Logical and Probabilistic Models of Belief Change

Eric Pacuit

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

August 10, 2017

Plan

- ✓ Introduction to belief revision, AGM
- ✓ Possible worlds models, Bayesian models
- ✓ Updating probabilities,The value of learning, Lottery Paradox, Preface Paradox, Review Paradox
- Day 4 Lottery Paradox, Preface Paradox, Review Paradox, Iterated belief revision, Context shifts, Becoming aware
- Day 5 Interactive epistemology (Agreement Theorems), Convergence Theorems

Taking Stock

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

Taking Stock

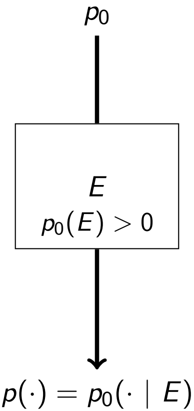
- Epistemic states: AGM, Plausibility Models, Bayesian Model (and the many variations)
- "Finding out that φ"
  - Learn that φ
  - Suppose that φ
  - Accept φ
  - ...

Taking Stock

- 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)
  - ...

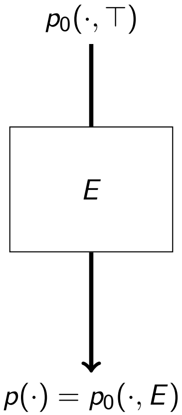

p_0
E
$p_0(E) > 0$
$p(\cdot) = p_0(\cdot \mid E)$

p_0

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

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

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

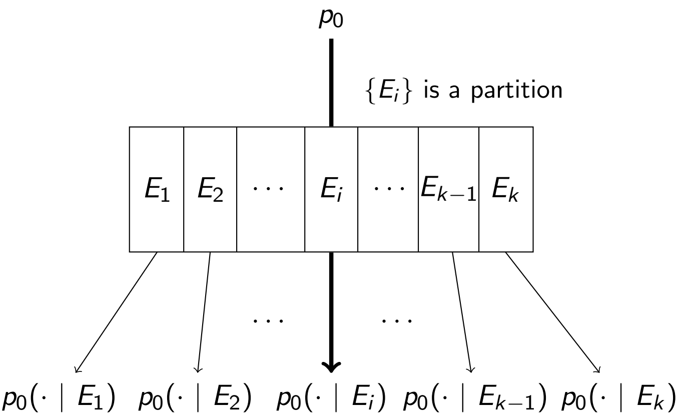

$p_0(\cdot, \top)$
E
$p(\cdot) = p_0(\cdot, E)$

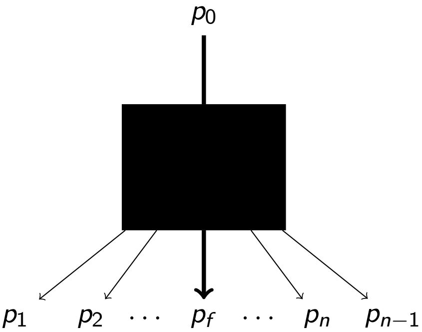
p_0

$\mathcal{C}$: set of constraints

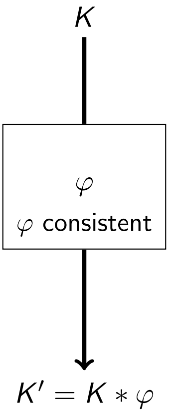
p satisfies $\mathcal{C}$

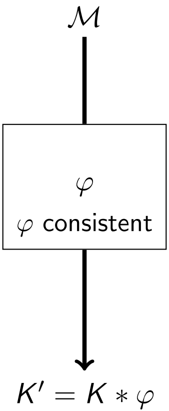
p minimizes $I(p, p_0) = \sum_i p_1(x_i) \log \frac{p_1(x_i)}{p_0(x_i)}$

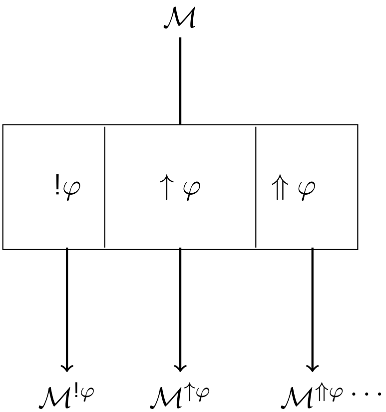

p_0
$\{E_i\}$ is a partition
E_1
E_2
$\cdots$
E_i
$\cdots$
E_{k-1}
E_k
$\cdots$
$\cdots$
$p_0(\cdot \mid E_1)$
$p_0(\cdot \mid E_2)$
$p_0(\cdot \mid E_i)$
$p_0(\cdot \mid E_{k-1})$
$p_0(\cdot \mid E_k)$



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


K
φ
φ consistent
$K' = K * \varphi$


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



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

Two important distinctions.

1. If Shakespeare had not written Hamlet, it would never have been written.
2. If Shakespeare didn't write Hamlet, someone else did.

1. is a causal counterfactual, and 2. is an expression of a belief revision policy.

1. General Smith is a shrewd judge of character—he knows (better than I) who is brave and who is not.
2. The general sends only brave men into battle.
3. Private Jones is cowardly.

I believe that (1) Jones would run away if he were sent into battle and (2) if Jones *is* sent into battle, then he won't run away.

1. Ann cheats — she has seen her opponent's cards.
2. Ann has a losing hand, since I have seen both her hand and her opponent's.
3. Ann is rational.

So, I conclude that she will not bet. But how should I revise my beliefs if I learn that Ann did bet?

1. Ann cheats — she has seen her opponent's cards.
2. Ann has a losing hand, since I have seen both her hand and her opponent's.
3. Ann is rational.

So, I conclude that she will not bet. But how should I revise my beliefs if I learn that Ann did bet?

It may be perfectly reasonable for me to be disposed to give up 2.

1. Ann cheats — she has seen her opponent's cards.
2. Ann has a losing hand, since I have seen both her hand and her opponent's.
3. Ann is rational.

So, I conclude that she will not bet. But how should I revise my beliefs if I learn that Ann did bet?

It may be perfectly reasonable for me to be disposed to give up 2.

I believe that (1) I Ann *were* to bet, she would lose (since she has a losing hand) and (2) If I were to learn that she *did* bet, I would conclude she will win.

Updating vs. Revising

Revision vs. Update

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

Revision vs. Update

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

An important distinction:

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

Revision vs. Update

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

An important distinction:

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

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

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

The logic of updating differs from that of revision. This can be seen from the following example:

To begin with, the agent knows that there is either a book on the table (p) or a magazine on the table (q), but not both.

- Case 1: The agent is told that there is a book on the table. She concludes that there is no magazine on the table. This is revision.
- Case 2: The agent is told that after the first information was given, a book has been put on the table. In this case she should not conclude that there is no magazine on the table. This is updating.

J. Lang. *Belief Update Revisited*. Proceedings of IJCAI-07.

N. Friedman and J. Halpern. *Modeling Belief in Dynamics Systems Part II: Revision and Update*. Journal of Artificial Intelligence Research, 10, pp. 117 - 167 (1999).

A. Herzig. *Belief Change Operations: A shorty history of nearly everything, told in dynamic logic of propositional assignments*. AAAI, 2014.

KM Postulates

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

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

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

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

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

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

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

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

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

Updating and Revising

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

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

In the literature on belief change the distinction between static and dynamic environment has become important....

In the literature on belief change the distinction between static and dynamic environment has become important....it seems right to say that belief change due to new information in an unchanging environment has come to be called belief revision (the static case, in the sense that the "world" remains unchanged), while it is fairly generally accepted to use the term belief update for belief change that is due to reported changes in the environment itself (the dynamic case, in the sense that the "world" changes; compare our analysis in the last subsection).

In the literature on belief change the distinction between static and dynamic environment has become important....it seems right to say that belief change due to new information in an unchanging environment has come to be called belief revision (the static case, in the sense that the "world" remains unchanged), while it is fairly generally accepted to use the term belief update for belief change that is due to reported changes in the environment itself (the dynamic case, in the sense that the "world" changes; compare our analysis in the last subsection). It has been held for some time that these cases support different logics (...)

In the literature on belief change the distinction between static and dynamic environment has become important....it seems right to say that belief change due to new information in an unchanging environment has come to be called belief revision (the static case, in the sense that the "world" remains unchanged), while it is fairly generally accepted to use the term belief update for belief change that is due to reported changes in the environment itself (the dynamic case, in the sense that the "world" changes; compare our analysis in the last subsection). It has been held for some time that these cases support different logics (...) The established tradition notwithstanding, it would be interesting to see a really convincing argument for tying AGM revision to static environments.

Hannes Leitgeb and Krister Segerberg. *Dynamic doxastic logic: why, how, and where to?*. Synthese, 155, pp. 167 - 190 (2007).

1. Beliefs in a dynamic world (ontic changes)
2. Supposing
3. Learning

Leitgeb & Segerberg

...given new evidence, we find that in the case of belief revision the agent tries to change his beliefs in a manner such that the worlds that he subsequently believes to be in comprise the *subjectively most plausible deviation* from the worlds he originally believed to inhabit.

Leitgeb & Segerberg

...given new evidence, we find that in the case of belief revision the agent tries to change his beliefs in a manner such that the worlds that he subsequently believes to be in comprise the *subjectively most plausible deviation* from the worlds he originally believed to inhabit.

However, when confronted with the same evidence in belief update, the agent tries to change his beliefs in a way such that the worlds that he subsequently believes to be in are as **objectively similar as possible** to the worlds he originally believed to be the most plausible candidates for being the actual world.

Supposing vs. Learning

Note that in this setting the difference between supposing and updating is mathematically clearcut. In a typical Bayesian updating situation one is uncertain about the chances, and so ones subjective probability distribution on the outcome space is a mixture of the possible chance distributions. Updating is an operation which typically takes one from one point in the interior of the convex closure of the chance distributions to another; supposing moves from one chance distribution to another.

B. Skyrms. *Updating, Supposing and MAXENT*. Theory and Decision, 22, pp. 225 - 246, 1987.

Such normative virtues suggest a psychological question. One way of formulating (1) is that *supposing* an event B should have the same impact on the credibility of an event A as *learning* B. Is this true for typical assessments of chance? For example, is the judged probability of a Democratic victory in 2012 supposing that Hilary Clinton is the vice presidential candidate the same as the judged probability of a Democratic victory in 2012 after learning that Clinton, as a matter of fact, is the vice presidential candidate?

Jiaying Zhao, Vincenzo Crupi, Katya Tentori, Branden Fitelson, and Daniel Osherson. *Updating: Learning versus supposing*. Cognition 124 (2012) 373378.

Learn that φ
Suppose that φ

$$K_0 \quad \Longrightarrow \quad K_t = K_0 * \varphi$$

$$\updownarrow \qquad\qquad\qquad \updownarrow$$

Learning experience

$$p_0 \quad \Longrightarrow \quad p_t = ???$$

Learn that φ
Suppose that φ

$$K_0 \Longrightarrow K_t = K_0 * \varphi$$

Learning experience

$$p_0 \Longrightarrow p_t = ???$$

Bridge Principles

Probability 1: $Bel(A)$ iff $P(A) = 1$

Bridge Principles

Probability 1: $Bel(A)$ iff $P(A) = 1$

The Lockean Thesis: $Bel(A)$ iff $P(A) > r$

Bridge Principles

Probability 1: $Bel(A)$ iff $P(A) = 1$

The Lockean Thesis: $Bel(A)$ iff $P(A) > r$

Decision-theoretic accounts: $Bel(A)$ iff
$\sum_{w \in W} P(\{w\}) \cdot u(bel\ A, w)$ has such-and-such property

Bridge Principles

Probability 1: $Bel(A)$ iff $P(A) = 1$

The Lockean Thesis: $Bel(A)$ iff $P(A) > r$

Decision-theoretic accounts: $Bel(A)$ iff $\sum_{w \in W} P(\{w\}) \cdot u(bel\ A, w)$ has such-and-such property

The Nihilistic proposal: "...no explication of belief is possible within the confines of the probability model."

Preface Paradox

D. Makinson. *The Paradox of the Preface*. Analysis, 25, 205 - 207, 1965.

Preface Paradox

Suppose that in the course of his book an author makes a great many assertions: $s_1, s_2, \ldots, s_n$.

Preface Paradox

Suppose that in the course of his book an author makes a great many assertions: $s_1, s_2, \ldots, s_n$.

Given each one of these, he believes that it is true (for each i, $B_A(s_i)$)

Preface Paradox

Suppose that in the course of his book an author makes a great many assertions: $s_1, s_2, \ldots, s_n$.

Given each one of these, he believes that it is true (for each i, $B_A(s_i)$)

If he has already written other books, and received corrections from readers and reviewers, he may also believe that not everything he has written in his latest book is true.

Preface Paradox

Suppose that in the course of his book an author makes a great many assertions: $s_1, s_2, \ldots, s_n$.

Given each one of these, he believes that it is true (for each i, $B_A(s_i)$)

If he has already written other books, and received corrections from readers and reviewers, he may also believe that not everything he has written in his latest book is true.

$B_A(\neg(s_1 \wedge s_2 \wedge \cdots \wedge s_n))$

Preface Paradox

Suppose that in the course of his book an author makes a great many assertions: $s_1, s_2, \ldots, s_n$.

Given each one of these, he believes that it is true (for each i, $B_A(s_i)$)

If he has already written other books, and received corrections from readers and reviewers, he may also believe that not everything he has written in his latest book is true.

$B_A(\neg(s_1 \wedge s_2 \wedge \cdots \wedge s_n))$

But $\{s_1, \ldots, s_n, \neg(s_1 \wedge \cdots \wedge s_n)\}$ is logically inconsistent.

Preface Paradox

A philosopher who asserts "all of my present philosophical positions are correct" would be regarded as rash and over-confident

A philosopher who asserts "at least some of my present philosophical beliefs will turn out to be incorrect" is simply being sensible and honest.

Preface Paradox

1. each belief from the set $\{s_1, \ldots, s_n, s_{n+1}\}$ is rational
2. the set $\{s_1, \ldots, s_n, s_{n+1}\}$ of beliefs is rational.

1\. does not necessarily imply 2.

Preface Paradox: The Problem

"The author of the book is being rational even though inconsistent. More than this: he is being rational even though he believes each of a certain collection of statements, which *he knows* are logically incompatible....this appears to present a living and everyday example of a situation which philosophers have commonly dismissed as absurd; that it is sometimes rational to hold incompatible beliefs."

D. Makinson. *The Paradox of the Preface.* Analysis, 25, 205 - 207, 1965.

H. Leitgeb. *The Review Paradox: On the Diachronic Costs of Not Closing Rational Belief Under Conjunction*. Nous, 2013.

Bel_t is the set of propositions believed at time t

P_t is the agent's degree of belief function at time t

$t' > t$

P1 If the degrees of belief that the agents assigns to two propositions are identical then either the agent believes both of them or neither of them.

For all X, Y: if $P_t(X) = P_t(Y)$, then $Bel_t(X)$ iff $Bel_t(Y)$.

P2 If the agent already believes X, then updating on the piece of evidence X does not change her system of (all-or-nothing) beliefs at all.

For all X: if the evidence that the agent obtains between t and $t' > t$ is the proposition X, but it holds already that $Bel_t(X)$, then for all Y:

$$Bel_{t'}(Y) \text{ iff } Bel_t(Y)$$

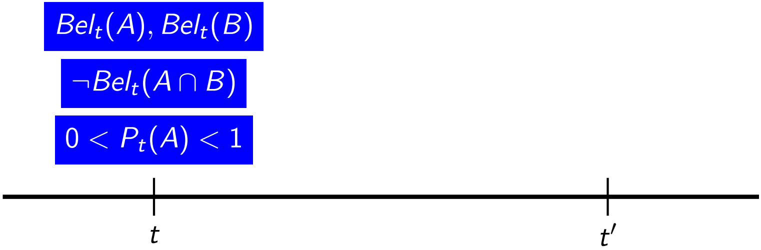
P3 When the agent learns, this is captured probabilistically by conditionalization.

For all X (with $P_t(X) > 0$): if the evidence that the agent obtains between t and $t' > t$ is the proposition X, but it holds already that $Bel_t(X)$, then for all Y:

$$P_{t'}(Y) = P_t(Y \mid X)$$

Assume $Bel_t(A), Bel_t(B)$ but not $Bel_t(A \cap B)$

- Suppose that the agent receive A as evidence.
- $P_{t'}(B) = P_t(B \mid A) = P_t(A \cap B \mid A) = P_{t'}(A \cap B)$.
- By P1, the agent must have the same doxastic attitude towards B and $A \cap B$.
- By P2, the agent's attitude towards B and $A \cap B$ must be the same at t' as at t.
- But, $Bel_t(B)$ and not $Bel_t(A \cap B)$



Assumption

$Bel_t(A), Bel_t(B)$

$\neg Bel_t(A \cap B)$

$0 < P_t(A) < 1$

$P_{t'}(B) = P_t(B \mid A)$

$P_{t'}(A \cap B) = P_t(A \cap B \mid A) = P_t(B \mid A)$

t Receives evidence A t'

By P3

$Bel_t(A), Bel_t(B)$

$\neg Bel_t(A \cap B)$

$0 < P_t(A) < 1$

$P_{t'}(B) = P_t(B \mid A)$

$P_{t'}(A \cap B) = P_t(A \cap B \mid A) = P_t(B \mid A)$

$Bel_{t'}(B)$ iff $Bel_{t'}(A \cap B)$

t — Receives evidence A — t'

By P1

$Bel_t(A)$, $Bel_t(B)$

$\neg Bel_t(A \cap B)$

$0 < P_t(A) < 1$

$P_{t'}(B) = P_t(B \mid A)$

$P_{t'}(A \cap B) = P_t(A \cap B \mid A) = P_t(B \mid A)$

$Bel_{t'}(B)$ iff $Bel_{t'}(A \cap B)$

t — Receives evidence A — t'

$Bel_t(A)$ iff $Bel_{t'}(A)$

$Bel_t(B)$ iff $Bel_{t'}(B)$

$Bel_t(A \cap B)$ iff $Bel_{t'}(A \cap B)$

By P2

$Bel_t(B)$

$\neg Bel_t(A \cap B)$

$0 < P_t(A) < 1$

$P_{t'}(B) = P_t(B \mid A)$

$P_{t'}(A \cap B) = P_t(A \cap B \mid A) = P_t(B \mid A)$

$Bel_{t'}(B)$ iff $Bel_{t'}(A \cap B)$

t Receives evidence A t'

$Bel_t(A)$ iff $Bel_{t'}(A)$

$Bel_t(B)$ iff $Bel_{t'}(B)$

$Bel_t(A \cap B)$ iff $Bel_{t'}(A \cap B)$

$Bel_t(B)$ iff $Bel_{t'}(B)$ iff $Bel_{t'}(A \cap B)$ iff $Bel_t(A \cap B)$

Lottery Paradox

H. Kyburg. *Probability and the Logic of Rational Belief*. Wesleyan University Press, 1961.

G. Wheeler. *A Review of the Lottery Paradox*. Probability and Inference: Essays in honor of Henry E. Kyburg, Jr., College Publications, 2007.

Lottery Paradox

Lottery Paradox

Consider a fair lottery with 1,000,000 tickets and one prize.

Lottery Paradox

Consider a fair lottery with 1,000,000 tickets and one prize.

The probability that a given ticket will win is 0.000001 $(1/1,000,000)$ and the probability that it will not win is 0.999999.

Lottery Paradox

Consider a fair lottery with 1,000,000 tickets and one prize.

The probability that a given ticket will win is 0.000001 $(1/1,000,000)$ and the probability that it will not win is 0.999999.

"Surely if a sheer probability is ever sufficient to warrant the acceptance of a hypothesis, this is a case"

Lottery Paradox

Consider a fair lottery with 1,000,000 tickets and one prize.

The probability that a given ticket will win is 0.000001 $(1/1,000,000)$ and the probability that it will not win is 0.999999.

"Surely if a sheer probability is ever sufficient to warrant the acceptance of a hypothesis, this is a case"

For each lottery ticket t_i $(i = 1, \ldots, 1000000)$, the agent believes that t_i will loose $B_A(\neg$'t_i will win'$)$

Lottery Paradox

A **rule of acceptance**: If S and T are acceptable statements, their conjunction is also acceptable.

Lottery Paradox

A **rule of acceptance**: If S and T are acceptable statements, their conjunction is also acceptable.

So, the conjunction $\bigwedge_{i=1}^{1000000}$ 't_i will not win' should be accepted. That is, the agent should rationally accept that *no lottery ticket will win*.

Lottery Paradox

A **rule of acceptance**: If S and T are acceptable statements, their conjunction is also acceptable.

So, the conjunction $\bigwedge_{i=1}^{1000000}$ 't_i will not win' should be accepted. That is, the agent should rationally accept that *no lottery ticket will win*.

But, this is a fair lottery, so at least one ticket is *guaranteed* to win!

The Lottery Paradox

Kyburg: The following are inconsistent,

1. It is rational to accept a proposition that is very likely true,
2. It is not rational to accept a propositional that you are aware is inconsistent
3. It is rational to accept a proposition P and it is rational to accept another proposition P' then it is rational to accept $P \wedge P'$

Learn that φ

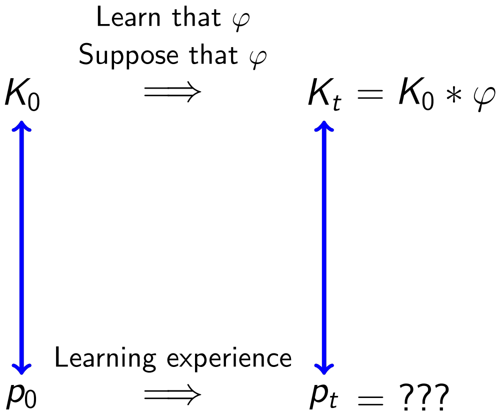
Suppose that φ

$$K_0 \Longrightarrow K_t = K_0 * \varphi$$

$$\updownarrow \qquad\qquad \updownarrow$$

Learning experience

$$p_0 \Longrightarrow p_t = ???$$


Learn that φ
Suppose that φ
$K_0 \Longrightarrow K_t = K_0 * \varphi$
Learning experience
$p_0 \Longrightarrow p_t = ???$

Bridge Principles

Probability 1: $Bel(A)$ iff $P(A) = 1$

Bridge Principles

Probability 1: $Bel(A)$ iff $P(A) = 1$

The Lockean Thesis: $Bel(A)$ iff $P(A) > r$

Bridge Principles

Probability 1: $Bel(A)$ iff $P(A) = 1$

The Lockean Thesis: $Bel(A)$ iff $P(A) > r$

Decision-theoretic accounts: $Bel(A)$ iff
$\sum_{w \in W} P(\{w\}) \cdot u(bel\ A, w)$ has such-and-such property

Bridge Principles

Probability 1: $Bel(A)$ iff $P(A) = 1$

The Lockean Thesis: $Bel(A)$ iff $P(A) > r$

Decision-theoretic accounts: $Bel(A)$ iff $\sum_{w \in W} P(\{w\}) \cdot u(bel\ A, w)$ has such-and-such property

The Nihilistic proposal: "...no explication of belief is possible within the confines of the probability model."

T. Shear, J. Weisberg and B. Fitelson. *Two Approaches to Belief Revision*. manuscript, 2016.

$$u(B(X), w) = \begin{cases} \mathfrak{r} & \text{if } X \text{ is true at } w \\ -\mathfrak{w} & \text{if } X \text{ is false at } w \end{cases}$$

$$1 \geq \mathfrak{w} > \left(\frac{1+\sqrt{5}}{2}\right) \cdot \mathfrak{r} > 0$$

$$EEU(B(X), p) := \sum_{w \in W} p(w) u(B(X), w)$$

$$EEU(B, p) := \sum_{X \in B} EEU(B(X), w)$$

Theorem (Dorst). An agent's belief set B maximizes EEU from the point of view of her credence function p if and only if, for every $X \in B$

$$p(X) > \frac{\mathfrak{w}}{\mathfrak{r} + \mathfrak{w}}$$

$$B \divideontimes E = \{X \mid p(X \mid E) > \frac{\mathfrak{w}}{\mathfrak{r} + \mathfrak{w}}\}$$

(P2) If an agent initially believes X (i.e., if $X \in B$), then updating B on X should *not change* B. [More formally, $X \in B$ implies that $B' = B \star X = B$

AGM Postulates

Closure $B * E = Cn(B * E)$

AGM Postulates

Closure $B * E = Cn(B * E)$

Success $E \in B * E$

AGM Postulates

Closure $B * E = Cn(B * E)$

Success $E \in B * E$

Inclusion $B * E \subseteq Cn(B \cup \{E\})$

AGM Postulates

Closure $B * E = Cn(B * E)$

Success $E \in B * E$

Inclusion $B * E \subseteq Cn(B \cup \{E\})$

Vacuity If E is consistent with B, then $B * E \supseteq Cn(B \cup \{E\})$

AGM Postulates

Closure $B * E = Cn(B * E)$

Success $E \in B * E$

Inclusion $B * E \subseteq Cn(B \cup \{E\})$

Vacuity If E is consistent with B, then $B * E \supseteq Cn(B \cup \{E\})$

Consistency If E is not self-contradictory, then $B * E$ is consistent

AGM Postulates

Closure $B * E = Cn(B * E)$

Success $E \in B * E$

Inclusion $B * E \subseteq Cn(B \cup \{E\})$

Vacuity If E is consistent with B, then $B * E \supseteq Cn(B \cup \{E\})$

Consistency If E is not self-contradictory, then $B * E$ is consistent

Extensionality If $X \equiv Y \in Cn(\emptyset)$, then $B * X = B * Y$

AGM Postulates

Closure $B * E = Cn(B * E)$

Success $E \in B * E$

Inclusion $B * E \subseteq Cn(B \cup \{E\})$

Vacuity If E is consistent with B, then $B * E \supseteq Cn(B \cup \{E\})$

Consistency If E is not self-contradictory, then $B * E$ is consistent

Extensionality If $X \equiv Y \in Cn(\emptyset)$, then $B * X = B * Y$

Superexpansion $B * (X \wedge Y) \subseteq Cn((B * X) \cup \{Y\})$

AGM Postulates

Closure $B * E = Cn(B * E)$

Success $E \in B * E$

Inclusion $B * E \subseteq Cn(B \cup \{E\})$

Vacuity If E is consistent with B, then $B * E \supseteq Cn(B \cup \{E\})$

Consistency If E is not self-contradictory, then $B * E$ is consistent

Extensionality If $X \equiv Y \in Cn(\emptyset)$, then $B * X = B * Y$

Superexpansion $B * (X \wedge Y) \subseteq Cn((B * X) \cup \{Y\})$

Subexpansion If Y is consistent with $Cn(B * X)$, then $B * (X \wedge Y) \supseteq Cn((B * X) \cup \{Y\})$

Claim. (P2) follows from the AGM postulates Closure, Inclusion and Vacuity.

Claim. (P2) follows from the AGM postulates Closure, Inclusion and Vacuity.

1. $X \in B$. Assumption

Claim. (P2) follows from the AGM postulates Closure, Inclusion and Vacuity.

1.	$X \in B$.	Assumption
2.	$B \not\vdash \neg X$.	B is consistent

Claim. (P2) follows from the AGM postulates Closure, Inclusion and Vacuity.

1.	$X \in B$.	Assumption
2.	$B \not\vdash \neg X$.	B is consistent
3.	$B * X = Cn(B \cup \{X\})$.	(2), Vacuity, Inclusion

Claim. (P2) follows from the AGM postulates Closure, Inclusion and Vacuity.

1.	$X \in B$.	Assumption
2.	$B \nvdash \neg X$.	B is consistent
3.	$B * X = Cn(B \cup \{X\})$.	(2), Vacuity, Inclusion
4.	$Cn(B \cup \{X\}) = Cn(B)$.	(1)

Claim. (P2) follows from the AGM postulates Closure, Inclusion and Vacuity.

1.	$X \in B$.	Assumption
2.	$B \not\vdash \neg X$.	B is consistent
3.	$B * X = Cn(B \cup \{X\})$.	(2), Vacuity, Inclusion
4.	$Cn(B \cup \{X\}) = Cn(B)$.	(1)
5.	$B * X = Cn(B) = B$	Closure

Claim. (P2) follows from the AGM postulates Closure, Inclusion and Vacuity.

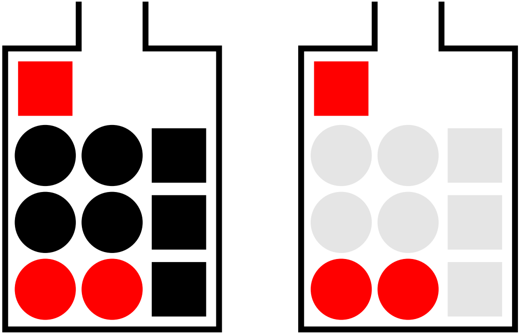
1.	$X \in B$.	Assumption
2.	$B \not\vdash \neg X$.	B is consistent
3.	$B * X = Cn(B \cup \{X\})$.	(2), Vacuity, Inclusion
4.	$Cn(B \cup \{X\}) = Cn(B)$.	(1)
5.	$B * X = Cn(B) = B$	Closure

Theorem. (Gärdenfors) Suppose $\mathfrak{r} = 0$, $\mathfrak{w} = 1$, B is synchronically coherent in the *EUT* sense, and that for all propositions X and Y that our agent might learn, $p(X \mid Y) > 0$. Then $\divideontimes$ satisfies all eight of the AGM postulates above.

EUT revision satisfies:

- Success.
- Inclusion.
- Extensionality.
- Superexpansion.

Proposition. Non-Extremal *EUT* Revision violates Vacuity — even if it is restricted to deductively cogent agents.

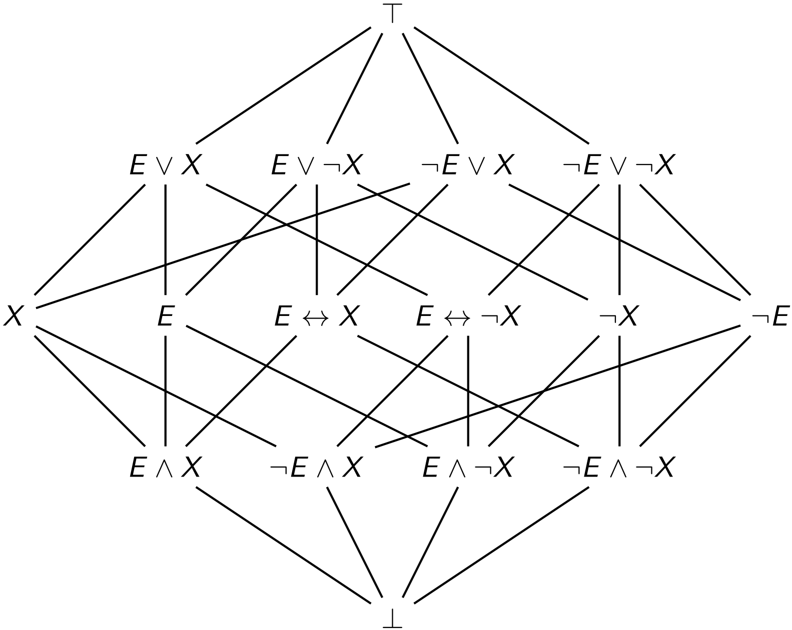


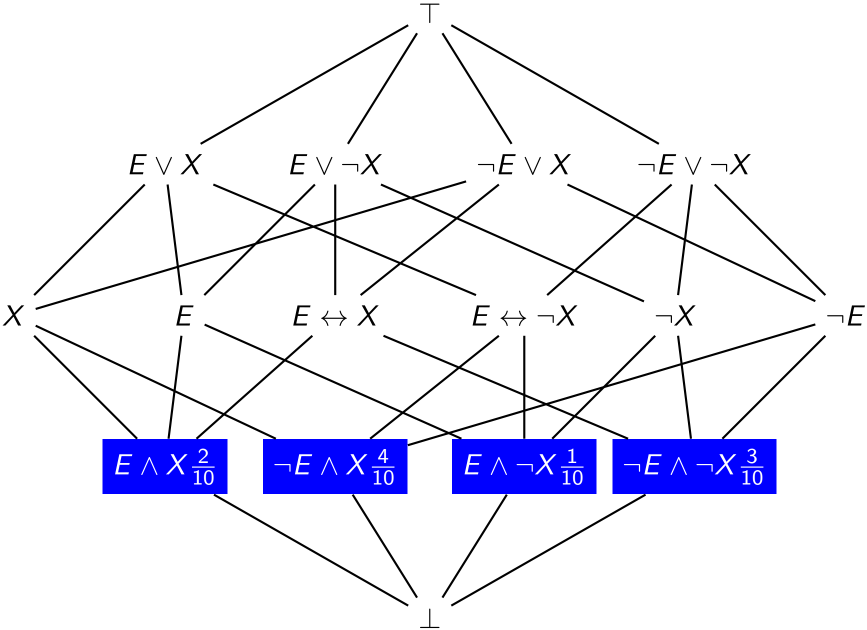
(A) Prior distribution (B) Posterior (given E)

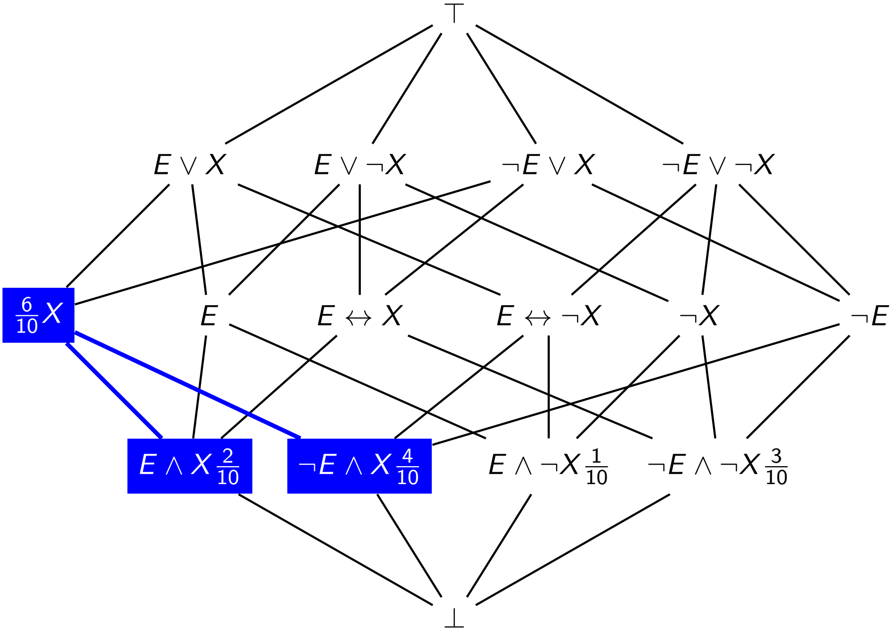
FIGURE 2. Visualization of counterexample to Vacuity for EUT Revision

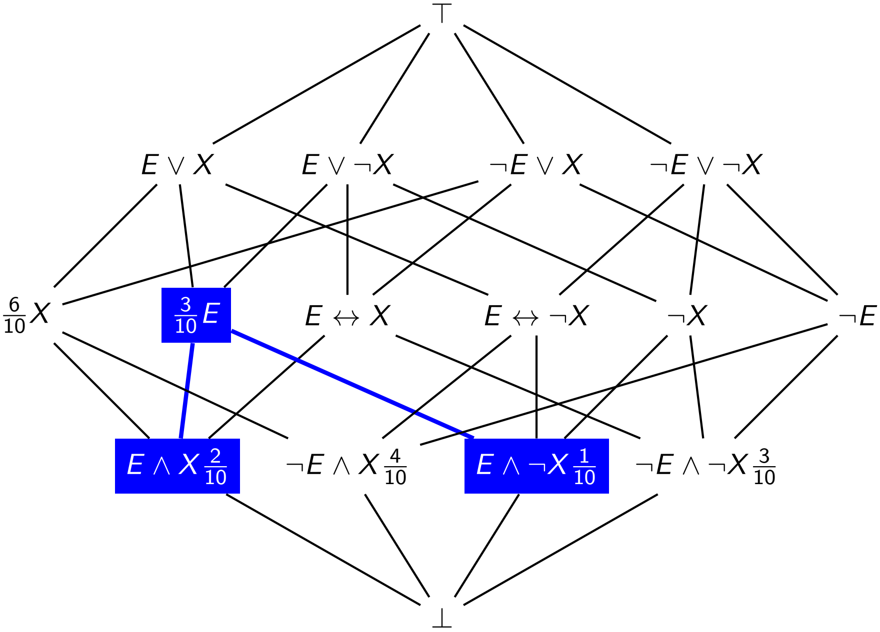
E :=‘The object sampled from the urn is red’

X := ‘The object sampled from the urn is a circle’.

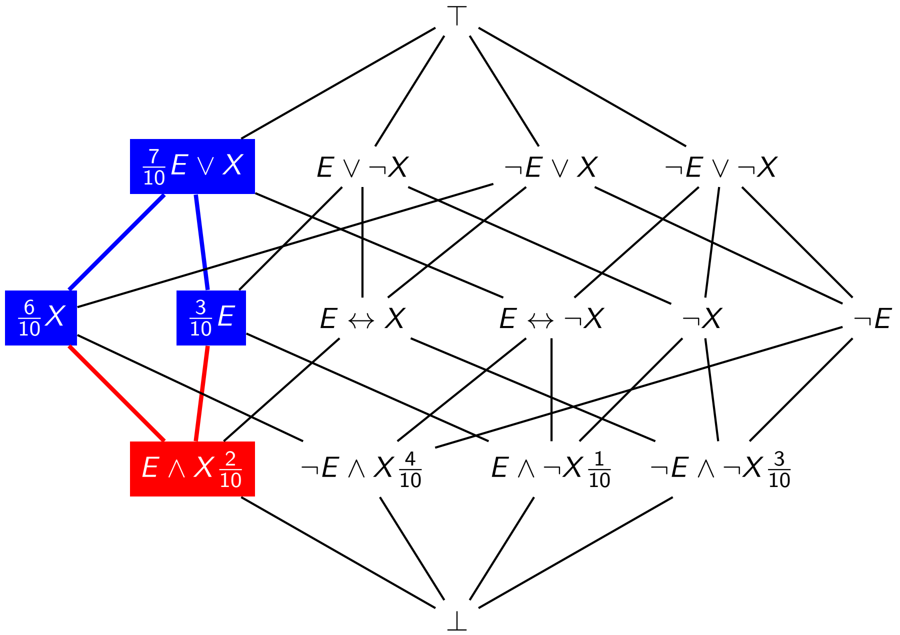

⊤
E ∨ X
E ∨ ¬X
¬E ∨ X
¬E ∨ ¬X
X
E
E ↔ X
E ↔ ¬X
¬X
¬E
E ∧ X
¬E ∧ X
E ∧ ¬X
¬E ∧ ¬X
⊥

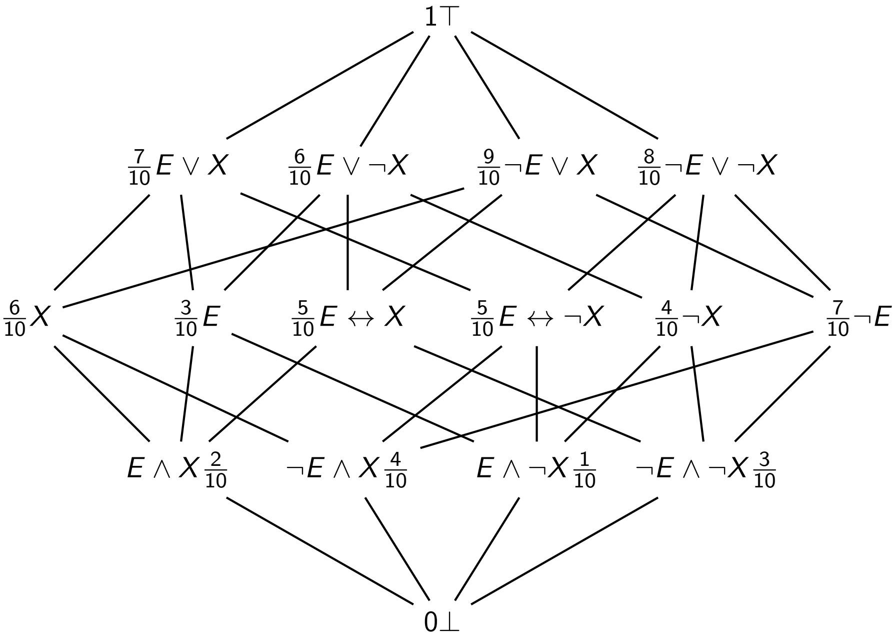

$\top$
$E \vee X$
$E \vee \neg X$
$\neg E \vee X$
$\neg E \vee \neg X$
X
E
$E \leftrightarrow X$
$E \leftrightarrow \neg X$
$\neg X$
$\neg E$
$E \wedge X\, \frac{2}{10}$
$\neg E \wedge X\, \frac{4}{10}$
$E \wedge \neg X\, \frac{1}{10}$
$\neg E \wedge \neg X\, \frac{3}{10}$
$\bot$


$\top$
$E \vee X$
$E \vee \neg X$
$\neg E \vee X$
$\neg E \vee \neg X$
$\frac{6}{10} X$
E
$E \leftrightarrow X$
$E \leftrightarrow \neg X$
$\neg X$
$\neg E$
$E \wedge X \, \frac{2}{10}$
$\neg E \wedge X \, \frac{4}{10}$
$E \wedge \neg X \, \frac{1}{10}$
$\neg E \wedge \neg X \, \frac{3}{10}$
$\bot$

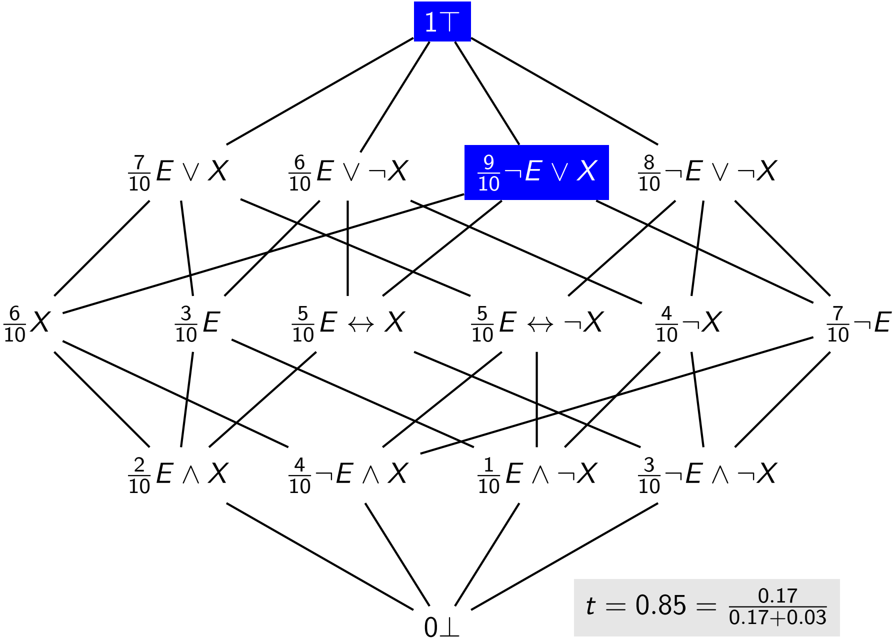


⊤
$E \vee X$
$E \vee \neg X$
$\neg E \vee X$
$\neg E \vee \neg X$
$\frac{6}{10} X$
$\frac{3}{10} E$
$E \leftrightarrow X$
$E \leftrightarrow \neg X$
$\neg X$
$\neg E$
$E \wedge X \frac{2}{10}$
$\neg E \wedge X \frac{4}{10}$
$E \wedge \neg X \frac{1}{10}$
$\neg E \wedge \neg X \frac{3}{10}$
⊥

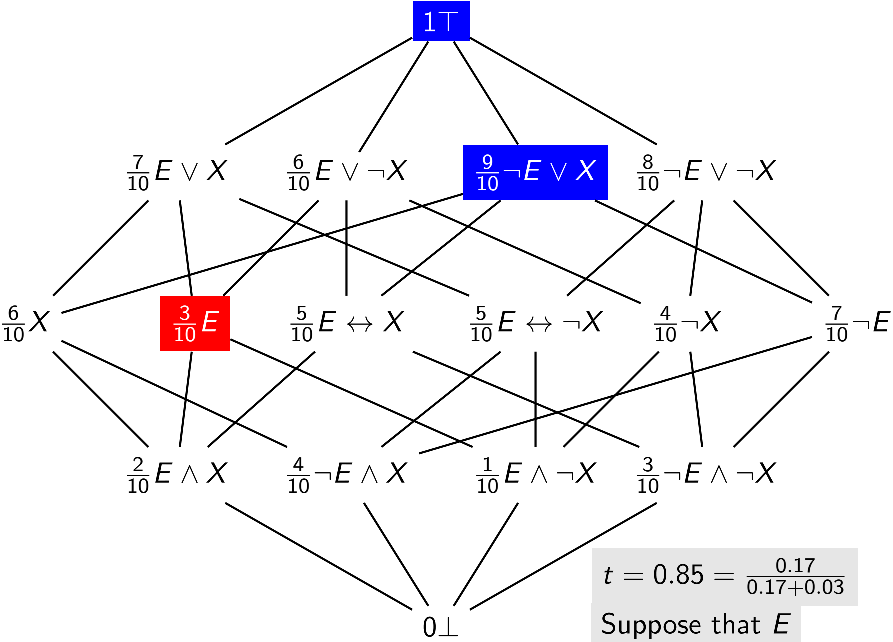

$\top$
$\frac{7}{10} E \vee X$
$E \vee \neg X$
$\neg E \vee X$
$\neg E \vee \neg X$
$\frac{6}{10} X$
$\frac{3}{10} E$
$E \leftrightarrow X$
$E \leftrightarrow \neg X$
$\neg X$
$\neg E$
$E \wedge X \frac{2}{10}$
$\neg E \wedge X \frac{4}{10}$
$E \wedge \neg X \frac{1}{10}$
$\neg E \wedge \neg X \frac{3}{10}$
$\bot$



$1 \top$
$\frac{7}{10} E \vee X$
$\frac{6}{10} E \vee \neg X$
$\frac{9}{10} \neg E \vee X$
$\frac{8}{10} \neg E \vee \neg X$
$\frac{6}{10} X$
$\frac{3}{10} E$
$\frac{5}{10} E \leftrightarrow X$
$\frac{5}{10} E \leftrightarrow \neg X$
$\frac{4}{10} \neg X$
$\frac{7}{10} \neg E$
$E \wedge X \frac{2}{10}$
$\neg E \wedge X \frac{4}{10}$
$E \wedge \neg X \frac{1}{10}$
$\neg E \wedge \neg X \frac{3}{10}$
$0 \bot$

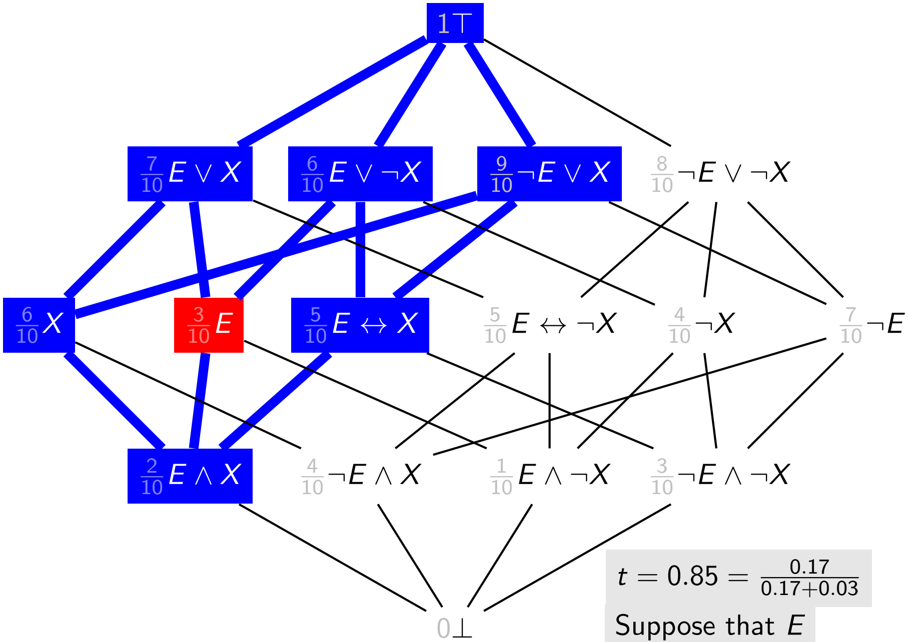


$$t = 0.85 = \frac{0.17}{0.17+0.03}$$



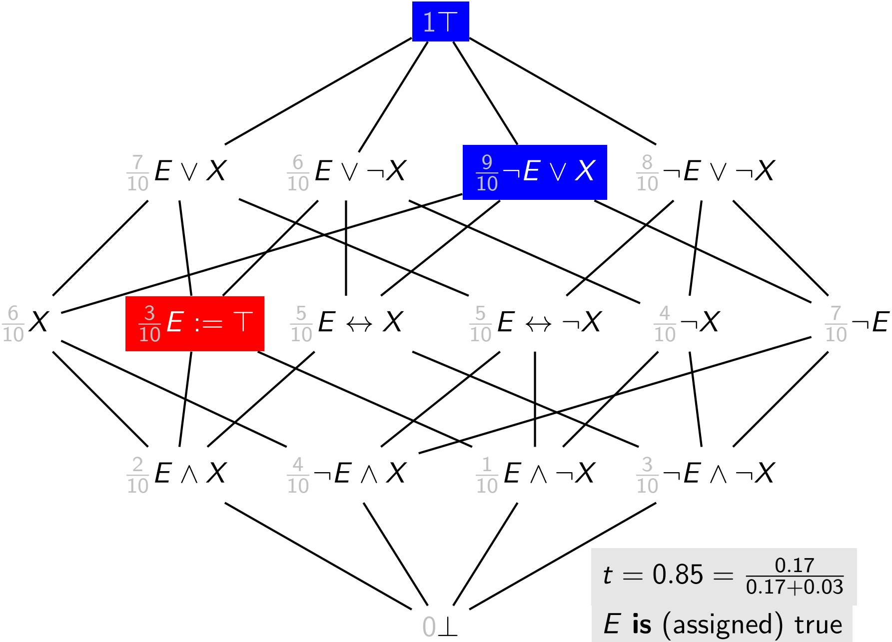
$t = 0.85 = \frac{0.17}{0.17+0.03}$

Suppose that E



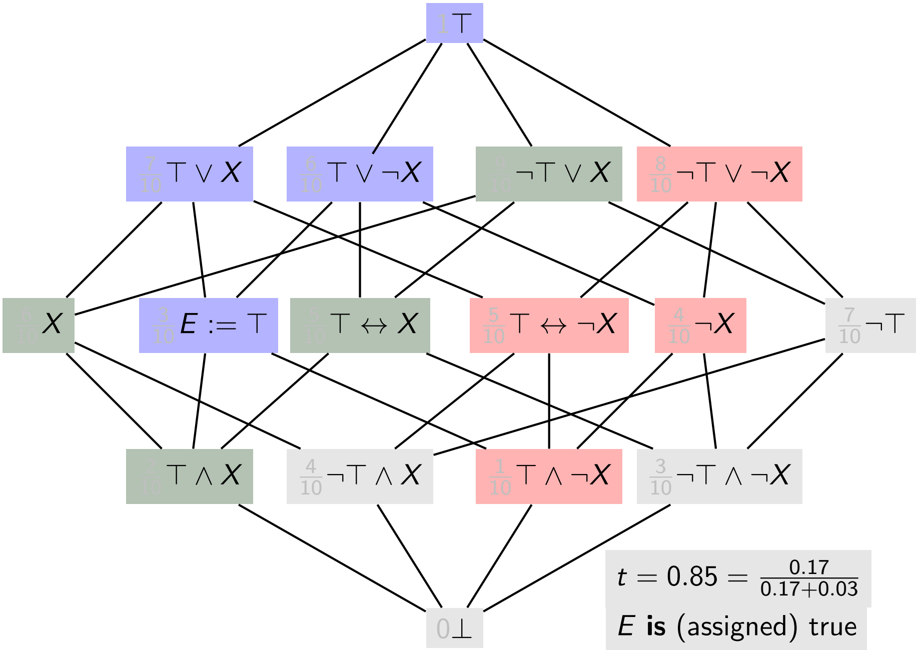
$t = 0.85 = \frac{0.17}{0.17+0.03}$

Suppose that E



$t = 0.85 = \frac{0.17}{0.17+0.03}$

E **is** (assigned) true



$t = 0.85 = \frac{0.17}{0.17+0.03}$

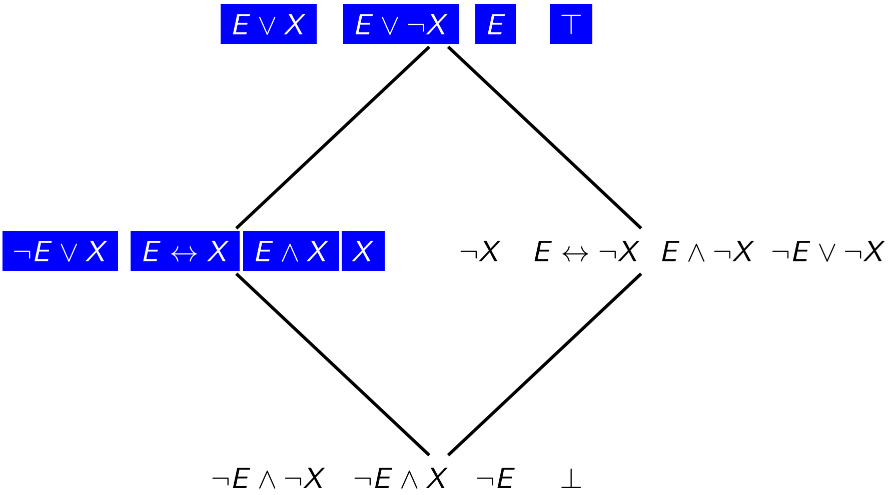
E **is** (assigned) true

$E \vee X$ $E \vee \neg X$ E $\top$

$\neg E \vee X$ $E \leftrightarrow X$ $E \wedge X$ X $\neg X$ $E \leftrightarrow \neg X$ $E \wedge \neg X$ $\neg E \vee \neg X$

$\neg E \wedge \neg X$ $\neg E \wedge X$ $\neg E$ $\bot$

E **is** (assigned) true



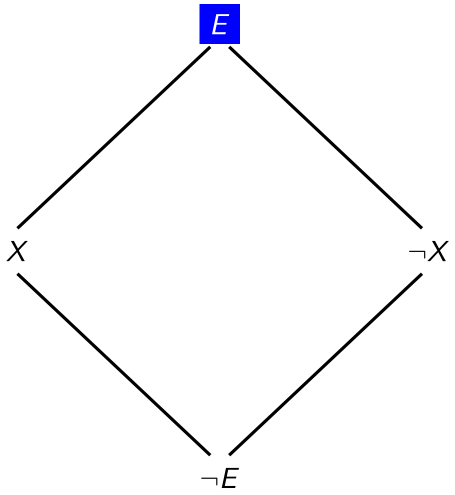
E **is** (assigned) true

$E \vee X$ $E \vee \neg X$ E $\top$

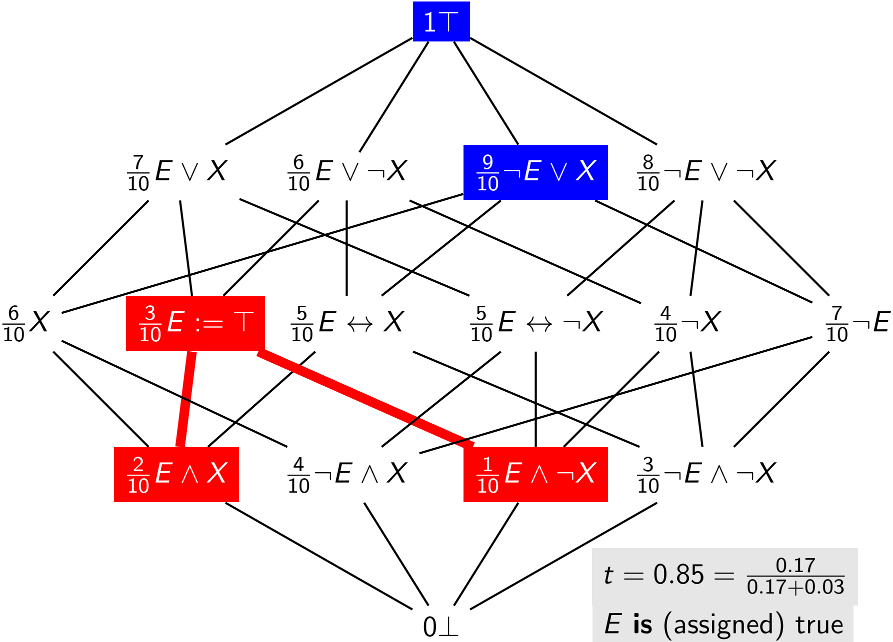
$\neg E \vee X$ $E \leftrightarrow X$ $E \wedge X$ X $\neg X$ $E \leftrightarrow \neg X$ $E \wedge \neg X$ $\neg E \vee \neg X$

$\neg E \wedge \neg X$ $\neg E \wedge X$ $\neg E$ $\bot$

E **is** (assigned) true

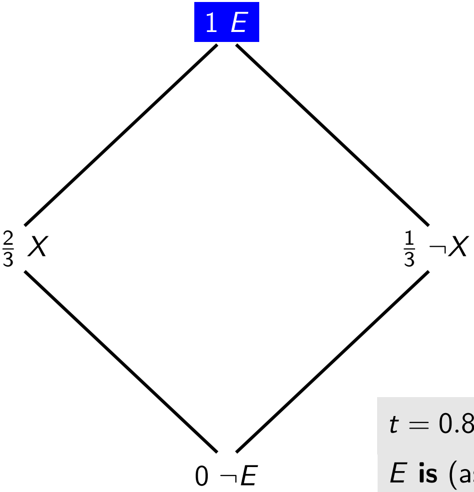


E **is** (assigned) true



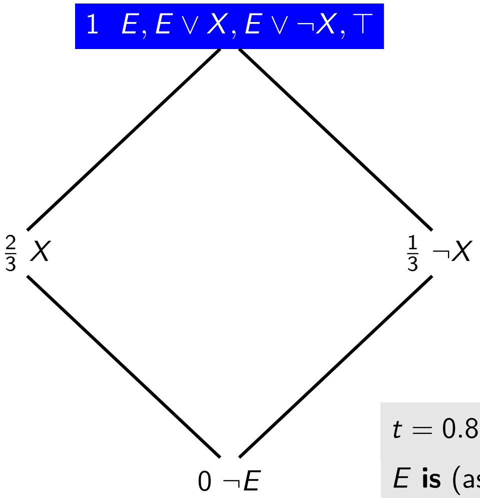
$t = 0.85 = \frac{0.17}{0.17+0.03}$

E **is** (assigned) true



$t = 0.85 = \frac{0.17}{0.17+0.03}$

E **is** (assigned) true



$t = 0.85 = \frac{0.17}{0.17+0.03}$

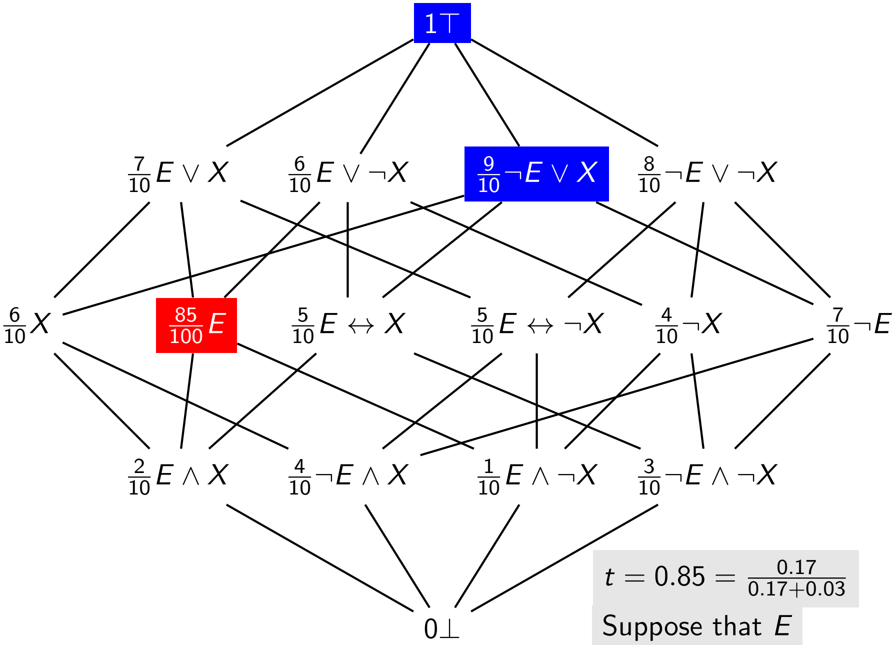
E **is** (assigned) true

Vacuity If E is consistent with B, then $B * E \supseteq Cn(B \cup \{E\})$

$B = Cn(\{\neg E \vee X\})$

$B \not\vdash \neg E$, $B * E = Cn(B \cup \{E\}) = Cn(\{E \wedge X\})$ So, $X \in B * E$.

$B \circledast E = Cn(\{E, E \vee X, E \vee \neg X\})$, so $X \not\in B \circledast E$.



$t = 0.85 = \frac{0.17}{0.17+0.03}$

Suppose that E

Non-Extremal EUT revision is more conservative than AGM revision (when the two approaches interestingly) diverge:

Theorem EUT violates Vacuity (wrt B, E) if and only if E is consistent with B and $B \divideontimes E \subset B * E$

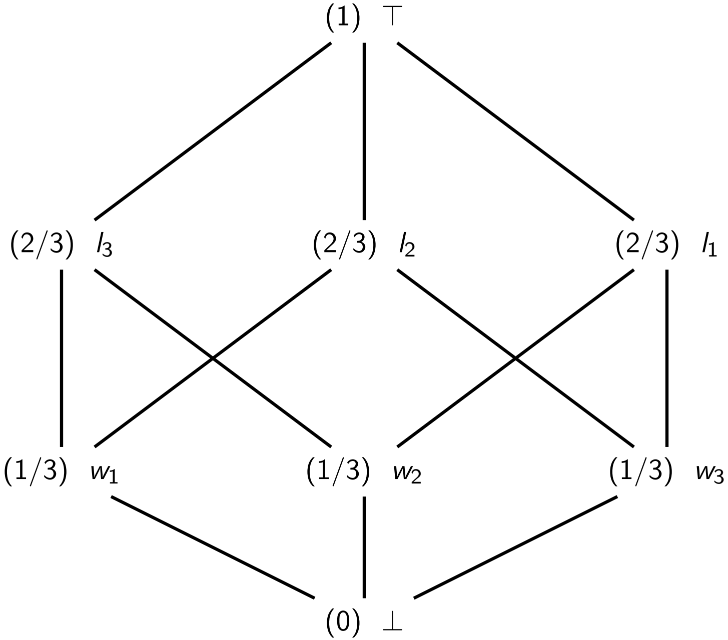
Non-Extremal EUT revision is more conservative than AGM revision (when the two approaches interestingly) diverge:

Theorem EUT violates Vacuity (wrt B, E) if and only if E is consistent with B and $B \circledast E \subset B * E$

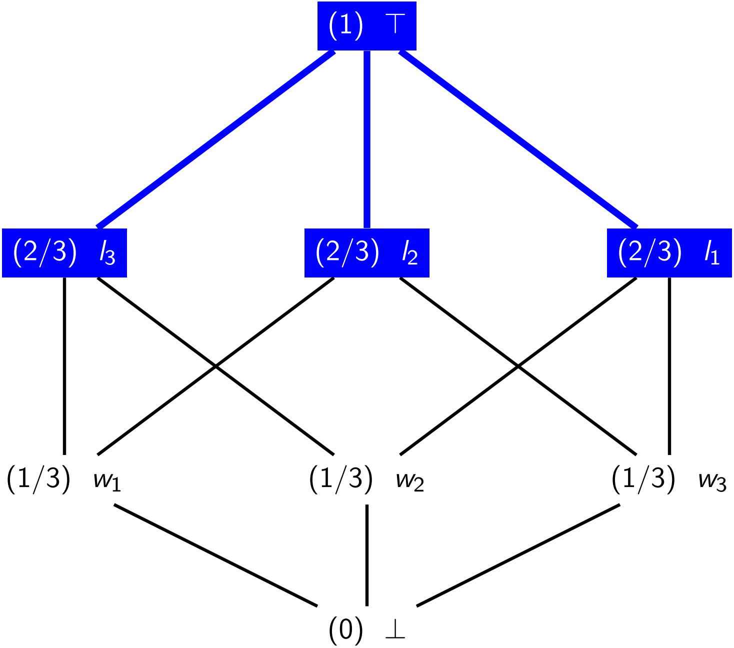
In other words, when EUT and AGM (interestingly) diverge, AGM will be more demanding on an agents beliefs (insofar as they are maintained via revision). Since AGM will require agents to maintain beliefs in the face of counter-evidence (such as in our counter-example to Vacuity), it may be seen as an epistemically risk-seeking policy for belief revision. On the other hand, EUT will recommend that agents suspend belief in many cases and so it may be seen as epistemically risk-averse.

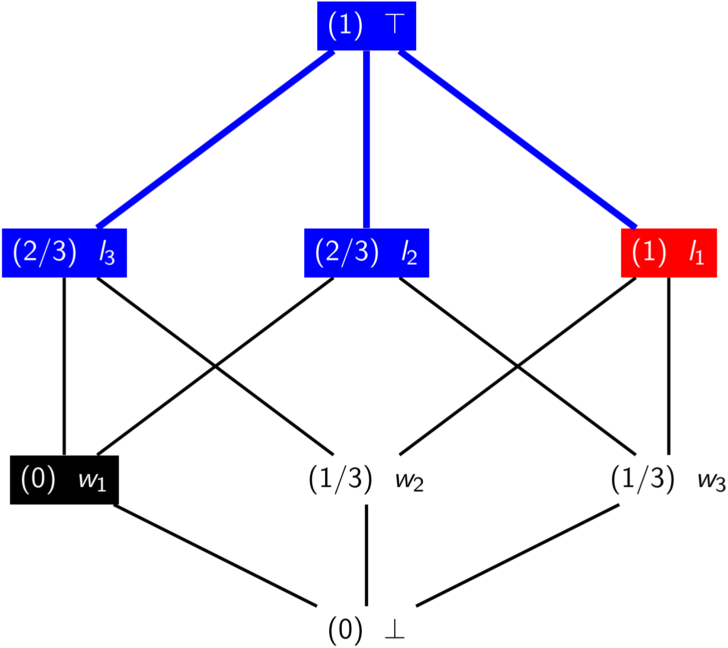
Beliefs that obey the Lockean thesis can be undermined by new evidence that is consistent with the agents current beliefs.

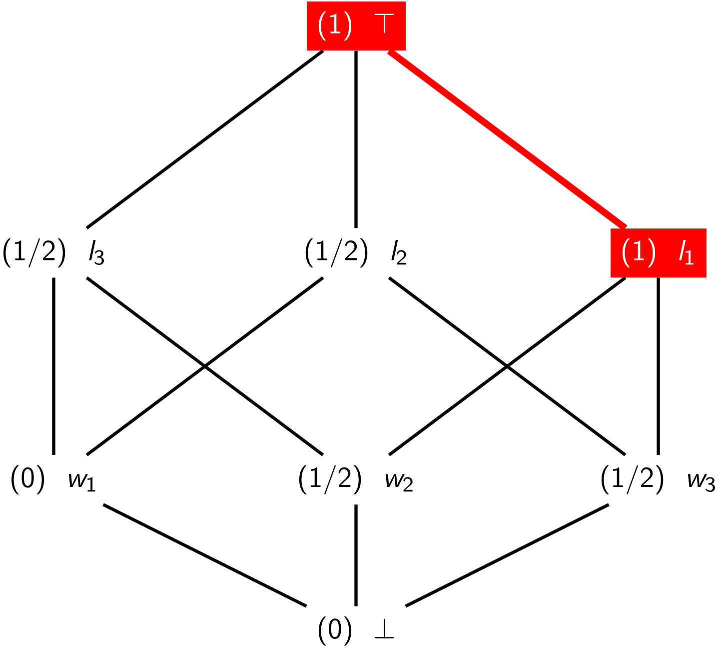
For each $i = 1, 2, 3$, let l_i be the proposition Ticket i won't win (and w_i is the proposition that "ticket i will win"). And let us set our threshold for Lockean belief at $r = 0.6$.

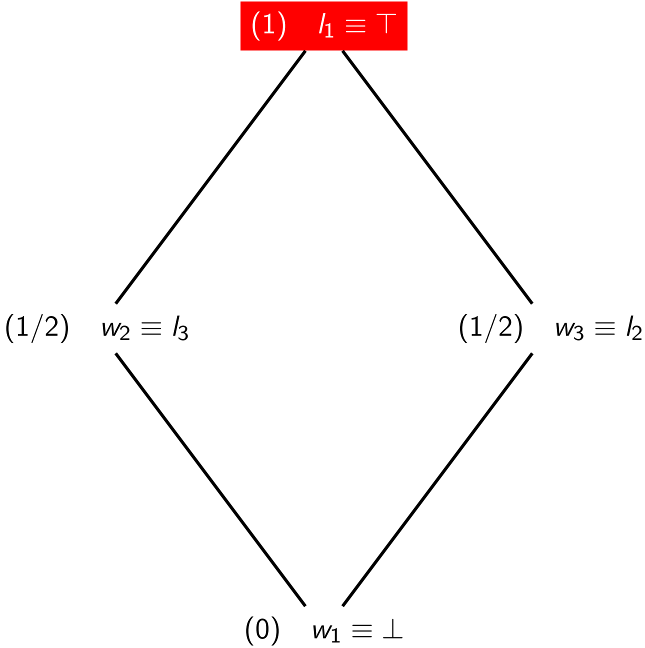


(1) $\top$
(2/3) l_3
(2/3) l_2
(2/3) l_1
(1/3) w_1
(1/3) w_2
(1/3) w_3
(0) $\bot$


(1) ⊤
(2/3) l_3
(2/3) l_2
(2/3) l_1
(1/3) w_1
(1/3) w_2
(1/3) w_3
(0) ⊥


(1) ⊤
(2/3) l_3
(2/3) l_2
(1) l_1
(0) w_1
(1/3) w_2
(1/3) w_3
(0) ⊥


(1) ⊤
(1/2) l_3
(1/2) l_2
(1) l_1
(0) w_1
(1/2) w_2
(1/2) w_3
(0) ⊥


(1) $l_1 \equiv \top$
(1/2) $w_2 \equiv l_3$
(1/2) $w_3 \equiv l_2$
(0) $w_1 \equiv \bot$

Resiliency, Robust Belief, Stable Belief

B. Skyrms. *Resiliency, propensities, and causal necessity*. Journal of Philosophy, 74:11, pgs. 704 - 713, 1977.

A. Baltag and S. Smets. *Probabilistic Belief Revision*. Synthese, 2008.

H. Leitgeb. *Reducing belief simpliciter to degrees of belief*. Annals of Pure and Applied Logic, 16:4, pgs. 1338 - 1380, 2013.

R. Stalnaker. *Belief revision in games: forward and backward induction*. Mathematical Social Sciences, 36, pgs. 31 - 56, 1998.

Probability

Let W be a set of states and $\mathfrak{A}$ a σ-algebra: $\mathfrak{A} \subseteq \wp(W)$ such that

- $W, \emptyset \in \mathfrak{A}$
- if $X \in \mathfrak{A}$ then $W - X \in \mathfrak{A}$
- if $X, Y \in \mathfrak{A}$ then $X \cup Y \in \mathfrak{A}$
- if $X_0, X_1, \ldots \in \mathfrak{A}$ then $\bigcup_{i \in \mathbb{N}} X_i \in \mathfrak{A}$.

Probability

$P : \mathfrak{A} \to [0,1]$ satisfying the usual constraints

- $P(W) = 1$
- (finite additivity) If $X_1, X_2 \in \mathfrak{A}$ are pairwise disjoint, then $P(X_1 \cup X_2) = P(X_1) + P(X_2)$

$P(Y|X) = \frac{P(Y \cap X)}{P(X)}$ whenever $P(X) > 0$. So, $P(Y|W)$ is $P(Y)$.

- P is countably additive (σ-additive): if $X_1, X_2, \ldots, X_n, \ldots$ are pairwise disjoint members of $\mathfrak{A}$, then $P(\bigcup_{n \in \mathbb{N}} X_n) = \sum_{n \in \mathbb{N}} P(X_n)$

P-stability^r

Definition. Let P be a probability measure on $\mathfrak{A}$ over W, let $0 \leq t < 1$. For all $X \in \mathfrak{A}$:

X is P-stable^t if and only if for all $Y \in \mathfrak{A}$ with $Y \cap X \neq \emptyset$ and $P(Y) > 0$: $P(X|Y) > t$.

P-stability^r

Definition. Let P be a probability measure on $\mathfrak{A}$ over W, let $0 \leq t < 1$. For all $X \in \mathfrak{A}$:

$$X \text{ is } P\text{-stable}^t \text{ if and only if for all } Y \in \mathfrak{A} \text{ with } Y \cap X \neq \emptyset \text{ and } P(Y) > 0\text{: } P(X|Y) > t.$$

- Trivially, the empty set of P-stable^t.

P-stability^r

Definition. Let P be a probability measure on $\mathfrak{A}$ over W, let $0 \leq t < 1$. For all $X \in \mathfrak{A}$:

$$X \text{ is } P\text{-stable}^t \text{ if and only if for all } Y \in \mathfrak{A} \text{ with } Y \cap X \neq \emptyset \text{ and } P(Y) > 0\text{: } P(X|Y) > t.$$

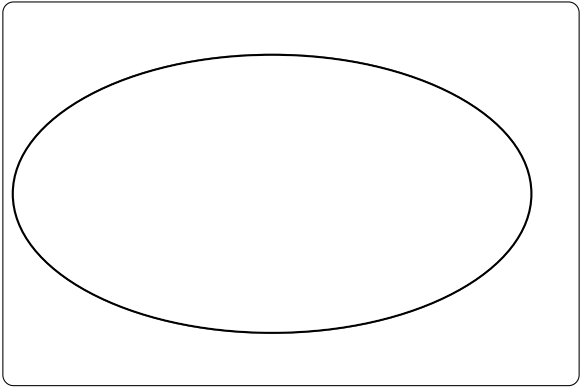
- Trivially, the empty set of P-stable^t.
- If $P(X) = 1$, then X is P-stable^t.

P-stability^r

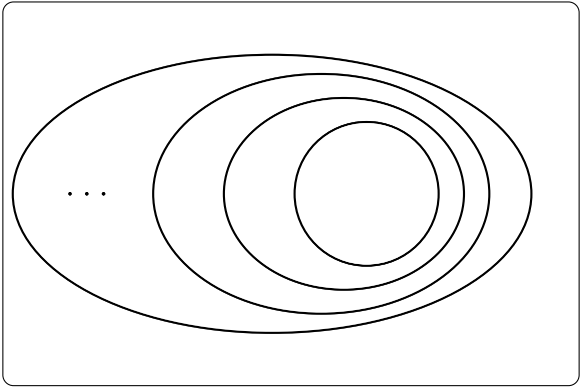
Definition. Let P be a probability measure on $\mathfrak{A}$ over W, let $0 \leq t < 1$. For all $X \in \mathfrak{A}$:

$$X \text{ is } P\text{-stable}^t \text{ if and only if for all } Y \in \mathfrak{A} \text{ with } Y \cap X \neq \emptyset \text{ and } P(Y) > 0\text{: } P(X|Y) > t.$$

- Trivially, the empty set of P-stable^t.
- If $P(X) = 1$, then X is P-stable^t.
- There are P-stable^t sets with $0 < P(X) < 1$.



- Assuming countable additivity and $t \geq \frac{1}{2}$, The class of P-stable^t propositions X in $\mathfrak{A}$ with $P(X) < 1$ is well-ordered with respect to the subset relation.
- If there is a non-empty P-stable^r $X \in \mathfrak{A}$ with $P(X) < 1$, then there is also a least such X.



- Assuming countable additivity and $t \geq \frac{1}{2}$, The class of P-stable^t propositions X in $\mathfrak{A}$ with $P(X) < 1$ is well-ordered with respect to the subset relation.
- If there is a non-empty P-stable^r $X \in \mathfrak{A}$ with $P(X) < 1$, then there is also a least such X.

$w \in SB(H)$ iff for all $E \in \mathfrak{A}(W)$ with $H \cap E \neq \emptyset$ and $P(E) \neq 0$: $P(H \mid E) \geq t$

$w \in SB(H)$ iff for all $E \in \mathfrak{A}(W)$ with $H \cap E \neq \emptyset$ and $P(E) \neq 0$: $P(H \mid E) \geq t_C$

1. The threshold t is determined contextually (the "cautiousness level")

$w \in SB(H)$ iff for all $E \in \mathfrak{A}_H(W)$ with $H \cap E \neq \emptyset$ and $P(E) \neq 0$: $P(H \mid E) \geq t_C$

1. The threshold t is determined contextually (the "cautiousness level")
2. The evidence "relevant" to H

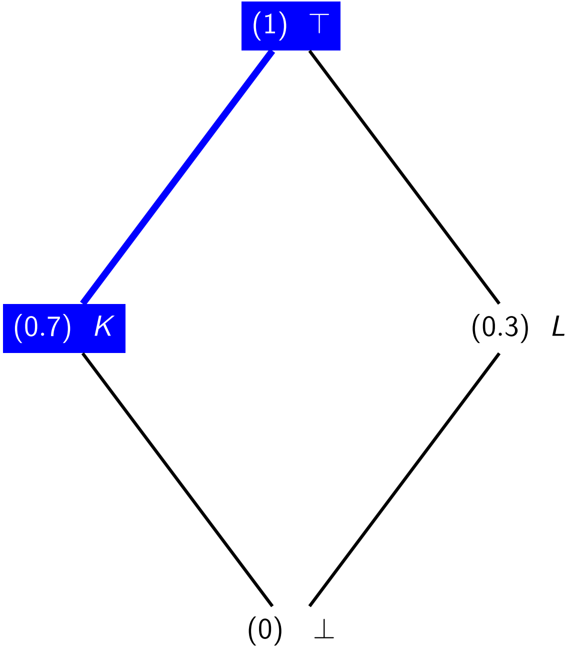
$w \in SB(H)$ iff for all $E \in \mathfrak{A}_H(W_\Pi)$ with $H \cap E \neq \emptyset$ and $P(E) \neq 0$: $P(H \mid E) \geq t_C$

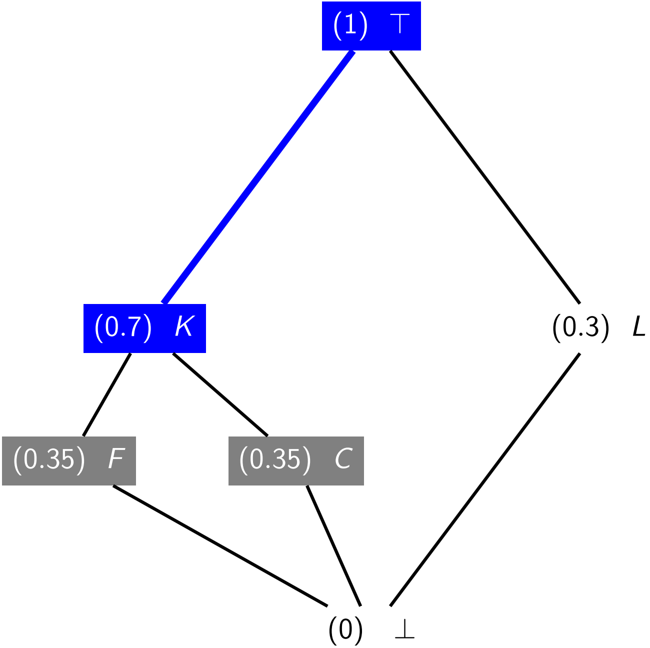
1. The threshold t is determined contextually (the "cautiousness level")
2. The evidence "relevant" to H
3. The states may be contextually determined (by a partition Π on a set W of "maximally specific worlds")

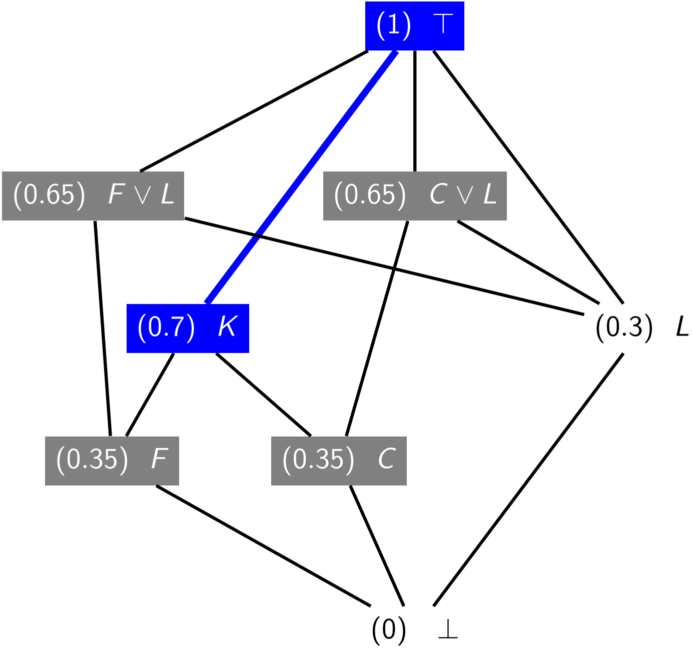
H. Leitgeb. *The Stability Theory of Belief*. The Philosophical Review 123/2, 131171, 2014.

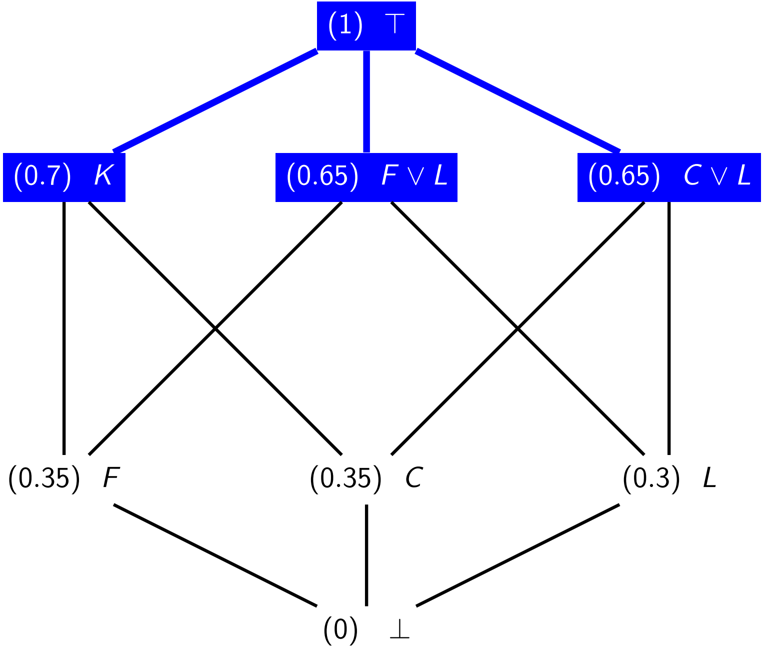
H. Leitgeb. *The Humean Thesis on Belief*. Proceedings of the Aristotelian Society of Philosophy 89(1), 143185, 2015.

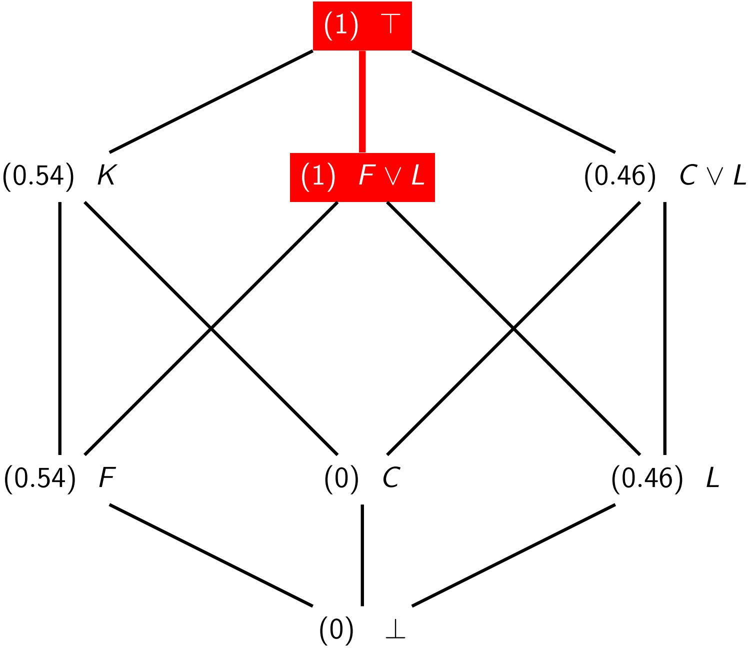
R. Pettigrew. *Pluralism about belief states*. Proceedings of the Aristotelian Society 89(1):187-204, 2015.


(1) ⊤
(0.7) K
(0.3) L
(0) ⊥


(1) ⊤
(0.7) K
(0.3) L
(0.35) F
(0.35) C
(0) ⊥


(1) ⊤
(0.65) $F \vee L$
(0.65) $C \vee L$
(0.7) K
(0.3) L
(0.35) F
(0.35) C
(0) ⊥


(1) ⊤
(0.7) K
(0.65) F ∨ L
(0.65) C ∨ L
(0.35) F
(0.35) C
(0.3) L
(0) ⊥


(1) ⊤
(0.54) K
(1) F ∨ L
(0.46) C ∨ L
(0.54) F
(0) C
(0.46) L
(0) ⊥

Thus, while **robust belief** is stable under acquisition of new (doxastically possible) evidence and Lockean belief is not, **robust belief** is not stable under fine-graining of possibilities while Lockean belief is.

Leitgeb's Solution to the Lottery Paradox

In a context in which the agent is interested in *whether ticket i will be drawn*; for example, for $i = 1$: Let Π be the corresponding partition:

$$\{\{w_1\}, \{w_2, \ldots, w_{1,000,000}\}\}$$

The resulting probability measure P_Π is given so that P is given by P so that:

$$P_\Pi(\{\{w_1\}\}) = \frac{1}{1,000,000} \qquad P_\Pi(\{\{w_2, \ldots, w_{1,000,000}\}\}) = \frac{999,999}{1,000,000}$$

There are two P_Π-stable sets, and one of the two possible choices for the strongest believed proposition $B_W^\Pi = \{\{w_2, \ldots, w_{1,000,000}\}\}$.

If B_W^Π is chosen as such, our perfectly rational agent believes of ticket $i = 1$ that it will not be drawn, (and of course P1 -P3 are satisfied).

For example, this might be a context in which a single ticket holder—the person holding ticket 1—would be inclined to say of his or her ticket: "I believe it won't win."

In a context in which the agent is interested in *which ticket will be drawn*: Let Π' be the corresponding partition that consists of all singleton subsets of W. The probability measure P^{Π} is the uniform probability on W.

The only P-stable set—and hence the only choice for the strongest believed proposition $B_W^{\Pi'}$—is W itself: our perfectly rational agent believes that some ticket will be drawn, but he or she does not believe of any ticket that it will not win

For example, this might be a context in which a salesperson of tickets in a lottery would be inclined to say of each ticket: "It might win" (that is, it is not the case that I believe that it won't win).

In either of the two contexts from before, the theory avoids the absurd conclusion of the Lottery Paradox; in each context, it preserves the closure of belief under conjunction; and in each context, it preserves the Lockean thesis for some threshold ($r = \frac{999{,}999}{1{,}000{,}000}$ in the first case, $r = 1$ in the second case)-all of this follows from P-stability and the theorem.

In the first Π-context, the intuition is preserved that, in some sense, one believes of ticket i that it will lose since it is so likely to lose.

In the second Π'-context, the intuition is preserved that, in a different sense, one should not believe of any ticket that it will lose since the situation is symmetric with respect to tickets, as expressed by the uniform probability measure, and of course some ticket must win.

Finally, by disregarding or mixing the contexts, it becomes apparent why one might have regarded all of the premises of the Lottery Paradox as true.

But according to the present theory, contexts should not be disregarded or mixed: partitions Π and Π' differ from each other, and different partitions may lead to different beliefs, as observed in the last section and as exemplified in the Lottery Paradox.

Accordingly, the thresholds in the Lockean thesis may have to be chosen differently in different contexts, and once again, this is what happens in the Lottery Paradox—which makes good sense: in the second Π'-context, by uniformity, the agent's degrees of belief do not give him or her much of a hint of what to believe. That is why the agent ought to be supercautious about her beliefs in that