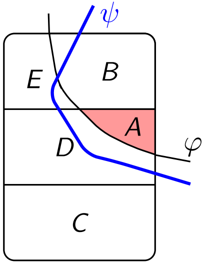
- $w_1 \sim w_2 \sim w_3$
- $w_1 \preceq w_2$ and $w_2 \preceq w_1$ (w_1 and w_2 are equi-plausbile)
- $w_1 \prec w_3$ ($w_1 \preceq w_3$ and $w_3 \not\preceq w_1$)
- $w_2 \prec w_3$ ($w_2 \preceq w_3$ and $w_3 \not\preceq w_2$)
- $\{w_1, w_2\} \subseteq Min_{\preceq}([w_i])$



More on Plausibility Structures

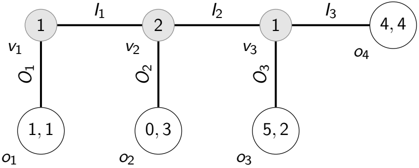


More on Plausibility Structures

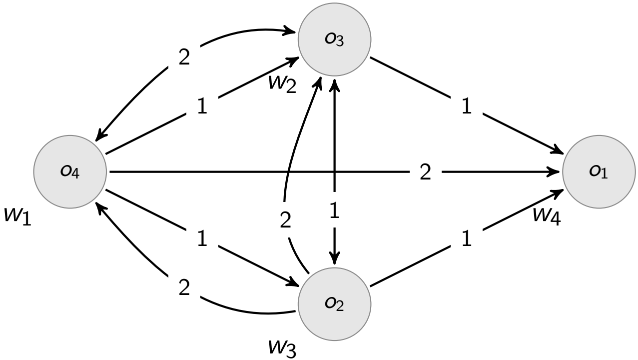


Conditional Belief: $B^{\varphi}\psi$

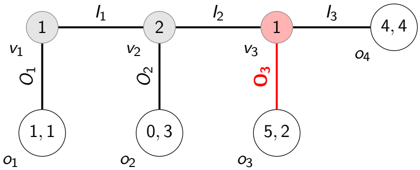
$$Min_{\preceq}(W \cap [\![\varphi]\!]_{\mathcal{M}}) \subseteq [\![\psi]\!]_{\mathcal{M}}$$



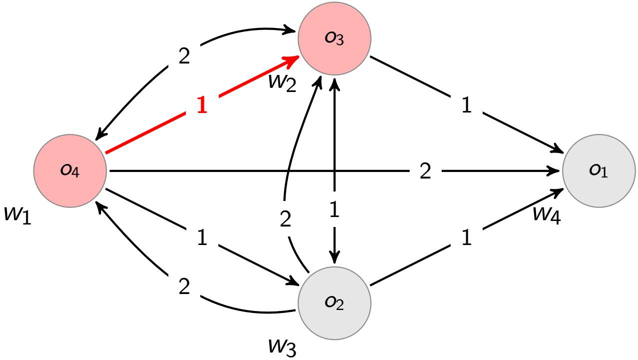
1
2
1
4, 4
l_1
l_2
l_3
v_1
v_2
v_3
o_4
O_1
O_2
O_3
1, 1
0, 3
5, 2
o_1
o_2
o_3



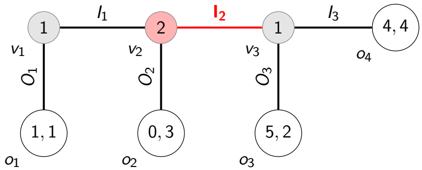
o_3
o_4
o_1
o_2
w_1
w_2
w_3
w_4
2
1
1
2
1
2
1
1
2
1



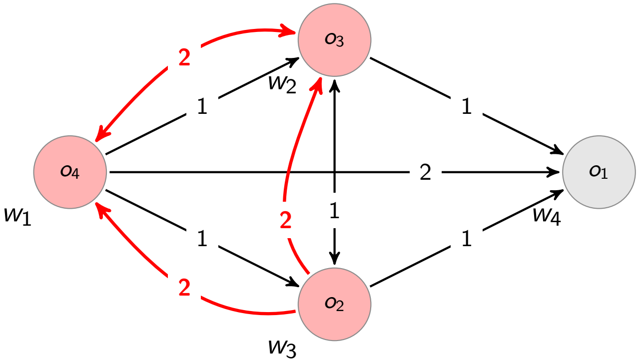
1
2
1
4, 4
l_1
l_2
l_3
v_1
v_2
v_3
o_4
O_1
O_2
O_3
1, 1
0, 3
5, 2
o_1
o_2
o_3



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



1
2
1
4, 4
l_1
l_2
l_3
v_1
v_2
v_3
o_4
O_1
O_2
O_3
1, 1
0, 3
5, 2
o_1
o_2
o_3



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

Game play as public announcemnets

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

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

The Dynamics of Rational Play

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

Hard vs. Soft Information in a Game

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

Hard vs. Soft Information in a Game

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

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

Hard vs. Soft Information in a Game

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

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

What belief revision policy leads to BI?

Dynamic Rationality: The event R that all players are *rational* changes during the play of the game.

Players are assumed to be "incurably optimistic" about the rationality of their opponents.

What belief revision policy leads to BI?

Dynamic Rationality: The event R that all players are *rational* changes during the play of the game.

Players are assumed to be "incurably optimistic" about the rationality of their opponents.

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

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

When is an example a counterexample?

EP, J.-W. Romeijn and P. Pedersen. *When is an Example a Counterexample?*. Proceedings of TARK, 2013.

Suppose that ESSLLI gave a final exam and you know that the conditional probability that you will complete your PhD and get a job in the next year given that you pass this exam is 60%.

Suppose that ESSLLI gave a final exam and you know that the conditional probability that you will complete your PhD and get a job in the next year given that you pass this exam is 60%.

If your friend tells you that she passed the exam, what is the probability that you assign to her completing her PhD and getting a job in the next year?

Suppose that ESSLLI gave a final exam and you know that the conditional probability that you will complete your PhD and get a job in the next year given that you pass this exam is 60%.

If your friend tells you that she passed the exam, what is the probability that you assign to her completing her PhD and getting a job in the next year? Is it 60%?

Suppose that ESSLLI gave a final exam and you know that the conditional probability that you will complete your PhD and get a job in the next year given that you pass this exam is 60%.

If your friend tells you that she passed the exam, what is the probability that you assign to her completing her PhD and getting a job in the next year? Is it 60%?

What if you are told by Patrick Blackburn that your friend passed the exam?

Suppose that ESSLLI gave a final exam and you know that the conditional probability that you will complete your PhD and get a job in the next year given that you pass this exam is 60%.

If your friend tells you that she passed the exam, what is the probability that you assign to her completing her PhD and getting a job in the next year? Is it 60%?

What if you are told by Patrick Blackburn that your friend passed the exam?

What if you read it on a list that was posted on the wall?

Evaluating counterexamples

. . . 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 dont 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.

Robert Stalnaker. *Iterated Belief Revision*. Erkenntnis 70, pp. 189209, 2009.

Belief Change

Belief Change

Consider the following beliefs of a rational agent:

- p_1 All Europeans swans are white.
- p_2 The bird caught in the trap is a swan.
- p_3 The bird caught in the trap comes from Sweden.
- p_4 Sweden is part of Europe.

Belief Change

Consider the following beliefs of a rational agent:

- p_1 All Europeans swans are white.
- p_2 The bird caught in the trap is a swan.
- p_3 The bird caught in the trap comes from Sweden.
- p_4 Sweden is part of Europe.

Thus, the agent believes:

- q The bird caught in the trap is white.

Belief Change

Consider the following beliefs of a rational agent:

- p_1 All Europeans swans are white.
- p_2 The bird caught in the trap is a swan.
- p_3 The bird caught in the trap comes from Sweden.
- p_4 Sweden is part of Europe.

Thus, the agent believes:

- q The bird caught in the trap is white.

Now suppose the rational agent—for example, You—learn that the bird caught in the trap is black ($\neg q$).

Belief Change, I

Consider the following beliefs of a rational agent:

- p_1 All Europeans swans are white.
- p_2 The bird caught in the trap is a swan.
- p_3 The bird caught in the trap comes from Sweden.
- p_4 Sweden is part of Europe.

Thus, the agent believes:

- q The bird caught in the trap is white.

Belief Change, I

Consider the following beliefs of a rational agent:

- p_1 All Europeans swans are white.
- p_2 The bird caught in the trap is a swan.
- p_3 The bird caught in the trap comes from Sweden.
- p_4 Sweden is part of Europe.

Thus, the agent believes:

- q The bird caught in the trap is white.

Question: How should the agent incorporate $\neg q$ into his belief state to obtain a consistent belief state?

Belief Change, I

Consider the following beliefs of a rational agent:

- p_1 All Europeans swans are white.
- p_2 The bird caught in the trap is a swan.
- p_3 The bird caught in the trap comes from Sweden.
- p_4 Sweden is part of Europe.

Thus, the agent believes:

- q The bird caught in the trap is white.

Question: How should the agent incorporate $\neg q$ into his belief state to obtain a consistent belief state?

Problem: Logical considerations alone are insufficient to answer this question! Why??

Belief Change, II

Consider the following beliefs of a rational agent:

- p_1 All Europeans swans are white.
- p_2 The bird caught in the trap is a swan.
- p_3 The bird caught in the trap comes from Sweden.
- p_4 Sweden is part of Europe.

Thus, the agent believes:

- q The bird caught in the trap is white.

Question: How should the agent incorporate $\neg q$ into his belief state to obtain a consistent belief state?
Problem: Logical considerations alone are insufficient to answer this question!
There are several logically distinct ways to incorporate $\neg q$!

Belief Change, II

What extralogical factors serve to determine what beliefs to give up and what beliefs to retain?

Belief Change, III

Belief revision is a matter of choice, and the choices are to be made in such a way that:

1. The resulting theory squares with the experience;
2. It is simple; and
3. The choices disturb the original theory as little as possible.

Belief Change, III

Belief revision is a matter of choice, and the choices are to be made in such a way that:

1. The resulting theory squares with the experience;
2. It is simple; and
3. The choices disturb the original theory as little as possible.

Research has relied on the following related guiding ideas:

1. When accepting a new piece of information, an agent should aim at a minimal change of his old beliefs.
2. If there are different ways to effect a belief change, the agent should give up those beliefs which are least entrenched.

Belief Revision

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

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

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

Counterexample to AGM 2

$$\varphi \in K * \varphi$$

Counterexample to AGM 2

$$\varphi \in K * \varphi$$

You are walking down a street and see someone holding a sign reading "The World will End Tomorrow", but you don't add this add this to your beliefs.

Counterexample to AGM 2

$$\varphi \in K * \varphi$$

You are walking down a street and see someone holding a sign reading "The World will End Tomorrow", but you don't add this add this to your beliefs.

Two people, Ann and Bob, are reliable sources of information on whether The Netherlands will win the world cup. They are equally reliable.

Counterexample to AGM 2

$$\varphi \in K * \varphi$$

You are walking down a street and see someone holding a sign reading "The World will End Tomorrow", but you don't add this add this to your beliefs.

Two people, Ann and Bob, are reliable sources of information on whether The Netherlands will win the world cup. They are equally reliable. AGM assumes that the most recent evidence that you received takes precedent. Ann says "yes" and a little bit later, Bob says "no".

Counterexample to AGM 2

$$\varphi \in K * \varphi$$

You are walking down a street and see someone holding a sign reading "The World will End Tomorrow", but you don't add this add this to your beliefs.

Two people, Ann and Bob, are reliable sources of information on whether The Netherlands will win the world cup. They are equally reliable. AGM assumes that the most recent evidence that you received takes precedent. Ann says "yes" and a little bit later, Bob says "no". Why should the, possibly arbitrary, order in which you receive the information give more weight to Bob's announcement?

Counterexample to AGM 2

$$\varphi \in K * \varphi$$

You are walking down a street and see someone holding a sign reading "The World will End Tomorrow", but you don't add this add this to your beliefs.

Two people, Ann and Bob, are reliable sources of information on whether The Netherlands will win the world cup. They are equally reliable. AGM assumes that the most recent evidence that you received takes precedent. Ann says "yes" and a little bit later, Bob says "no". Why should the, possibly arbitrary, order in which you receive the information give more weight to Bob's announcement? Is this a counterexample to AGM 1?

Counterexample to AGM 2

$$\varphi \in K * \varphi$$

You are walking down a street and see someone holding a sign reading "The World will End Tomorrow", but you don't add this add this to your beliefs.

Two people, Ann and Bob, are reliable sources of information on whether The Netherlands will win the world cup. They are equally reliable. AGM assumes that the most recent evidence that you received takes precedent. Ann says "yes" and a little bit later, Bob says "no". Why should the, possibly arbitrary, order in which you receive the information give more weight to Bob's announcement? Is this a counterexample to AGM 1? No (Why?)

Rott's Counterexample

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

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

So, if $\psi \in Cn(\{\varphi\})$, then $K * \varphi = Cn(K * \varphi \cup \{\psi\})$

Rott's Counterexample

There is an appointment to be made in a philosophy department. The position is a metaphysics position, and there are three main candidates: Andrew, Becker and Cortez.

1. Andrew is clearly the best metaphysician, but is weak in logic.
2. Becker is a very good metaphysician, also good in logic.
3. Cortez is a brilliant logician, but weak in metaphysics.

Rott's Counterexample

There is an appointment to be made in a philosophy department. The position is a metaphysics position, and there are three main candidates: Andrew, Becker and Cortez.

1. Andrew is clearly the best metaphysician, but is weak in logic.
2. Becker is a very good metaphysician, also good in logic.
3. Cortez is a brilliant logician, but weak in metaphysics.

Scenario 1: Paul is told by the dean, that the chosen candidate is either Andrew or Becker. Since Andrew is clearly the better metaphysician of the two, Paul concludes that the winning candidate will be Andrew.

Rott's Counterexample

1. Andrew is clearly the best metaphysician, but is weak in logic.
2. Becker is a very good metaphysician, also good in logic.
3. Cortez is a brilliant logician, but weak in metaphysics.

Scenario 2: Paul is told by the dean that the chosen candidate is either Andrew, Becker or Cortez.

Rott's Counterexample

1. Andrew is clearly the best metaphysician, but is weak in logic.
2. Becker is a very good metaphysician, also good in logic.
3. Cortez is a brilliant logician, but weak in metaphysics.

Scenario 2: Paul is told by the dean that the chosen candidate is either Andrew, Becker or Cortez.

" This piece of information sets off a rather subtle line of reasoning. Knowing that Cortez is a splendid logician, but that he can hardly be called a metaphysician, Paul comes to realize that his background assumption that expertise in the field advertised is the decisive criterion for the appointment cannot be upheld. Apparently, competence in logic is regarded as a considerable asset by the selection committee." Paul concludes Becker will be hired.

“...Rott seems to take the point about meta-information to explain why the example conflicts with the theoretical principles,

"...Rott seems to take the point about meta-information to explain why the example conflicts with the theoretical principles, whereas I want to conclude that it shows why the example does not conflict with the theoretical principles, since I take the relevance of the meta-information to show that the conditions for applying the principles in question are not met by the example.

"...Rott seems to take the point about meta-information to explain why the example conflicts with the theoretical principles, whereas I want to conclude that it shows why the example does not conflict with the theoretical principles, since I take the relevance of the meta-information to show that the conditions for applying the principles in question are not met by the example.... I think proper attention to the relation between concrete examples and the abstract models will allow us to reconcile some of the beautiful properties with the complexity of concrete reasoning."

(Stalnaker, 204)

Iterated Revision

Iterated Revision

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

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

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.

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 $L_1 \to L_2$, 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 $L_1 \to L_2$, 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 $L_1 \to L_2$, 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 $L_1 \to L_2$, 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

- 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

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

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 $K' * (T_1 \wedge T_2) * H_1 = K' * H_1$.

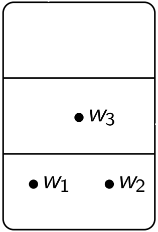
  Yet, since $H_1 \wedge H_2 \in K'$ and H_1 is consistent with H_2, we must have $H_1 \wedge H_2 \in K' * H_1$, which yields a conflict with the assumption that $H_1 \wedge T_2 \in K' * (T_1 \wedge T_2) * H_1$.

...[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.*

(pg. 207–208)

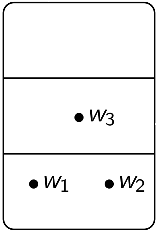
Informative Actions

- $w_1 \sim w_2 \sim w_3$



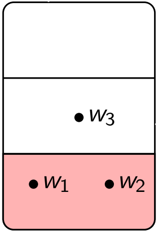
Informative Actions

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

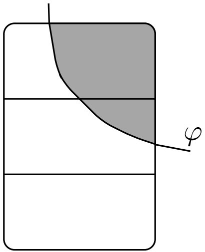


Informative Actions

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

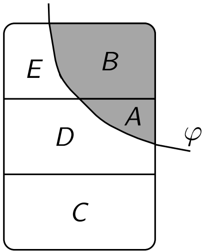


Informative Actions



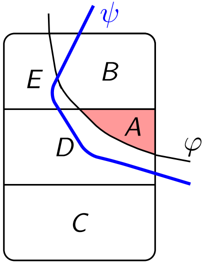
Incorporate the new information φ

Informative Actions



Incorporate the information that φ

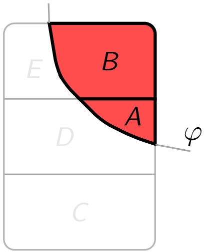
Informative Actions



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

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

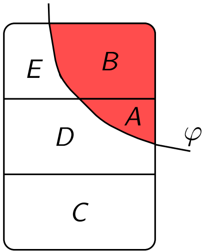
Informative Actions



Public Announcement: Information from an infallible source

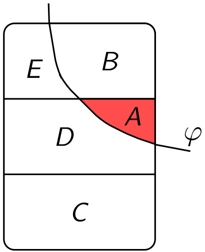
$(!\varphi)$: $A \prec_i B$

Informative Actions



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

Informative Actions



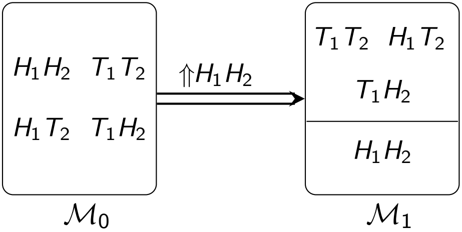
Conservative Upgrade: $(\uparrow\varphi)$: $A \prec_i C \prec_i D \prec_i B \cup E$

Heuristic Diagnosis of Stalnaker's Example

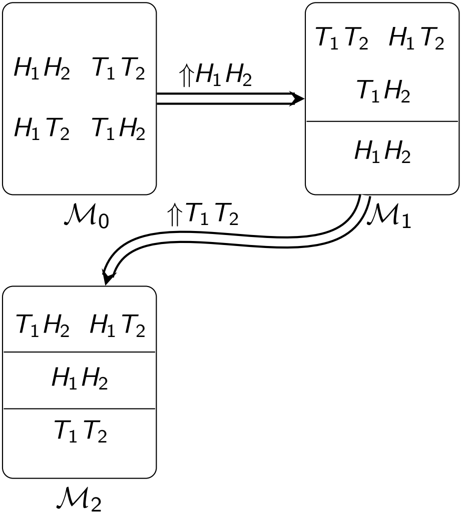
$$\boxed{\begin{array}{cc} H_1 H_2 & T_1 T_2 \\ H_1 T_2 & T_1 H_2 \end{array}}$$

$\mathcal{M}_0$

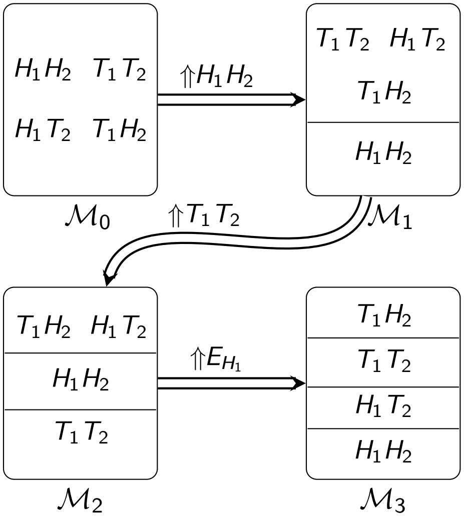
Heuristic Diagnosis of Stalnaker's Example



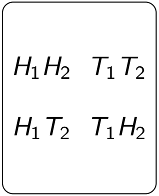
Heuristic Diagnosis of Stalnaker's Example



Heuristic Diagnosis of Stalnaker's Example

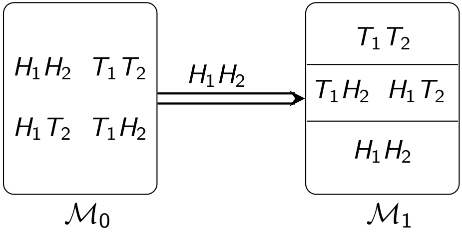


Heuristic Diagnosis of Stalnaker's Example

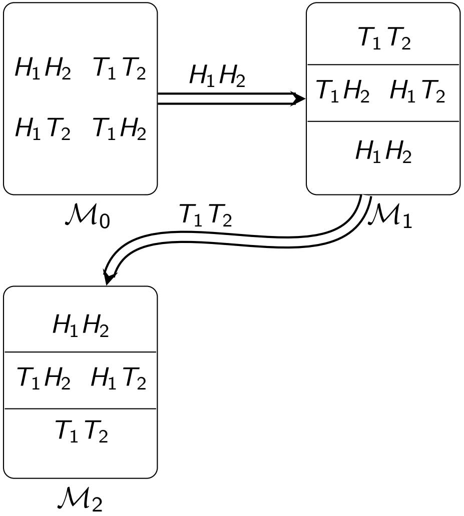


$\mathcal{M}_0$

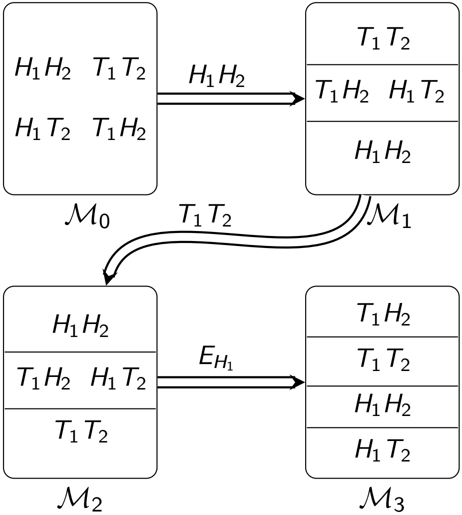
Heuristic Diagnosis of Stalnaker's Example



Heuristic Diagnosis of Stalnaker's Example



Heuristic Diagnosis of Stalnaker's Example



What do the Examples Demonstrate?

1. There is no suitable way to formalize the scenario in such a way that the AGM postulates (possibly including postulates of iterated belief revision) can be saved;
2. The AGM framework can be made to agree with the scenario but does not furnish a systematic way to formalize the relevant meta-information; or
3. There is a suitable and systematic way to make the meta-information explicit, but this is something that the AGM framework cannot properly accommodate. Our interest in this paper is the third response, which is concerned with the absence of guidelines for applying the theory of belief revision.

There are different kinds of independence—conceptual, causal and epistemic—that interact, and one might be able to say more about constraints on rational belief revision if one had a model theory in which causal-counterfactual and epistemic information could both be represented. There are familiar problems, both technical and philosophical, that arise when one tries to make meta-information explicit, since it is self-locating (and auto-epistemic) information, and information about changing states of the world. (pg. 208)

A Bayesian Model

1. The reports are independent, the content of the reports are very probable, and the content of subsequent reports are even more probable, thereby canceling out the impact of preceding reports.
2. The meta-information in the example may be such that earlier reports are dependent in a weak sense, so that Elmers report also encourages the agent to change her mind about the coin in the second box.
3. With some imagination, we can also provide a model in which the pairs of reports are independent in the strictest sense, and in which Elmers report is fully responsible for the belief change regarding both coins.

Discussion, I

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

Discussion, II

- Belief change by conditioning: There seems to be a trade-off between a rich set of states and event structure, and a rich theory of doxastic actions. How should we resolve this trade-off when analyzing counterexamples to postulates of belief changes over time?

Discussion, III

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

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