Logical and Probabilistic Models of Belief Change

Eric Pacuit

Department of Philosophy
University of Maryland, College Park
pacuit.org

July 15, 2016

Plan

Day 1 Introduction to belief revision, AGM, possible worlds models, Bayesian models (time permitted)

Day 2 Bayesian models (continued), Justifying Bayesian models (Dutch books, Accuracy-based arguments), Updating probabilities

Day 3 The value of learning, Lottery Paradox, Preface Paradox, Review Paradox, Context shifts, Becoming aware

Day 4 The value of learning, Lottery Paradox, Preface Paradox, Review Paradox, Context shifts, Becoming aware (continued)

Day 5 Iterated Belief Revision, Agreement Theorems pacuit.org/nasslli2016/belrev/

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

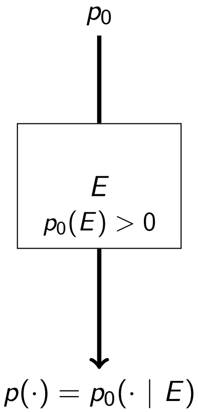
- Epistemic states: AGM, Plausibility Models, Bayesian Model (and the many variations)

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

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

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

- *How* did you find out that φ?
  - Directly observed φ
  - Indirectly observed φ
  - Told 'φ' (by an epistemic peer, by an expert, by a trusted individual)
  - ...
- Belief change over time


p_0
E
$p_0(E) > 0$
$p(\cdot) = p_0(\cdot \mid E)$

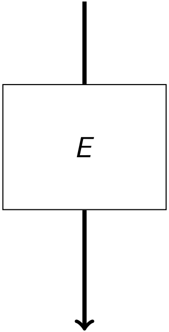
p_0

$$(E_1 : q_1, \ldots, E_k : q_k)$$

$\{E_i\}$ is a partition, $\sum_i q_i = 1$

$$p(\cdot) = \sum_i q_i * p_0(\cdot \mid E_i)$$

$$p_0(\cdot, \top)$$



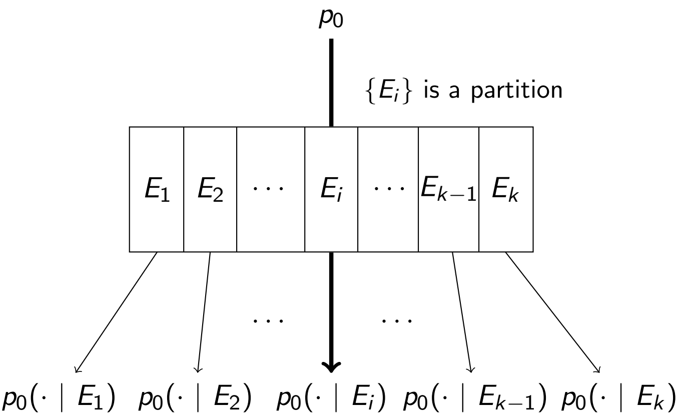
$$p(\cdot) = p_0(\cdot, E)$$

p_0

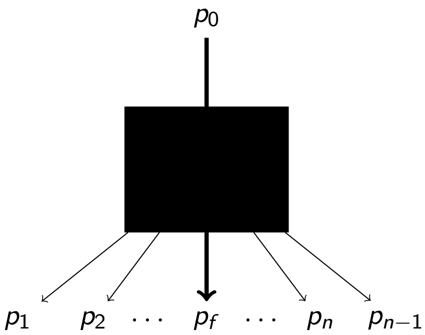
$\mathcal{C}$: set of constraints

p satisfies $\mathcal{C}$

p minimizes $I(p, p_0) = \sum_i p_1(x_i) \log \frac{p_1(x_i)}{p_0(x_i)}$



p_0
$\{E_i\}$ is a partition
E_1 E_2 $\cdots$ E_i $\cdots$ E_{k-1} E_k
$\cdots$ $\cdots$
$p_0(\cdot \mid E_1)$ $p_0(\cdot \mid E_2)$ $p_0(\cdot \mid E_i)$ $p_0(\cdot \mid E_{k-1})$ $p_0(\cdot \mid E_k)$



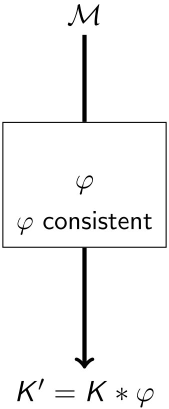
(Martingale Property) $\quad p_0(A \mid p_f) = p_f(A)$

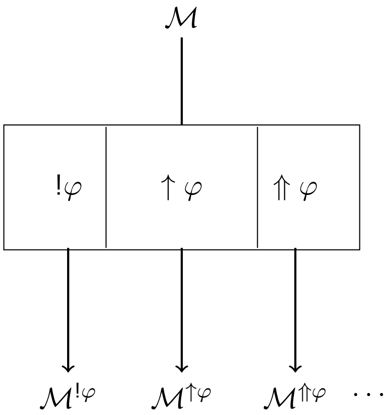
K

φ

φ consistent

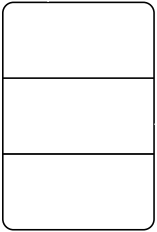
$K' = K * \varphi$


$\mathcal{M}$
φ
φ consistent
$K' = K * \varphi$

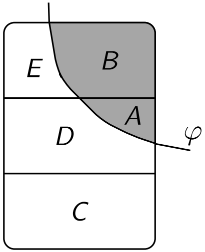

$\mathcal{M}$
$!\varphi$
$\uparrow\varphi$
$\Uparrow\varphi$
$\mathcal{M}^{!\varphi}$
$\mathcal{M}^{\uparrow\varphi}$
$\mathcal{M}^{\Uparrow\varphi}$
$\cdots$

Iterated revision

Informative Actions

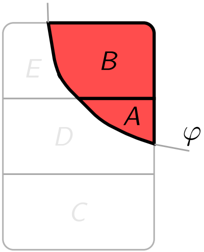


Informative Actions



Incorporate the new information φ

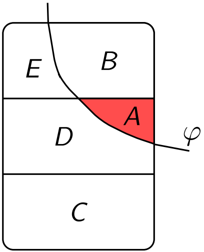
Informative Actions



Public Announcement: Information from an infallible source

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

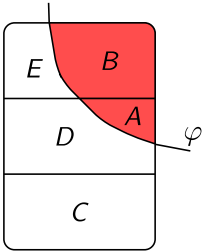
Informative Actions



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

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

Informative Actions



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

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

$$\mathcal{M}_0 \overset{!\varphi_1}{\Longrightarrow} \mathcal{M}_1 \overset{!\varphi_2}{\Longrightarrow} \mathcal{M}_2 \overset{!\varphi_3}{\Longrightarrow} \cdots \overset{!\varphi_n}{\Longrightarrow} \mathcal{M}_f$$

initial model

fixed-point

$$\underset{\text{initial model}}{\mathcal{M}_0} \overset{\Uparrow\varphi_1}{\Longrightarrow} \mathcal{M}_1 \overset{\Uparrow\varphi_2}{\Longrightarrow} \mathcal{M}_2 \overset{\Uparrow\varphi_3}{\Longrightarrow} \cdots \overset{\Uparrow\varphi_n}{\Longrightarrow} \underset{\text{fixed-point}}{\mathcal{M}_f}$$

$$\mathcal{M}_0 \overset{!\varphi_1}{\Longrightarrow} \mathcal{M}_1 \overset{\Uparrow\varphi_2}{\Longrightarrow} \mathcal{M}_2 \overset{\uparrow\varphi_3}{\Longrightarrow} \cdots \overset{\Uparrow\varphi_n}{\Longrightarrow} \mathcal{M}_f$$

initial model

fixed-point

$$\underset{\text{initial model}}{\mathcal{M}_0} \overset{\tau(\varphi_1)}{\Longrightarrow} \mathcal{M}_1 \overset{\tau(\varphi_2)}{\Longrightarrow} \mathcal{M}_2 \overset{\tau(\varphi_3)}{\Longrightarrow} \cdots \overset{\tau(\varphi_n)}{\Longrightarrow} \underset{\text{fixed-point}}{\mathcal{M}_f}$$

Where do the φ_k come from?

Dynamic Characterization of Informational Attitudes

$!\varphi_1, !\varphi_2, !\varphi_3, \ldots, !\varphi_n$
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.

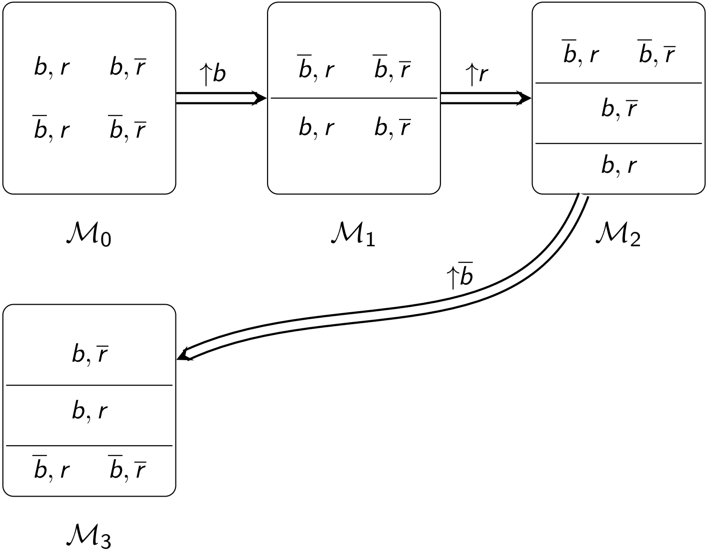
Suppose that you are in the forest and happen to a see strange-looking animal.

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.

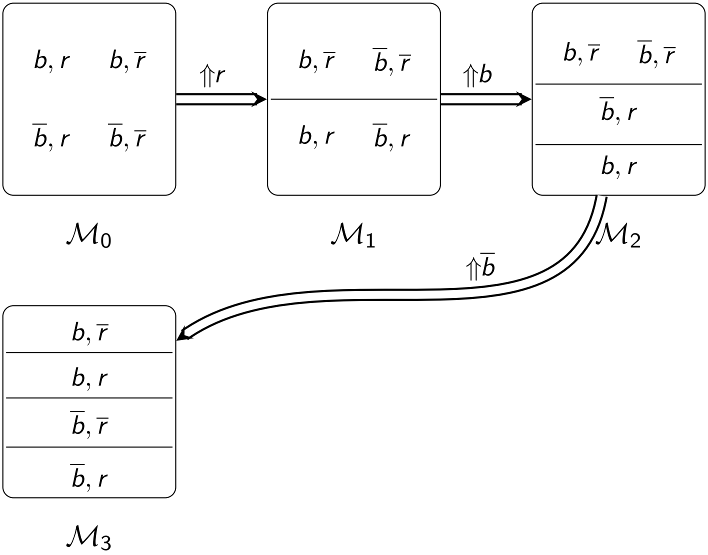
Suppose that you are in the forest and happen to a see strange-looking animal. You consult your animal guidebook and find a picture that seems to match the animal you see. The guidebook says that the animal is a type of bird, so that is what you conclude: The animal before you is a bird. After looking 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
$b, \overline{r}$
$\overline{b}, r$
$\overline{b}, \overline{r}$
$\uparrow b$
$\overline{b}, r$
$\overline{b}, \overline{r}$
b, r
$b, \overline{r}$
$\uparrow r$
$\overline{b}, r$
$\overline{b}, \overline{r}$
$b, \overline{r}$
b, r
$\mathcal{M}_0$
$\mathcal{M}_1$
$\mathcal{M}_2$
$\uparrow \overline{b}$
$b, \overline{r}$
b, r
$\overline{b}, r$
$\overline{b}, \overline{r}$
$\mathcal{M}_3$

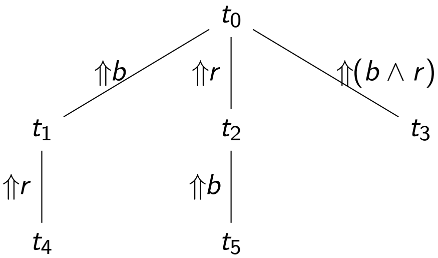
Note that in the last model, $\mathcal{M}_3$, the agent does not believe that the bird is red.

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

Two Postulates of Iterated Revision

I1 If $B \in Cn(\{A\})$ then $(K * B) * A = K * A$.

I2 If $\neg B \in Cn(\{A\})$ then $(K * A) * B = K * B$

Two Postulates of Iterated Revision

I1 If $B \in Cn(\{A\})$ then $(K * B) * A = K * A$.

I2 If $\neg B \in Cn(\{A\})$ then $(K * A) * B = K * B$

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

Two Postulates of Iterated Revision

I1 If $B \in Cn(\{A\})$ then $(K * B) * A = K * A$.

I2 If $\neg B \in Cn(\{A\})$ then $(K * A) * B = K * B$

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

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

I3 If $B \in K * A$ then $B \in (K * B) * A$.

I4 If $\neg B \notin K * A$ then $\neg B \notin (K * B) * A$.

Robert Stalnaker. *Iterated Belief Revision*. Erkenntnis 70, pp. 189 - 209, 2009.

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 $L_1 \rightarrow L_2$, but not the converse)
- Suppose I receive $L_1 \wedge L_2$, this does not change the story.
- Suppose I learn that L_2. This is irrelevant to Carla's report, but it means either Ann or Bob is wrong.

Stalnaker's Counterexample to I1

UUU	DDD
UUD	DDU
UDU	DUD
UDD	DUU

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

Stalnaker's Counterexample to I1

UUU	DDD
UUD	DDU
UDU	DUD
UDD	DUU

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

Stalnaker's Counterexample to I1

UUU	DDD
UUD	DDU
UDU	DUD
UDD	DUU

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

EP, P. Pedersen and J.-W. Romeijn. *When is an example and counterexample?*. Proceedings of TARK, 2013.

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.

What Do the Examples Demonstrate?

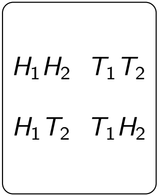
1. There is no suitable way to formalize the scenario in such a way that the AGM postulates (possibly including postulates of iterated belief revision) can be saved;

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.

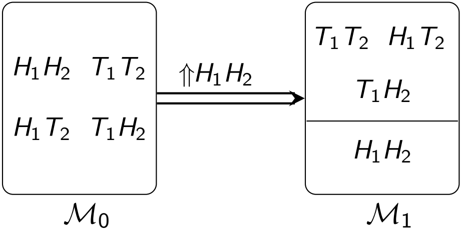
Our interest in this paper is the third response, which is concerned with the absence of guidelines for applying the theory of belief revision.

Heuristic Diagnosis of Stalnaker's Example

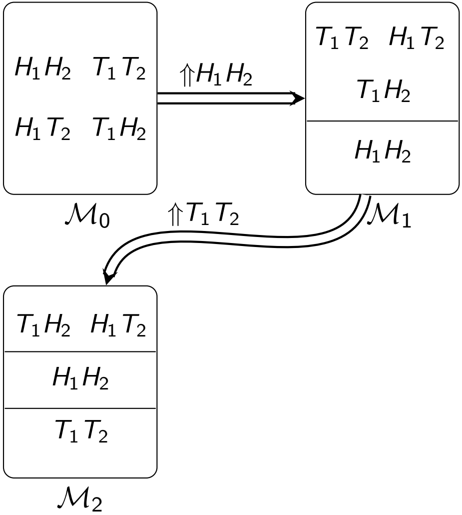


$\mathcal{M}_0$

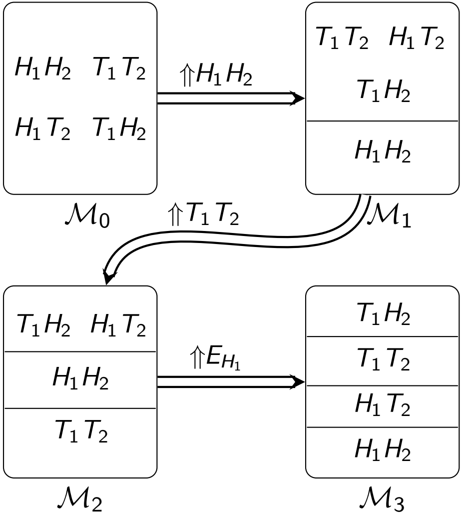
Heuristic Diagnosis of Stalnaker's Example



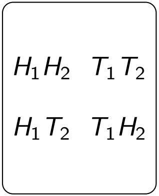
Heuristic Diagnosis of Stalnaker's Example



Heuristic Diagnosis of Stalnaker's Example

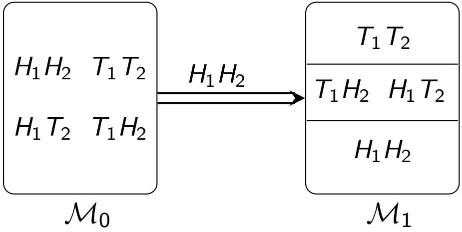


Heuristic Diagnosis of Stalnaker's Example

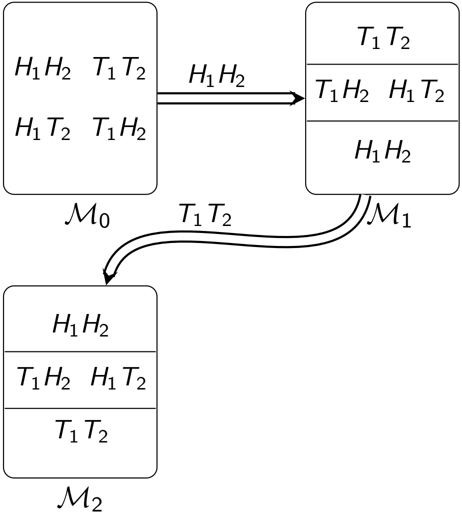


$\mathcal{M}_0$

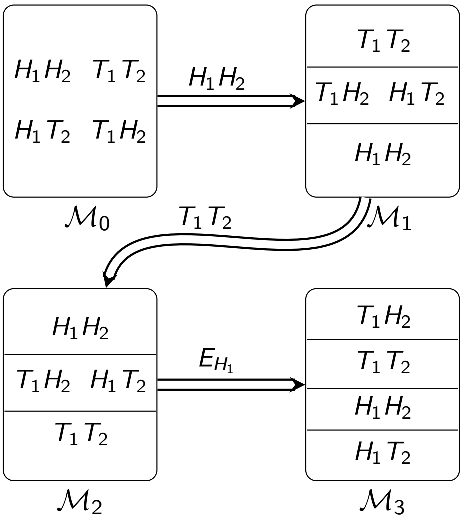
Heuristic Diagnosis of Stalnaker's Example



Heuristic Diagnosis of Stalnaker's Example



Heuristic Diagnosis of Stalnaker's Example



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

We accommodate the counterexamples in two steps:

A Bayesian Model

We accommodate the counterexamples in two steps:

1. We provide a Bayesian model in which presuppositions on order and dependence of the reports can be made explicit.
2. The qualitative and diachronic character of belief revision can be replicated by an extension to nonstandard probability assignments.

A Bayesian Model

We accommodate the counterexamples in two steps:

1. We provide a Bayesian model in which presuppositions on order and dependence of the reports can be made explicit.
2. The qualitative and diachronic character of belief revision can be replicated by an extension to nonstandard probability assignments.

Apart from this we refined the event structure of reports and states.

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

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

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

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

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

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

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

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

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

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

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

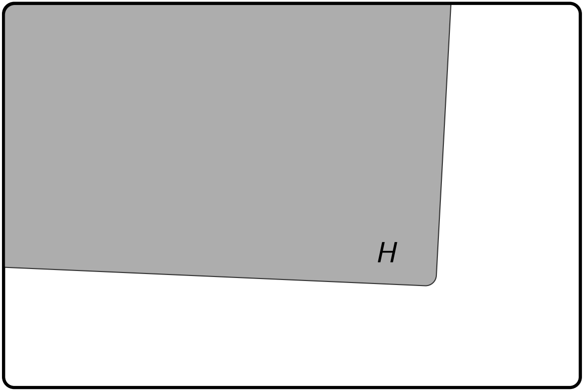
EP. *Dynamics for Probabilistic Common Belief*. Studies in Logic, 2015.

Starting with the same premises, using (for example) first-order logic, two agents cannot disagree about a conclusion.

Starting with the same probability, using (for example) strict conditionalization, two agents cannot disagree about their posterior probability given the same evidence.

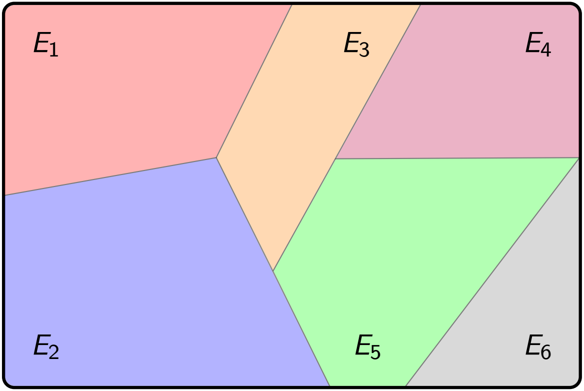
Aumann's Agreeing to Disagree Theorem. Suppose that n agents share a common prior and have different private information. If there is common knowledge of the posteriors of a fixed event, then the posteriors must be equal.

Robert Aumann. *Agreeing to Disagree*. Annals of Statistics **4(6)**, pgs. 1236-1239 (1976).

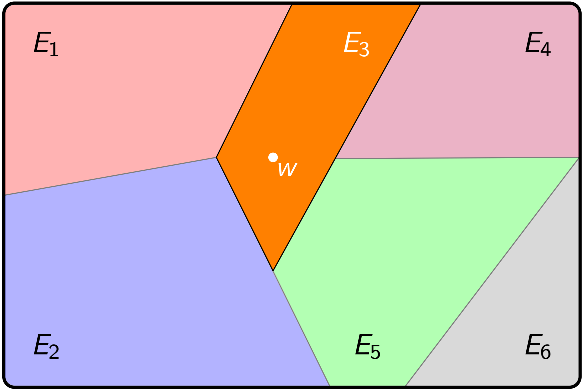


An **event/proposition** is a (definable) subset $H \subseteq W$.

A σ-**algebra** is the collection of events/propositions (closed under countable unions and complementation)



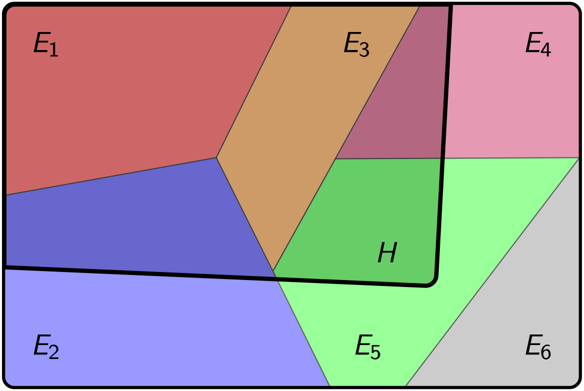
An **experiment/question/set of signals** is a partition $\mathcal{E}$ on W.



An **experiment/question/set of signals** is a partition $\mathcal{E}$ on W.

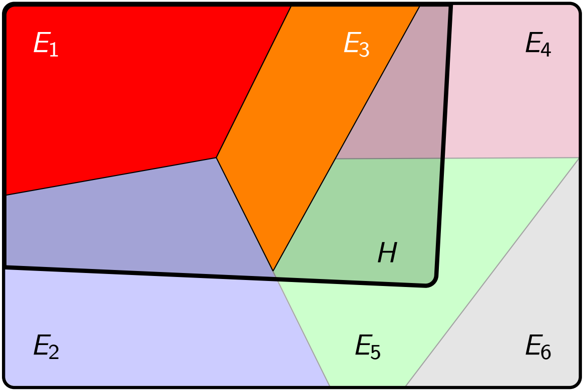
If $w \in W$, let $\mathcal{E}[w] = E$ where $w \in E \in \mathcal{E}$.

E.g, if $\mathcal{E} = \{E_1, E_2, E_3, E_4, E_5, E_6\}$, then $\mathcal{E}[w] = E_3$

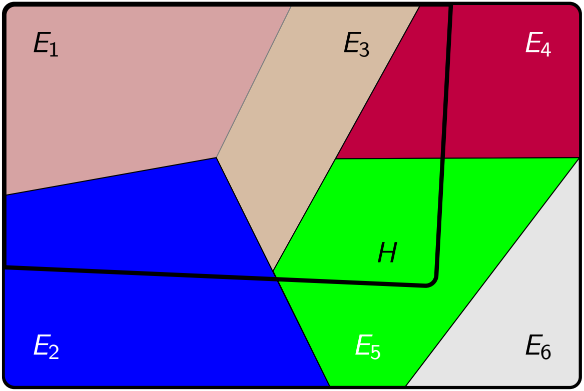


$K_{\mathcal{E}} : \wp(W) \to \wp(W)$, where for $H \subseteq W$,

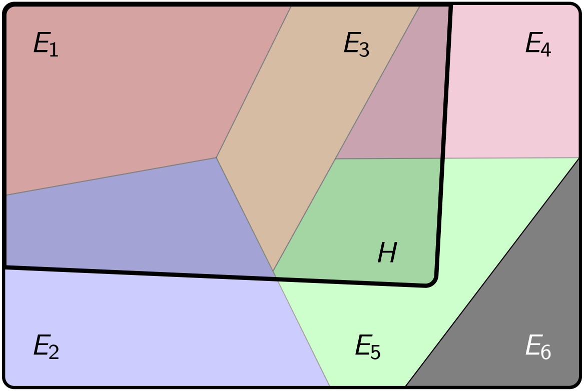
$$K_{\mathcal{E}}(H) = \{w \mid \mathcal{E}[w] \subseteq H\}$$



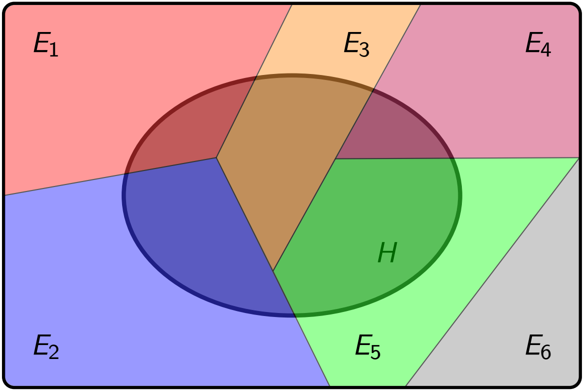
$$K_{\mathcal{E}}(H) \quad = \quad E_1 \cup E_3$$



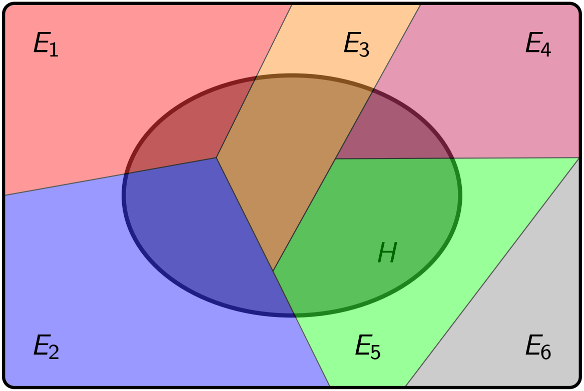
$$
\begin{aligned}
K_{\mathcal{E}}(H) &= E_1 \cup E_3 \\
-K_{\mathcal{E}}(H) \cap -K_{\mathcal{E}}(-H) &= E_2 \cup E_4 \cup E_5
\end{aligned}
$$



$$
\begin{aligned}
K_{\mathcal{E}}(H) &= E_1 \cup E_3 \\
-K_{\mathcal{E}}(H) \cap -K_{\mathcal{E}}(-H) &= E_2 \cup E_4 \cup E_5 \\
K_{\mathcal{E}}(-H) &= E_6
\end{aligned}
$$

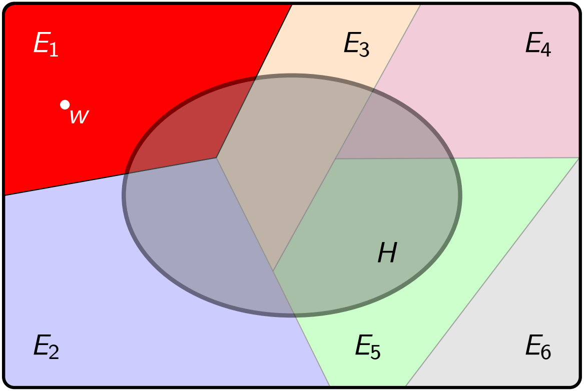


If p is a probability on W (with respect to a σ-algebra $\mathcal{F}$)



If p is a probability on W (with respect to a σ-algebra $\mathcal{F}$)

The posterior at w with respect to $\mathcal{E}$ is $p_{\mathcal{E},w}(H) = p(H \mid \mathcal{E}[w])$



If p is a probability on W (with respect to a σ-algebra $\mathcal{F}$)

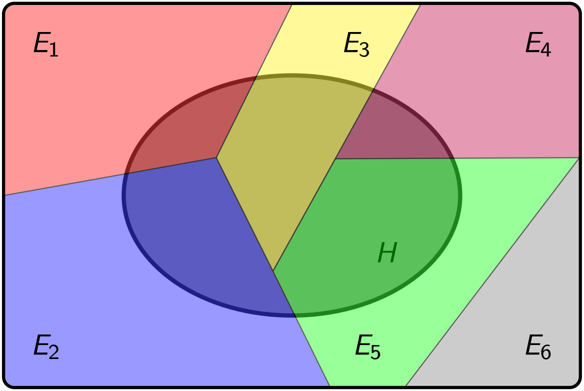
The posterior at w with respect to $\mathcal{E}$ is $p_{\mathcal{E},w}(H) = p(H \mid \mathcal{E}[w])$

E.g., $p_{\mathcal{E},w}(H) = p(H \mid E_1)$

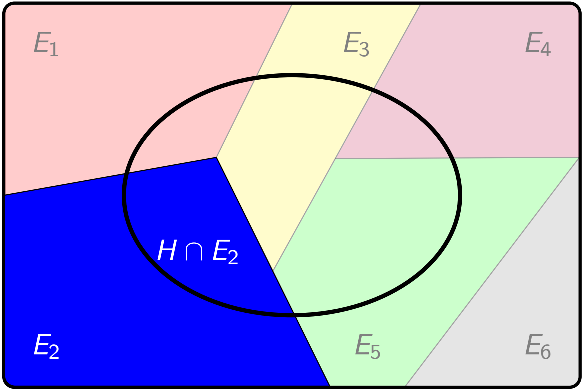
A basic result about probabilities.

For any finite partition $\mathcal{E} = \{E_i\}$ of W and an event H,

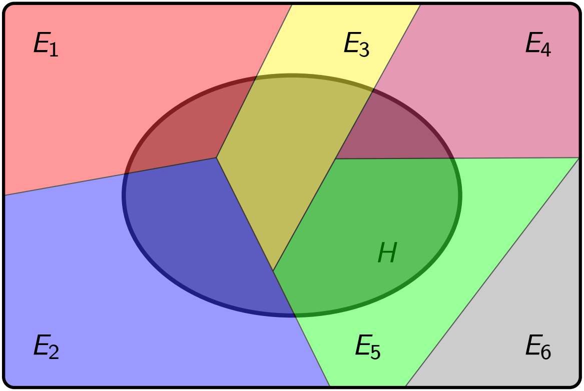
$$p(H) = \sum_i p(E_i)p(H \mid E_i)$$



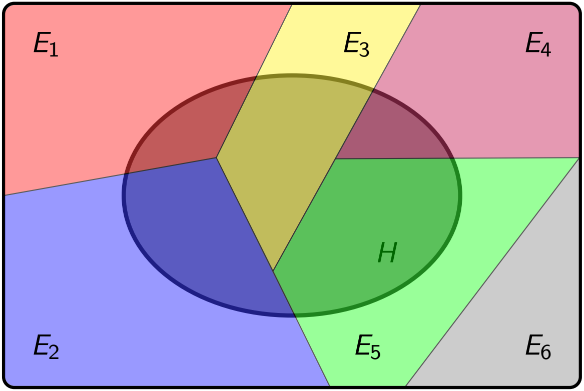
$$p(H) \quad = \quad p(H \cap E_1) + \cdots + p(H \cap E_6)$$



$$p(H) \quad = \quad p(H \cap E_1) + p(H \cap E_2) + \cdots + p(H \cap E_6)$$



$$
\begin{aligned}
p(H) &= p(H \cap E_1) + \cdots + p(H \cap E_6) \\
&= \frac{p(E_1)}{p(E_1)} p(H \cap E_1) + \cdots + \frac{p(E_6)}{p(E_6)} p(H \cap E_6)
\end{aligned}
$$

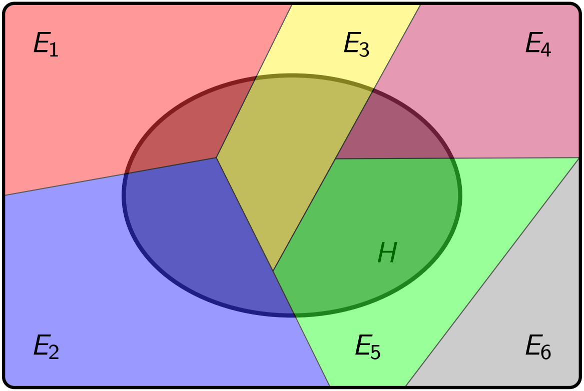


$$
\begin{aligned}
p(H) &= p(H \cap E_1) + \cdots + p(H \cap E_6) \\
&= \tfrac{p(E_1)}{p(E_1)} p(H \cap E_1) + \cdots + \tfrac{p(E_6)}{p(E_6)} p(H \cap E_6) \\
&= \textstyle\sum_i p(E_i) P(H \mid E_i)
\end{aligned}
$$

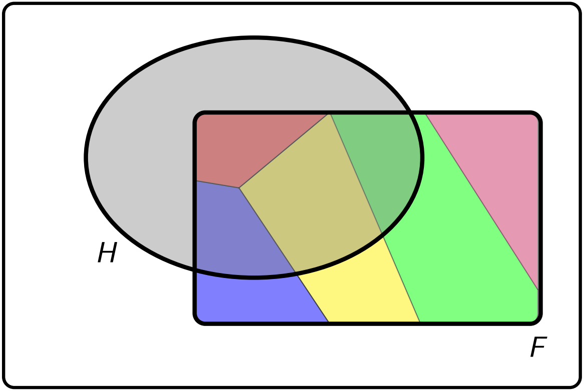
A basic result about probabilities.

For any finite partition $\mathcal{E} = \{E_i\}$ of F and and event H,

$$p(H \mid F) = \sum_i p(E_i \mid F)p(H \mid E_i)$$



$$\begin{aligned} p(H \mid W) &= \textstyle\sum_i p(E_i \mid W) p(H \mid E_i \cap W) \\ &= \textstyle\sum_i p(E_i \mid W) p(H \mid E_i) \end{aligned}$$



$$\begin{aligned} p(H \mid F) &= \textstyle\sum_i p(E_i \mid F)p(H \mid E_i \cap F) \\ &= \textstyle\sum_i p(E_i \mid F)p(H \mid E_i) \end{aligned}$$

Common Knowledge

"*Common Knowledge*" is informally described as what any fool would know: It encompasses what is relevant, agreed upon, established by precedent, assumed, being attended to, salient, or in the conversational record.

"*Common Knowledge*" is informally described as what any fool would know: It encompasses what is relevant, agreed upon, established by precedent, assumed, being attended to, salient, or in the conversational record.

It is not common knowledge who defined 'Common Knowledge'...

The first formal definition of common knowledge?

M. Friedell. *On the Structure of Shared Awareness*. Behavioral Science (1969).

R. Aumann. *Agreeing to Disagree*. Annals of Statistics (1976).

The first formal definition of common knowledge?

M. Friedell. *On the Structure of Shared Awareness*. Behavioral Science (1969).

R. Aumann. *Agreeing to Disagree*. Annals of Statistics (1976).

The first rigorous analysis of common knowledge

D. Lewis. *Convention, A Philosophical Study*. 1969.

The first formal definition of common knowledge?

M. Friedell. *On the Structure of Shared Awareness*. Behavioral Science (1969).

R. Aumann. *Agreeing to Disagree*. Annals of Statistics (1976).

The first rigorous analysis of common knowledge

D. Lewis. *Convention, A Philosophical Study*. 1969.

Fixed-point definition: $\gamma :=$ *i* and *j* know that (φ and γ)

G. Harman. *Review of Linguistic Behavior*. Language (1977).

J. Barwise. *Three views of Common Knowledge*. TARK (1987).

The first formal definition of common knowledge?

M. Friedell. *On the Structure of Shared Awareness*. Behavioral Science (1969).

R. Aumann. *Agreeing to Disagree*. Annals of Statistics (1976).

The first rigorous analysis of common knowledge

D. Lewis. *Convention, A Philosophical Study*. 1969.

Fixed-point definition: $\gamma :=$ i and j know that (φ and γ)

G. Harman. *Review of Linguistic Behavior*. Language (1977).

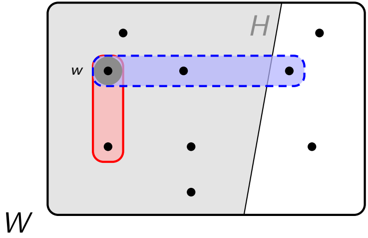
J. Barwise. *Three views of Common Knowledge*. TARK (1987).

Shared situation: There is a *shared situation s* such that (1) *s* entails φ, (2) *s* entails everyone knows φ, plus other conditions

H. Clark and C. Marshall. *Definite Reference and Mutual Knowledge*. 1981.

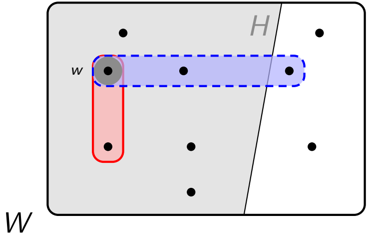
M. Gilbert. *On Social Facts*. Princeton University Press (1989).

P. Vanderschraaf and G. Sillari. *"Common Knowledge", The Stanford Encyclopedia of Philosophy (2009)*.
http://plato.stanford.edu/entries/common-knowledge/.



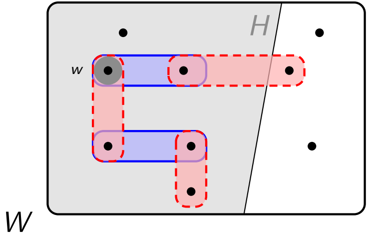
Each agent i is associated with a partition $\mathcal{E}_i$.

$K_i : \wp(W) \rightarrow \wp(W)$ where $K_i(H) = \{w \mid \mathcal{E}_i[w] \subseteq H\}$



Each agent i is associated with a partition $\mathcal{E}_i$.

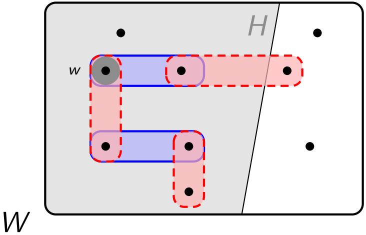
$$w \notin K_1(H) \quad \text{and} \quad w \in K_2(H)$$



Everyone Knows: $K(H) = \bigcap_{i\in\mathcal{A}} K_i(H)$

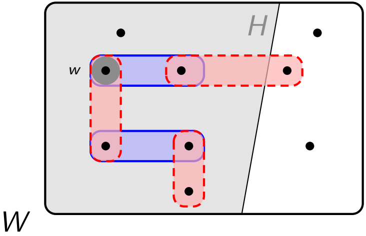
$K^m(H)$ for all $m \geq 0$ is defined as:

$K^0(H) = H \qquad K^m(H) = K(K^{m-1}(H))$

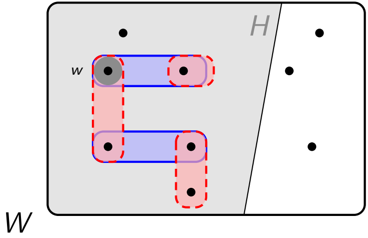


Common Knowledge: $C : \wp(W) \to \wp(W)$ with

$$C(H) = \bigcap_{m \geq 0} K^m(H)$$



$w \in K(H)$ $w \notin C(H)$

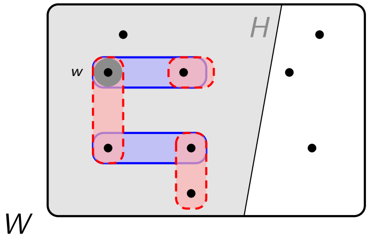


$$w \in C(H)$$

Fact. $w \in C(H)$ if every **finite path** starting at w ends in a state in H

There is a finite path from w to v if there is $v_1, \ldots, v_m$ such that $w = v_1$, $v = v_m$ and there are $E_1, \ldots, E_{m-1} \in \cup\mathcal{E}_i$ such that $\{v_1, v_2\} \subseteq E_1, \ldots, \{v_{m-1}, v_m\} \subseteq E_{m-1}$.

$I_C(w) = \{v \mid \text{there is a finite path from } w \text{ to } v\}$, so $C(H) = \{w \mid I_C(w) \subseteq H\}$.



Fact. $w \in C(H)$ if every **finite path** starting at w ends in a state in H

Theorem. Suppose that n agents share a common prior and have different private information. If there is common knowledge in the group of the posterior probabilities, then the posteriors must be equal.

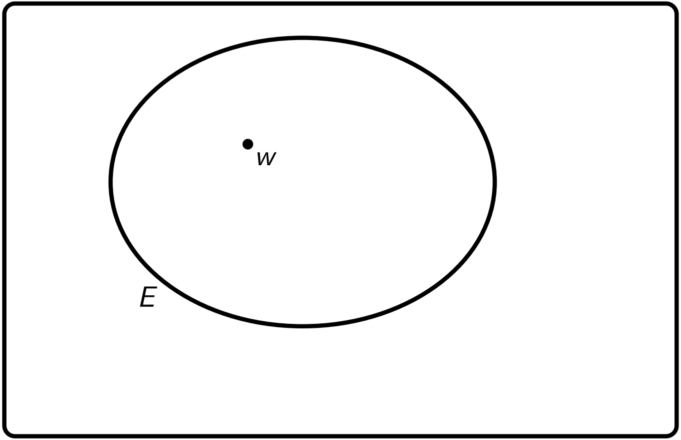
Robert Aumann. *Agreeing to Disagree*. Annals of Statistics **4** (1976).

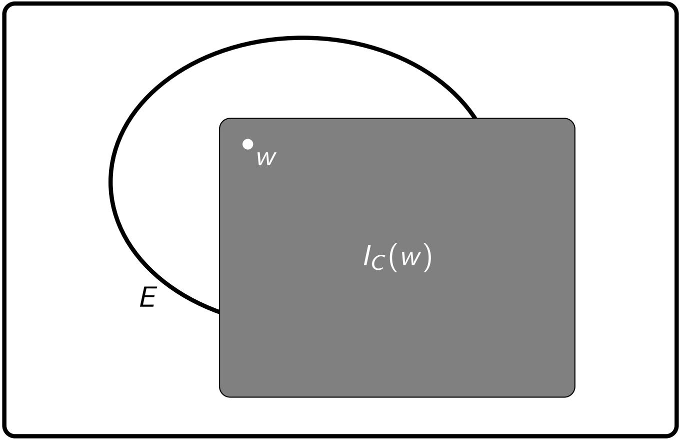
Suppose that W is, $E \subseteq W$ is an event, and two (or more) agents with partitions $\mathcal{E}_i$. Let p be the **common prior**.

The agent's posterior probabilities of the event E are *random variables*: $P_i^E : W \to [0, 1]$, $P_i^E(w) = p(E \mid \mathcal{E}_i[w])$.

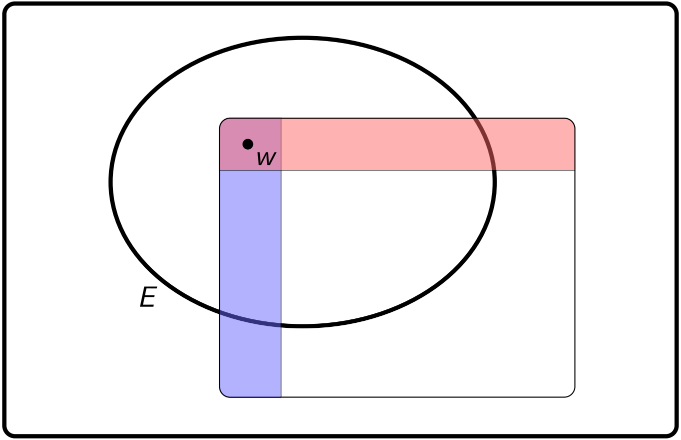
So, $[\![P_i^E = r]\!] = \{w \mid P_i^E(w) = r\}$

Assume that $w \in C([\![P_1^E = r \wedge P_2^E = q]\!])$.

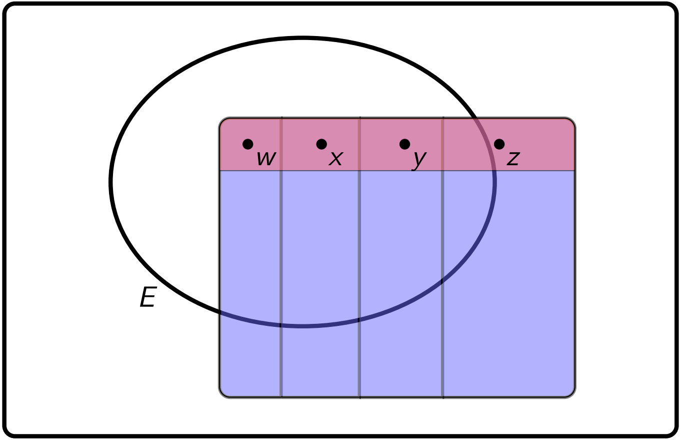

w
E



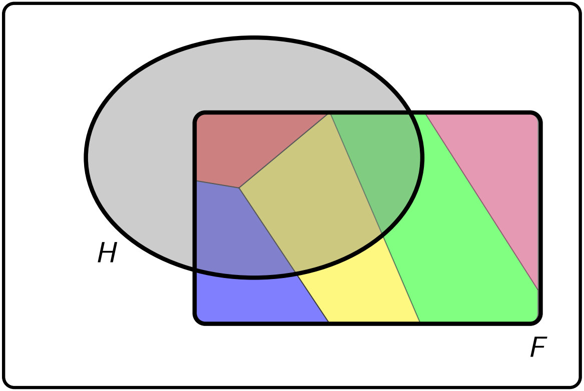
$$I_C(w) \subseteq [\![P_1^E = r \wedge P_2^E = q]\!]$$



$$p(E \mid \mathcal{E}_1[w]) = q,\ p(E \mid \mathcal{E}_2[w]) = r$$



$$p(E \mid \mathcal{E}_1[w]) = p(E \mid \mathcal{E}_1[x]) = p(E \mid \mathcal{E}_1[y]) = p(E \mid \mathcal{E}_1[z]) = q$$



$$p(H \mid F) = \sum_i p(E_i \mid F) p(H \mid E_i)$$

Fact. If $p(H \mid E_i) = q$ for all i, then $p(H \mid F) = q$.

Fact. Suppose that $\mathcal{E} = \{E_1, \ldots, E_m\}$ partitions F. If $p(H \mid E_i) = q$ for all i, then $p(H \mid F) = q$.

$$\begin{aligned} p(H \mid F) &= \sum_i p(E_i \mid F)p(H \mid E_i) \\ &= \sum_i p(E_i \mid F)q \\ &= q\sum_i p(E_i \mid F) \\ &= q \end{aligned}$$

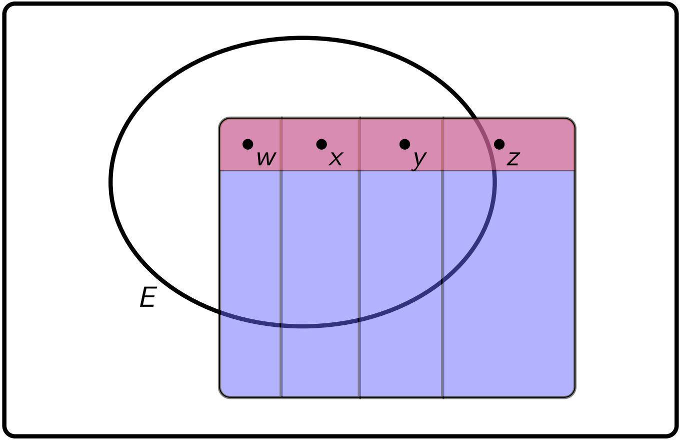
Fact. Suppose that $\{F_i\}$ is a partition of F (so $F = \bigcup_i F_i$ and $F_i \cap F_j \neq \emptyset$ for $i \neq j$). If $p(E \mid F_i) = q$ for all i, then $p(E \mid F) = q$.

If $p(E \mid F_i) = q$, then $p(E \cap F_i) = qp(F_i)$.

$$p(E \mid F) = \frac{p(E \cap F)}{p(F)} = \frac{p((E \cap F_1) \cup \cdots \cup (E \cap F_n))}{p(F)}$$

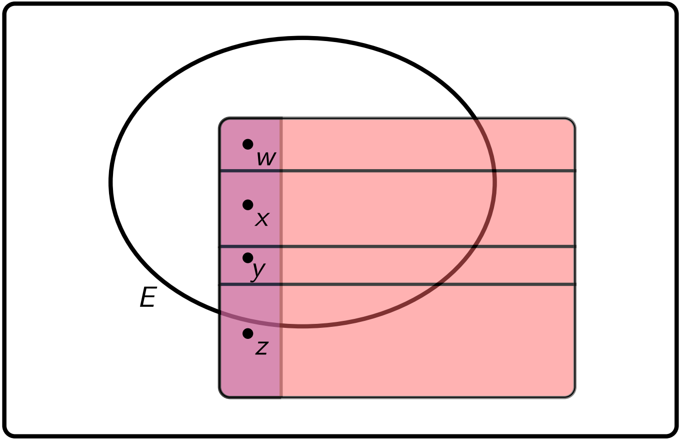
$$= \frac{p(E \cap F_1) + \cdots + p(E \cap F_n)}{p(F)} = \frac{qp(F_1) + \cdots + qp(F_n)}{p(F)}$$

$$= \frac{q(p(F_1) + \cdots + p(F_n))}{p(F)} = \frac{qp(F)}{p(F)} = q$$



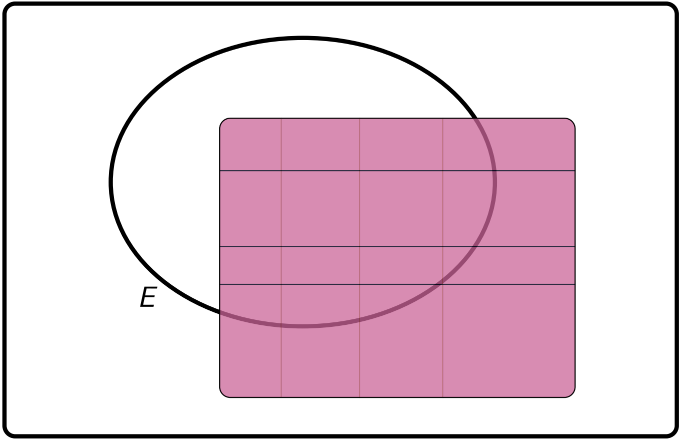
$$p(E \mid \mathcal{E}_1[w]) = p(E \mid \mathcal{E}_1[x]) = p(E \mid \mathcal{E}_1[y]) = p(E \mid \mathcal{E}_1[z]) = q$$

So, $p(E \mid I_C(w)) = q$.



$$p(E \mid \mathcal{E}_2[w]) = p(E \mid \mathcal{E}_2[x]) = p(E \mid \mathcal{E}_2[y]) = p(E \mid \mathcal{E}_2[z]) = r$$

So, $p(E \mid I_C(w)) = r$.



Thus, $q = p(E \mid I_C(w)) = r$.

Qualitative versions

like-minded individuals cannot agree to make different decisions.

M. Bacharach. *Some Extensions of a Claim of Aumann in an Axiomatic Model of Knowledge*. Journal of Economic Theory, 37(1), pgs. 167-190, 1985.

J.A.K. Cave. *Learning to Agree*. Economic Letters, 12(2), pgs. 147 - 152, 1983.

D. Samet. *Agreeing to disagree: The non-probabilistic case*. Games and Economic Behavior, 69, pgs. 169-174, 2010.

Rational Disagreement

M. Caie. *Agreement Theorems for Self-Locating Belief*. Review of Symbolic Logic, 2016.

J. Halpern and W. Kets. *Ambiguous Language and Common Priors*. Games and Economic Behavior, 2014.

H. Lederman. *People with common priors can agree to disagree*. Review of Symbolic Logic, 8(1), pp. 1145, 2015.

A. Rubinstein and A. Wolinsky. *On the logic of 'agreeing to disagree' type results*. Journal of Economic Theory, 1, 184193, 1990.

Dynamic characterization of Aumann's Theorem

- How do the posteriors *become* common knowledge?

J. Geanakoplos and H. Polemarchakis. *We Can't Disagree Forever*. Journal of Economic Theory (1982).

Dynamic characterization of Aumann's Theorem

- How do the posteriors *become* common knowledge?

J. Geanakoplos and H. Polemarchakis. *We Can't Disagree Forever*. Journal of Economic Theory (1982).

- What happens when communication is not the the whole group, but pairwise?

R. Parikh and P. Krasucki. *Communication, Consensus and Knowledge*. Journal of Economic Theory (1990).

$t = 0 \qquad \langle W, \mathcal{E}_{0,a}, \mathcal{E}_{0,b}, p \rangle$

$t = 0 \qquad \langle W, \mathcal{E}_{0,a}, \mathcal{E}_{0,b}, p\rangle$

$$P^E_{0,a}(w) = r_0 \quad P^E_{0,b}(w) = q_0$$

$$t = 0 \qquad \langle W, \mathcal{E}_{0,a}, \mathcal{E}_{0,b}, p \rangle$$

$$P^E_{0,a}(w) = r_0 \quad P^E_{0,b}(w) = q_0$$

$$t = 1 \qquad \langle W, \mathcal{E}_{1,a}, \mathcal{E}_{1,b}, p \rangle$$

$t = 0 \qquad \langle W, \mathcal{E}_{0,a}, \mathcal{E}_{0,b}, p \rangle$

$$P^E_{0,a}(w) = r_0 \quad P^E_{0,b}(w) = q_0$$

$t = 1 \qquad \langle W, \mathcal{E}_{1,a}, \mathcal{E}_{1,b}, p \rangle$

$$P^E_{1,a}(w) = r_1 \quad P^E_{1,b}(w) = q_1$$

$t = 0 \qquad \langle W, \mathcal{E}_{0,a}, \mathcal{E}_{0,b}, p\rangle$

$$P^E_{0,a}(w) = r_0 \quad P^E_{0,b}(w) = q_0$$

$t = 1 \qquad \langle W, \mathcal{E}_{1,a}, \mathcal{E}_{1,b}, p\rangle$

$$P^E_{1,a}(w) = r_1 \quad P^E_{1,b}(w) = q_1$$

$t = 2 \qquad \langle W, \mathcal{E}_{2,a}, \mathcal{E}_{2,b}, p\rangle$

$$P^E_{2,a}(w) = r_2 \quad P^E_{2,b}(w) = q_2$$

$t = 3 \qquad \langle W, \mathcal{E}_{3,a}, \mathcal{E}_{3,b}, p\rangle$

$\vdots \qquad \vdots$

Geanakoplos and Polemarchakis

- Assuming that the information partitions are finite, given an event A, the revision process converges in finitely many steps.

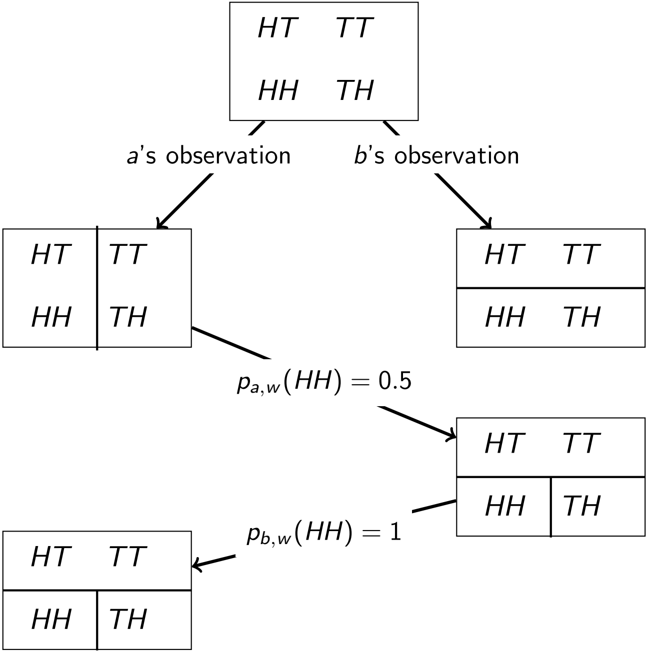
Geanakoplos and Polemarchakis

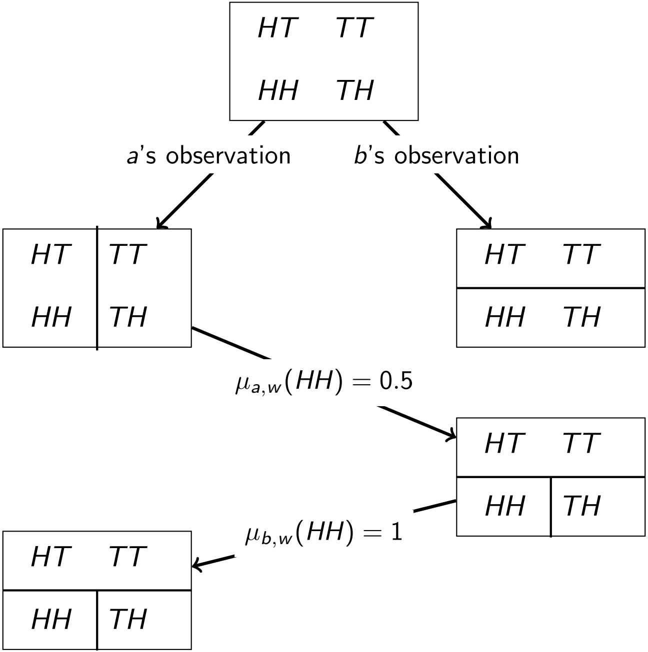
- Assuming that the information partitions are finite, given an event A, the revision process converges in finitely many steps.

- For each n, there are examples where the process takes n steps.

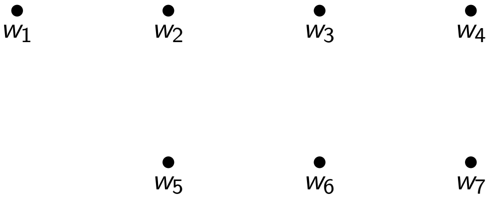
Geanakoplos and Polemarchakis

- Assuming that the information partitions are finite, given an event A, the revision process converges in finitely many steps.
- For each n, there are examples where the process takes n steps.
- An *indirect communication* equilibrium is not necessarily a *direct communication* equilibrium.


HT TT
HH TH
a's observation
b's observation
HT TT
HH TH
HT TT
HH TH
$p_{a,w}(HH) = 0.5$
HT TT
HH TH
$p_{b,w}(HH) = 1$
HT TT
HH TH

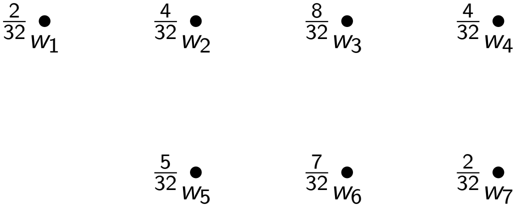

HT TT
HH TH
a's observation
b's observation
HT TT
HH TH
HT TT
HH TH
$\mu_{a,w}(HH) = 0.5$
HT TT
HH TH
$\mu_{b,w}(HH) = 1$
HT TT
HH TH

2 Scientists Perform an Experiment



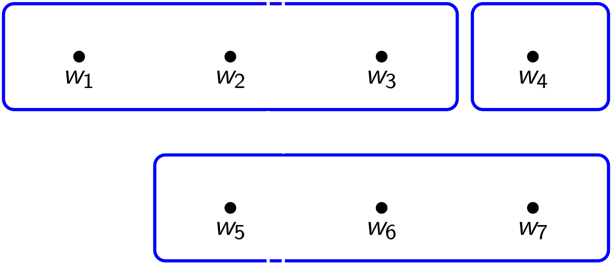
They agree the true state is one of seven different states.

2 Scientists Perform an Experiment



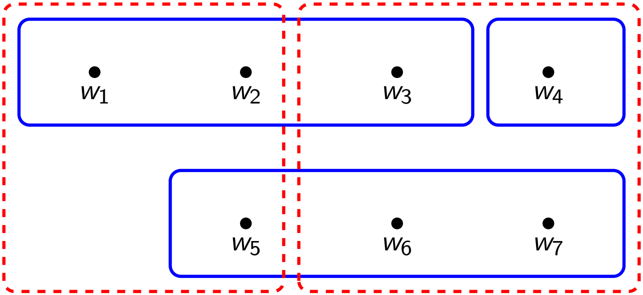
They agree on a common prior.

2 Scientists Perform an Experiment



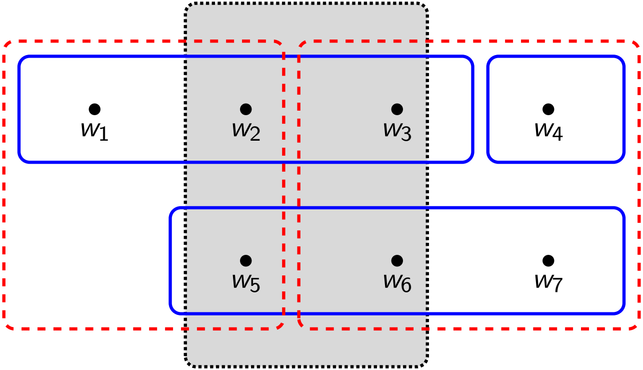
They agree that Experiment 1 would produce the blue partition.

2 Scientists Perform an Experiment



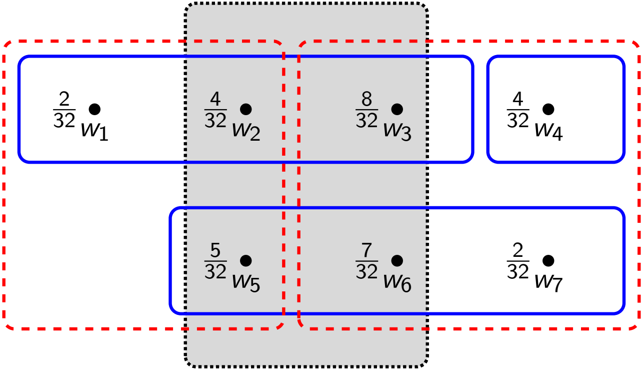
They agree that Experiment 1 would produce the blue partition and Experiment 2 the red partition.

2 Scientists Perform an Experiment



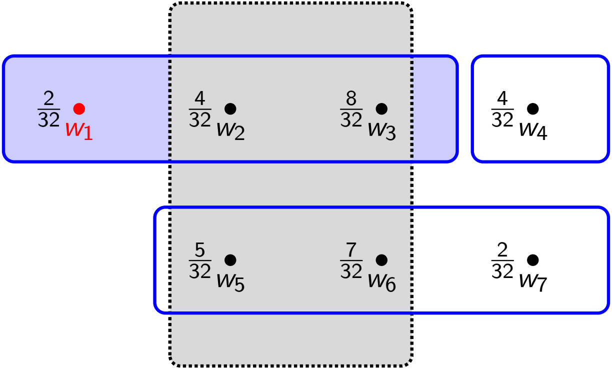
They are interested in the truth of $E = \{w_2, w_3, w_5, w_6\}$.

2 Scientists Perform an Experiment



So, they agree that $P(E) = \frac{24}{32}$.

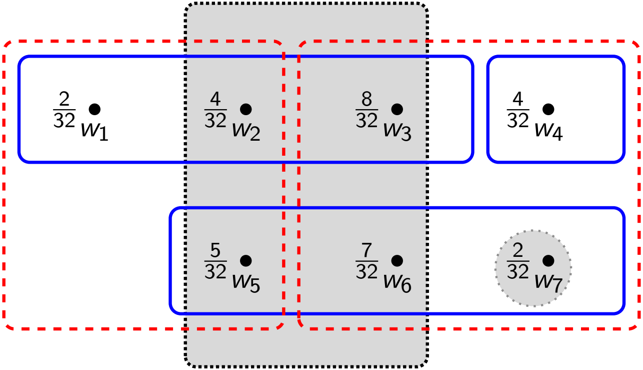
2 Scientists Perform an Experiment



Also, that if the true state is w_1, then Experiment 1 will yield

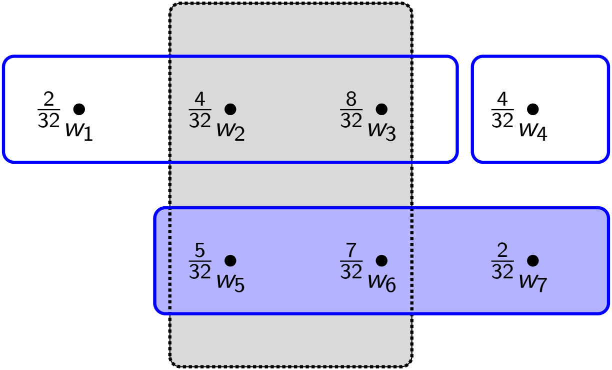
$$P(E|I) = \frac{P(E \cap I)}{P(I)} = \frac{12}{14}$$

2 Scientists Perform an Experiment



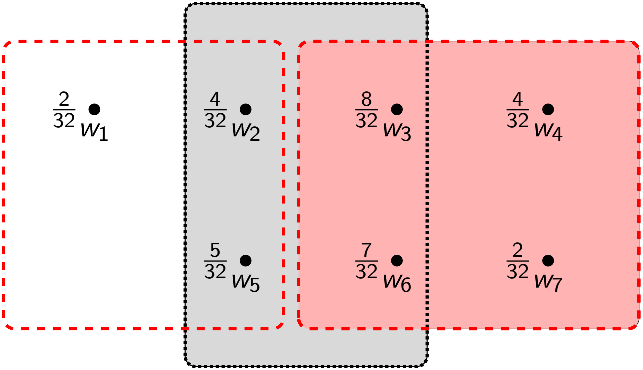
Suppose the true state is w_7 and the agents preform the experiments.

2 Scientists Perform an Experiment



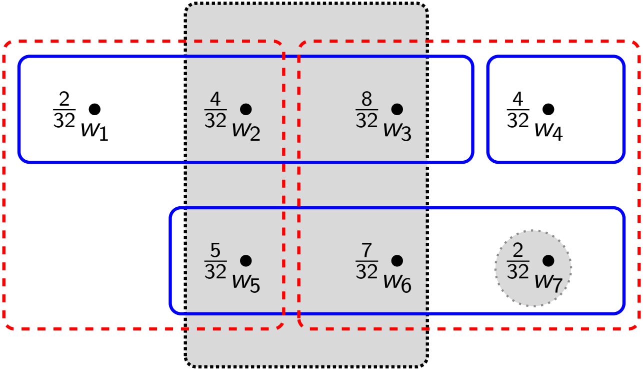
Suppose the true state is w_7, then $Pr_1(E) = \frac{12}{14}$

2 Scientists Perform an Experiment



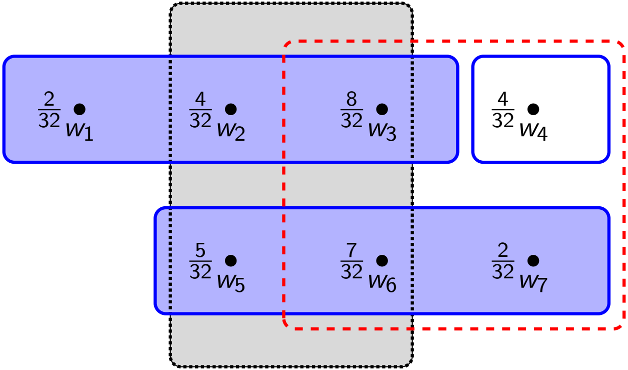
Then $Pr_1(E) = \frac{12}{14}$ and $Pr_2(E) = \frac{15}{21}$

2 Scientists Perform an Experiment



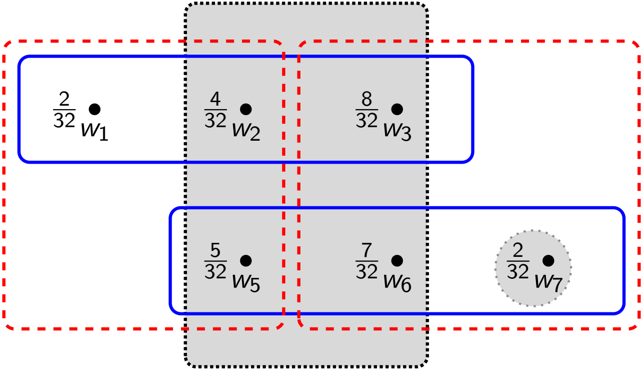
Suppose they exchange emails with the new subjective probabilities: $Pr_1(E) = \frac{12}{14}$ and $Pr_2(E) = \frac{15}{21}$

2 Scientists Perform an Experiment



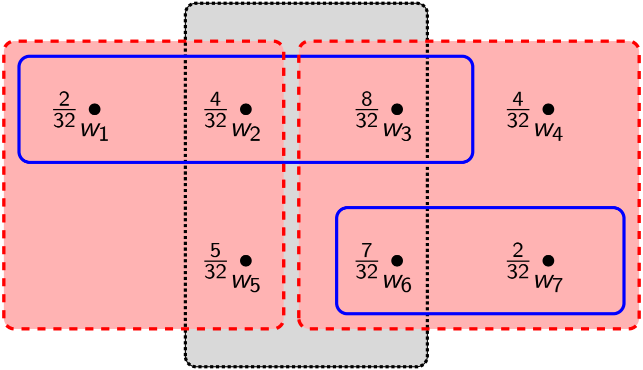
Agent 2 learns that w_4 is **NOT** the true state (same for Agent 1).

2 Scientists Perform an Experiment



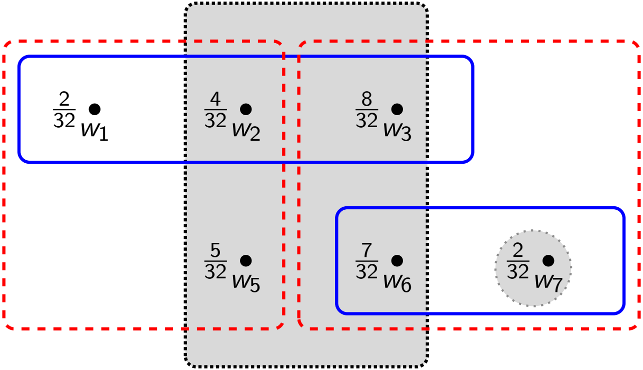
Agent 2 learns that w_4 is **NOT** the true state (same for Agent 1).

2 Scientists Perform an Experiment



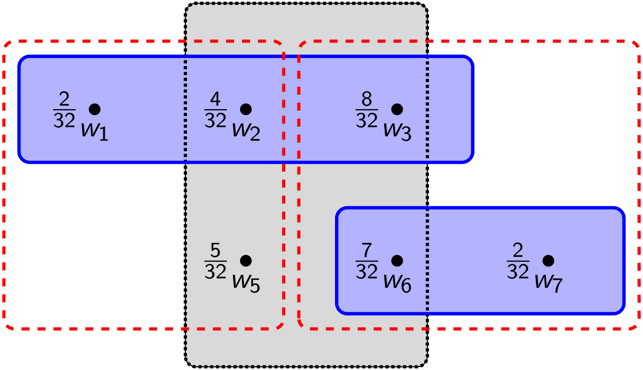
Agent 1 learns that w_5 is **NOT** the true state (same for Agent 1).

2 Scientists Perform an Experiment



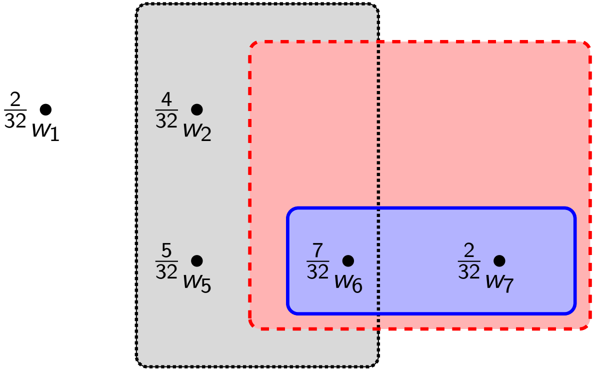
The new probabilities are $Pr_1(E|I') = \frac{7}{9}$ and $Pr_2(E|I') = \frac{15}{17}$

2 Scientists Perform an Experiment

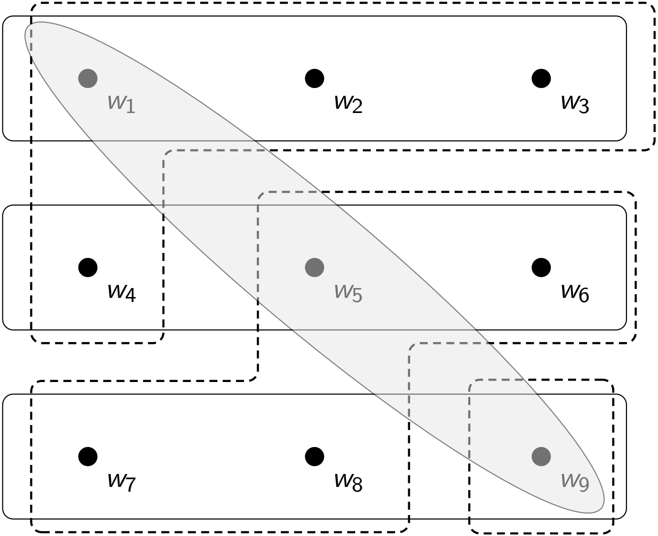


After exchanging this information ($Pr_1(E|I') = \frac{7}{9}$ and $Pr_2(E|I') = \frac{15}{17}$), Agent 2 learns that w_3 is **NOT** the true state.

2 Scientists Perform an Experiment



No more revisions are possible and the agents agree on the posterior probabilities.



for all $j = 1, \ldots, 9$, $p(w_j) = \frac{1}{9}$

Revisions

1. Ann announces that her probability of E is $\frac{1}{3}$ and Bob announces that his is $\frac{1}{4}$.

Revisions

1. Ann announces that her probability of E is $\frac{1}{3}$ and Bob announces that his is $\frac{1}{4}$.

2. Ann announces that her probability of E is $\frac{1}{3}$ and Bob announces that his is $\frac{1}{4}$.

Revisions

1. Ann announces that her probability of E is $\frac{1}{3}$ and Bob announces that his is $\frac{1}{4}$.
2. Ann announces that her probability of E is $\frac{1}{3}$ and Bob announces that his is $\frac{1}{4}$.
3. Ann announces that her probability of E is $\frac{1}{3}$ and Bob announces that his is $\frac{1}{4}$.

Revisions

1. Ann announces that her probability of E is $\frac{1}{3}$ and Bob announces that his is $\frac{1}{4}$.

2. Ann announces that her probability of E is $\frac{1}{3}$ and Bob announces that his is $\frac{1}{4}$.

3. Ann announces that her probability of E is $\frac{1}{3}$ and Bob announces that his is $\frac{1}{4}$.

4. Ann announces that her probability of E is $\frac{1}{3}$ and Bob announces that his is $\frac{1}{4}$. As a result of this announcement, Bob further refines his partition so that $\mathcal{E}_b(w_4) = \{w_4\}$ and $\mathcal{E}_b(w_1) = \{w_1, w_2, w_3\}$. Now Ann and Bob both assign probability $\frac{1}{3}$ to the event E.

Common r-belief

The typical example of an event that creates common knowledge is a **public announcement**.

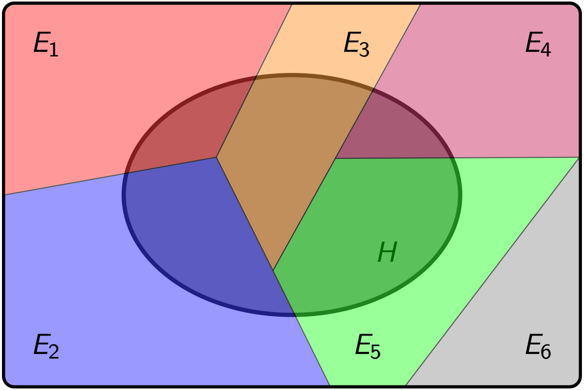
Common r-belief

The typical example of an event that creates common knowledge is a **public announcement**.

Shouldn't one always allow for some small probability that a participant was absentminded, not listening, sending a text, checking Facebook, proving a theorem, asleep, ...

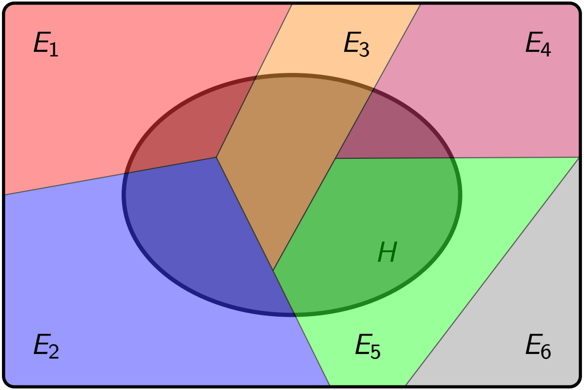
D. Monderer and D. Samet. *Approximating Common Knowledge with Common Beliefs*. Games and Economic Behavior (1989).

From Knowledge to r-Belief



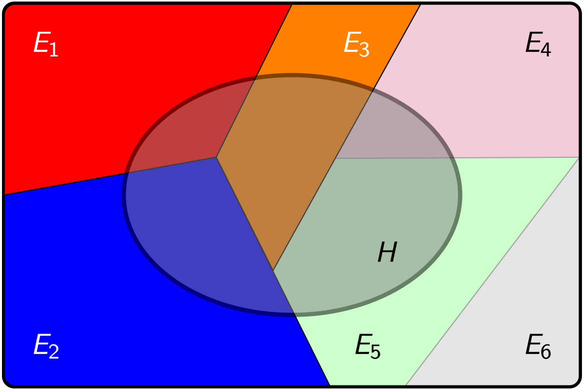
Given a partition $\mathcal{E}$, define $K_{\mathcal{E}} : \wp(W) \rightarrow \wp(W)$ as:
$$K_{\mathcal{E}}(H) = \{w \mid \mathcal{E}[w] \subseteq H\}$$

From Knowledge to r-Belief



Given $r \in [0,1]$ and a partition $\mathcal{E}$, define $B^r_{\mathcal{E}} : \wp(W) \rightarrow \wp(W)$ as:
$B^r_{\mathcal{E}}(H) = \{w \mid p_{\mathcal{E},w}(H) \geq r\}$

From Knowledge to r-Belief

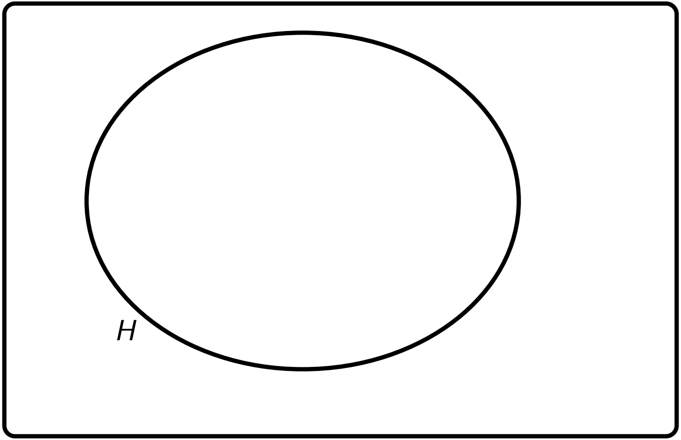


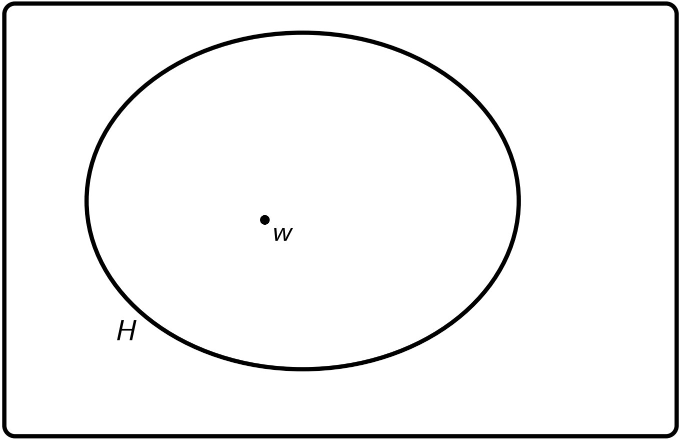
Given $r \in [0,1]$ and a partition $\mathcal{E}$, define $B^r_{\mathcal{E}} : \wp(W) \to \wp(W)$ as:
$$B^r_{\mathcal{E}}(H) = \{w \mid p_{\mathcal{E},w}(H) \geq r\}$$

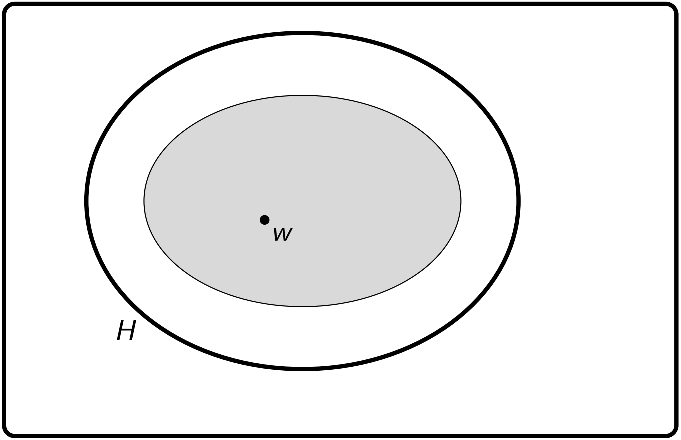
From Common Knowledge to Common r-Belief

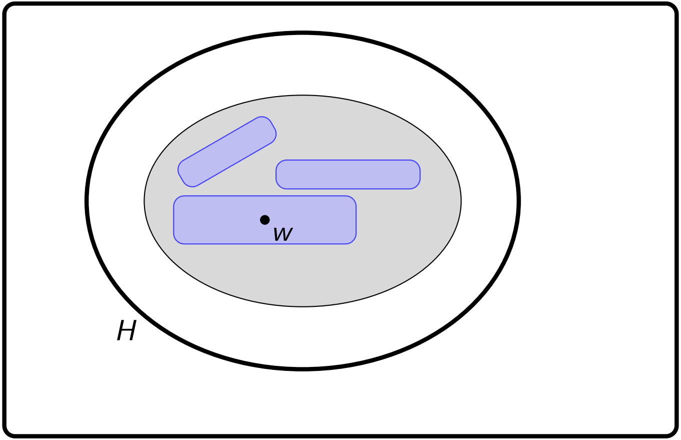
Suppose that $C : \wp(W) \rightarrow \wp(W)$ is a common knowledge operator. TFAE

1. $w \in C(H) = \bigcap_{m \geq 0} K^m(H)$
2. $I_c(w) \subseteq H$
3. There is a set $F \subseteq W$ such that
   3.1 $w \in F \subseteq K(F) = \bigcap_i K_i(F)$
   3.2 $F \subseteq H$

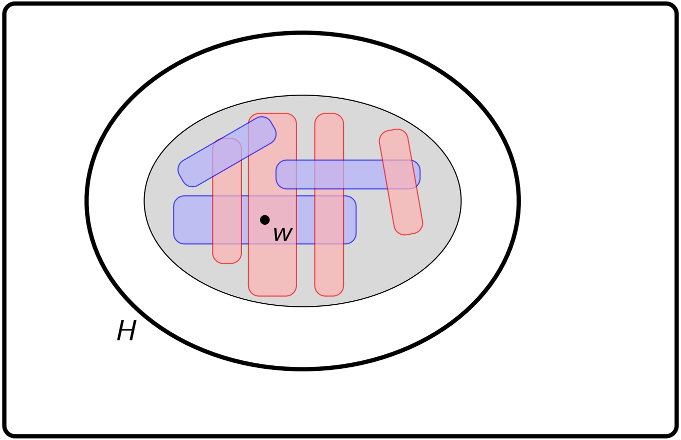

H


w
H


w
H



w
H


w
H

From Common Knowledge to Common r-Belief

$$B_i^r(E) = \{w \mid p(E \mid \mathcal{E}_i[w]) \geq r\}$$

From Common Knowledge to Common r-Belief

$$B_i^r(E) = \{w \mid p(E \mid \mathcal{E}_i[w]) \geq r\}$$

F is an **evident** r**-belief** if for each $i \in \mathcal{A}$, $F \subseteq B_i^r(F)$

From Common Knowledge to Common r-Belief

$$B_i^r(E) = \{w \mid p(E \mid \mathcal{E}_i[w]) \geq r\}$$

F is an **evident** r-**belief** if for each $i \in \mathcal{A}$, $F \subseteq B_i^r(F)$

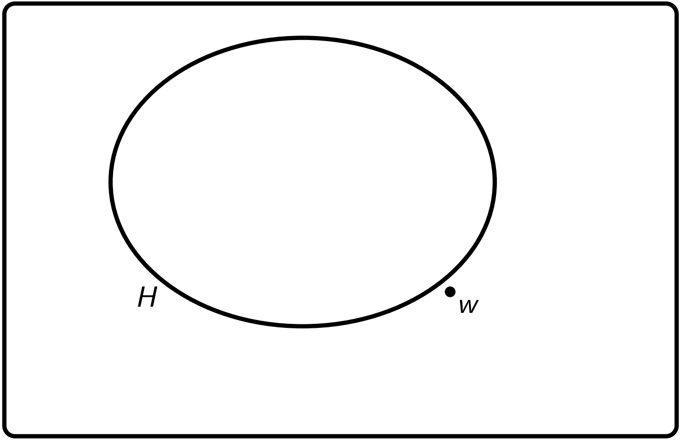
An event H is **common** r-**belief** at w if there exists and evident r-belief event F such that $w \in F$ and for all $i \in \mathcal{A}$, $F \subseteq B_i^r(H)$

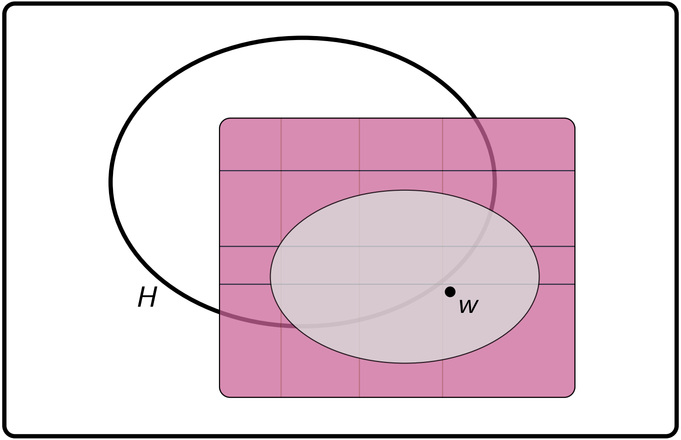
$w \in C(H)$ iff there is an event $F \subseteq W$ such that

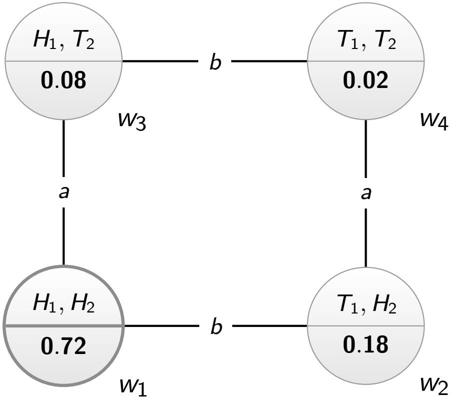
1. $w \in F \subseteq K(F) = \bigcap_i K_i(F)$
2. $F \subseteq H$

$w \in C^r(H)$ iff there is an event $F \subseteq W$ such that

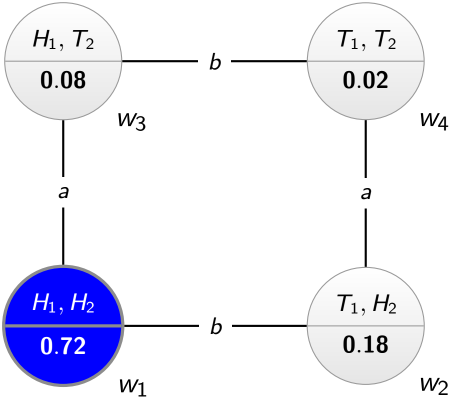
1. $w \in F \subseteq B^r(F) = \bigcap_i B_i^r(F)$
2. $F \subseteq B^r(H)$


H
w

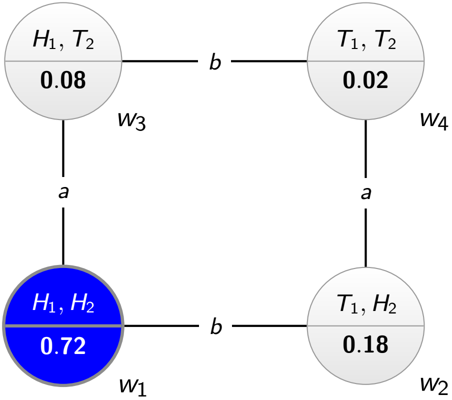

H
w



H_1, T_2
0.08
w_3
T_1, T_2
0.02
w_4
H_1, H_2
0.72
w_1
T_1, H_2
0.18
w_2
a
a
b
b



- $\{w_1\} = B_a^{0.9}(H_1 \cap H_2) \cap B_b^{0.8}(H_1 \cap H_2)$.
- $X = \{w_1\}$ is an evident 0.8-belief for both Ann and Bob.

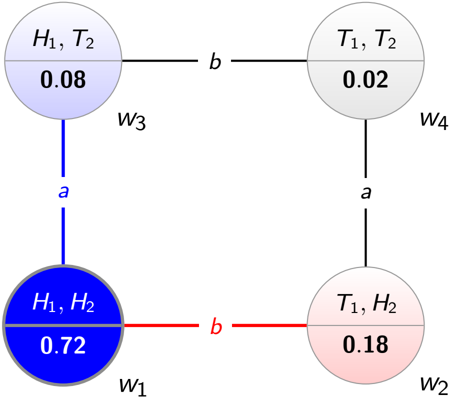


- $\{w_1\} = B_a^{0.9}(H_1 \cap H_2) \cap B_b^{0.8}(H_1 \cap H_2)$.
- $X = \{w_1\}$ is an evident 0.8-belief for both Ann and Bob.
- $X \subseteq B_a^{0.8}(H_1 \cap H_2) \cap B_b^{0.8}(H_1 \cap H_2)$.
- $w_1 \in C_{a,b}^{0.8}(H_1 \cap H_2)$.

Generalizing Aumann's Theorem

Theorem. If the posteriors of an event E are common r-belief at some state w, then any two posteriors can differ by at most $1 - r$.

D. Samet and D. Monderer. *Approximating Common Knowledge with Common Beliefs*. Games and Economic Behavior, Vol. 1, No. 2, 1989.



- $w_1 \in C^{0.8}_{a,b}(H_1 \cap H_2)$.
- $|p_{a,w_1}(H_1 \cap H_2) - p_{b,w_1}(H_1 \cap H_2)| = |0.9 - 0.8| = 0.1$

Questions

- How do the posteriors *become* commonly p-believed?

- What happens when communication is not between the whole group, but pairwise?

That is, what information should the agents exchange so that the dynamic process of information exchange converges with common r-belief (for some $0.5 < r < 1$) of the agents' probabilities of E?

- Exchanging current probabilities of an event leads to common knowledge of the posteriors

- Exchanging current probabilities of an event leads to common knowledge of the posteriors

- Announcements resolve uncertainty about what the other agents' information

- Exchanging current probabilities of an event leads to common knowledge of the posteriors

- Announcements resolve uncertainty about what the other agents' information

- Announcements of posterior probabilities with some "error" term (Jeffrey Updating)

- Exchanging current probabilities of an event leads to common knowledge of the posteriors
- Announcements resolve uncertainty about what the other agents' information
- Announcements of posterior probabilities with some "error" term (Jeffrey Updating)
- Which events can become commonly p-believed by exchanging current probabilities? (Common Learning Theorem)

Rational Disagreement

M. Caie. *Agreement Theorems for Self-Locating Belief*. Review of Symbolic Logic, 2016.

J. Halpern and W. Kets. *Ambiguous Language and Common Priors*. Games and Economic Behavior, 2014.

H. Lederman. *People with common priors can agree to disagree*. Review of Symbolic Logic, 8(1), pp. 1145, 2015.

A. Rubinstein and A. Wolinsky. *On the logic of 'agreeing to disagree' type results*. Journal of Economic Theory, 1, 184193, 1990.

Rational disagreement over time

D. Blackwell and L. Dubins. *Merging of opinions with increasing information*. The Annals of Mathematical Statistics, 33, pp. 882 - 886.

D. Samet and D. Monderer. *Stochastic Common Learning*. Games and Economic Behavior, 9, pgs. 161 - 171, 1995.

S. Huttegger. *Merging of Opinions and Probability Kinematics*. Review of Symbolic Logic, 2015.

Thank you!

EP. *Dynamics for Probabilistic Common Belief*. Studies in Logic, 2015.

Soft Jeffrey Shifts

$\mathcal{E}_1 = \{E_1, E_2, E_3, E_4\}$ and consider the learning experience given by:

$$(\frac{1}{5} : E_1, \frac{3}{10} : E_2, \frac{1}{2} : E_3, 0 : E_n)$$

Soft Jeffrey Shifts

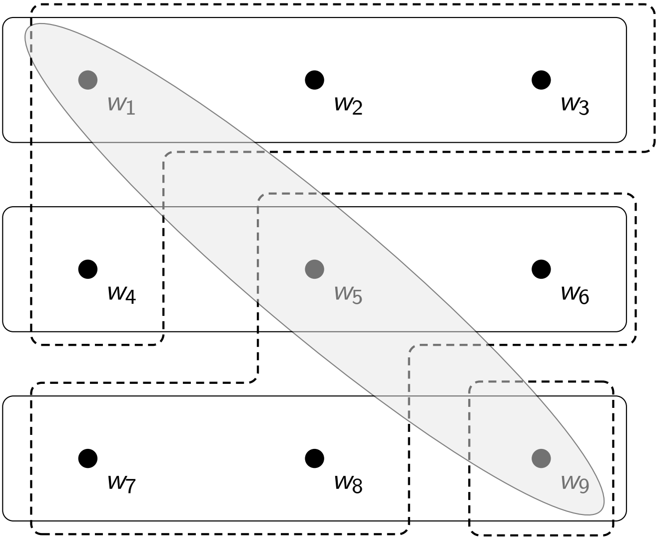
$\mathcal{E}_1 = \{E_1, E_2, E_3, E_4\}$ and consider the learning experience given by:

$$(\frac{1}{5} : E_1, \frac{3}{10} : E_2, \frac{1}{2} : E_3, 0 : E_n)$$

Consider instead:

- $p_1(E_1) = 2 \cdot p(E_1)$;
- $p_1(E_2) = \frac{1}{2} \cdot p(E_2)$;
- $p_1(E_3) = 5 \cdot p(E_3)$; and
- $p_1(E_4) = 0 \cdot p(E_4)$

If $p(E_1) = \frac{1}{10}$, $p(E_2) = \frac{3}{5}$, $p(E_3) = \frac{1}{10}$ and $p(E_4) = \frac{1}{5}$, then probability kinematics will lead to the same result whether or not it is a hard or "soft" Jeffrey shift.



$w_1 \in C^{\frac{1}{4}}(E)$, and this does not change during the information exchange – until the process converges, when we have $w \in C^{\frac{1}{3}}(E)$.

Common Learning Theorem

- Let $(W, \mathcal{F}, \mu)$ be an initial probability space.
- Each agent i receives private information represented by partitions Π_i^0 on W.
- All agents observe the outcome of a discrete random variable. Based on these observations, the agents refine their initial information: $\Pi_i^0, \Pi_i^1, \Pi_i^2, \ldots, \Pi_i^t, \ldots$.
- For each $t \geq 0$, $B_{i,t}^p$ C_t^p (for $p \in [0,1]$) are well-defined.

D. Samet and D. Monderer. *Stochastic Common Learning*. Games and Economic Behavior, 9, pgs. 161 - 171, 1995.

Common Learning Theorem

Common Learning Theorem: Suppose that $(E_t)_{t\geq 1}$ is any nondecreasing sequence of events such that $\lim_{t\rightarrow\infty}\mu(E_t) = 1$. Then, for all $0 \leq p < 1$, and almost all $w \in W$, there is a time t_w (depending on w) such that for all $t' \geq t_w$ the agents commonly p-believe E_t at time t (i.e., $w \in C^p_t(E_t)$).

D. Samet and D. Monderer. *Stochastic Common Learning*. Games and Economic Behavior, 9, pgs. 161 - 171, 1995.

D. Blackwell and L. Dubins. *Merging of opinions with increasing information*. The Annals of Mathematical Statistics, 33, pp. 882 - 886.

Suppose that W is a set of atomic events and $\mathcal{F}$ is a σ-field on W.

The elements of can be thought of as possible worlds and the members of $\mathcal{F}$ as propositions.

E.g., W can be the set of all infinite sequences of coin tosses and $\mathcal{F}$ contains all propositions about coin tossing events of interest. (It may including limiting events such as $\lim_{n\to\infty} S_n = 1/2$ where S_n is the total number of heads in the first n flips of the coin.

Let $\mathcal{E}_1, \mathcal{E}_2, \ldots, \mathcal{E}_n, \ldots$ be an infinite sequence of partitions on W such that $\mathcal{E}_{n+1}$ *refines* $\mathcal{E}_n$.

$\mathcal{E}_n$ is the information that the agent receives at time n.

$\mathcal{E}_n[w]$ is the element of $\mathcal{E}_n$ containing w.

For each n, let $\mathcal{F}_n$ be the σ-algebra generated by $\mathcal{E}_n$. We assume that $\mathcal{F} = \cup_n \mathcal{F}_n$.

For $A \in \mathcal{F}$, $P(A \mid \mathcal{E}_n[w]) = \frac{P(A \cap \mathcal{E}_n[w])}{P(\mathcal{E}_n[w])}$

Martingale Convergence Theorem

A martingale is an infinite sequence of fair gambles (a sequence of gambles for which there is no gambling system you could use to your own advantage).

This is expressed by saying that your expected total fortune after the next trial is equal to your present total fortune. On average you neither lose nor win.

Martingales are important because they lead to very general laws of large numbers (martingale convergence theorems) that do not depend on the quite stringent conditions required for the standard strong law of large numbers.

$$P(A \mid \mathcal{E}_n[w]) \to \begin{cases} 1 & w \in A \\ 0 & w \notin A \end{cases}$$

for **almost every** w with regard to the prior probability P.

The set of w for which the above does not hold has measure 0.

Convergence to certainty can be viewed as a consequence of dynamic coherence in the following sense.

- At each trial n the agent updates her probabilities on an element of the partition $\mathcal{E}_n$.
- This is plain vanilla Bayesian conditioning, which can be justified by any dynamic Dutch book argument (or epistemic utility argument, or...)
- Updating by Bayesian conditioning embeds the sequence of conditional probabilities in the convergence to certainty theorem.
- Thus, a dynamically coherent agent expects her future degrees of belief to converge to certainty under the appropriate conditions.

Convergence to certainty yields a first pass on merging of opinions.

Suppose that Eve's degrees of beliefs are represented by P and Adam's by Q, and let $P_n[A](w) = P(A \mid \mathcal{E}_n[w])$ and $Q_n[A](w) = Q(A \mid \mathcal{E}_n[w])$.

Then $P_n[A]$ and $Q_n[A]$ both converge to zero or to one almost surely with respect to the priors P and Q, respectively.

Now, Eve believes with certainty that $P_n[A]$ and $Q_n[A]$ will agree in the limit *whenever she assigns probability one to any set to which Adam assigns probability one*.

For then, since $Q_n[A]$ goes to certainty a.e. (Q), Eve also believes with probability one that her's and Adam's conditional probabilities for any event A are the same in the limit.

This result applies to particular events A. But it does not say anything about Eve's and Adam's overall conditional probabilities.

This result applies to particular events A. But it does not say anything about Eve's and Adam's overall conditional probabilities.

The Blackwell-Dubins theorem fills this gap.

As Adam and Eve observe more coin tosses and update by Bayesian conditioning, their degrees of belief will become close *uniformly* in all events.

Moreover, the Blackwell-Dubins theorem does not require that conditional probabilities converge (as in convergence to certainty). Eve's and Adam's conditional probabilities may get closer even if they don't converge.

Variational Distance

Suppose that μ and ν are two measures over all events in $\mathcal{F}$.

$$d(\mu, \nu) = \sup_{A \in \mathcal{F}} |\mu(A) - \nu(A)|$$

Variational Distance

Suppose that μ and ν are two measures over all events in $\mathcal{F}$.

$$d(\mu, \nu) = \sup_{A \in \mathcal{F}} |\mu(A) - \nu(A)|$$

P is said to **merge to** Q if for Q almost every w,
$d(P_n(w), P_n(w)) \to 0$ as $n \to \infty$.

Almost everywhere (a.e. Q), given any $\epsilon > 0$, there is an n_0 such that

$$|P(A \mid \mathcal{E}_n[w]) - Q(A \mid \mathcal{E}_n[w])| < \epsilon$$

for all $n > n_0$ and for all $A \in \mathcal{F}$. The number n_0 may depend on ϵ and on w but not on A.

If P merges to Q, then Adam (with probability Q) believes with probability one that his conditional degrees of belief for all propositions A will get arbitrarily close.

Absolute Continuity

Q is absolutely continuous relative to P ($Q \ll P$) if for all $A \in \mathcal{F}$,

$$Q(A) > 0 \Longrightarrow P(A) > 0$$

Blackwell & Dubbins Theorem. If $Q \ll P$, then P merges to Q.

Common Learning Theorem

- Let $(W, \mathcal{F}, \mu)$ be an initial probability space.
- Each agent i receives private information represented by partitions Π_i^0 on W.
- All agents observe the outcome of a discrete random variable. Based on these observations, the agents refine their initial information: $\Pi_i^0, \Pi_i^1, \Pi_i^2, \ldots, \Pi_i^t, \ldots$.
- For each $t \geq 0$, $B_{i,t}^p$ C_t^p (for $p \in [0, 1]$) are well-defined.

D. Samet and D. Monderer. *Stochastic Common Learning*. Games and Economic Behavior, 9, pgs. 161 - 171, 1995.

Common Learning Theorem

Common Learning Theorem: Suppose that $(E_t)_{t\geq 1}$ is any nondecreasing sequence of events such that $\lim_{t\rightarrow\infty}\mu(E_t) = 1$. Then, for all $0 \leq p < 1$, and almost all $w \in W$, there is a time t_w (depending on w) such that for all $t' \geq t_w$ the agents commonly p-believe E_t at time t (i.e., $w \in C^p_t(E_t)$).

D. Samet and D. Monderer. *Stochastic Common Learning*. Games and Economic Behavior, 9, pgs. 161 - 171, 1995.

S. Huttegger. *Merging of Opinions and Probability Kinematics*. Review of Symbolic Logic, 2015.

In general, no set of initial beliefs is more, or less, justified than another. This implies that conditioning on the same information can lead two agents to have different posterior beliefs.

...disagreement often turns out to be transient and disappears as one gets more information. In other words, as more evidence becomes available ca consensus may emerge...diverging opinions are just a sign that not enough evidence has accumulated.

Long run consensus is a consequence of dynamic coherence under the special assumption of absolute continuity.

In this case, group agreement is a consequence of individual rationality.

However, if long run consensus fails we may be able to trace it back, not to the individual irrationality of an agent, but to the fact that the agents' initial beliefs did not observe the absolute continuity requirement.

If the absolute continuity requirement holds, another possible explanation is that merging takes place very slowly; beliefs would merge if agents were given more evidence, but this might not always be possible. The long run may be too long.

Agents assumes to undergo learning experiences that yield a sequence $P_1, P_2, \ldots$ of probability measures on $(W, \mathcal{F}_1), (W, \mathcal{F}_2), \ldots$.

Each agent believes with probability one that she will revise her probabilities by performing *probability kinematics* with $p_1, p_2, \ldots$. Each probability measure P_n is fully determined by attaching probability values to members of $\mathcal{E}_n$.

$$P_n(A) = \sum_{E \in \mathcal{E}_n} P(A \mid E) p_n(E)$$

(AC) If $P_n \ll P_{\ |\mathcal{F}_n}$, then for every $\epsilon > 0$, there is a $\delta_n > 0$ such that

$$P(B) < \delta_n \Longrightarrow P_n(B) < \epsilon$$

for all $B \in \mathcal{F}_n$.

Uniformly Absolutely Continuous

1. $P_n \ll P|_{\mathcal{F}_n}$ for each n and
2. for every $\epsilon > 0$ there is a $\delta > 0$ such that for all n,

$$P(B) < \delta \Longrightarrow P_n(B) < \epsilon$$

for all $B \in \mathcal{F}_n$

- Condition 1. requires that events which are assigned probability zero now are expected to have probability zero in the future.
- Condition 2. says that this also holds in the limit. Otherwise, there may be events $F_1, F_2, \ldots$ such that $P(F_n) \to 0$ as $n \to \infty$, while the sequence $P_1(F_1), P_2(F_2), \ldots$ is bounded away from zero.

Suppose now that there are two prior probability measures P and Q which are updated successively by probability kinematics on $\mathcal{E}_1, \mathcal{E}_2, \ldots$ using the distributions $p_1, p_2, \ldots$ and $q_1, \ldots$, respectively.

Using the Jeffrey update rule this leads to the new probability measures P_n and Q_n on $\mathcal{F}$ for $n \geq 1$.

It is quite obvious that arbitrary choices of sequences $p_1, p_2, \ldots$ and $q_1, q_2, \ldots$ need not lead to merging. But this is also true for conditioning.

Recall that one requirement of the Blackwell-Dubins theorem is that agents condition on the *same factual evidence*.

Thus, the important question is whether beliefs merge for probability kinematics whenever p_n and q_n represent the same uncertain information.

But what does it mean to get the same uncertain evidence?

Hard Jeffrey Shift

A **hard Jeffrey shift** sets values for p_n regardless of the prior probability P_{n-1}, and so may destroy any information about the partition that was encoded in the prior.

As an example, consider a measurement instrument that makes noisy observations of a physical process, such as coin flips. Let's call this setup a 'mechanical observer'.

The probability space $(W, \mathcal{F})$ represents the set of states and events of the process. At each stage n, the output of the mechanical observer is a probability distribution over the partition $\mathcal{E}_n$.

The probabilities for members of the partition are determined by repeated previous observations under symmetric conditions in order to specify measurement error.

More generally, a hard Jeffrey shift can be viewed as a noisy signal where the noise has the form of a probability distribution over a partition such that the distribution is known to every observer.

In terms of hard Jeffrey shifts, having the same uncertain evidence at stage n means that $p_n = q_n \ \mathcal{F}_n$.

Suppose, for example, that Adam and Eve are two scientists observing coin flips with the help of a mechanical measurement instrument. They might not feel comfortable approximating their learning process by conditionalization if their measurements are not precise enough. Instead, they plan to update by the same hard Jeffrey shifts at each stage of their experiment.

Can they be certain to have similar beliefs after having taken

$$\text{(M)} \qquad P_n(F) = P_{n-1}(F) \text{ for all } F \in \mathcal{F}_{n-1}$$

Theorem. Suppose that $q_n = p_n$, that the sequence Q_n, $n = 1, 2, \ldots$ is uniformaly absolutely continuous relative to Q, and that $Q \ll P$ If condition (M) holds, then $d(P_n, Q_n) \to 0$ and $n \to \infty$.

Soft Jeffrey Shifts

$\mathcal{E}_1 = \{E_1, E_2, E_3, E_4\}$ and consider the learning experience given by:

$$(\frac{1}{5} : E_1, \frac{3}{10} : E_2, \frac{1}{2} : E_3, 0 : E_n)$$

Soft Jeffrey Shifts

$\mathcal{E}_1 = \{E_1, E_2, E_3, E_4\}$ and consider the learning experience given by:

$$(\frac{1}{5} : E_1, \frac{3}{10} : E_2, \frac{1}{2} : E_3, 0 : E_n)$$

Consider instead:

- $p_1(E_1) = 2 \cdot P(E_1)$;
- $p_1(E_2) = \frac{1}{2} \cdot P(E_2)$;
- $p_1(E_3) = 5 \cdot P(E_3)$; and
- $p_1(E_4) = 0 \cdot P(E_4)$

If $P(E_1) = \frac{1}{10}$, $P(E_2) = \frac{3}{5}$, $P(E_3) = \frac{1}{10}$ and $P(E_4) = \frac{1}{5}$, then probability kinematics will lead to the same result whether or not it is a hard or "soft" Jeffrey shift.

Do beliefs merge when agents have the same soft uncertain evidence?

We are going to see that this need not be the case. If Adam and Eve start with different (but mutually absolutely continuous) prior probabilities for infinite sequences of coin flips, and if both observe principle (M) as well as undergo the same soft Jeffrey shifts, their posterior degrees of beliefs may not get close to each other in the long run. [Theorem 6.3]

Our results lead to the conclusion that, even under otherwise favorable circumstances, a soft kind of information allows individual rationality to be consistent with sustained disagreement. I don't think that this is a weakness of the broadly Bayesian approach advocated in this essay. Merging of beliefs happens when it should, i.e., under conditions which may, for example, hold for certain carefully designed scientific investigations. But the claim of merging is not a no-brainer that can be used across the board.

Plan

Day 1 Introduction to belief revision, AGM, possible worlds models, Bayesian models (time permitted)

Day 2 Bayesian models (continued), Justifying Bayesian models (Dutch books, Accuracy-based arguments), Updating probabilities

Day 3 The value of learning, Lottery Paradox, Preface Paradox, Review Paradox, Context shifts, Becoming aware

Day 4 The value of learning, Lottery Paradox, Preface Paradox, Review Paradox, Context shifts, Becoming aware (continued)

Day 5 Iterated Belief Revision, Agreement Theorems

Thank You!

epacuit@umd.edu

pacuit.org/nasslli2016/belrev/