Logical and Probabilistic Models of Belief Change

Eric Pacuit

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

July 14, 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, Iterated belief revision, Context shifts, Becoming aware

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

Day 5 Interactive epistemology (Agreement Theorems, Belief Revision in Games)

pacuit.org/nasslli2016/belrev/

Recap

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

Recap

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

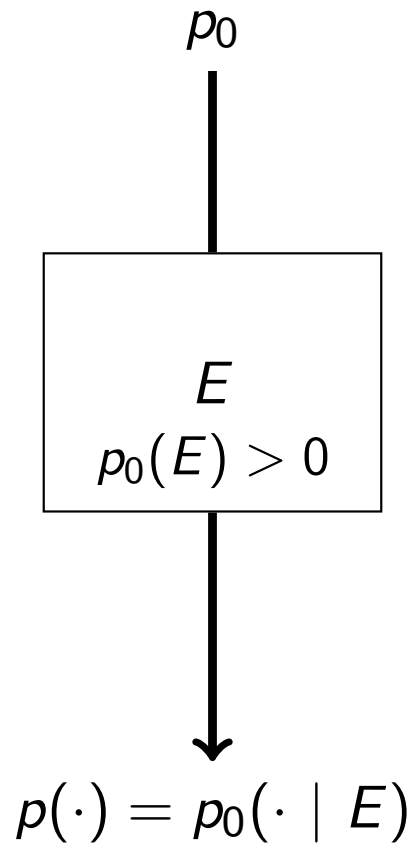
- "Finding out that φ"
  - Learn that φ
  - Suppose that φ
  - Accept φ
  - ...

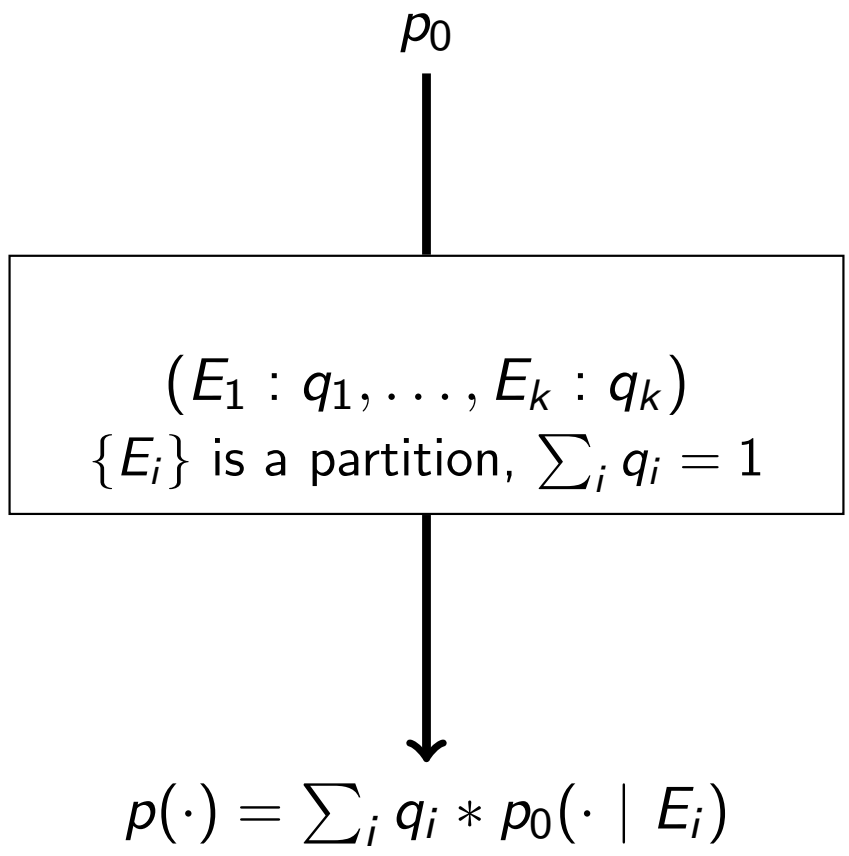
Recap

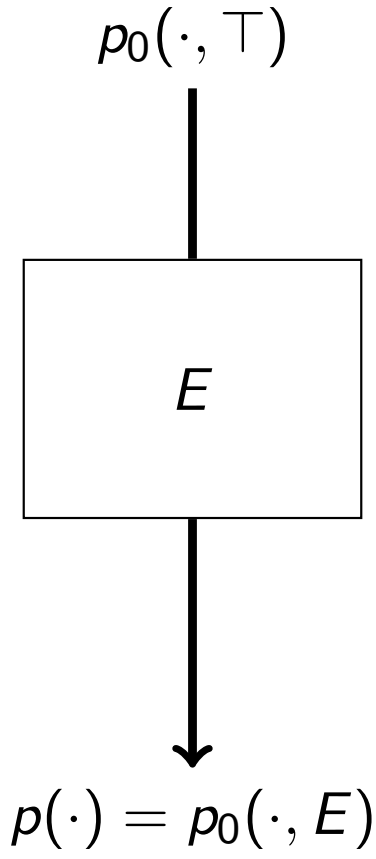
- Epistemic states: AGM, Plausibility Models, Bayesian Model (and the many variations)

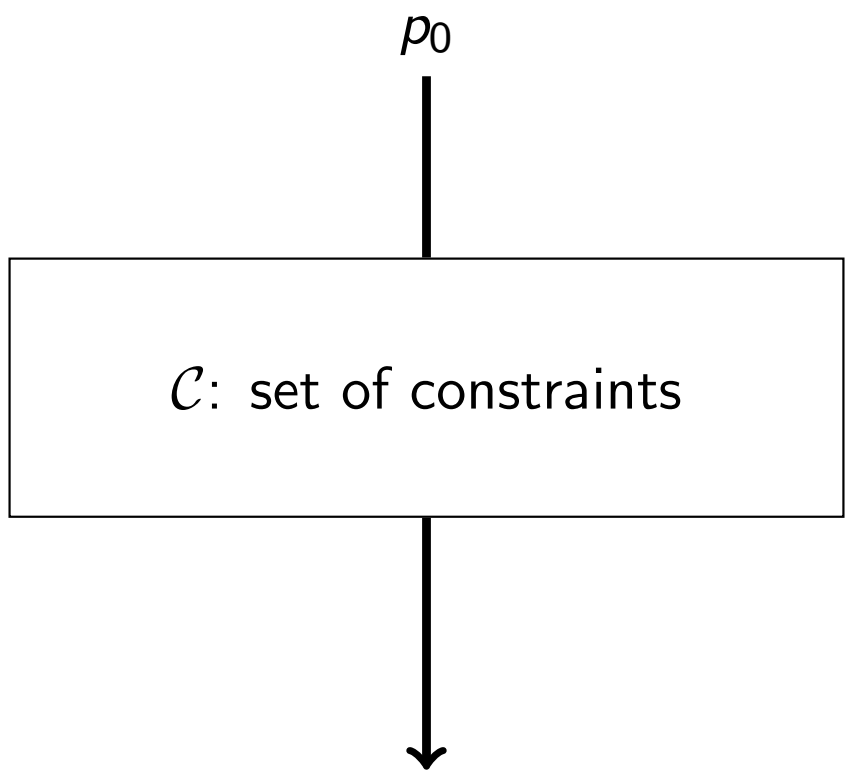
- "Finding out that φ"
  - Learn that φ
  - Suppose that φ
  - Accept φ
  - ...

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

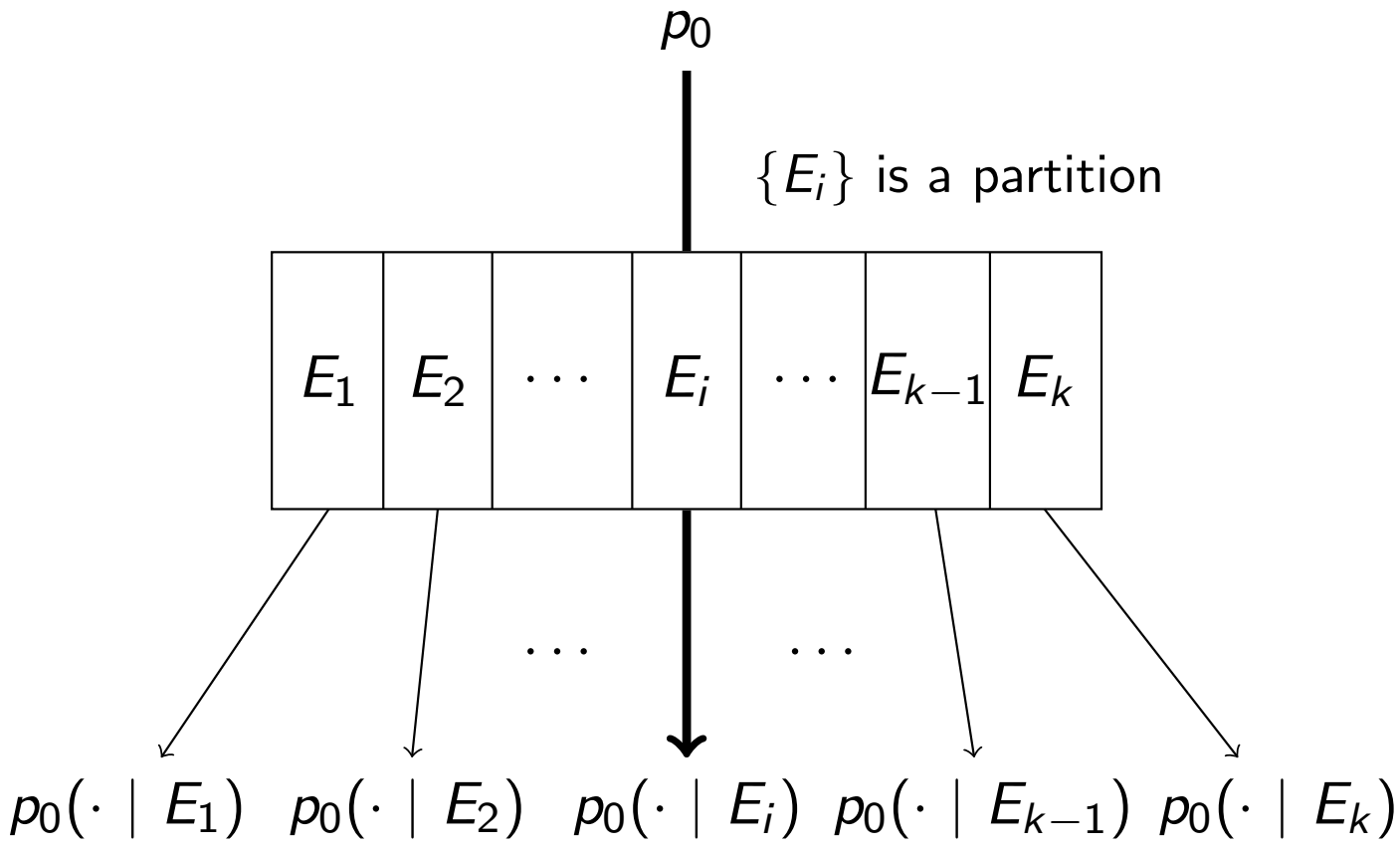

p_0
$(E_1 : q_1, \ldots, E_k : q_k)$
$\{E_i\}$ is a partition, $\sum_i q_i = 1$
$p(\cdot) = \sum_i q_i * p_0(\cdot \mid E_i)$

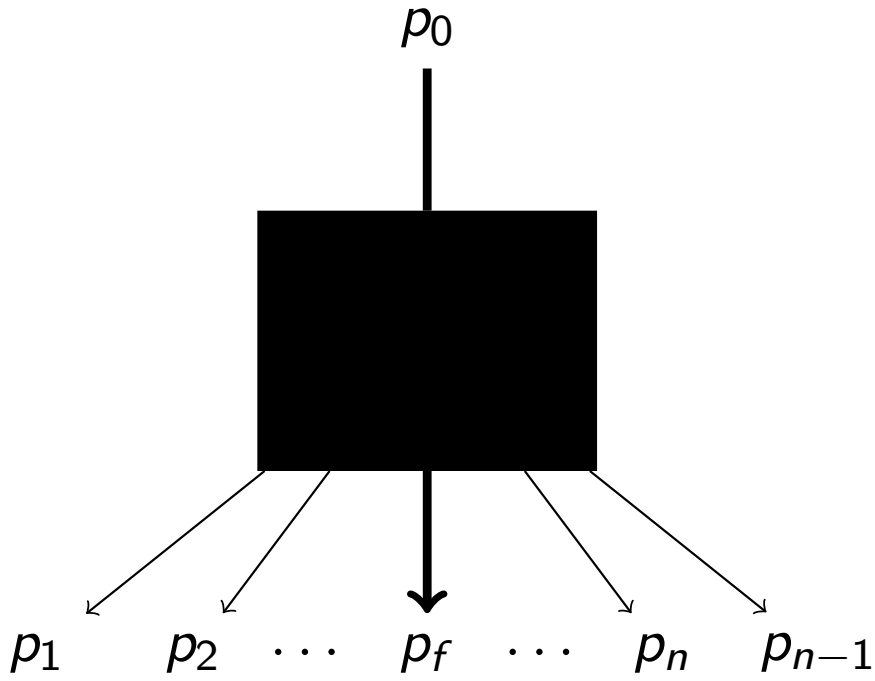

$p_0(\cdot, \top)$
E
$p(\cdot) = p_0(\cdot, E)$



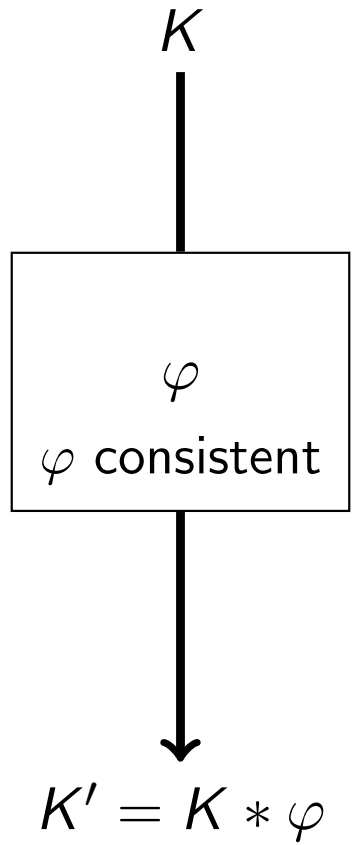
p satisfies $\mathcal{C}$

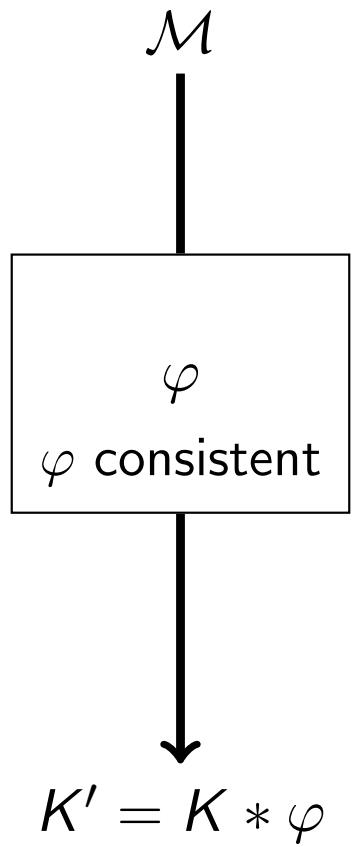
p minimizes $I(p, p_0) = \sum_i p_1(x_i) \log \frac{p_1(x_i)}{p_0(x_i)}$

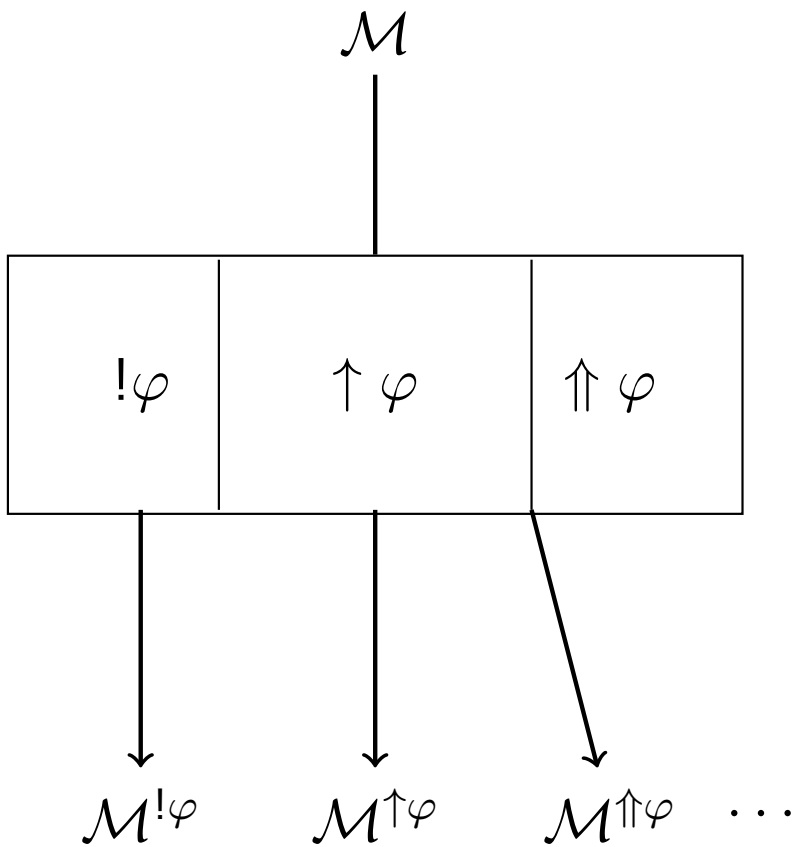

p_0
$\{E_i\}$ is a partition
E_1
E_2
$\cdots$
E_i
$\cdots$
E_{k-1}
E_k
$\cdots$
$\cdots$
$p_0(\cdot \mid E_1)$
$p_0(\cdot \mid E_2)$
$p_0(\cdot \mid E_i)$
$p_0(\cdot \mid E_{k-1})$
$p_0(\cdot \mid E_k)$

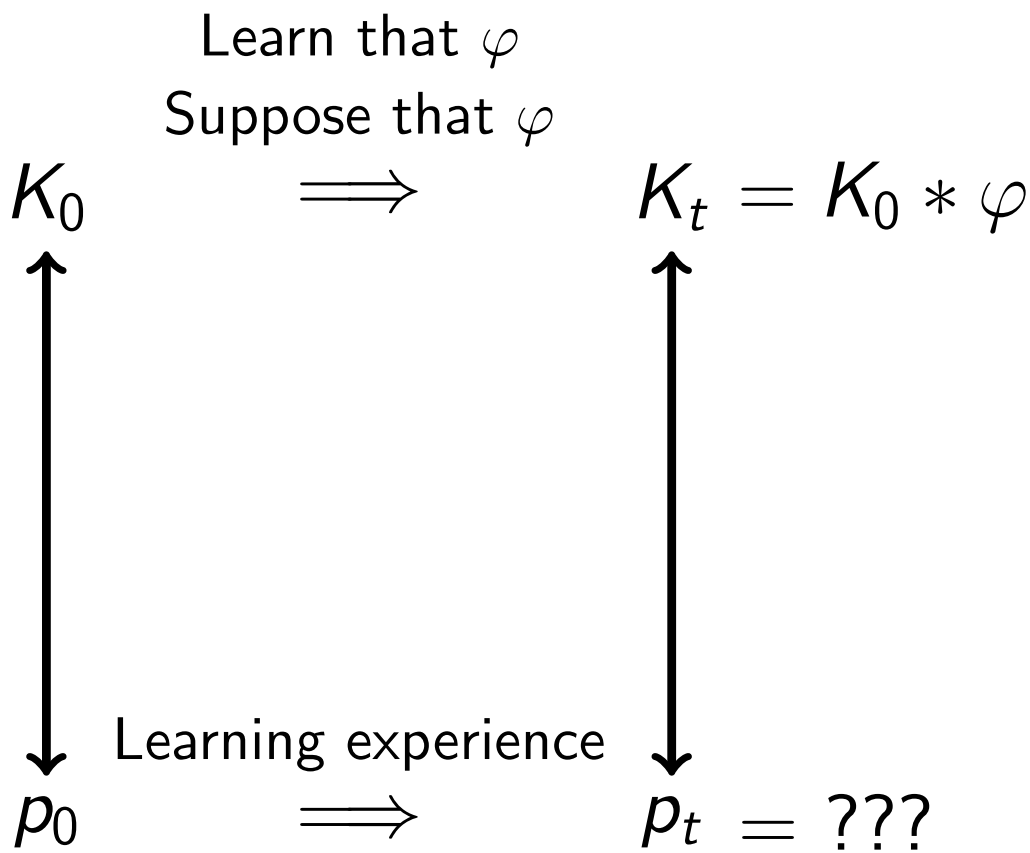


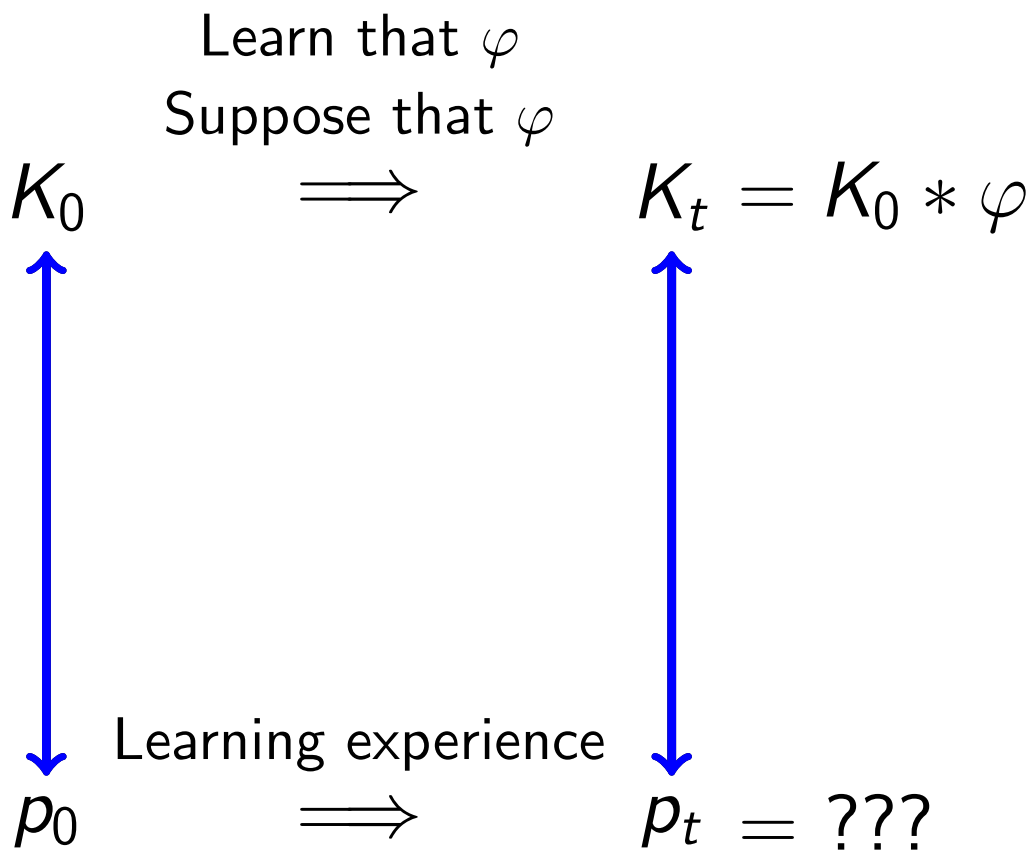
(Martingale Property) $\quad p_0(A \mid p_f) = p_f(A)$


K
φ
φ consistent
$K' = K * \varphi$


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


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


Learn that φ
Suppose that φ
K_0
$\Longrightarrow$
$K_t = K_0 * \varphi$
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."

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).

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

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

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

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

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

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 $X \in B$

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

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

(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 $\quad B * E = Cn(B * E)$

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.	$B \nvdash \neg X$.	B is consistent

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

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.	$B \nvdash \neg X$.	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.

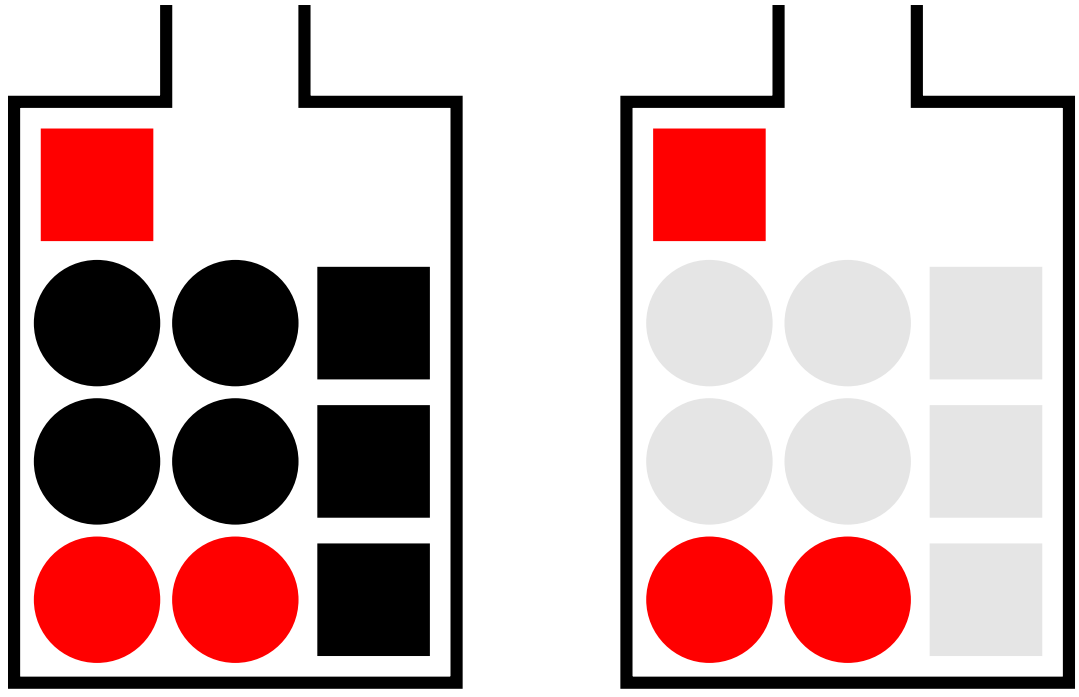
1.	$X \in B$.	Assumption
2.	$B \nvdash \neg X$.	B is consistent
3.	$B * X = Cn(B \cup \{X\})$.	(2), Vacuity, Inclusion
4.	$Cn(B \cup \{X\}) = Cn(B)$.	(1)
5.	$B * X = Cn(B) = B$	Closure

Theorem. (Gärdenfors) Suppose r $= 0$, 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.

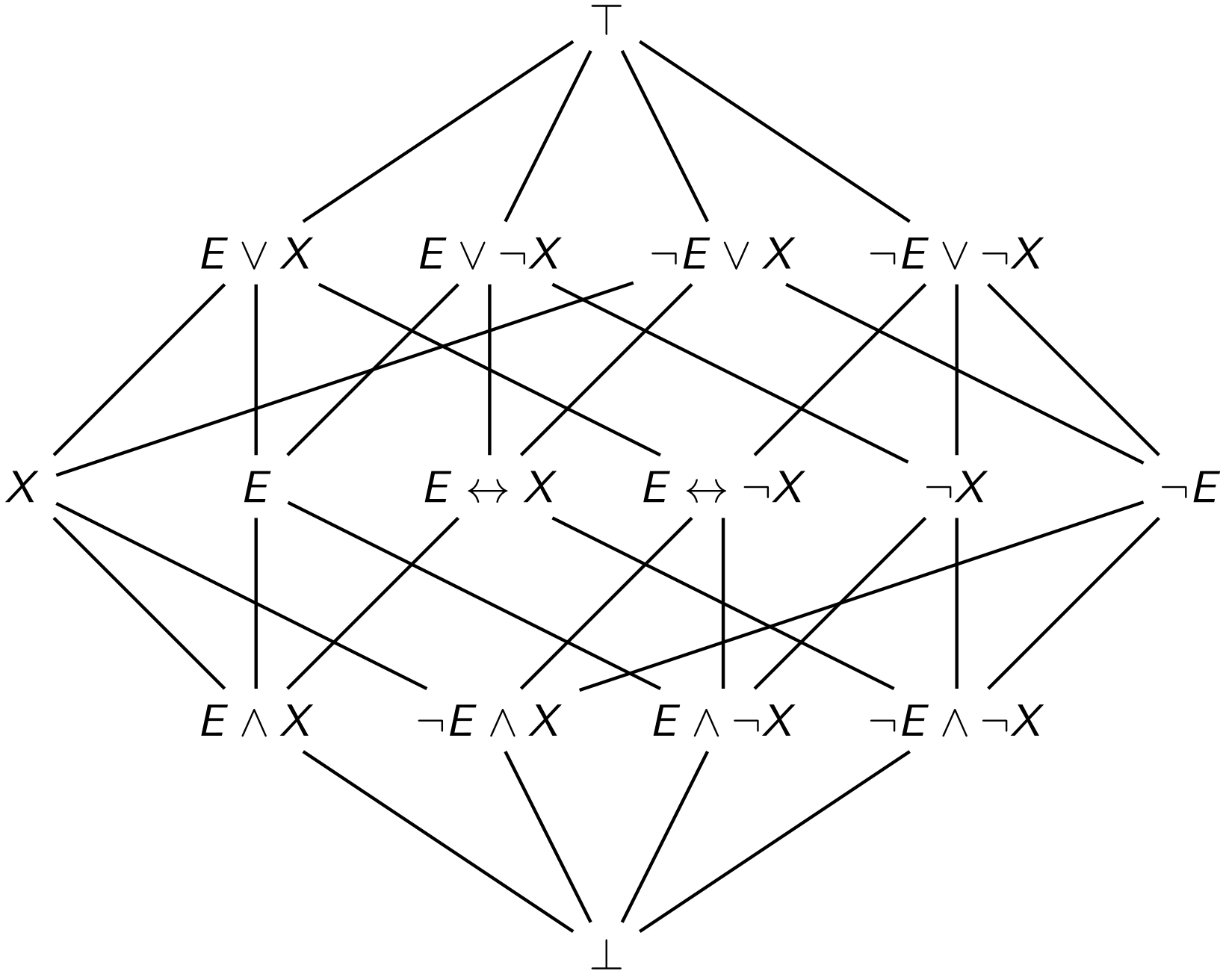


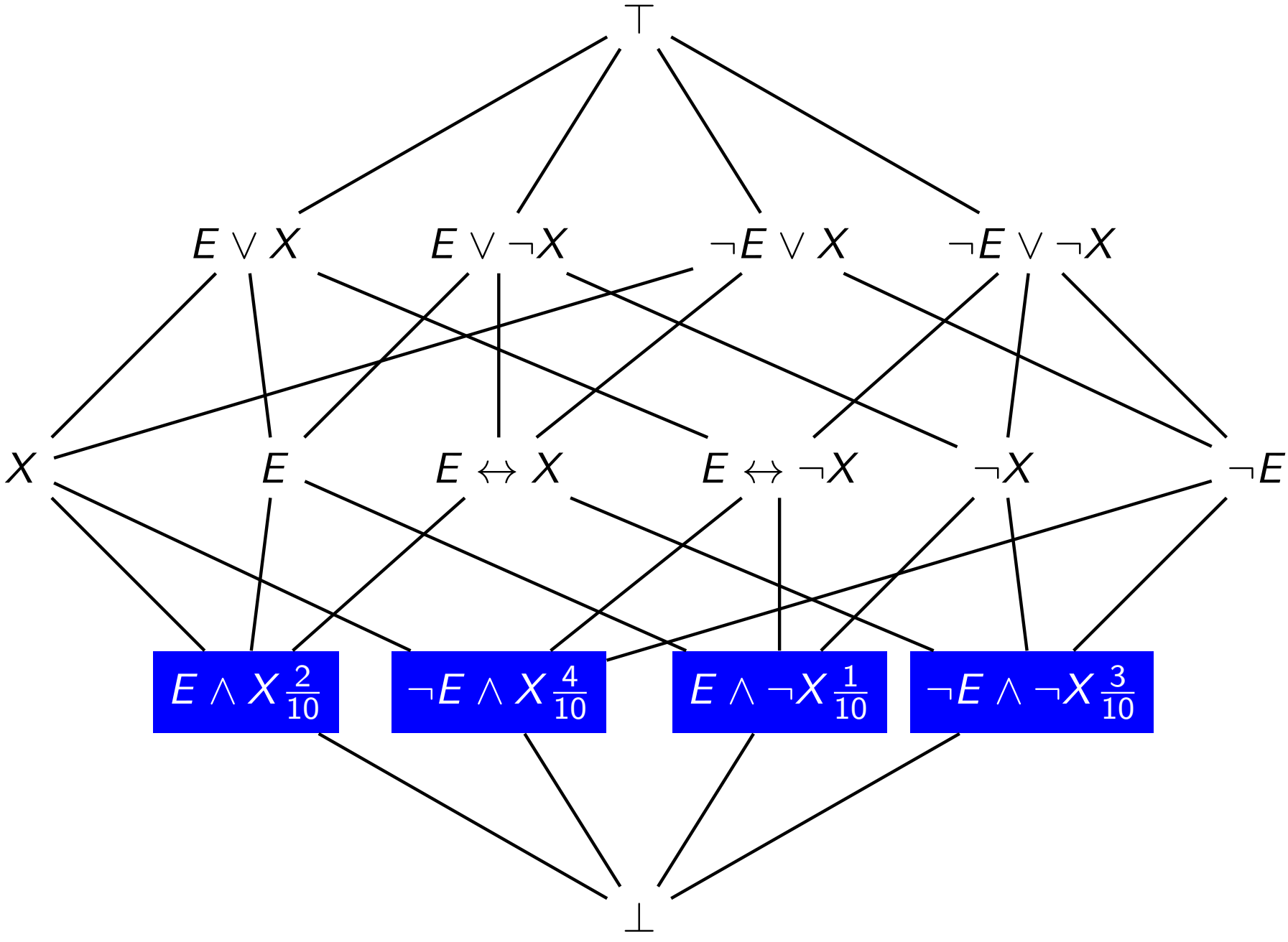
(A) Prior distribution (B) Posterior (given E)

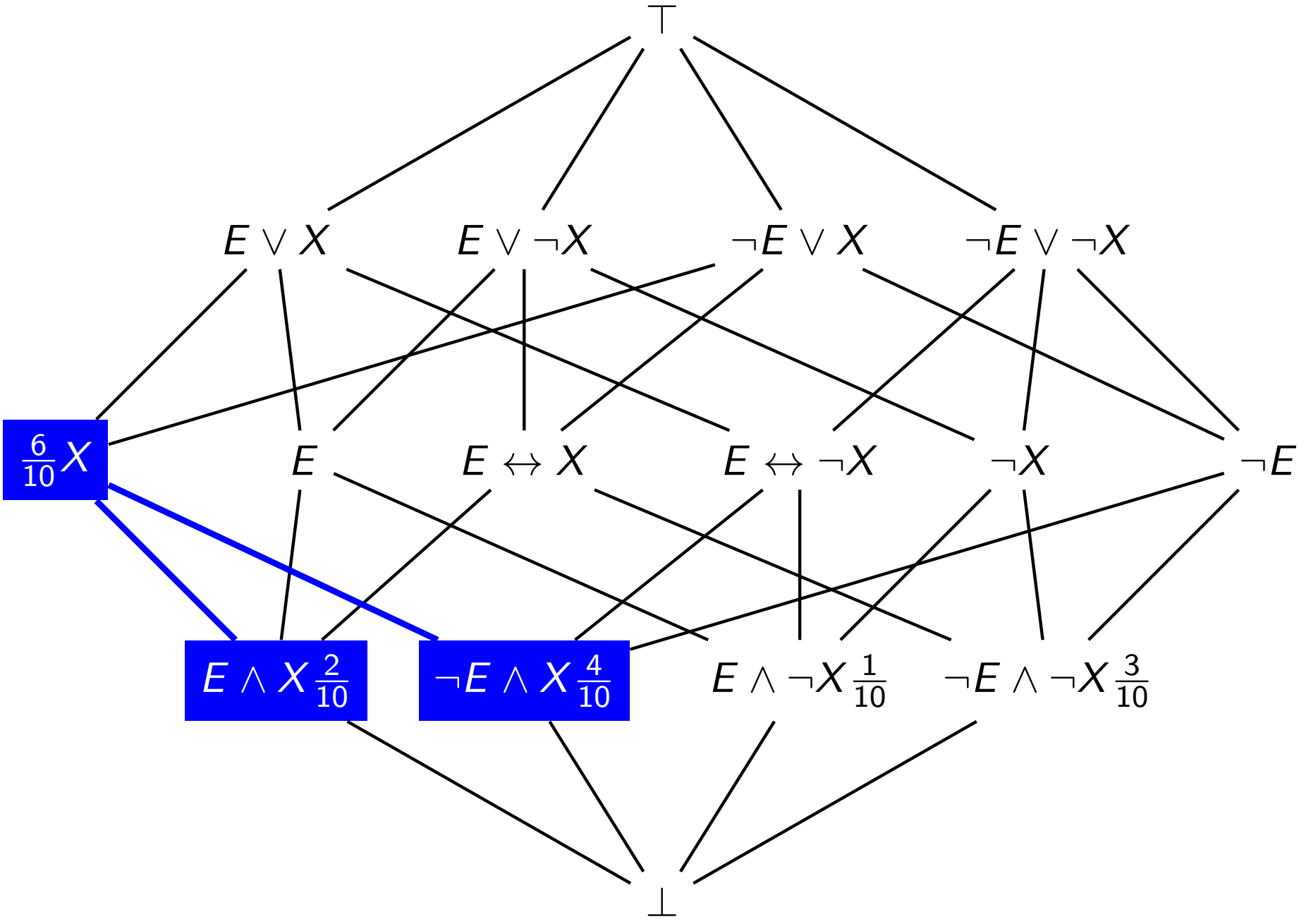
FIGURE 2. Visualization of counterexample to Vacuity for EUT Revision

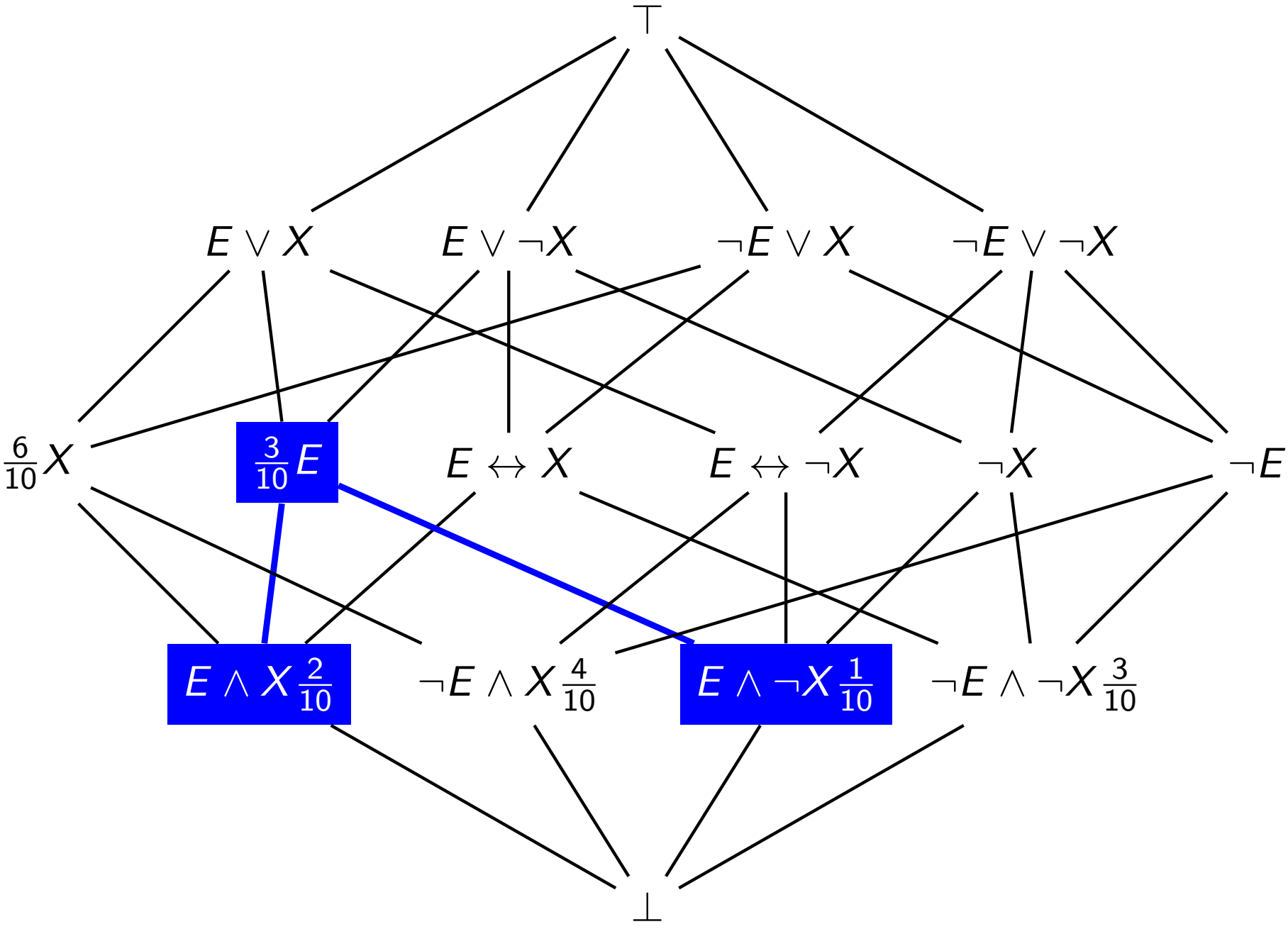
E :='The object sampled from the urn is red'

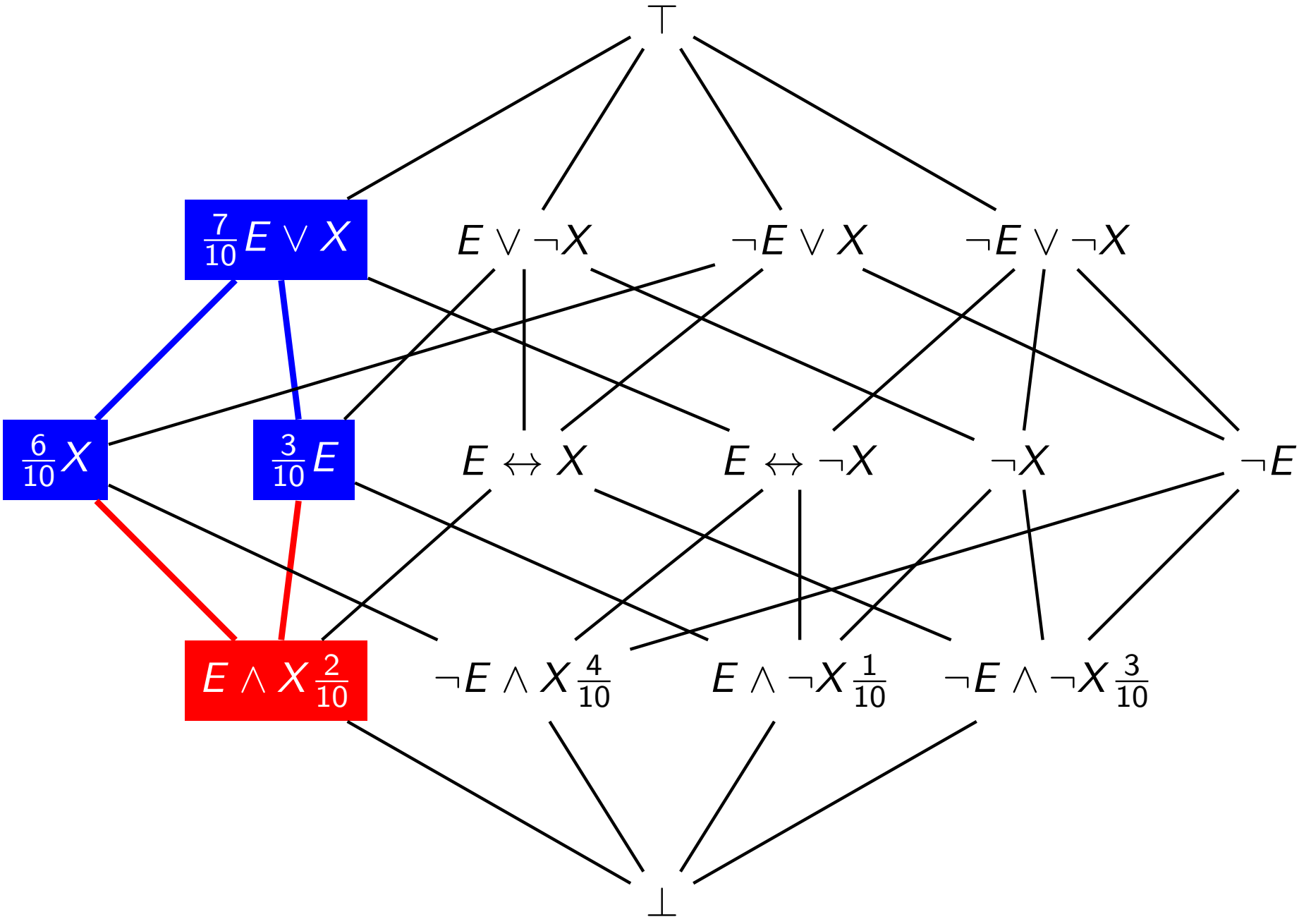
X := 'The object sampled from the urn is a circle'.

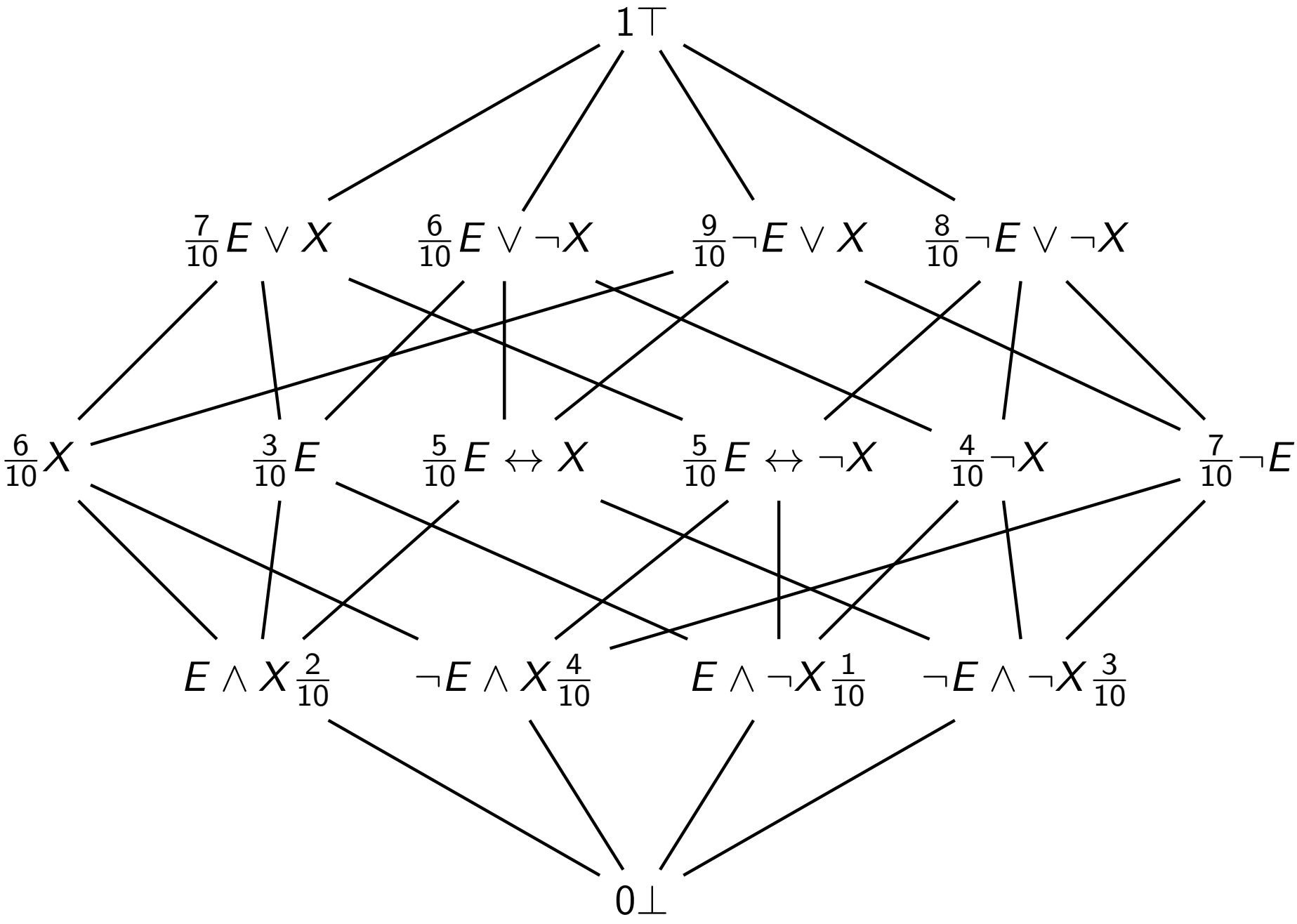

$\top$
$E \vee X$
$E \vee \neg X$
$\neg E \vee X$
$\neg E \vee \neg X$
X
E
$E \leftrightarrow X$
$E \leftrightarrow \neg X$
$\neg X$
$\neg E$
$E \wedge X$
$\neg E \wedge X$
$E \wedge \neg X$
$\neg E \wedge \neg X$
$\bot$

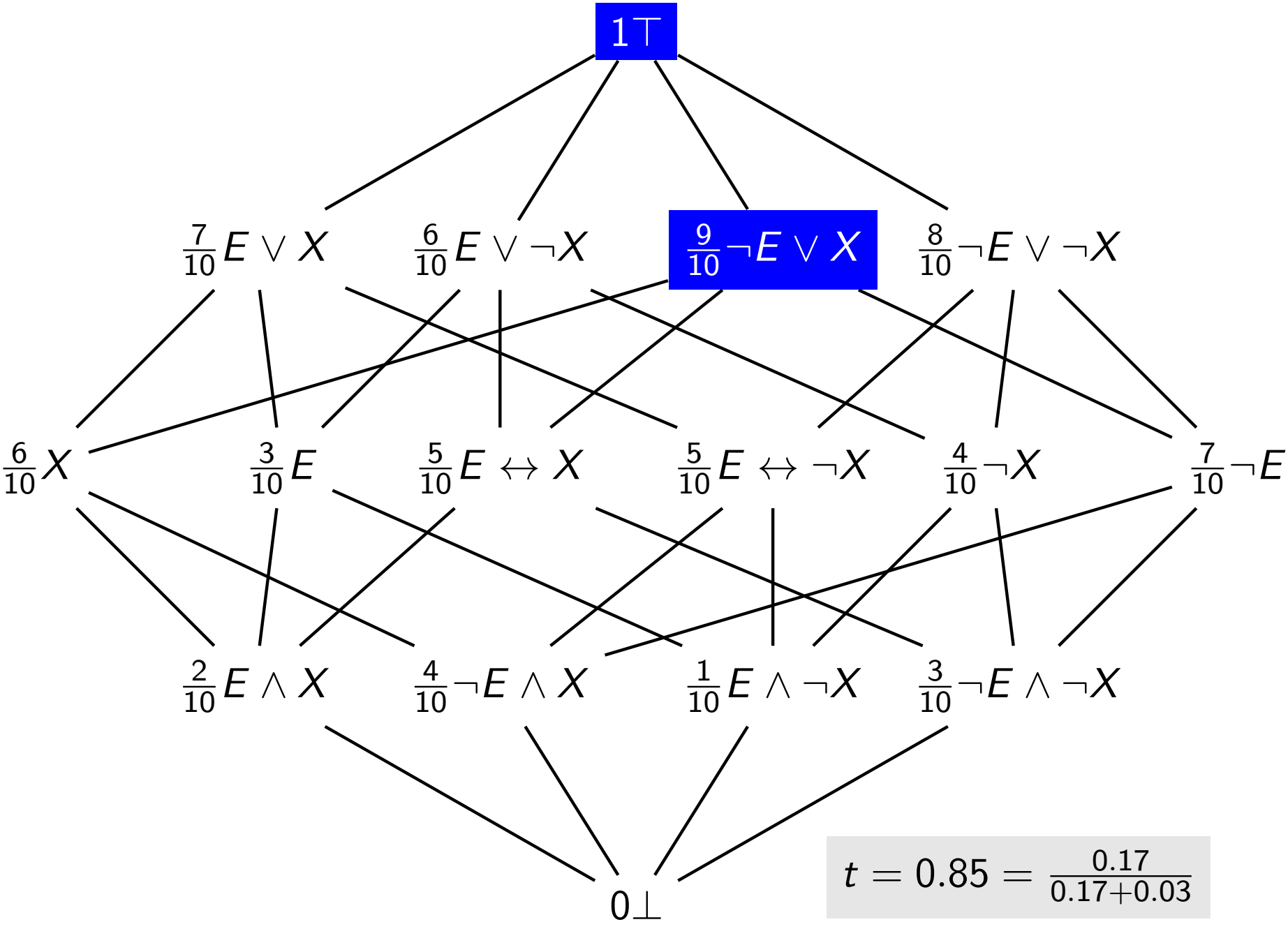

$\top$
$E \vee X$
$E \vee \neg X$
$\neg E \vee X$
$\neg E \vee \neg X$
X
E
$E \leftrightarrow X$
$E \leftrightarrow \neg X$
$\neg X$
$\neg E$
$E \wedge X \frac{2}{10}$
$\neg E \wedge X \frac{4}{10}$
$E \wedge \neg X \frac{1}{10}$
$\neg E \wedge \neg X \frac{3}{10}$
$\bot$


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

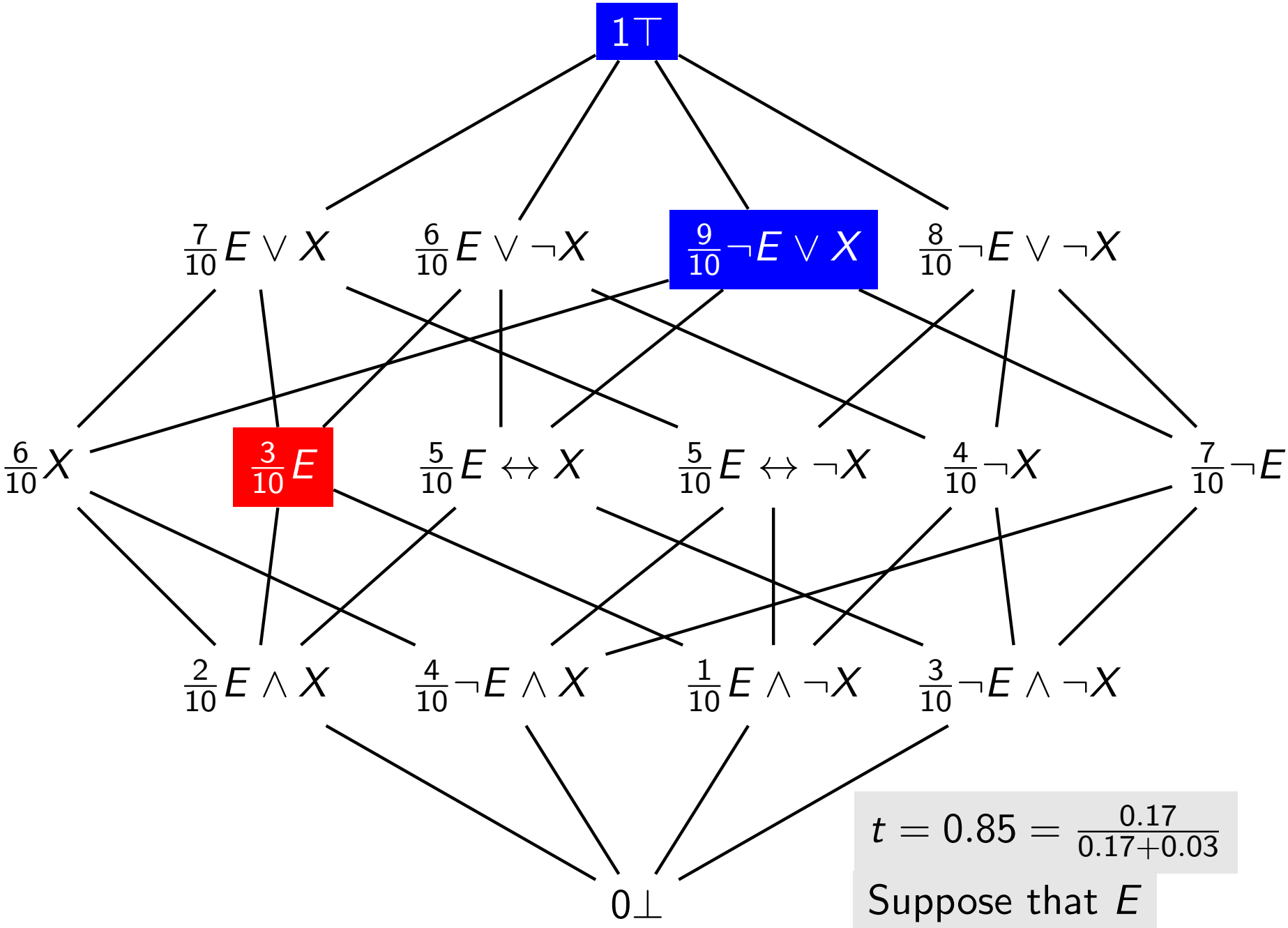

$\top$
$E \vee X$
$E \vee \neg X$
$\neg E \vee X$
$\neg E \vee \neg X$
$\frac{6}{10} X$
$\frac{3}{10} E$
$E \leftrightarrow X$
$E \leftrightarrow \neg X$
$\neg X$
$\neg E$
$E \wedge X \frac{2}{10}$
$\neg E \wedge X \frac{4}{10}$
$E \wedge \neg X \frac{1}{10}$
$\neg E \wedge \neg X \frac{3}{10}$
$\bot$


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


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

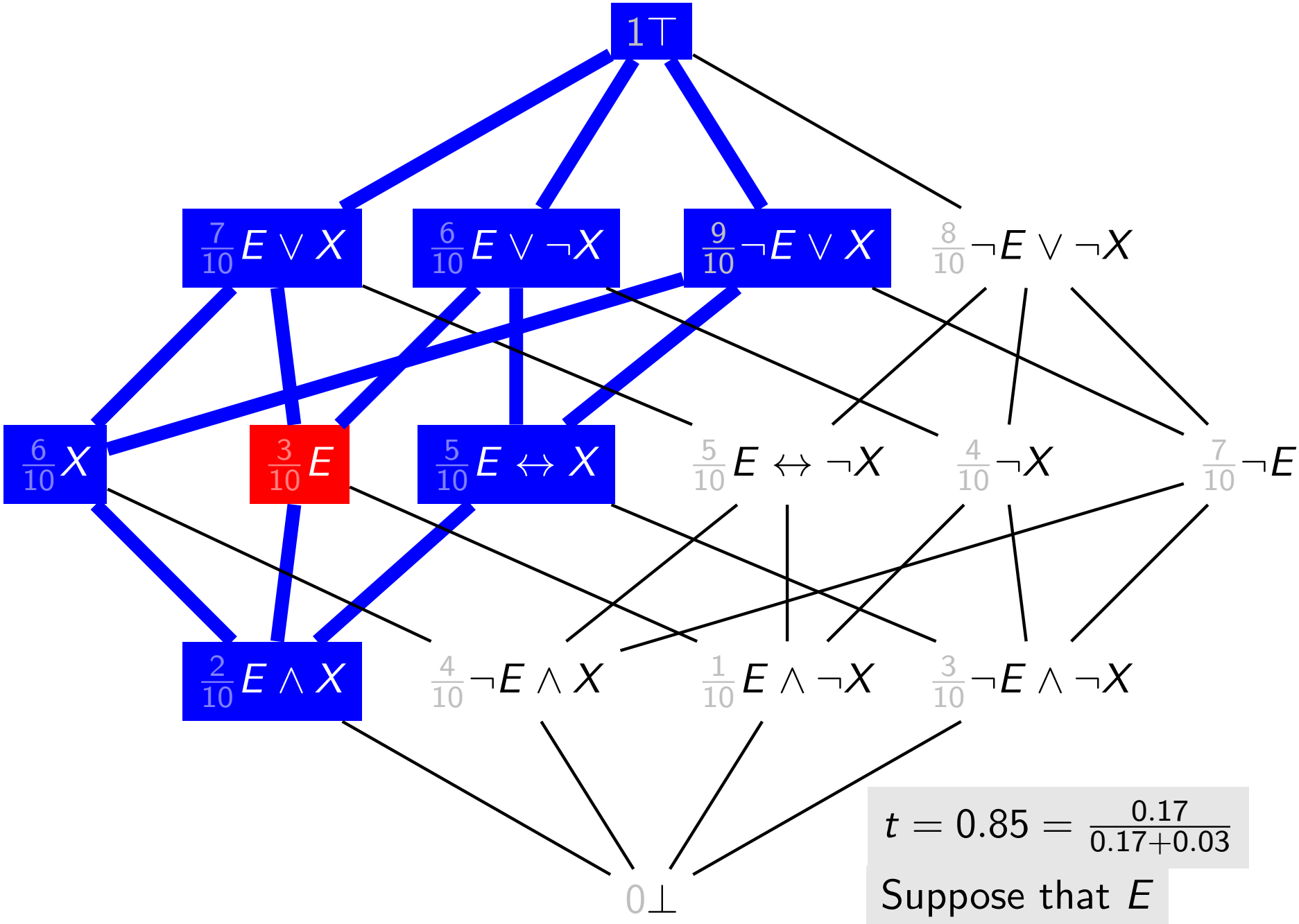


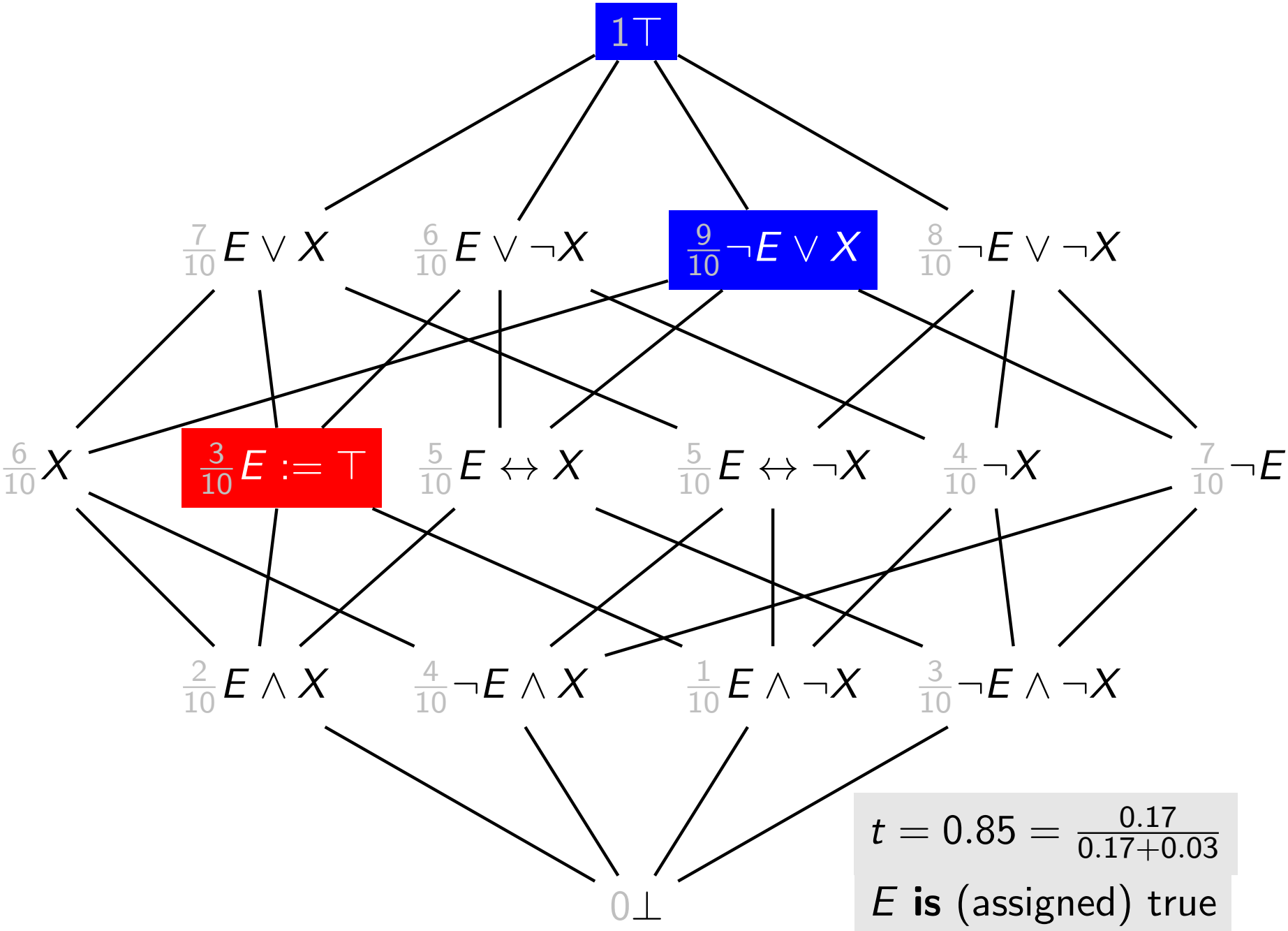
$$t = 0.85 = \frac{0.17}{0.17+0.03}$$



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

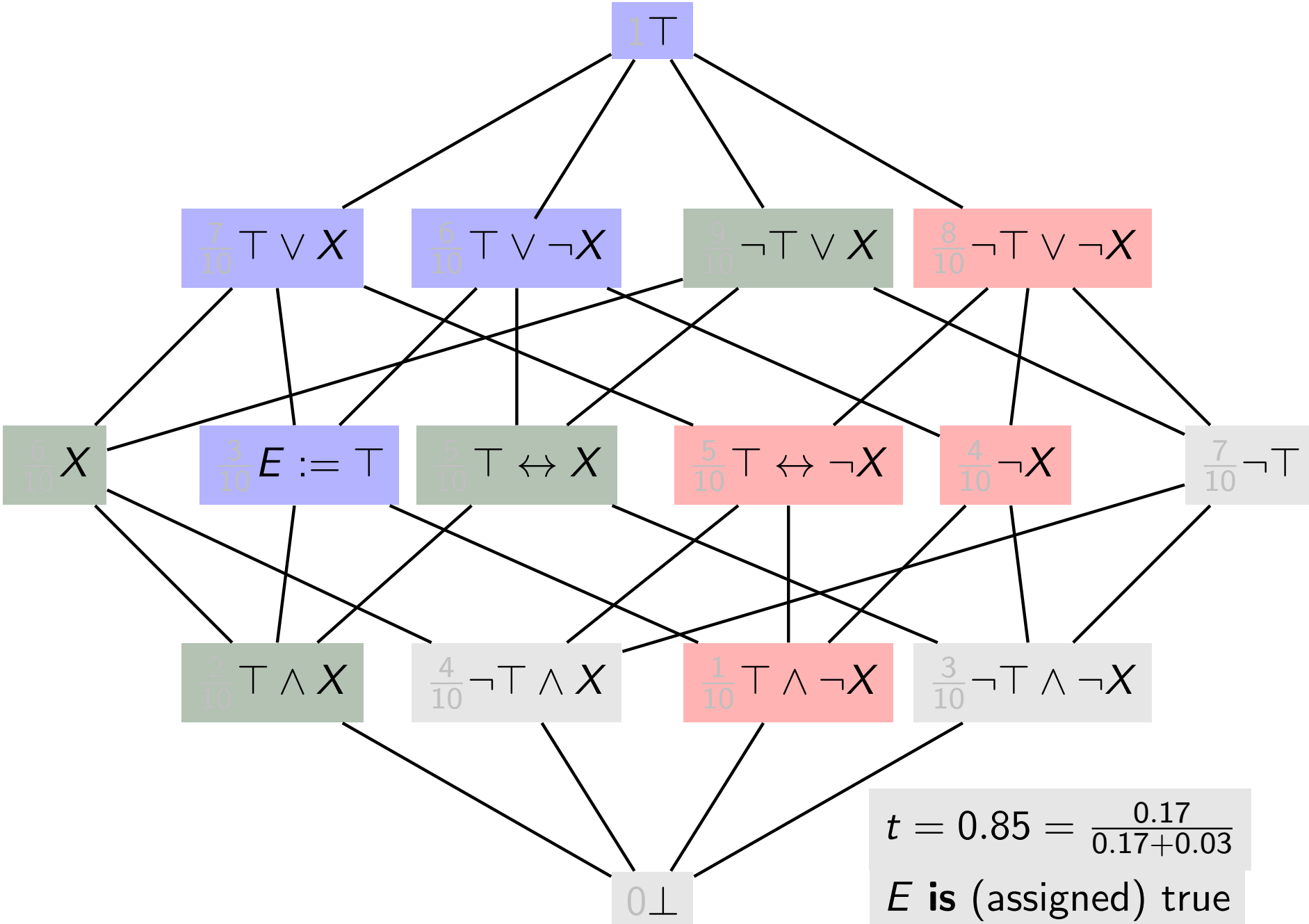
Suppose that E

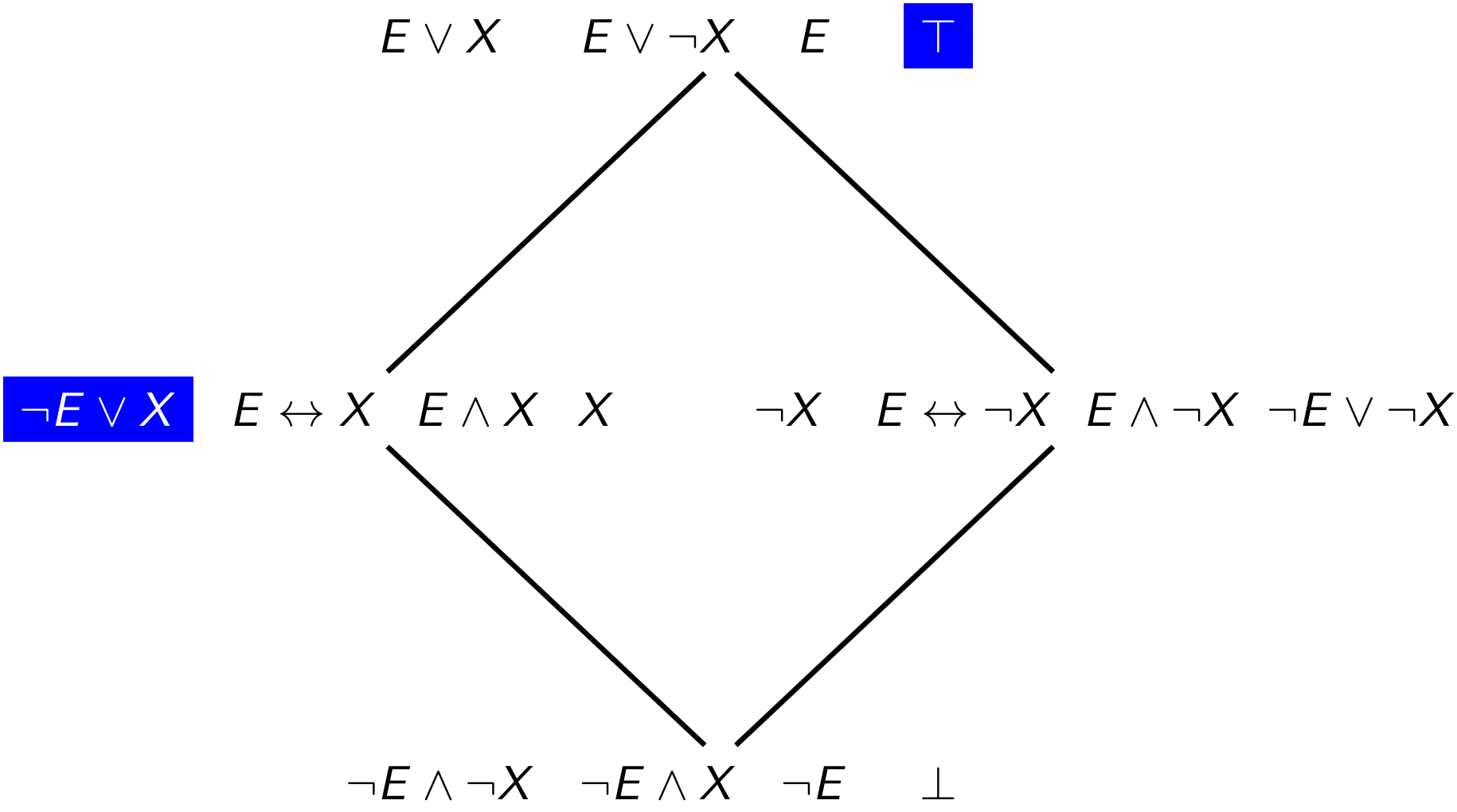

$1\top$
$\frac{7}{10} E \vee X$
$\frac{6}{10} E \vee \neg X$
$\frac{9}{10} \neg E \vee X$
$\frac{8}{10} \neg E \vee \neg X$
$\frac{6}{10} X$
$\frac{3}{10} E$
$\frac{5}{10} E \leftrightarrow X$
$\frac{5}{10} E \leftrightarrow \neg X$
$\frac{4}{10} \neg X$
$\frac{7}{10} \neg E$
$\frac{2}{10} E \wedge X$
$\frac{4}{10} \neg E \wedge X$
$\frac{1}{10} E \wedge \neg X$
$\frac{3}{10} \neg E \wedge \neg X$
$0\bot$
$t = 0.85 = \frac{0.17}{0.17+0.03}$
Suppose that E



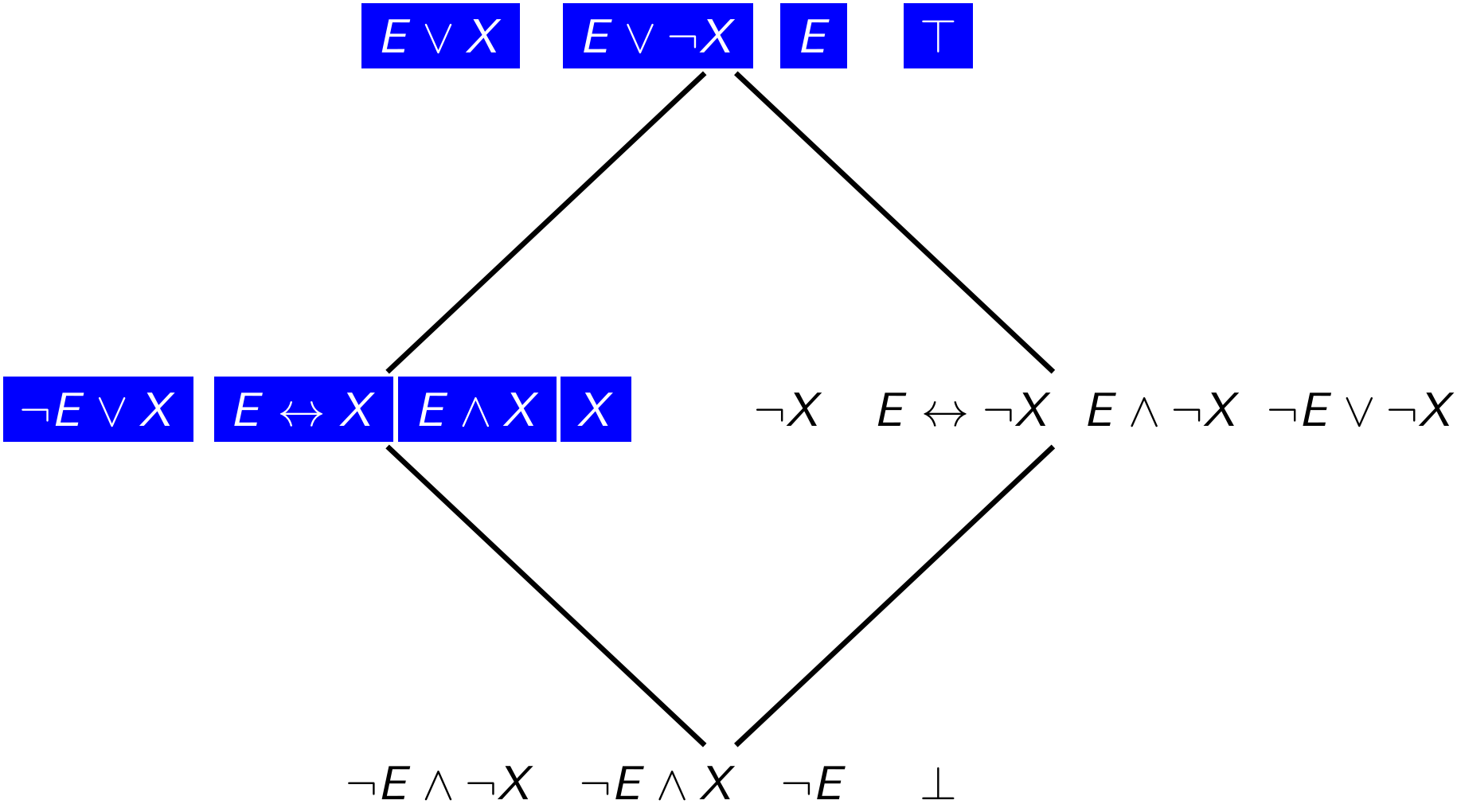
$t = 0.85 = \frac{0.17}{0.17+0.03}$

E **is** (assigned) true

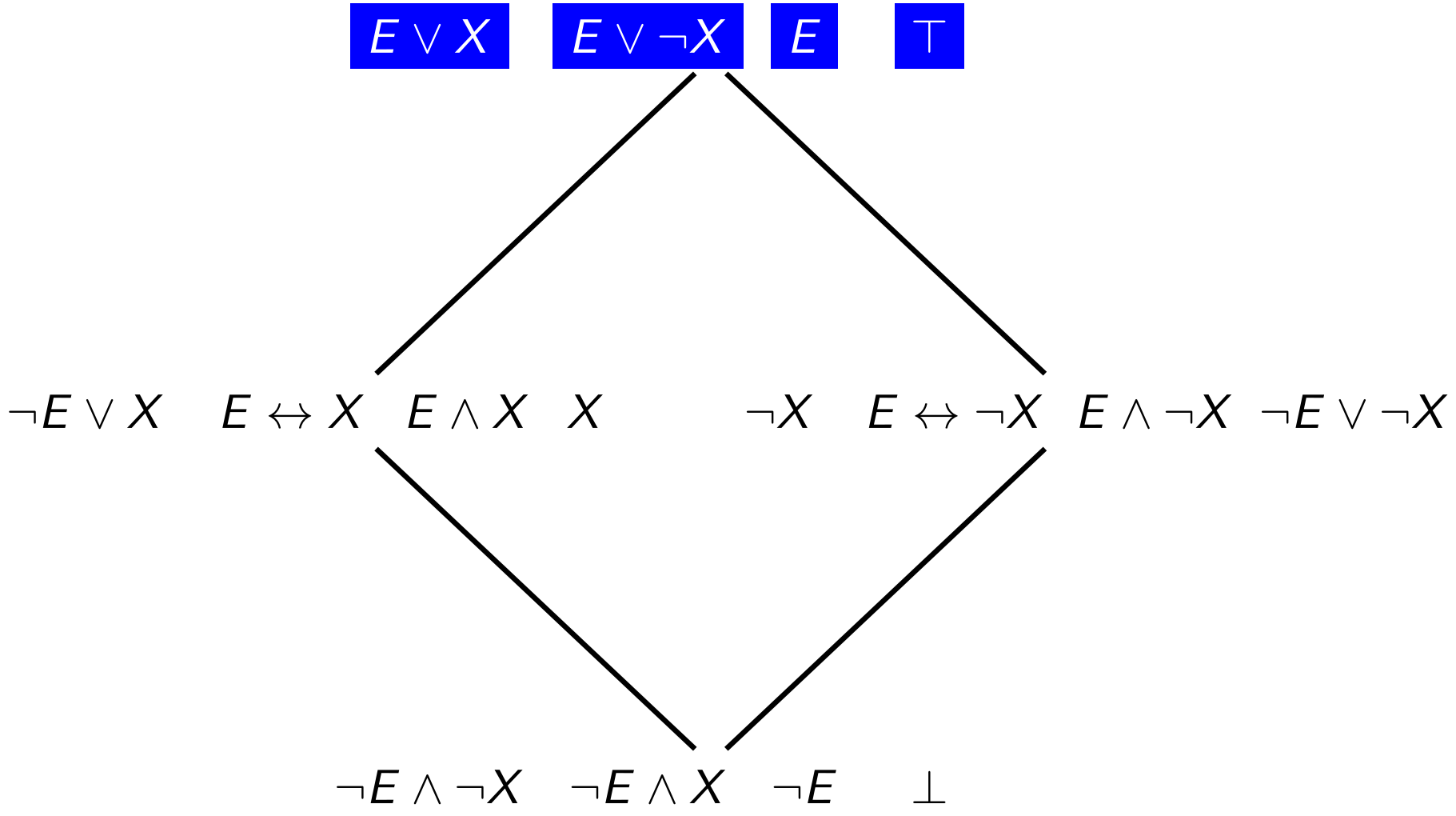

$\frac{1}{10}\top$
$\frac{7}{10}\top \vee X$
$\frac{6}{10}\top \vee \neg X$
$\frac{9}{10}\neg\top \vee X$
$\frac{8}{10}\neg\top \vee \neg X$
$\frac{6}{10}X$
$\frac{3}{10}E := \top$
$\frac{5}{10}\top \leftrightarrow X$
$\frac{5}{10}\top \leftrightarrow \neg X$
$\frac{4}{10}\neg X$
$\frac{7}{10}\neg\top$
$\frac{2}{10}\top \wedge X$
$\frac{4}{10}\neg\top \wedge X$
$\frac{1}{10}\top \wedge \neg X$
$\frac{3}{10}\neg\top \wedge \neg X$
$0\bot$
$t = 0.85 = \frac{0.17}{0.17+0.03}$
E **is** (assigned) true



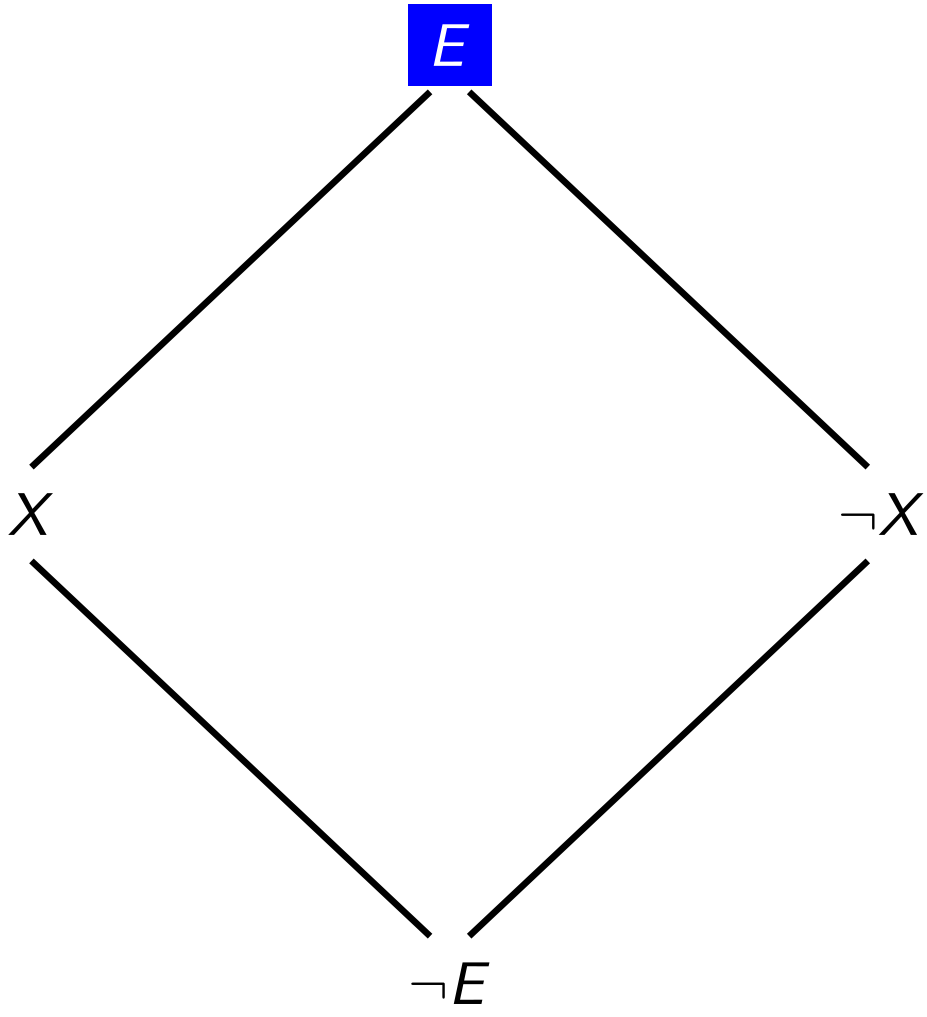
E **is** (assigned) true



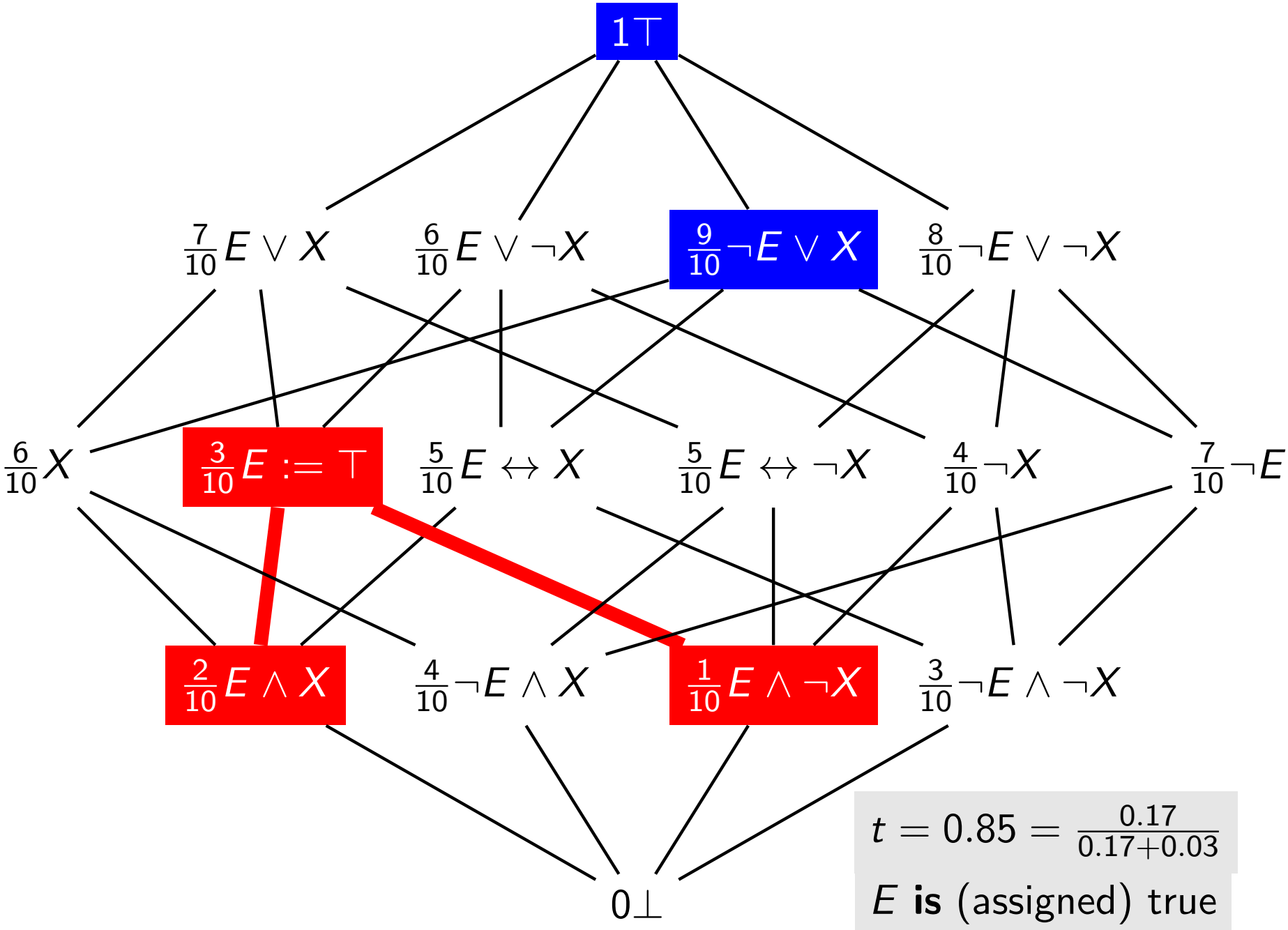
E **is** (assigned) true



E **is** (assigned) true

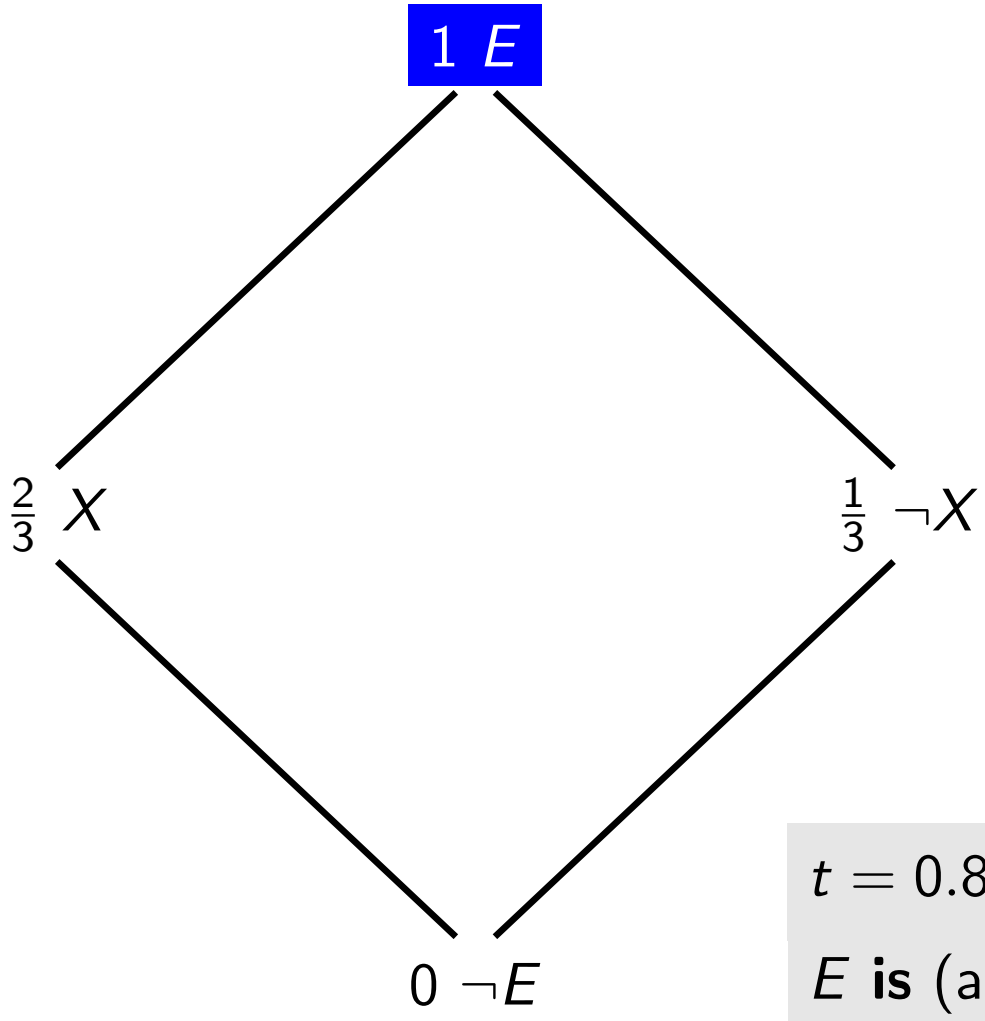


E **is** (assigned) true



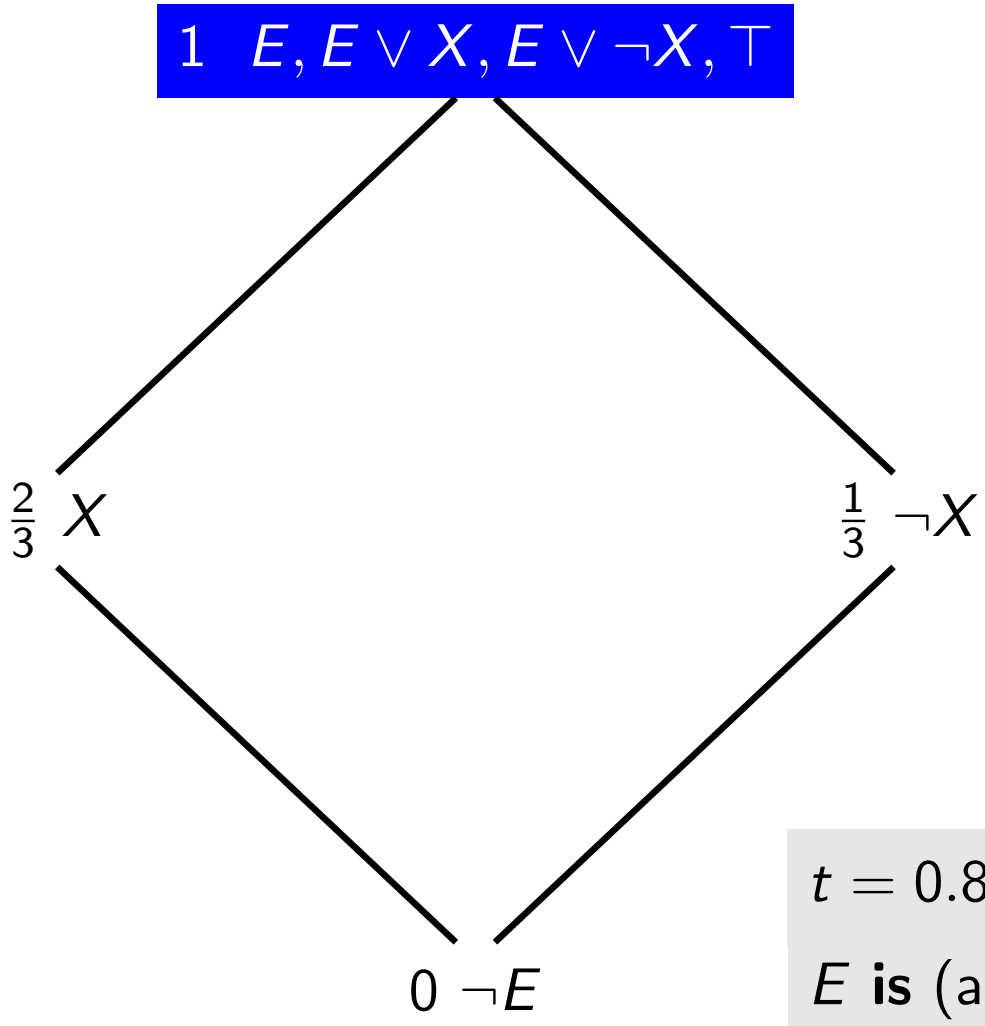
$t = 0.85 = \frac{0.17}{0.17+0.03}$

E **is** (assigned) true



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

E **is** (assigned) true



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

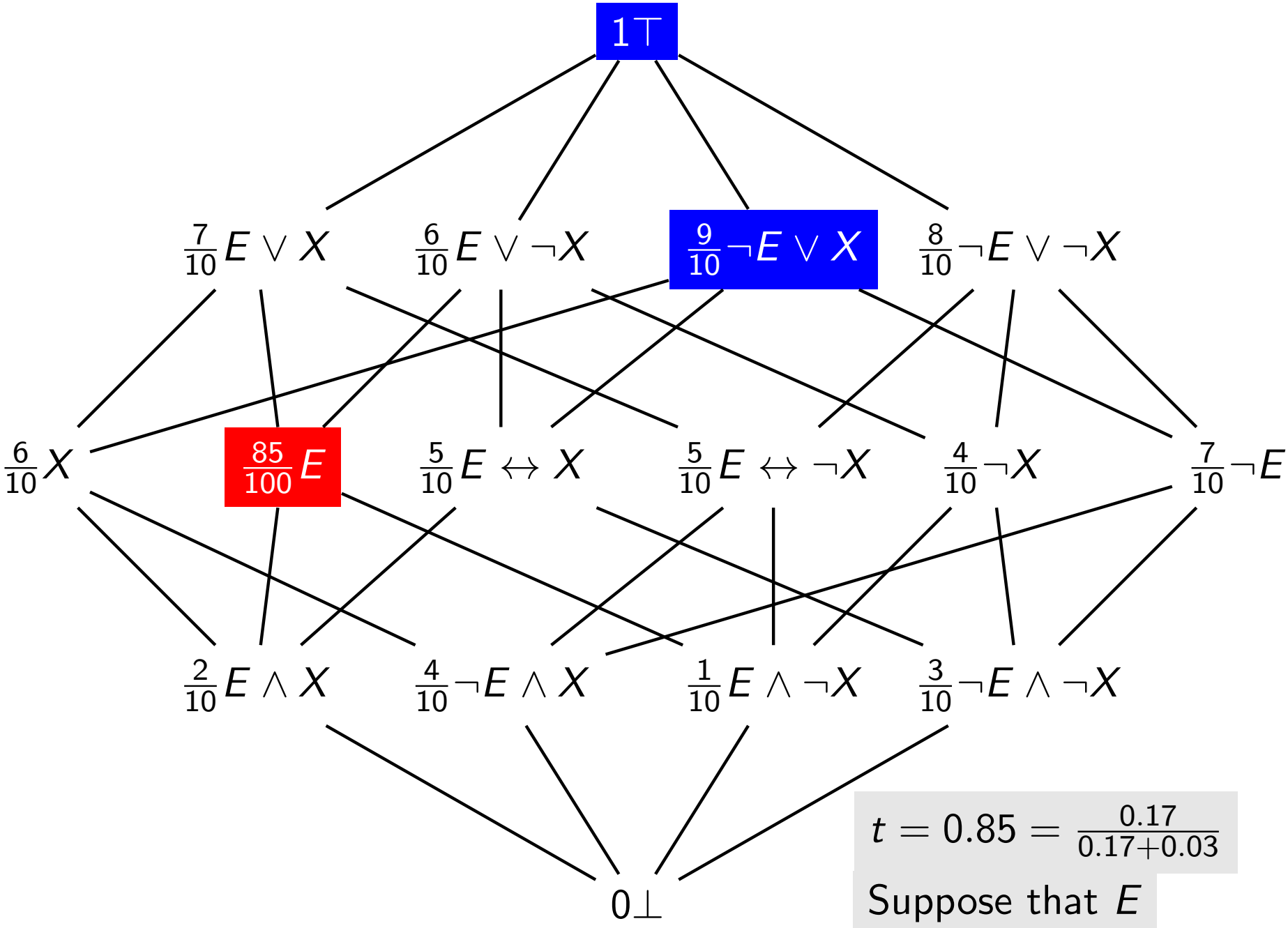
E **is** (assigned) true

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

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

$B \nvdash \neg E$, $B * E = Cn(B \cup \{E\}) = Cn(\{E \wedge X\})$ So, $X \in B * E$.

$B \divideontimes E = Cn(\{E, E \vee X, E \vee \neg X\})$, so $X \notin B \divideontimes E$.



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

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 $B \divideontimes E \subset B * E$

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.

Proposition. If an EUT/Lockean agent is deductively cogent (at all times), then they can only violate Vacuity (via learning some E that they do not already believe) if their Lockean threshold is on the half-open interval $[\varphi - 1, 1)$.

Leitgeb & Segerberg: Belief Update vs. Belief Revsions

...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: Belief Update vs. Belief Revsions

...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.

Leitgeb & Segerberg: Belief Update vs. Belief Revsions

It is tempting to relate these different views on belief change to the traditional distinction of indicative and subjective conditionals. Using the stock example: everyone considers the indicative 'If Oswald did not kill Kennedy somebody else did' as acceptable, but many regard the subjunctive 'If Oswald had not killed Kennedy somebody else would have' as false.

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.

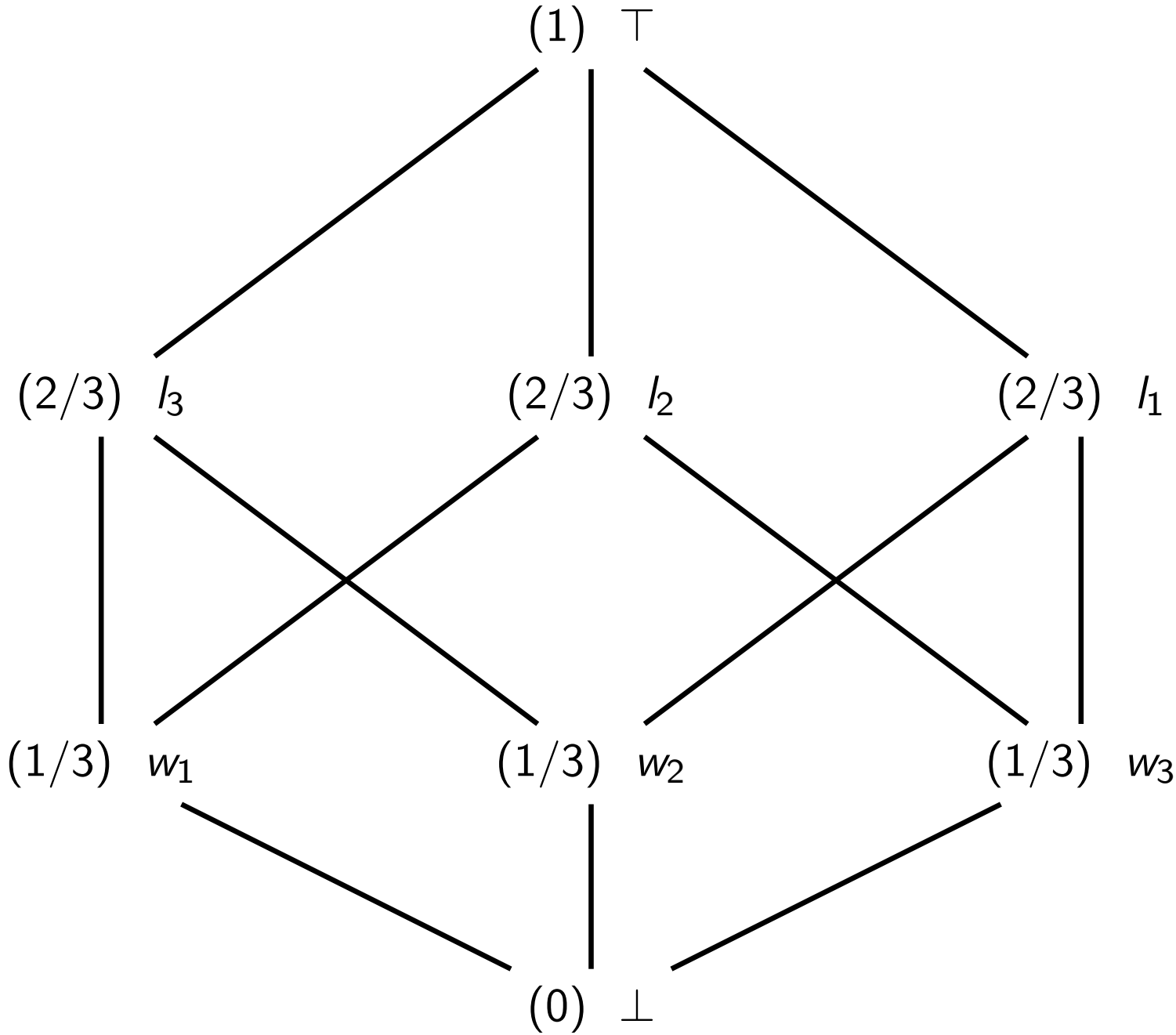
B. Skyrms. *Updating, Supposing and MAXENT*. Theory and Decision, 22, pp. 225 - 246, 1987.

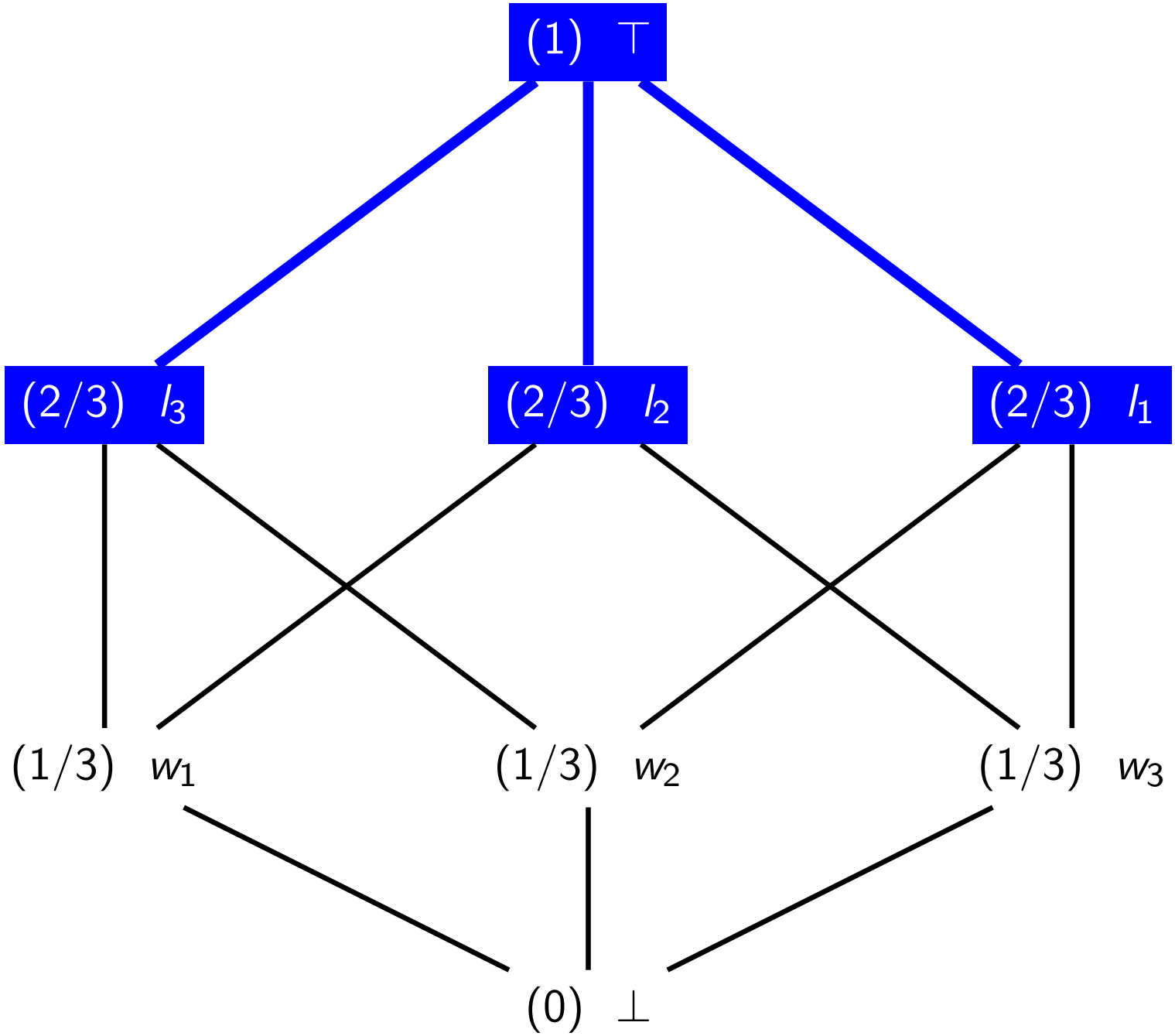
Such normative virtues suggest a psychological question. One way of formulating (1) is that *supposing* an event B should have the same impact on the credibility of an event A as *learning* B. Is this true for typical assessments of chance? For example, is the judged probability of a Democratic victory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

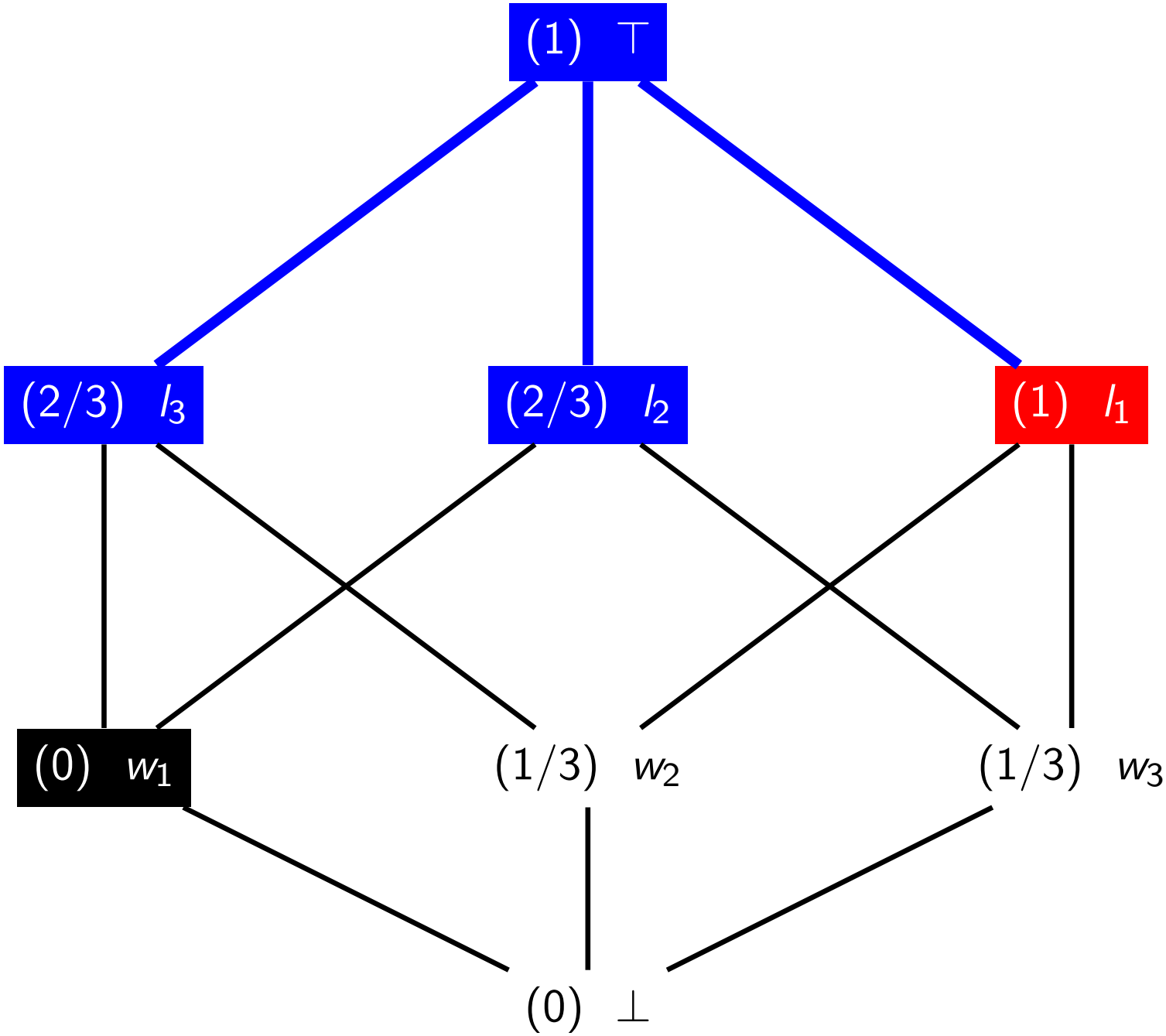
Jiaying Zhao, Vincenzo Crupi, Katya Tentori, Branden Fitelson, and Daniel Osherson. *Updating: Learning versus supposing*. Cognition 124 (2012) 373378.

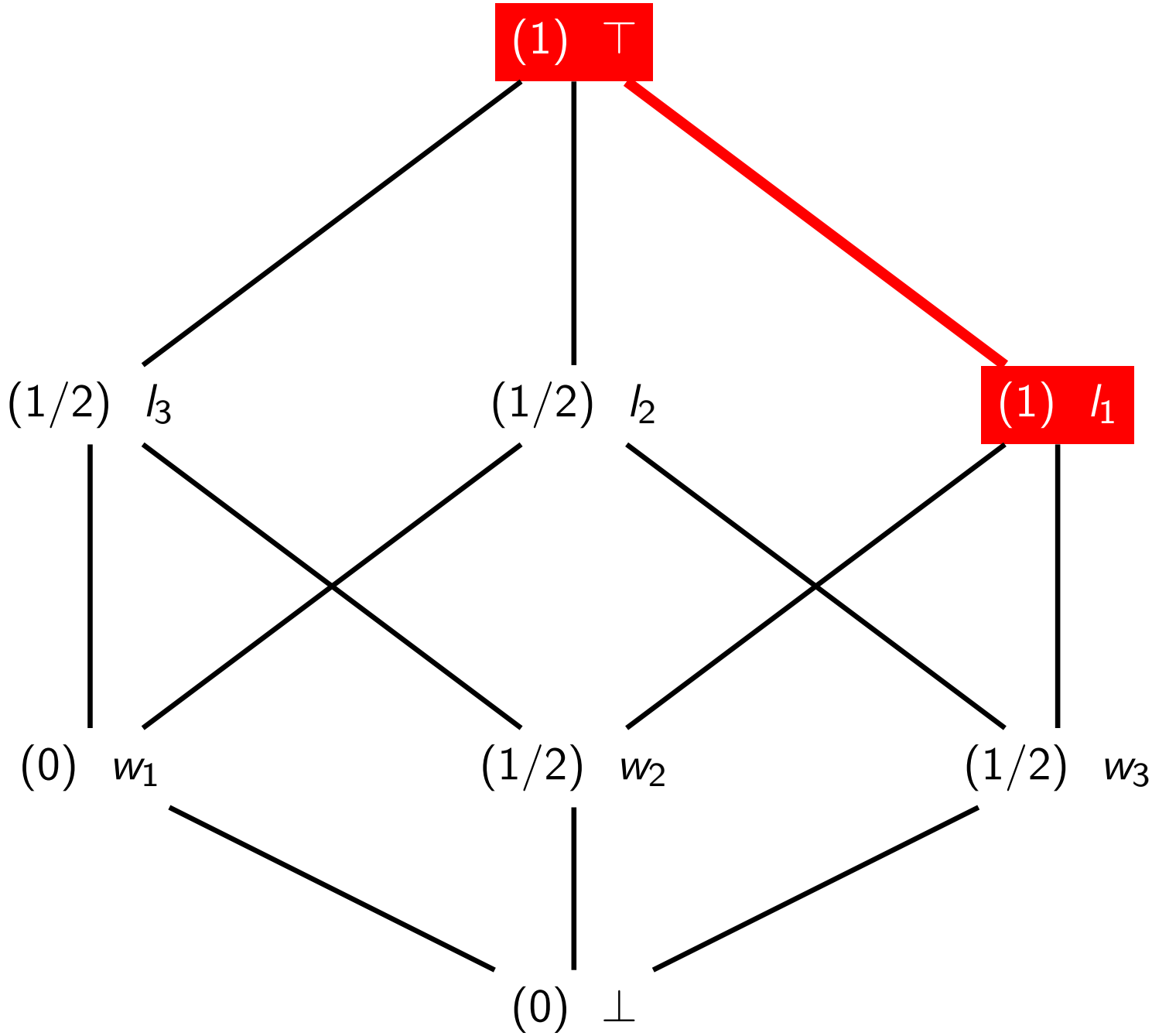
Beliefs that obey the Lockean thesis can be undermined by new evidence that is consistent with the agents current beliefs.

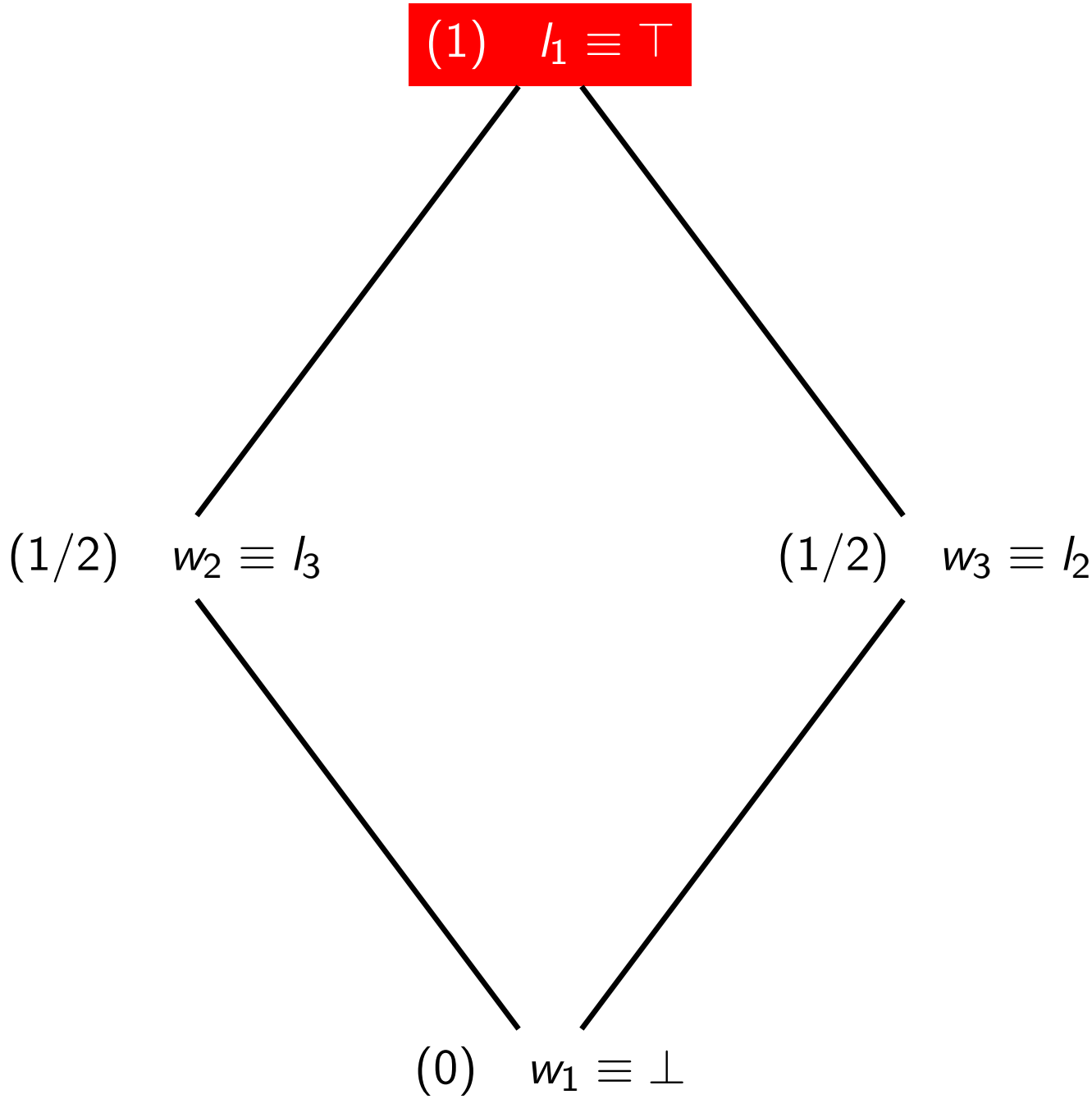
For each $i = 1, 2, 3$, let l_i be the proposition Ticket i won't win (and w_i is the proposition that "ticket i will win"). And let us set our threshold for Lockean belief at $r = 0.6$.


(1) ⊤
(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
(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 is P-stable^t if and only if for all $Y \in \mathfrak{A}$ with $Y \cap X \neq \emptyset$ and $P(Y) > 0$: $P(X|Y) > t$.

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

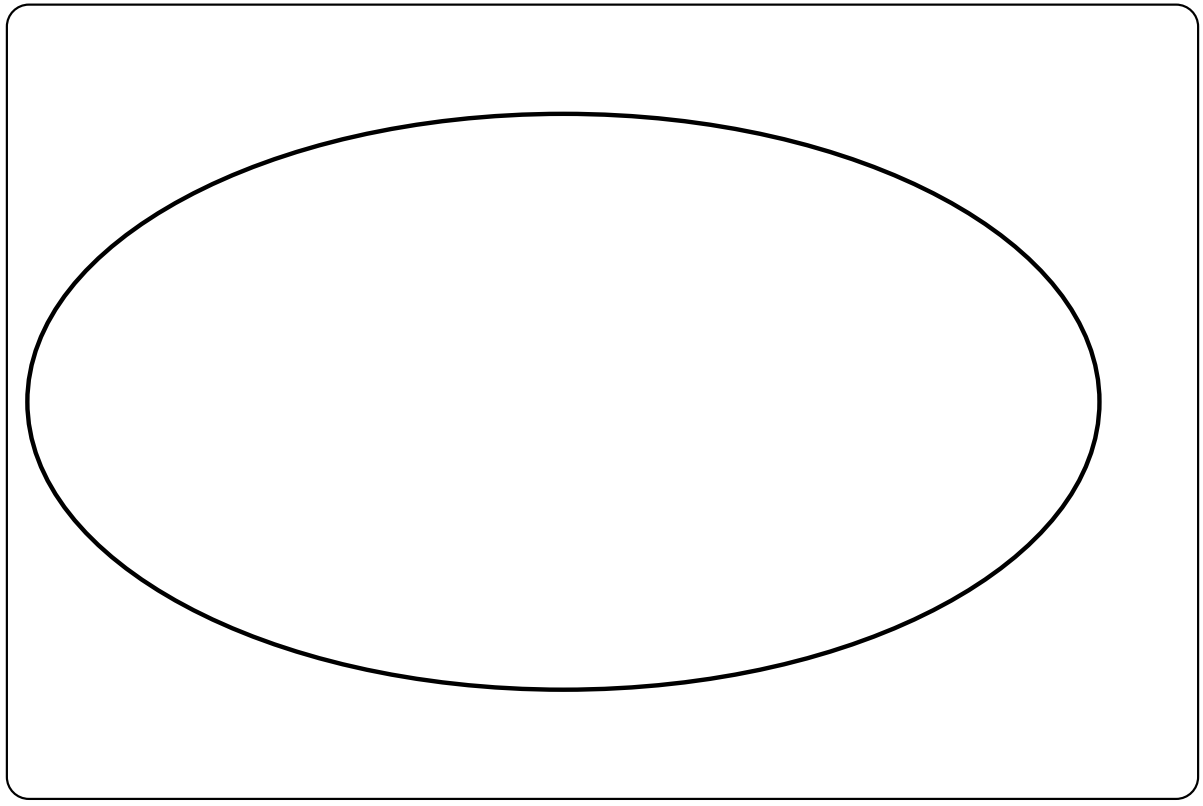
- Trivially, the empty set of P-stable^t.
- If $P(X) = 1$, then X is P-stable^t.

P-stability^r

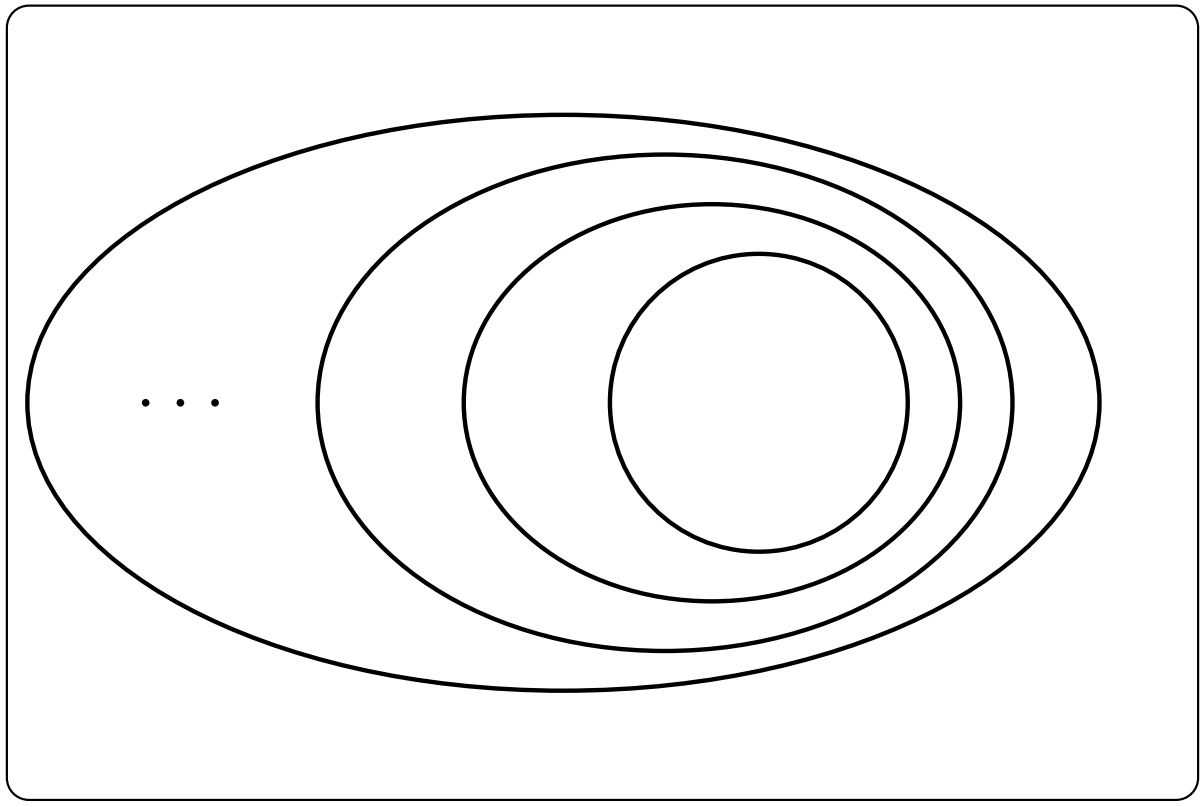
Definition. Let P be a probability measure on $\mathfrak{A}$ over W, let $0 \leq t < 1$. For all $X \in \mathfrak{A}$:

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

- 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

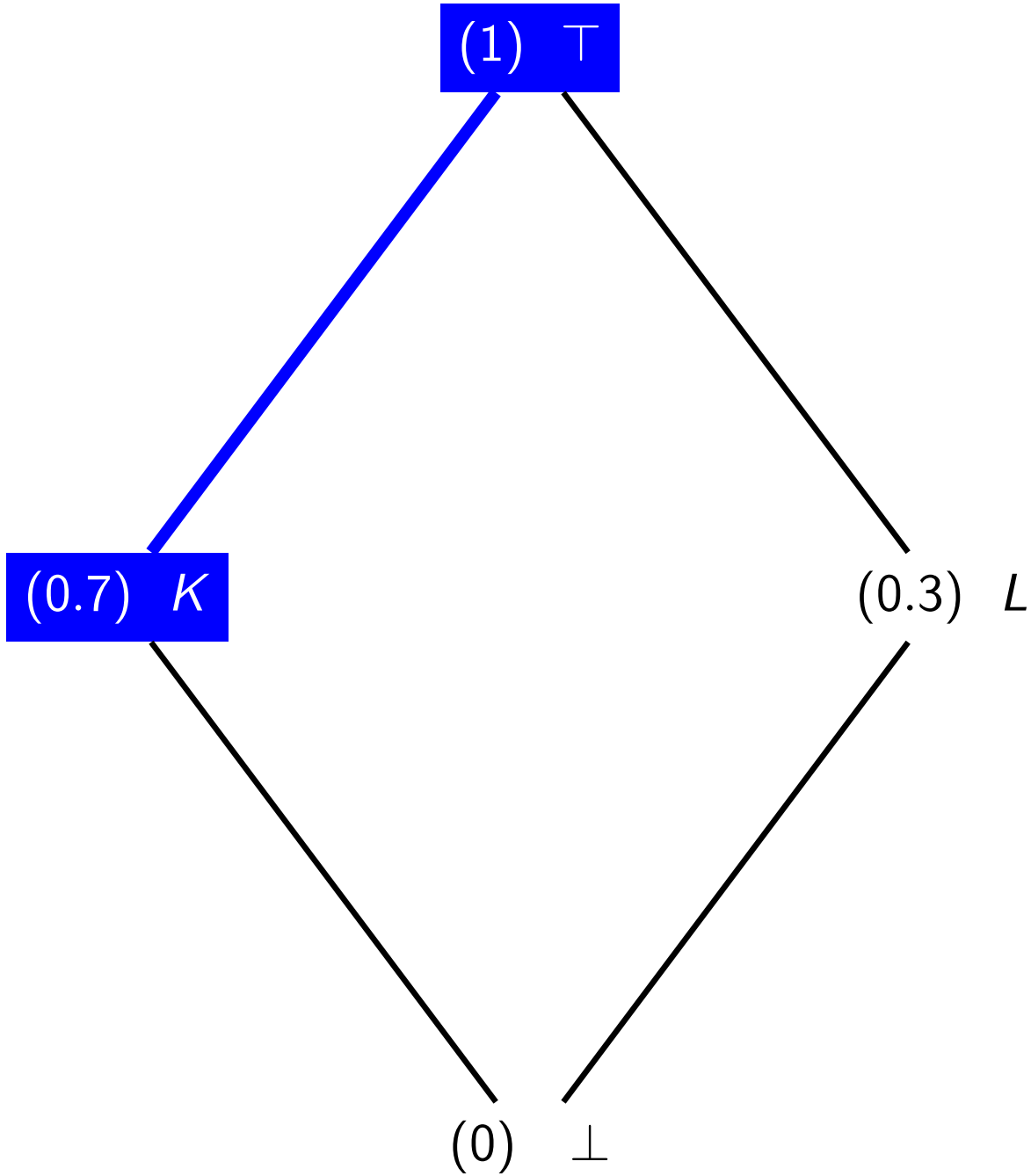
$w \in SB(H)$ iff for all $E \in \mathfrak{A}_H(W_\Pi)$ with $H \cap E \neq \emptyset$ and $P(E) \neq 0$: $P(H \mid E) \geq t_C$

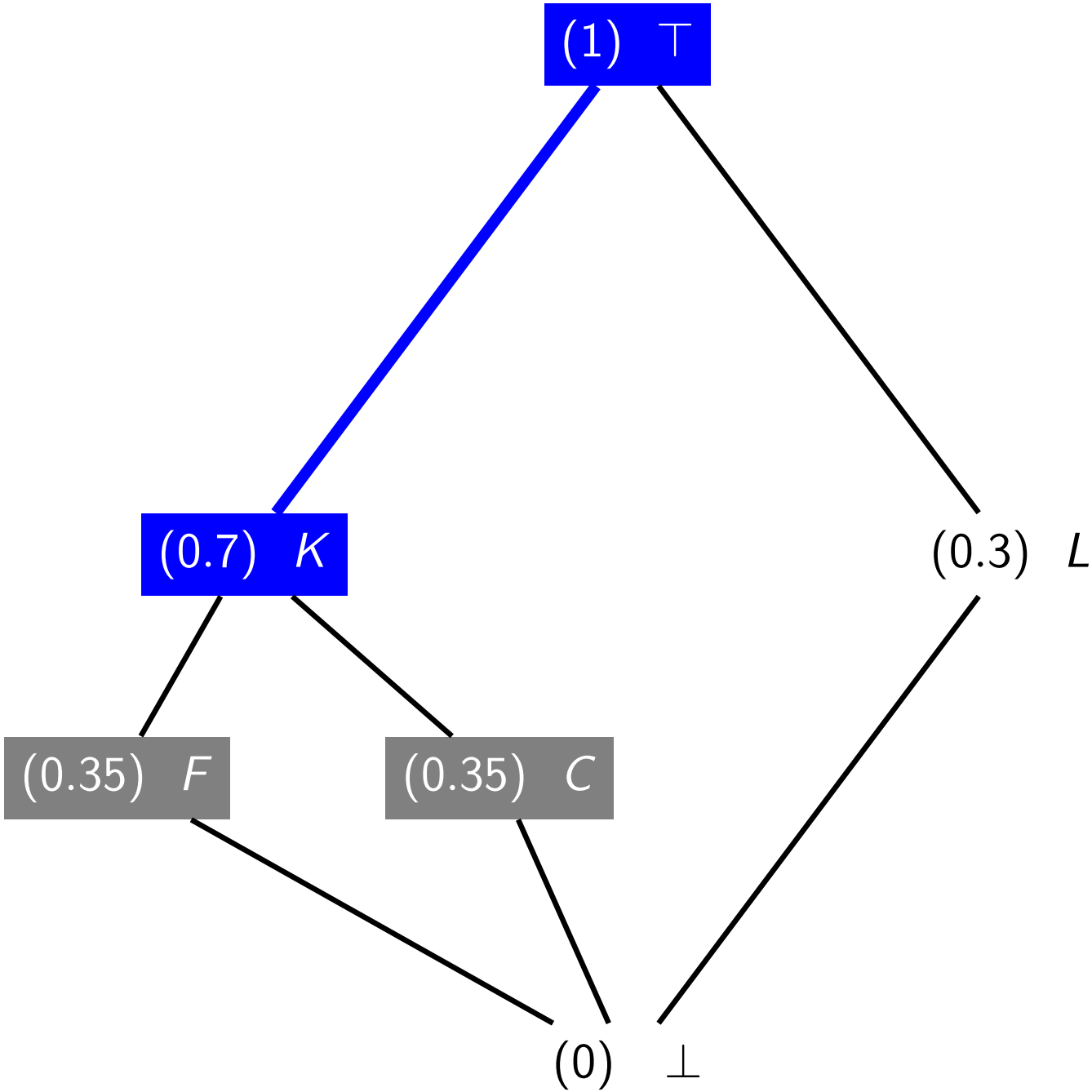
1. The threshold t is determined contextually (the "cautiousness level")
2. The evidence "relevant" to H
3. The states may be contextually determined (by a partition Π on a set W of "maximally specific worlds")

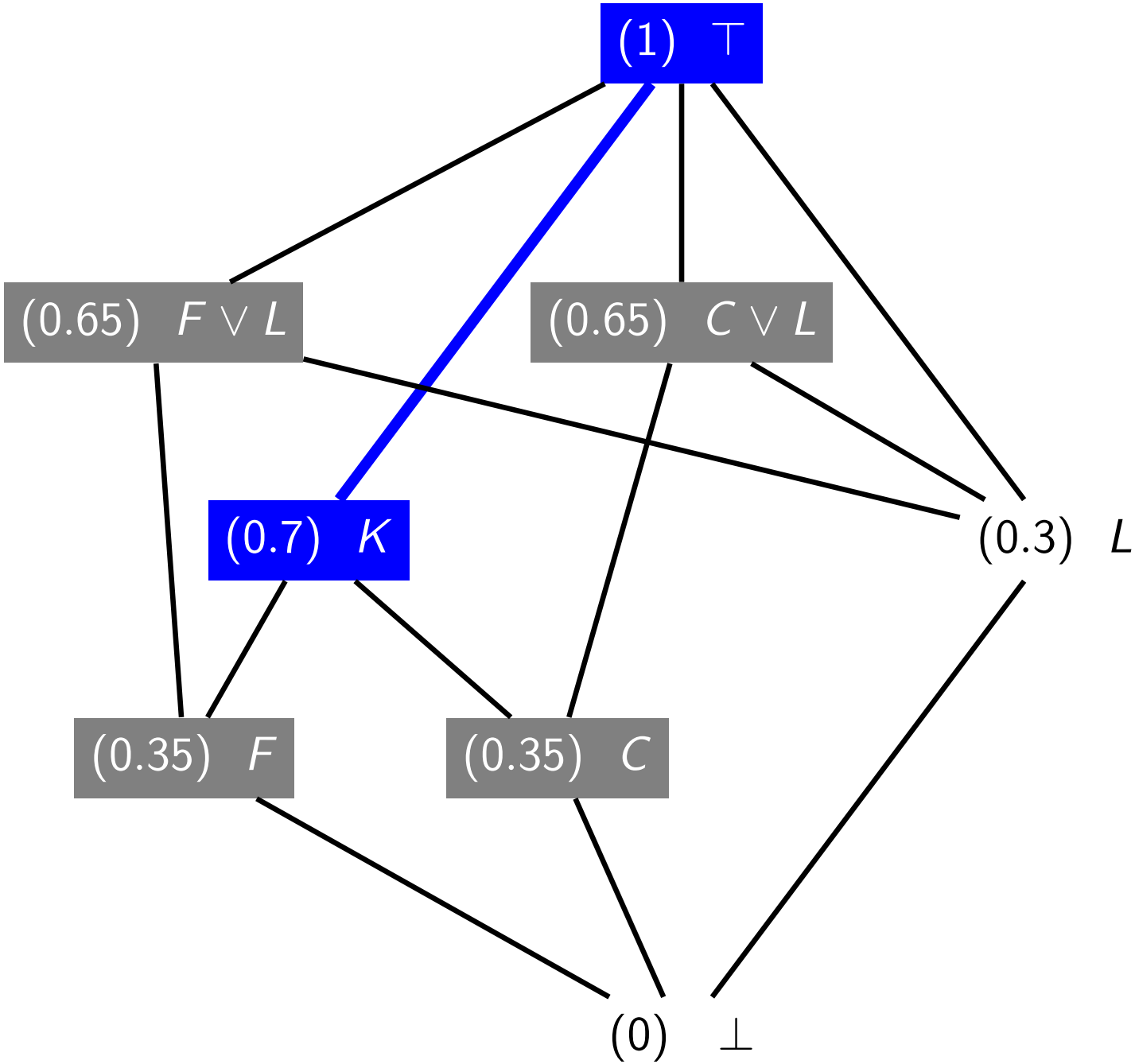
H. Leitgeb. *The Stability Theory of Belief*. The Philosophical Review 123/2, 131171, 2014.

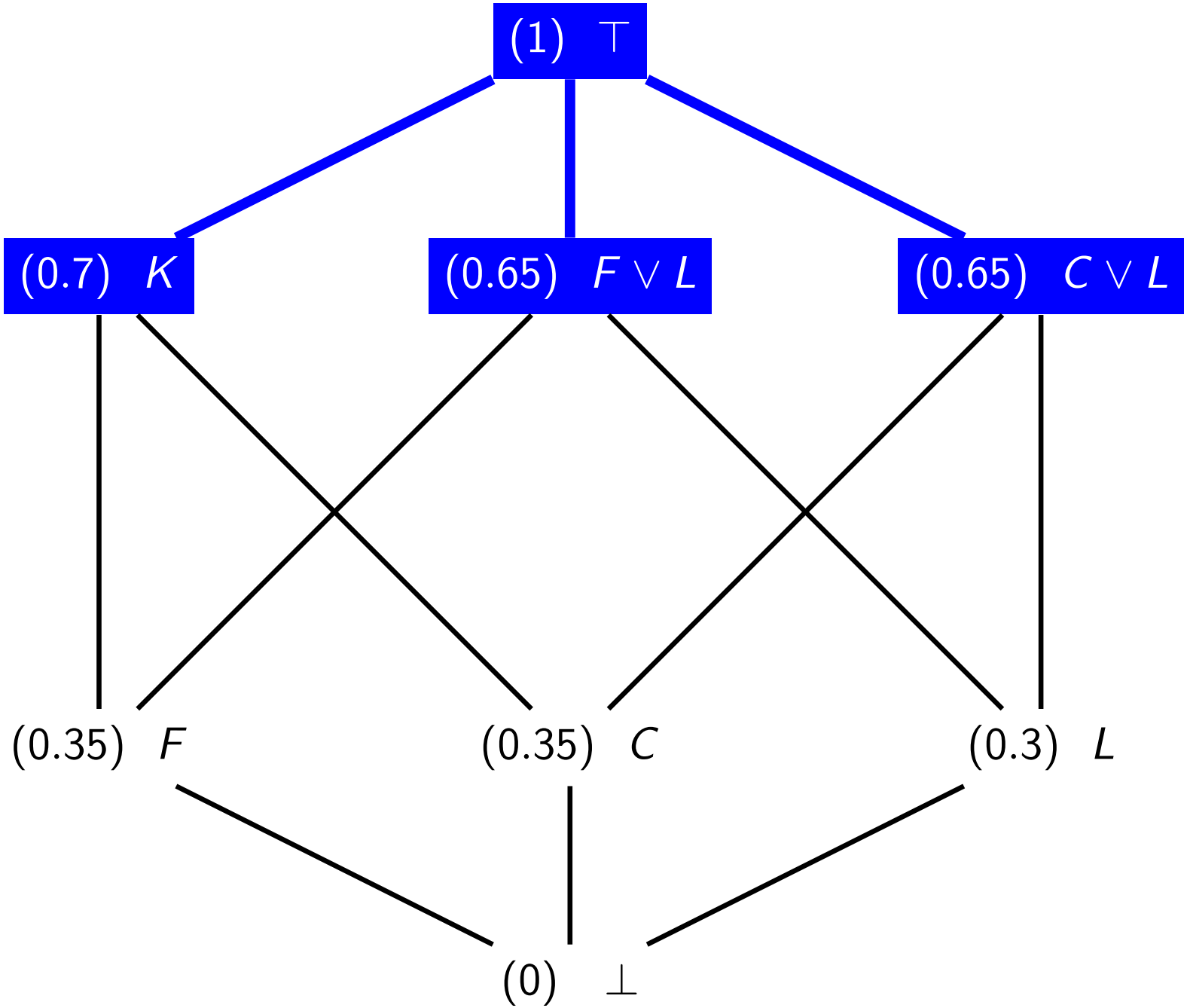
H. Leitgeb. *The Humean Thesis on Belief*. Proceedings of the Aristotelian Society of Philosophy 89(1), 143185, 2015.

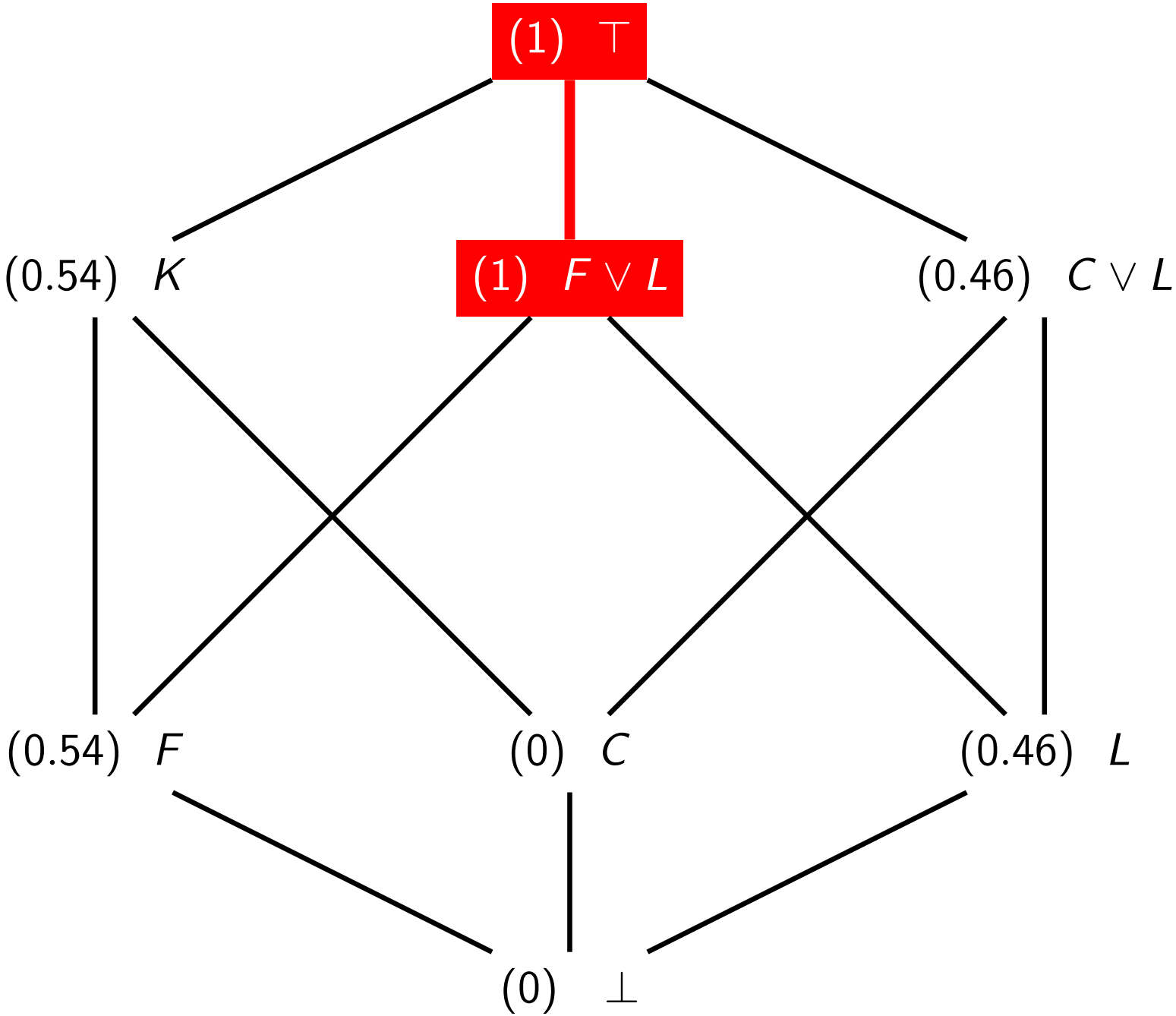
R. Pettigrew. *Pluralism about belief states*. Proceedings of the Aristotelian Society 89(1):187-204, 2015.


(1) $\top$
(0.7) K
(0.3) L
(0) $\bot$


(1) $\top$
(0.7) K
(0.3) L
(0.35) F
(0.35) C
(0) $\bot$


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


(1) $\top$
(0.7) K
(0.65) $F \vee L$
(0.65) $C \vee L$
(0.35) F
(0.35) C
(0.3) L
(0) $\bot$


(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 $B^\Pi_W = \{\{w_2, \ldots, w_{1,000,000}\}\}$.

If B^Π_W 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 wont 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 agents 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

F. Dietrich, C. List and R. Bradley. *Belief revision generalized: A joint characterization of Bayes's and Jeffrey's rules*. manuscript.

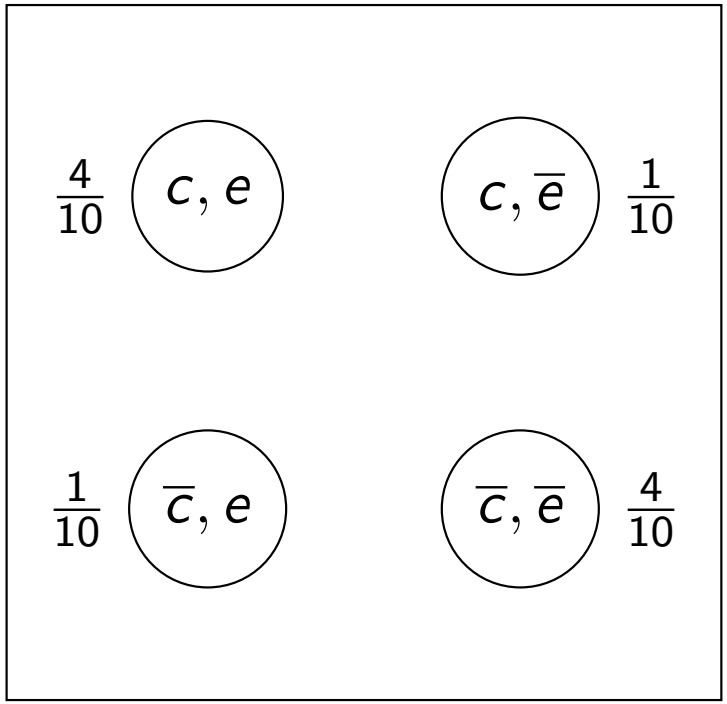
Characterization Result

Responsiveness: The agent's revised belief state *respects the constraint* given by the input.

Conservativeness: For all belief-input pairs (p, I), if I is "silent" on the probability of a (relevant) event A given another B, this conditional probability is preserved.

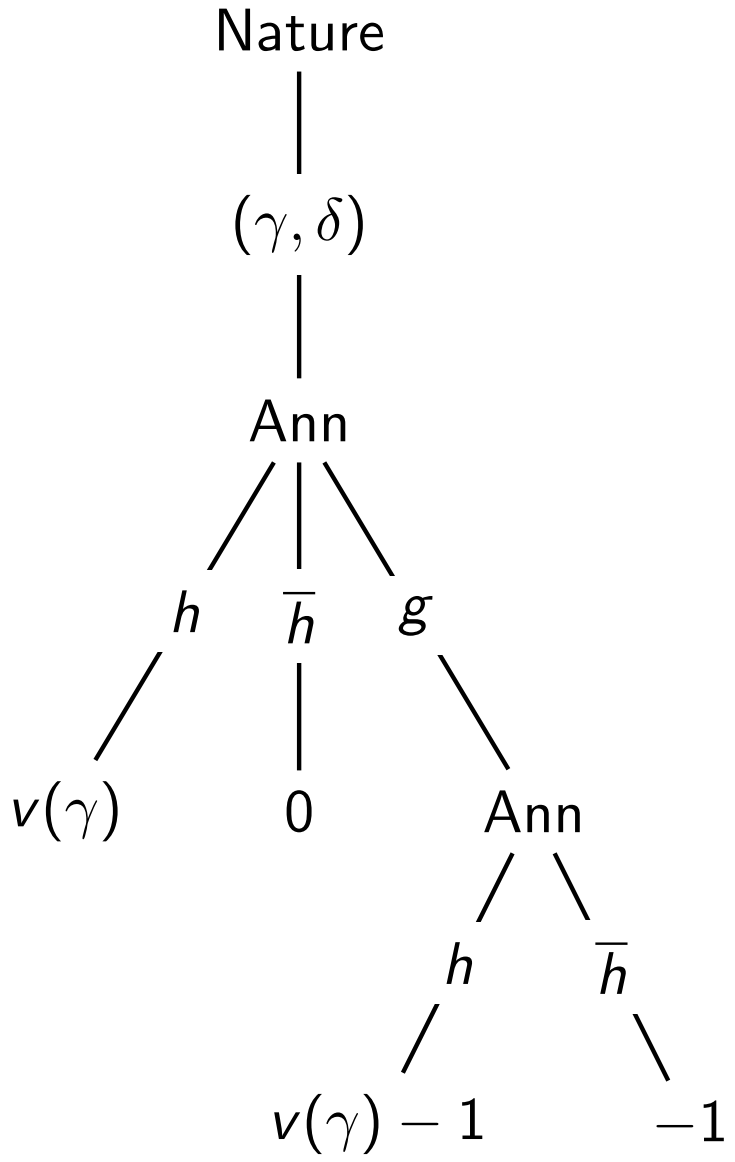
A decision-theoretic example

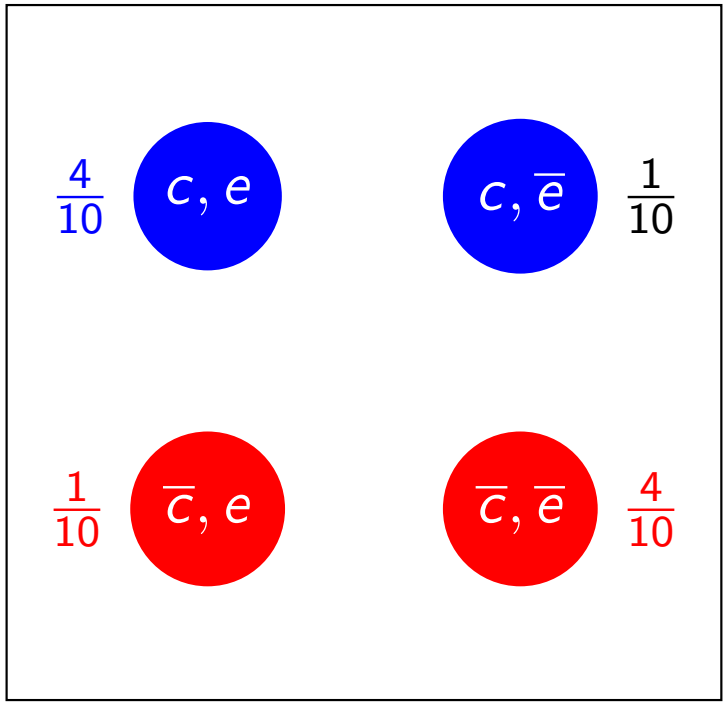
Ann, an employer, must decide whether to hire Bob, a job candidate. There is no time for a job interview, since a quick decision is needed. Ann is uncertain about whether Bob is competent or not; both possibilities have prior probability $\frac{1}{2}$. It would help Ann to know whether Bob has previous work experience, since this is positively correlated with competence, but gathering this information takes time.

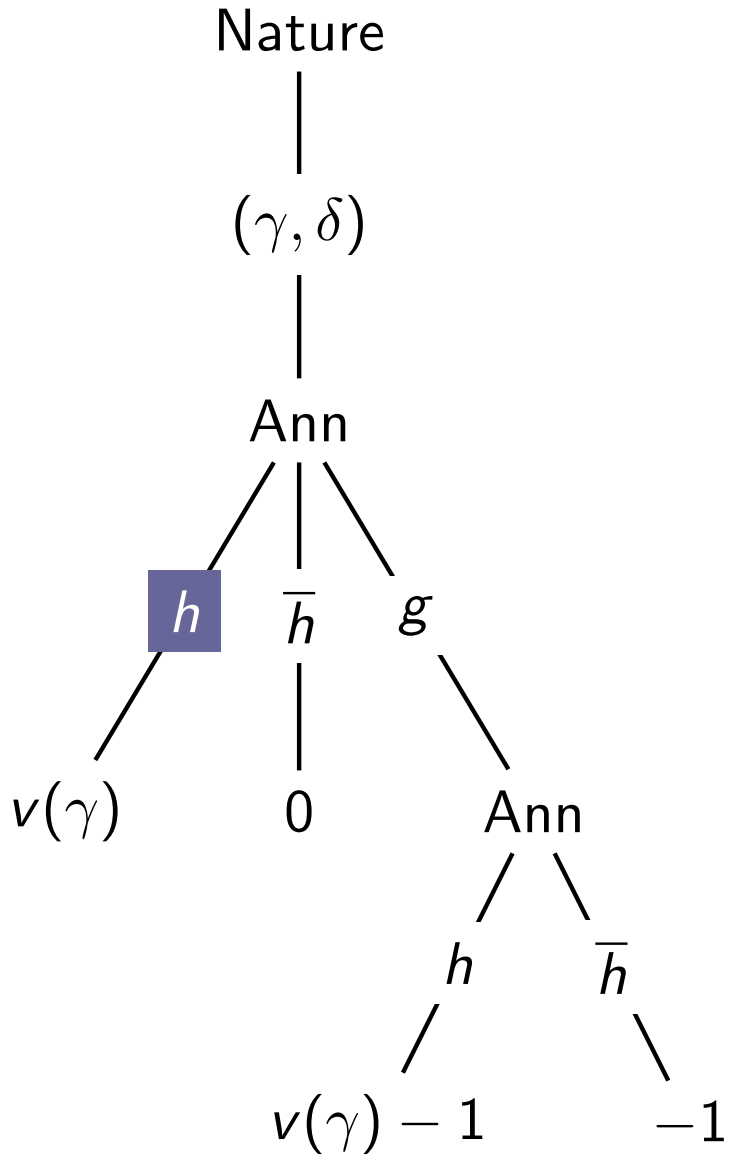


if $\gamma = c$, then $v(\gamma) = 5$

if $\gamma = \overline{c}$, then $v(\gamma) = -5$







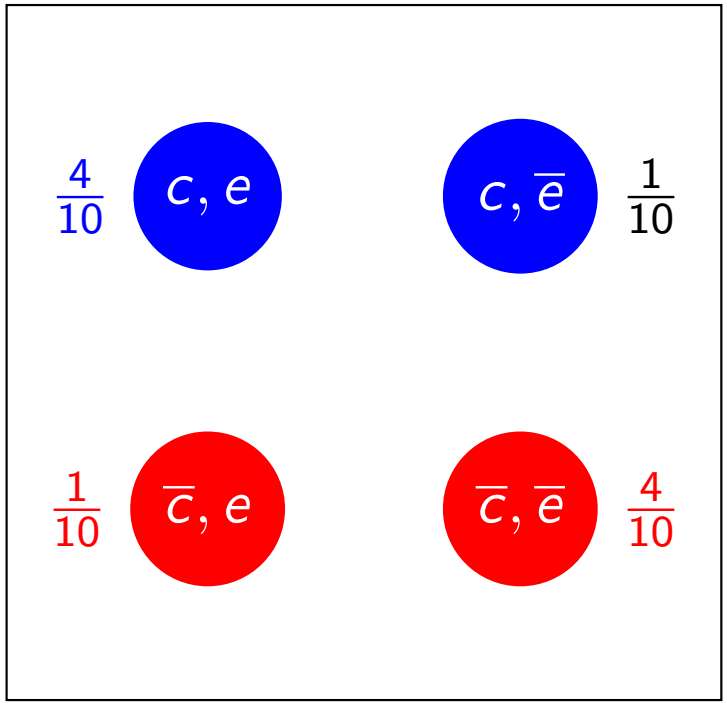
if $\gamma = c$, then $v(\gamma) = 5$

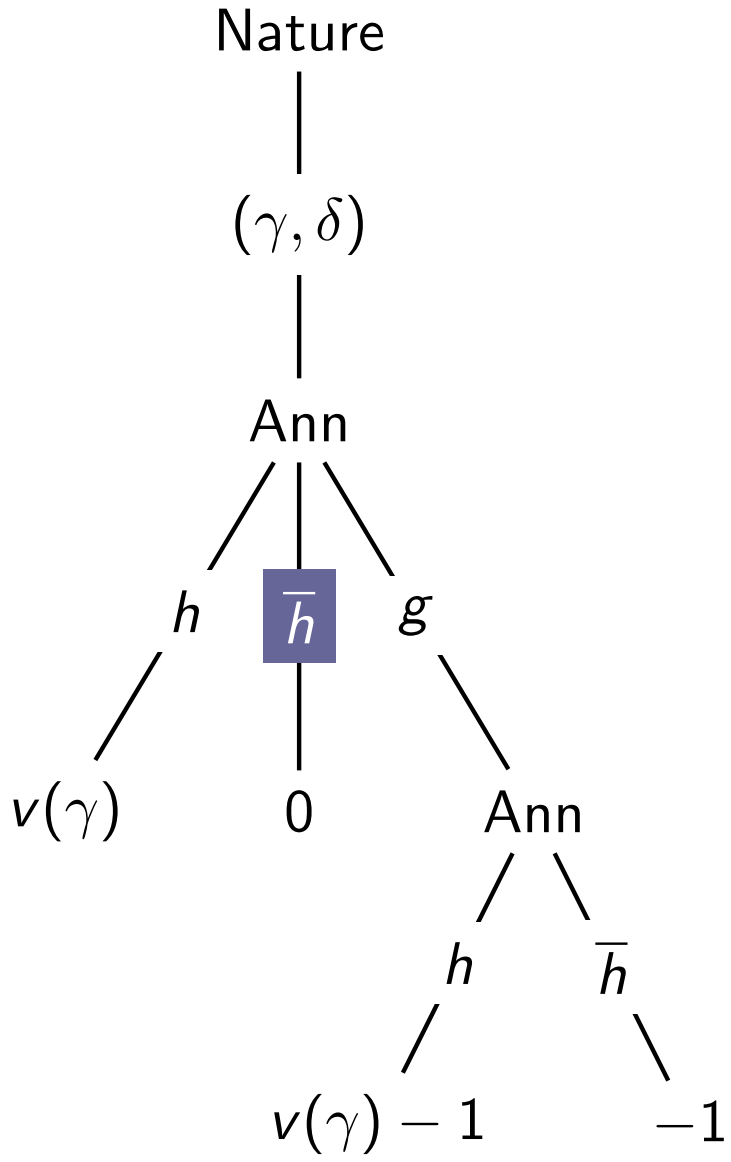
if $\gamma = \overline{c}$, then $v(\gamma) = -5$

$EU(h) =$

$p(C)u(C) + p(\overline{C})u(\overline{C}) =$

$0.5 * 5 + 0.5 * -5 = 0$





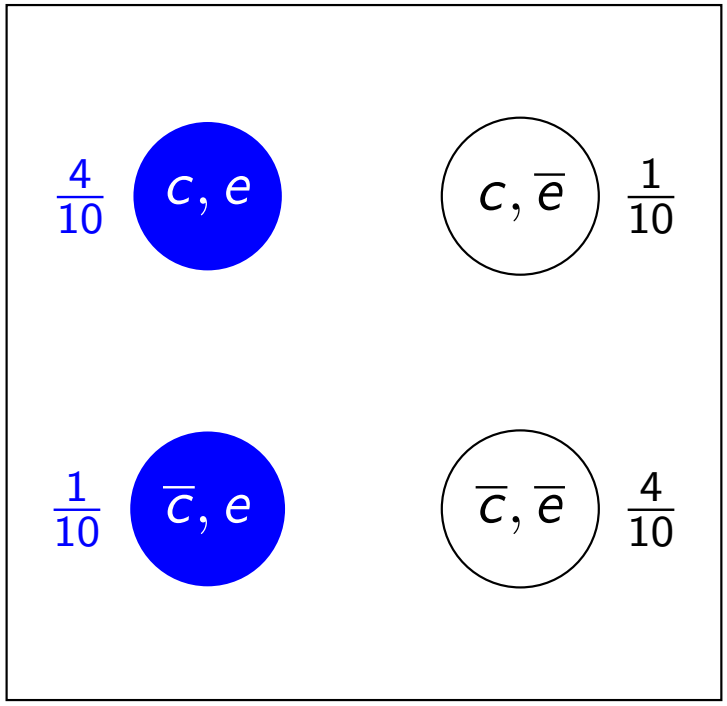
if $\gamma = c$, then $v(\gamma) = 5$

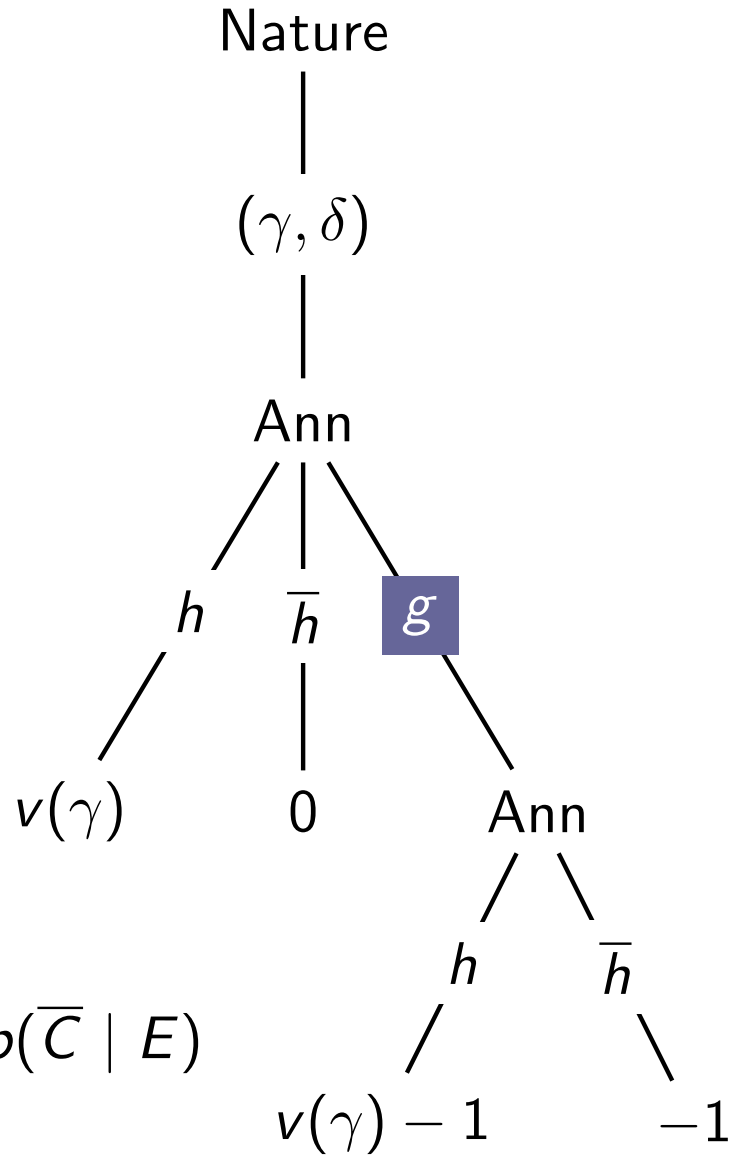
if $\gamma = \overline{c}$, then $v(\gamma) = -5$

$EU(\overline{h}) =$

$p(C)u(C) + p(\overline{C})u(\overline{C}) =$

$0.5 * 0 + 0.5 * 0 = 0$



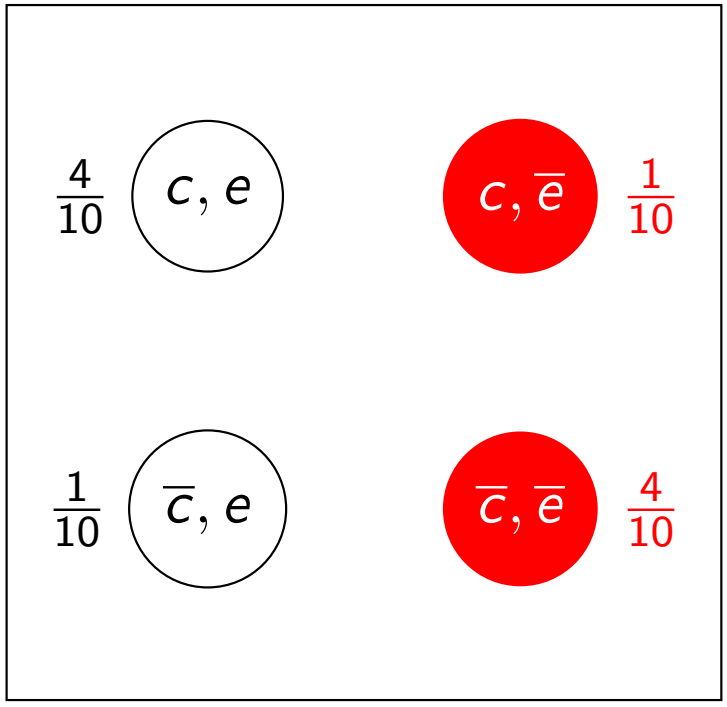


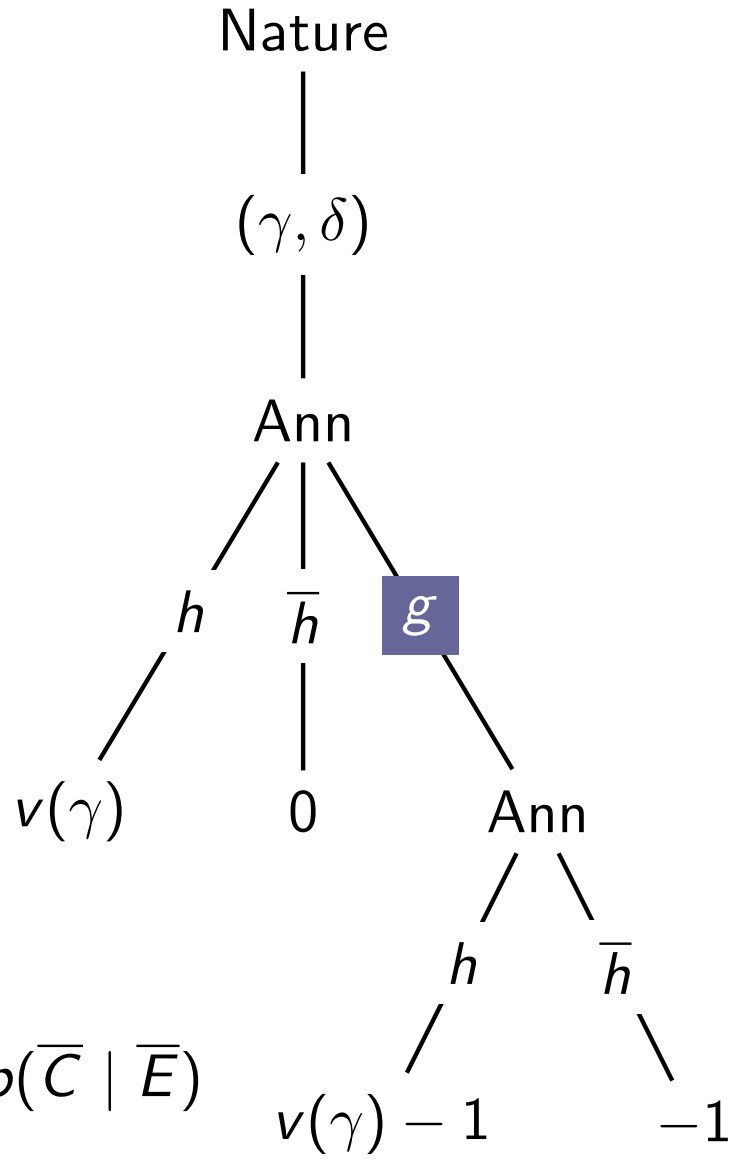
$$p(C \mid E) = \frac{\frac{4}{10}}{\frac{5}{10}} = \frac{4}{5}$$

$$(v(c) - 1)p(C \mid E) + (v(\overline{c}) - 1)p(\overline{C} \mid E)$$

$$EU_E(h) = 4\frac{4}{5} + -6\frac{1}{5} = 2$$

$$EU_E(\overline{h}) = -1$$



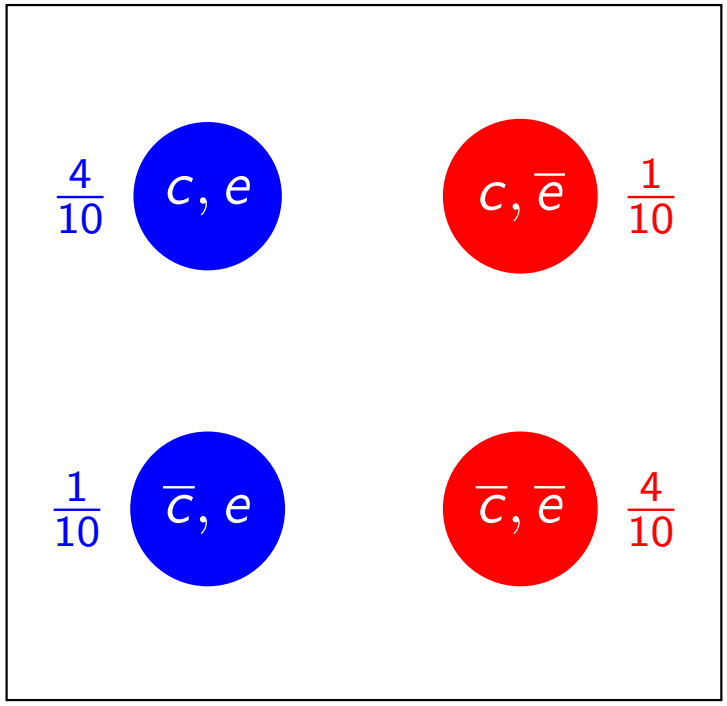


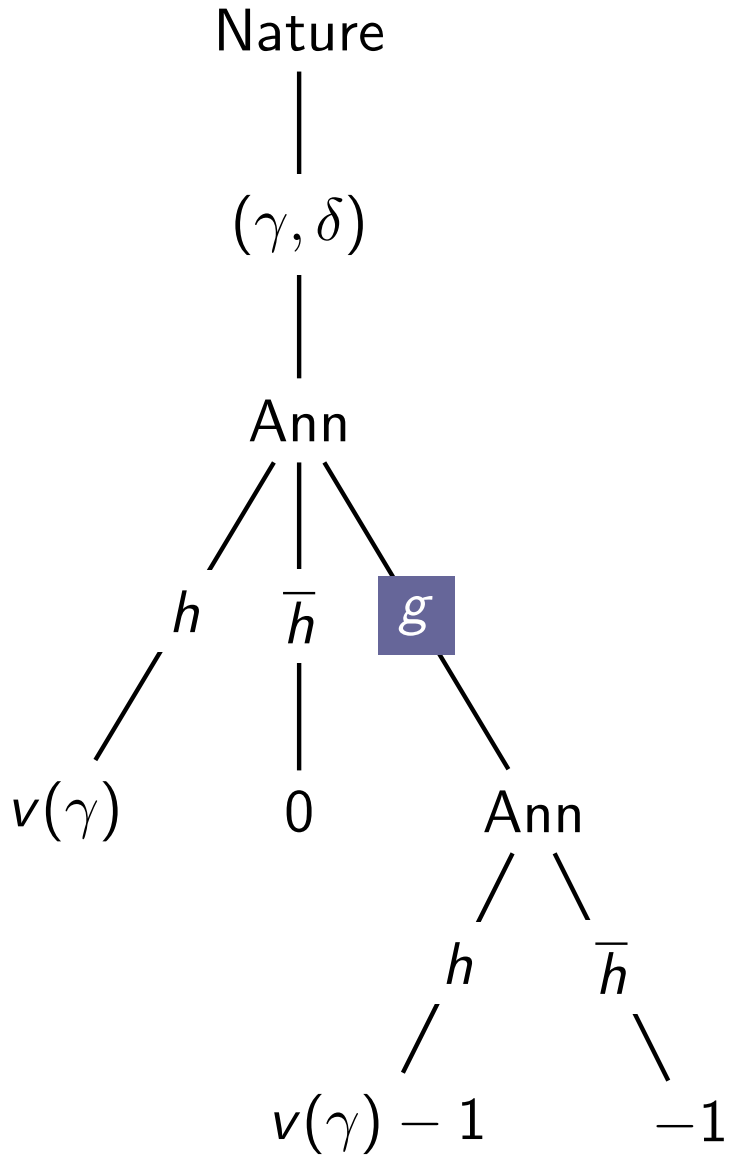
$$p(C \mid \overline{E}) = \frac{\frac{1}{10}}{\frac{5}{10}} = \frac{1}{5}$$

$$(v(c) - 1)p(C \mid \overline{E}) + (v(\overline{c}) - 1)p(\overline{C} \mid \overline{E})$$

$$EU_{\overline{E}}(h) = 4\frac{1}{5} + -6\frac{4}{5} = -4$$

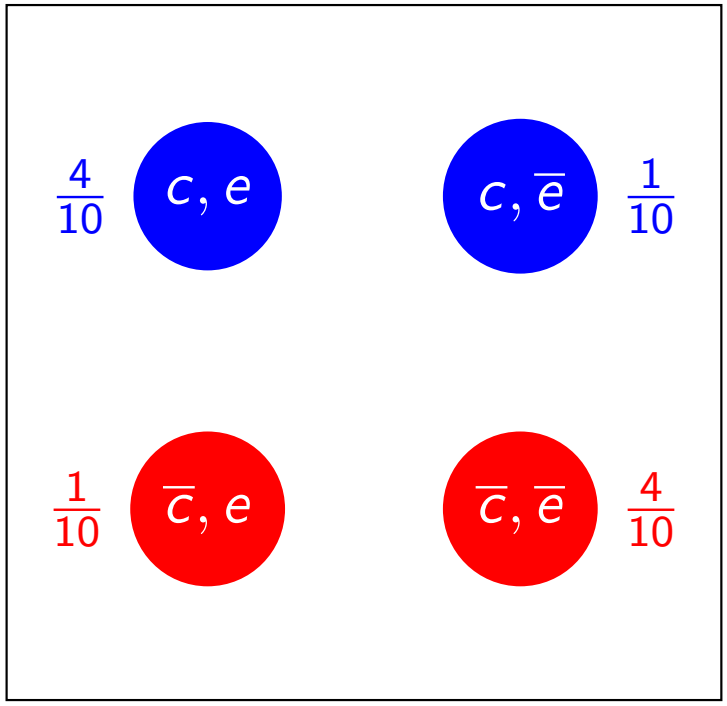
$$EU_{\overline{E}}(\overline{h}) = 4\frac{1}{5} + -6\frac{4}{5} = -1$$

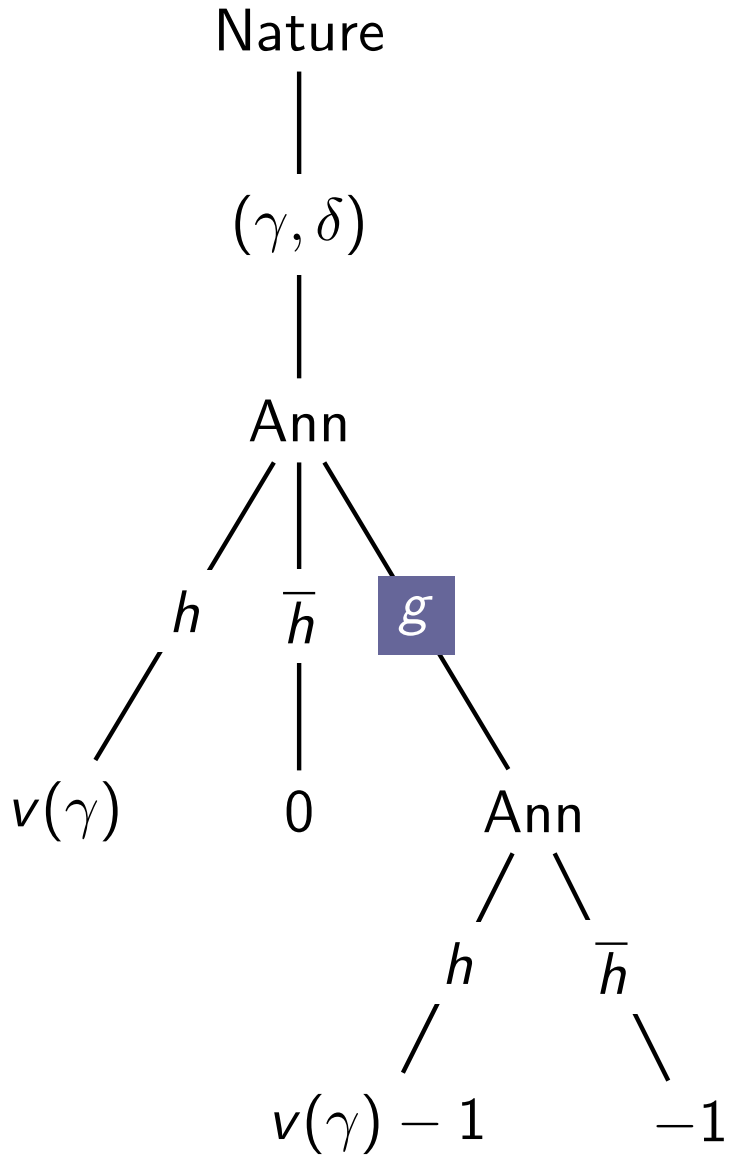




$$EU(g) =$$
$$p(E) * \max(EU_E(h), EU_E(\overline{h})) +$$
$$p(\overline{E}) * \max(EU_{\overline{E}}(h), EU_{\overline{E}}(\overline{h}))$$
$$= 0.5 * 2 + 0.5 * -1 = 0.5$$

Now suppose Ann follows her rational strategy. She writes to Bob to ask whether he has work experience. At this point, however, something surprising happens. Bob's answer reveals right from the beginning that his written English is poor. Ann notices this even before figuring out what Bob says about his work experience. In response to this unforeseen learnt input, Ann lowers her probability that Bob is competent from $\frac{1}{2}$ to $\frac{1}{8}$.





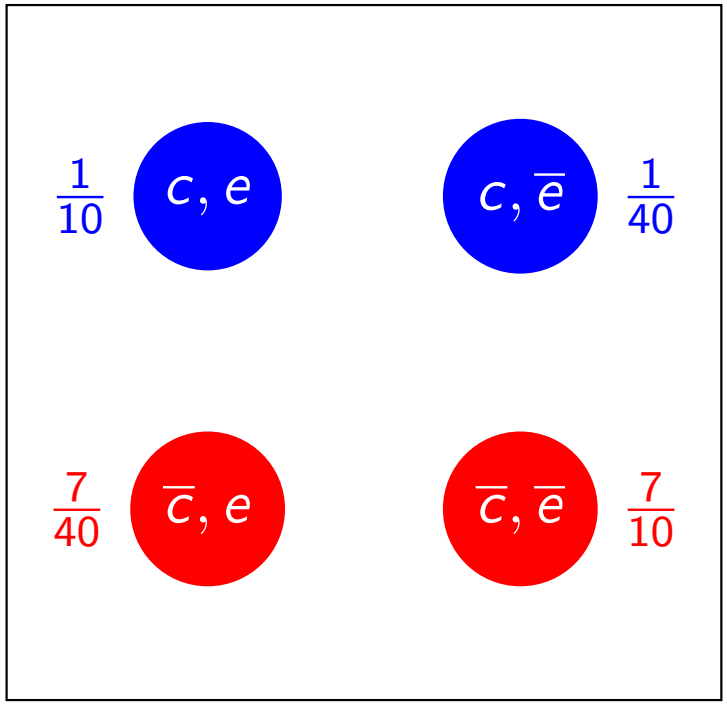
$$p_0 + (C : \tfrac{1}{8}, \overline{C} : \tfrac{7}{8}) = p_1$$

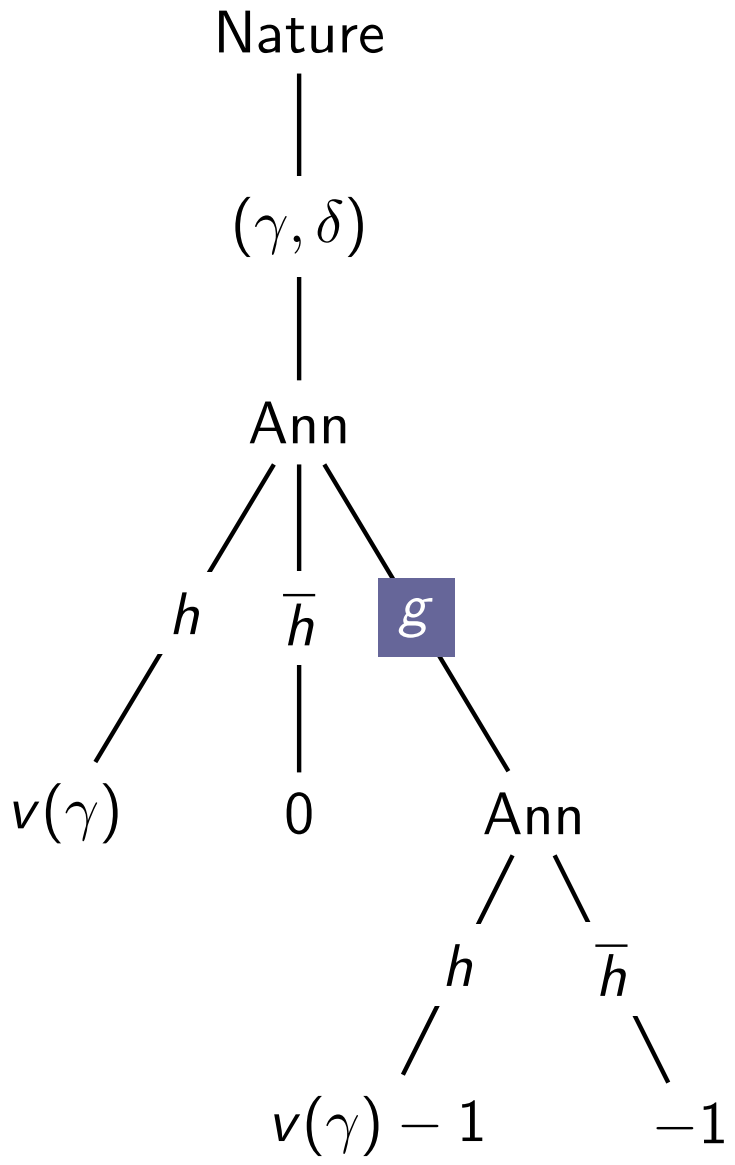
$$p_1(c, e) = \tfrac{4}{5} * \tfrac{1}{8} + 0 * \tfrac{7}{8} = \tfrac{1}{10}$$

$$p_1(c, \overline{e}) = \tfrac{1}{5} * \tfrac{1}{8} + 0 * \tfrac{7}{8} = \tfrac{1}{40}$$

$$p_1(\overline{c}, e) = 0 * \tfrac{1}{8} + \tfrac{1}{5} * \tfrac{7}{8} = \tfrac{7}{40}$$

$$p_1(\overline{c}, \overline{e}) = 0 * \tfrac{1}{8} + \tfrac{4}{5} * \tfrac{7}{8} = \tfrac{7}{10}$$





$$p_0 + (C : \tfrac{1}{8}, \overline{C} : \tfrac{7}{8}) = p_1$$

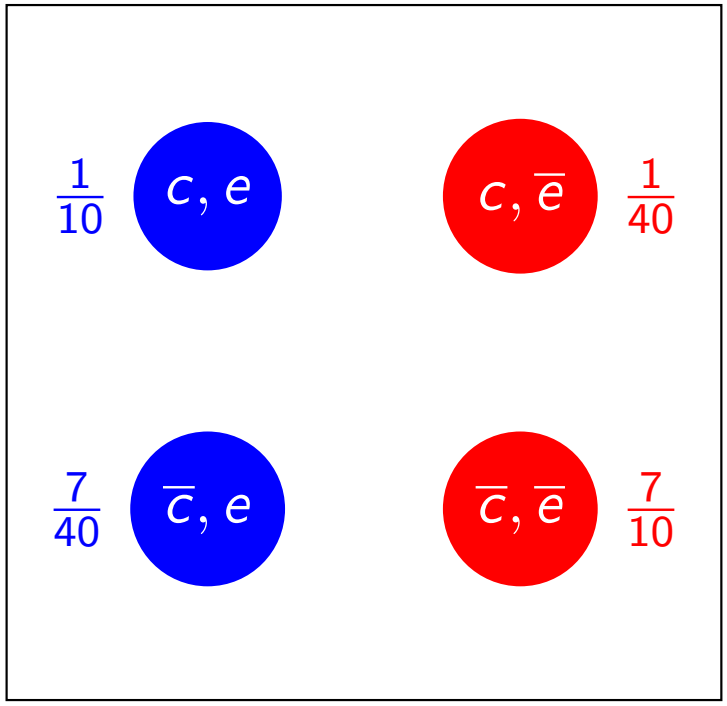
$$p_1(c, e) = \tfrac{4}{5} * \tfrac{1}{8} + 0 * \tfrac{7}{8} = \tfrac{1}{10}$$

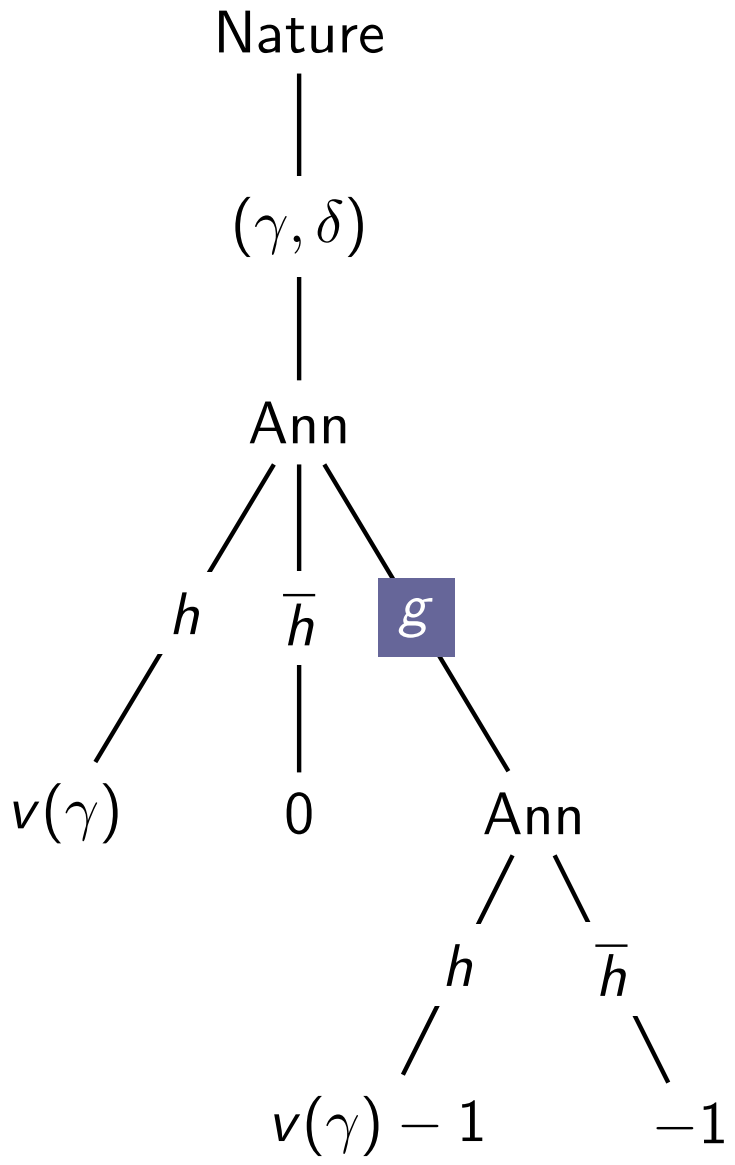
$$p_1(c, \overline{e}) = \tfrac{1}{5} * \tfrac{1}{8} + 0 * \tfrac{7}{8} = \tfrac{1}{40}$$

$$p_1(\overline{c}, e) = 0 * \tfrac{1}{8} + \tfrac{1}{5} * \tfrac{7}{8} = \tfrac{7}{40}$$

$$p_1(\overline{c}, \overline{e}) = 0 * \tfrac{1}{8} + \tfrac{4}{5} * \tfrac{7}{8} = \tfrac{7}{10}$$

As she reads the rest of Bob's letter, Ann eventually learns that he has previous work experience, which prompts a Bayesian belief revision...

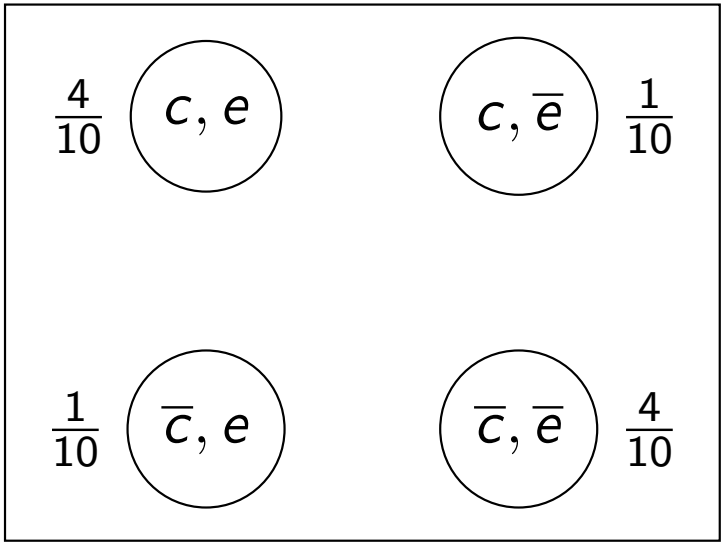


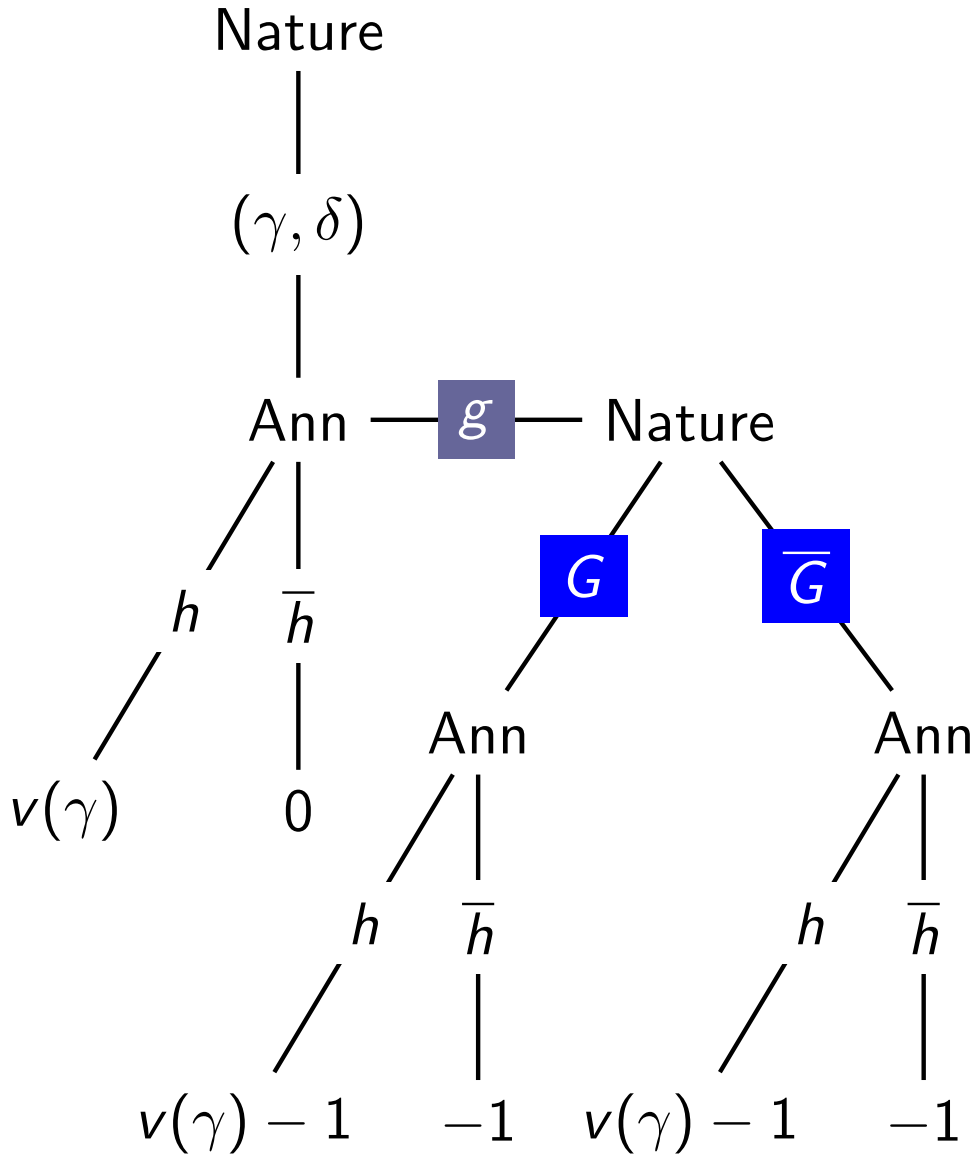


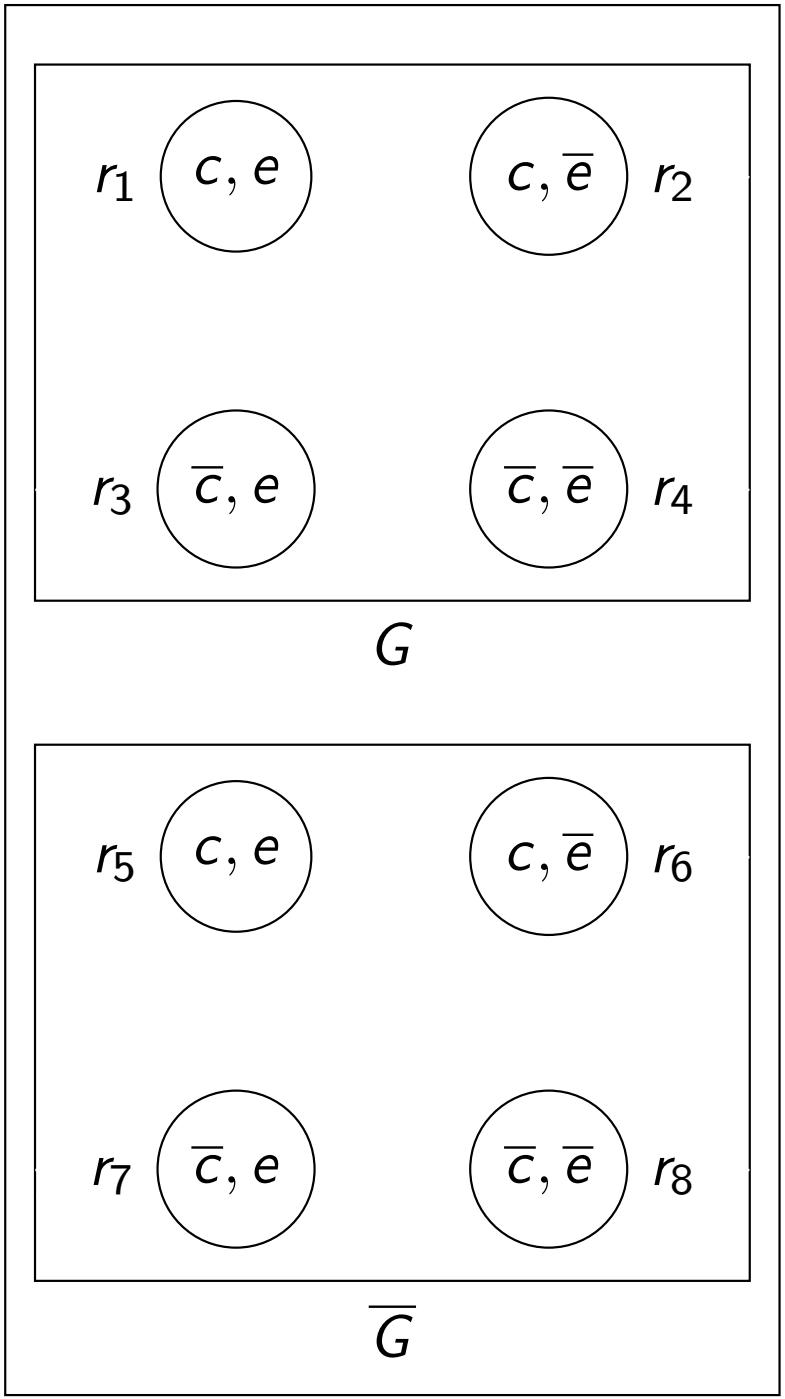
$EU_E(h) = 4\frac{4}{11} + -6\frac{7}{11} = -\frac{26}{11}$

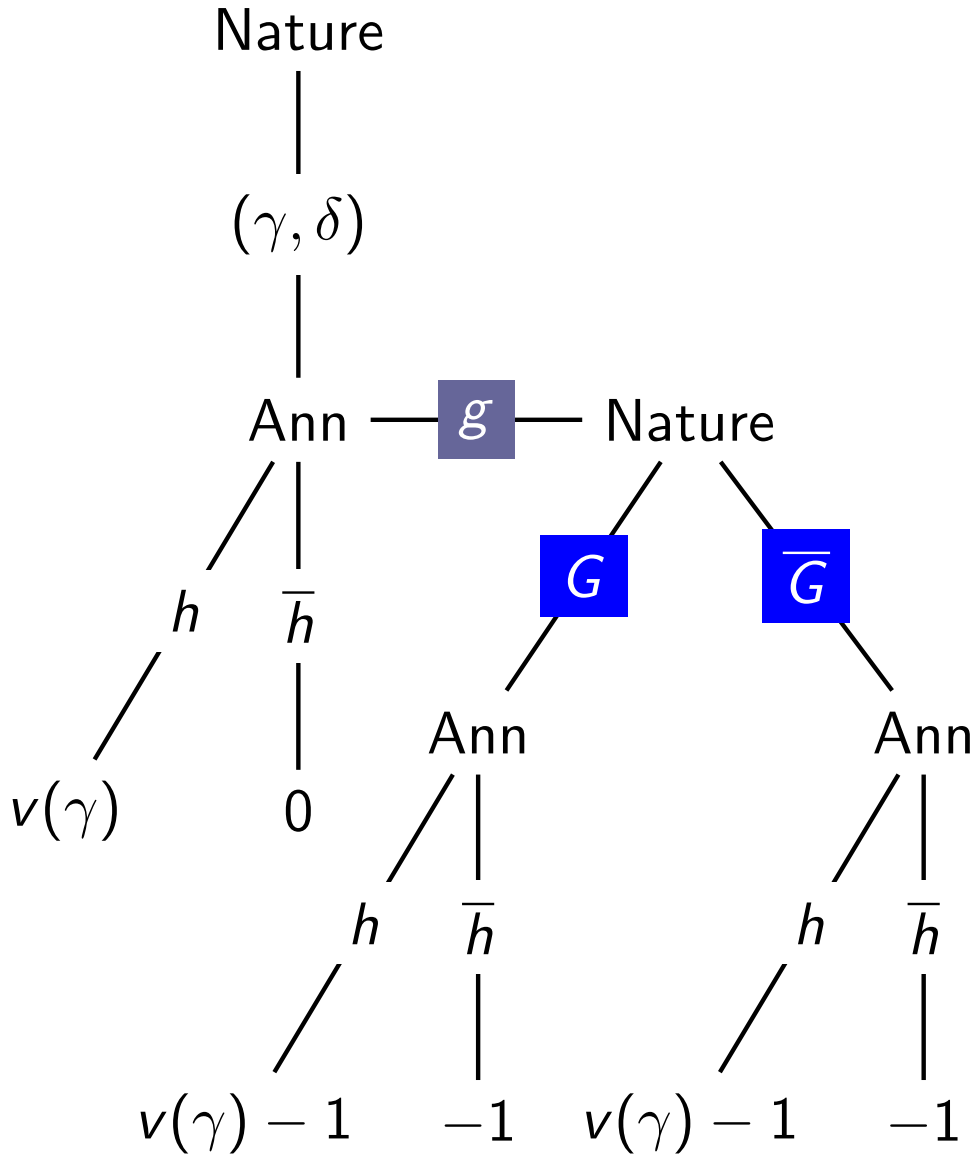
$EU_E(\overline{h}) = -1$

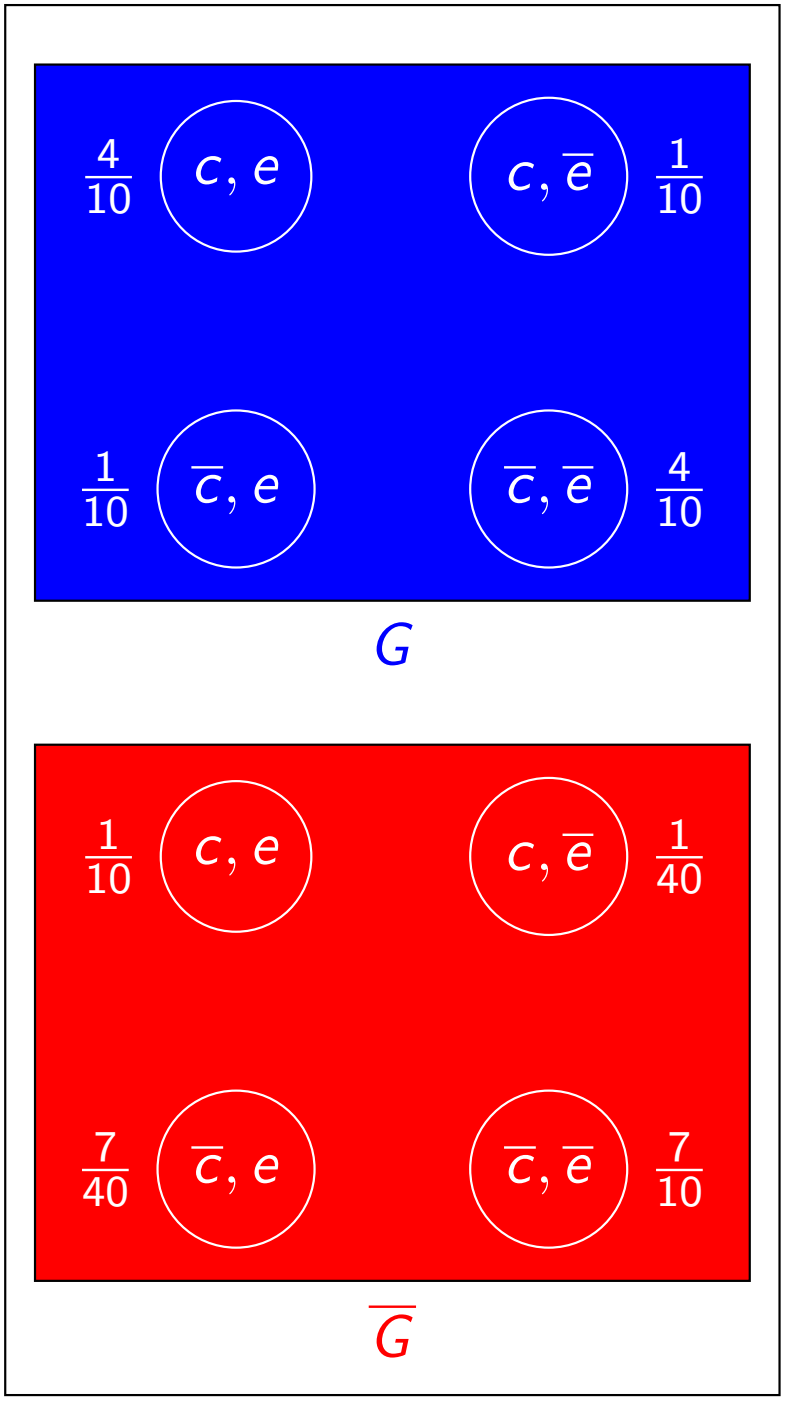
Ann does not hire Bob.

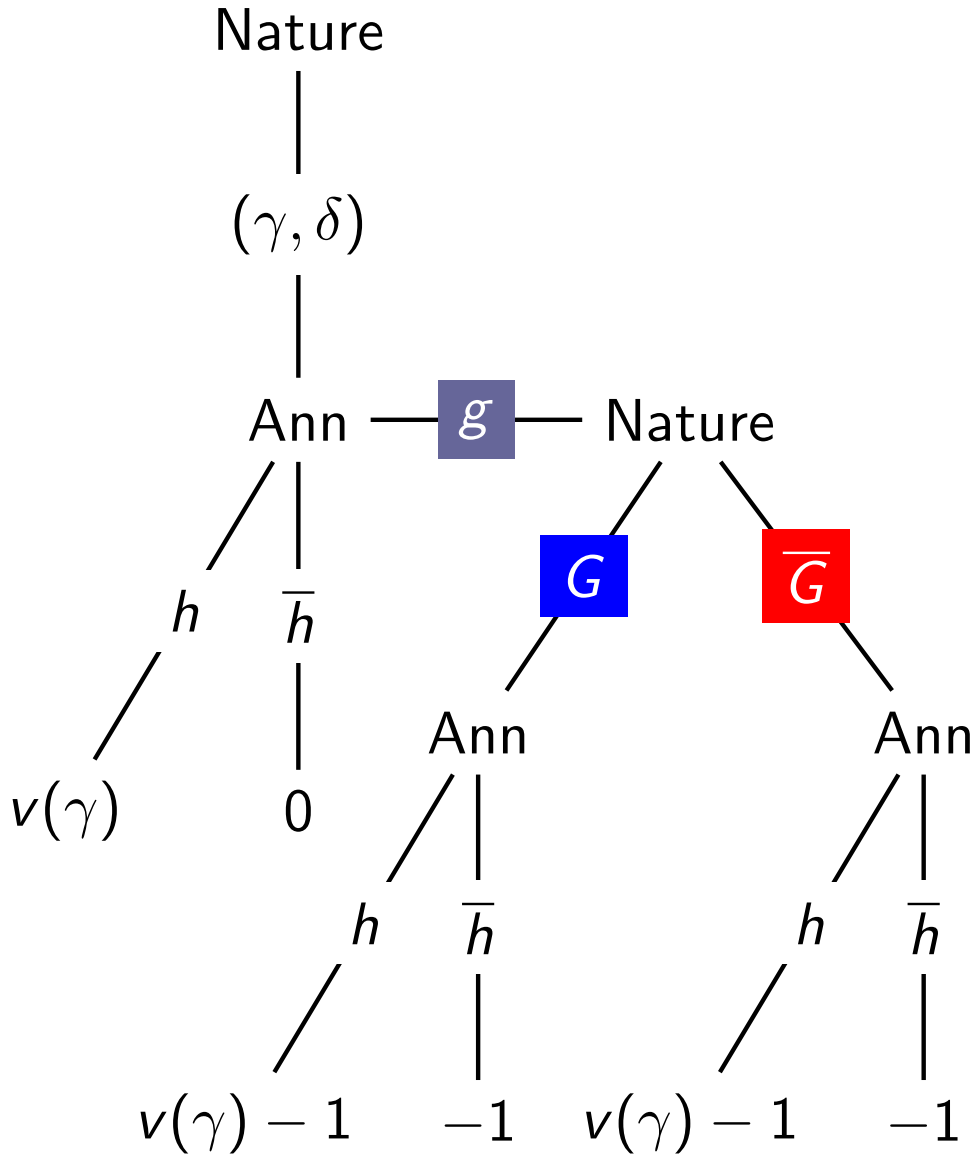

$\frac{4}{10}$
c, e
$c, \overline{e}$
$\frac{1}{10}$
$\frac{1}{10}$
$\overline{c}, e$
$\overline{c}, \overline{e}$
$\frac{4}{10}$


Nature
(γ, δ)
Ann
g
Nature
h
$\overline{h}$
G
$\overline{G}$
Ann
Ann
$v(\gamma)$
0
h
$\overline{h}$
h
$\overline{h}$
$v(\gamma) - 1$
-1
$v(\gamma) - 1$
-1


r_1
c, e
$c, \overline{e}$
r_2
r_3
$\overline{c}, e$
$\overline{c}, \overline{e}$
r_4
G
r_5
c, e
$c, \overline{e}$
r_6
r_7
$\overline{c}, e$
$\overline{c}, \overline{e}$
r_8
$\overline{G}$


Nature
(γ, δ)
Ann
g
Nature
h
$\overline{h}$
G
$\overline{G}$
$v(\gamma)$
0
Ann
Ann
h
$\overline{h}$
h
$\overline{h}$
$v(\gamma) - 1$
-1
$v(\gamma) - 1$
-1

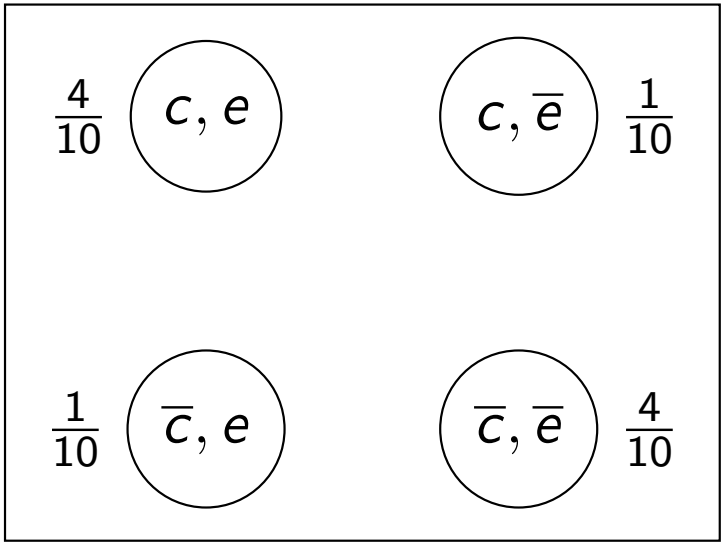

$\frac{4}{10}$
c, e
$c, \overline{e}$
$\frac{1}{10}$
$\frac{1}{10}$
$\overline{c}, e$
$\overline{c}, \overline{e}$
$\frac{4}{10}$
G
$\frac{1}{10}$
c, e
$c, \overline{e}$
$\frac{1}{40}$
$\frac{7}{40}$
$\overline{c}, e$
$\overline{c}, \overline{e}$
$\frac{7}{10}$
$\overline{G}$


Nature
(γ, δ)
Ann
g
Nature
h
$\overline{h}$
G
$\overline{G}$
$v(\gamma)$
0
Ann
Ann
h
$\overline{h}$
h
$\overline{h}$
$v(\gamma) - 1$
-1
$v(\gamma) - 1$
-1

1. In her first decision (between h, $\overline{h}$, and g), Ann is falsely taken to foresee the possibilities of learning G or learning $\overline{G}$...This artificially complicates her expected-utility maximization exercise...

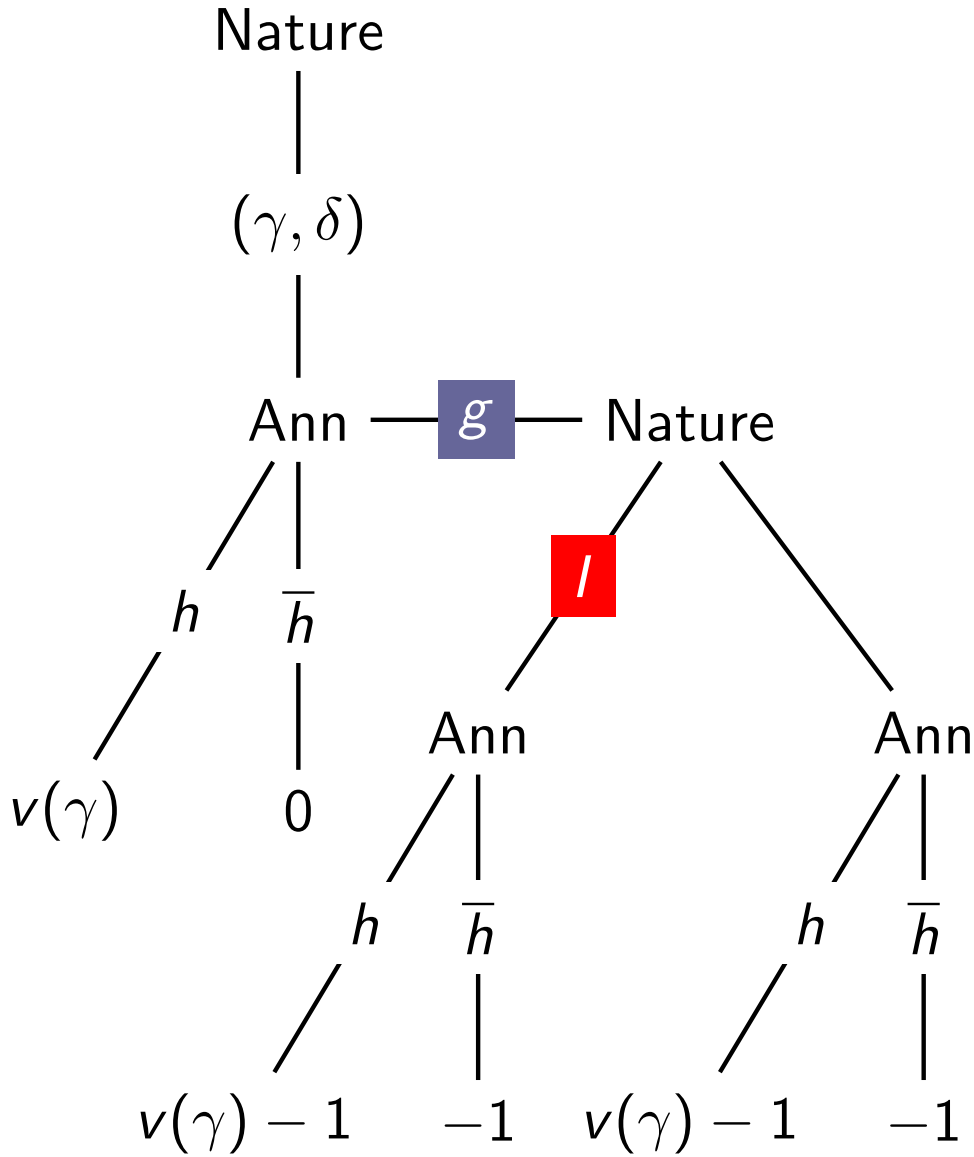
2. The additional conditionalization on G misrepresents Ann's beliefs, since the absence of linguistic errors in Bob's letter goes unnoticed.

3. Although it is true that the unforeseen news that Bob's written English is poor implies that Ann cannot uphold her original conceptualization of the decision problem, it does not follow that Ann re-conceptualizes her decision problem in line with the above model.



Nature does not reveal G or $\overline{G}$.

There is the possibility of a surprise move by Nature: Ann receives a particular unforeseen (Jeffrey) input I



Iterated revision

Current *dynamic* logics of belief revision and information update focus on two key aspects of informative actions:

Current *dynamic* logics of belief revision and information update focus on two key aspects of informative actions:

1. The agents' *observational* powers.

Current *dynamic* logics of belief revision and information update focus on two key aspects of informative actions:

1. The agents' *observational* powers.

2. The *type* of change triggered by the event.

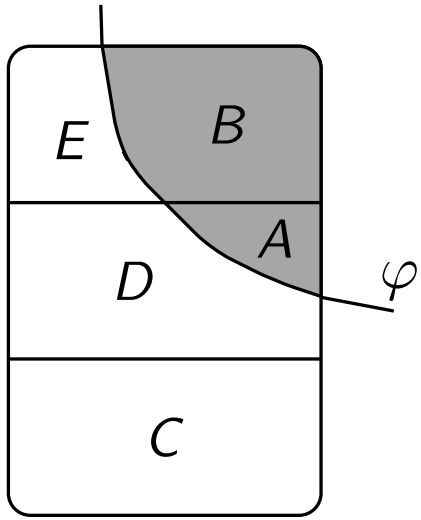
Current *dynamic* logics of belief revision and information update focus on two key aspects of informative actions:

1. The agents' *observational* powers.

2. The *type* of change triggered by the event. Agents may differ in precisely how they incorporate new information into their epistemic states. These differences are based, in part, on the agents' perception of the *source* of the information. For example, an agent may consider a particular source of information *infallible* (not allowing for the possibility that the source is mistaken) or merely *trustworthy* (accepting the information as reliable, though allowing for the possibility of a mistake).

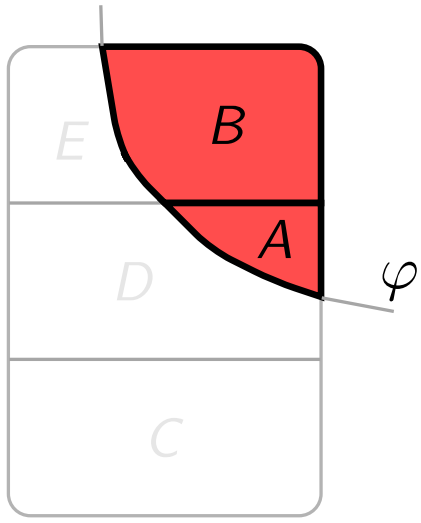
Informative Actions

Informative Actions



Incorporate the new information φ

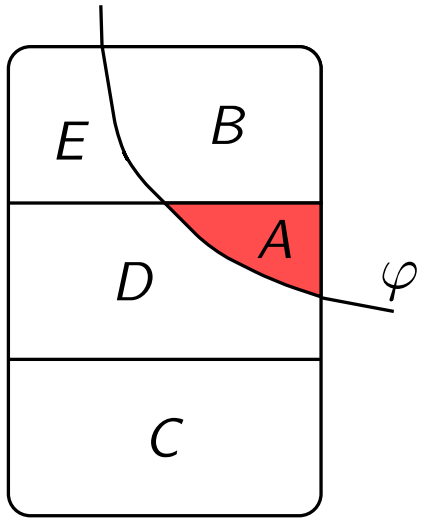
Informative Actions



Public Announcement: Information from an infallible source

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

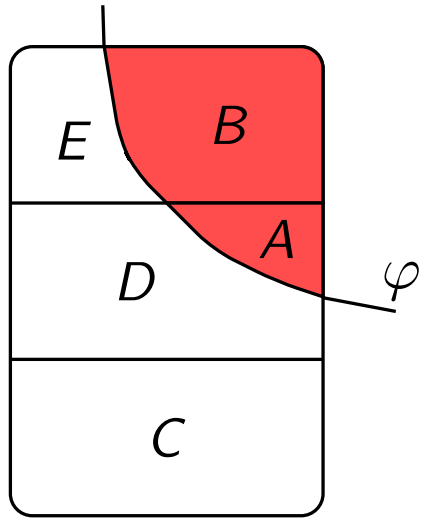
Informative Actions



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

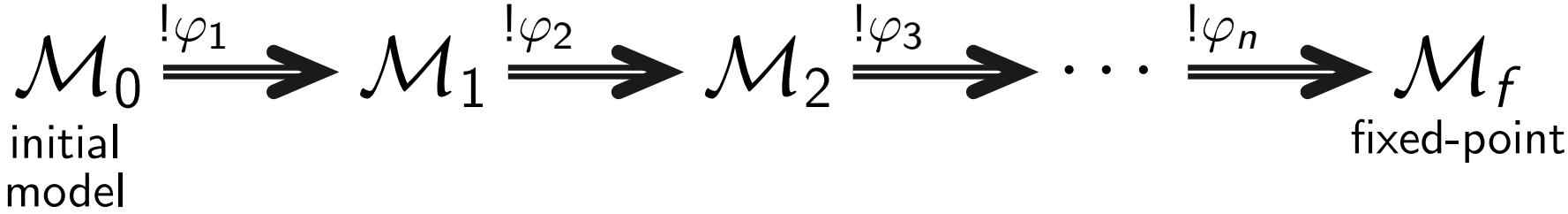


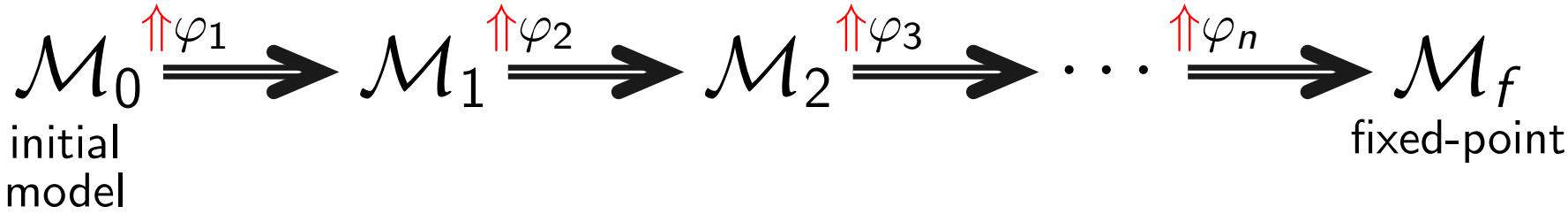
Public Announcement: Information from an infallible source
$(!\varphi)$: $A \prec_i B$

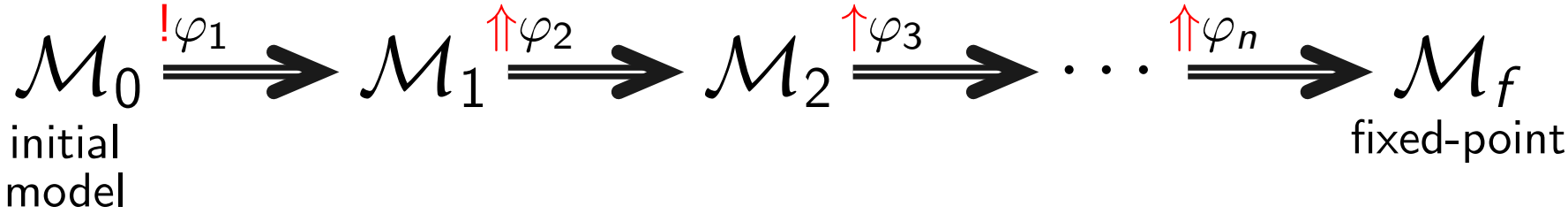
Conservative Upgrade: Information from a trusted source
$(\uparrow\varphi)$: $A \prec_i C \prec_i D \prec_i B \cup E$

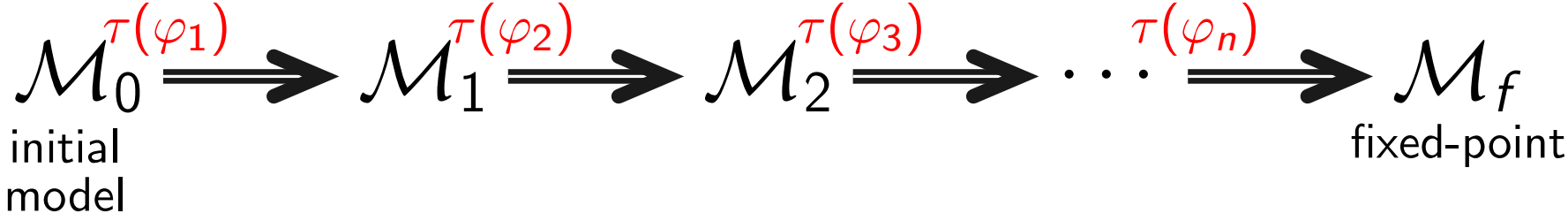
Radical Upgrade: Information from a strongly trusted source
$(\Uparrow\varphi)$: $A \prec_i B \prec_i C \prec_i D \prec_i E$

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


$\mathcal{M}_0$
$!\varphi_1$
$\mathcal{M}_1$
$!\varphi_2$
$\mathcal{M}_2$
$!\varphi_3$
$\cdots$
$!\varphi_n$
$\mathcal{M}_f$
initial model
fixed-point


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


$\mathcal{M}_0$
$!\varphi_1$
$\mathcal{M}_1$
$\Uparrow\varphi_2$
$\mathcal{M}_2$
$\uparrow\varphi_3$
$\cdots$
$\Uparrow\varphi_n$
$\mathcal{M}_f$
initial model
fixed-point



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.

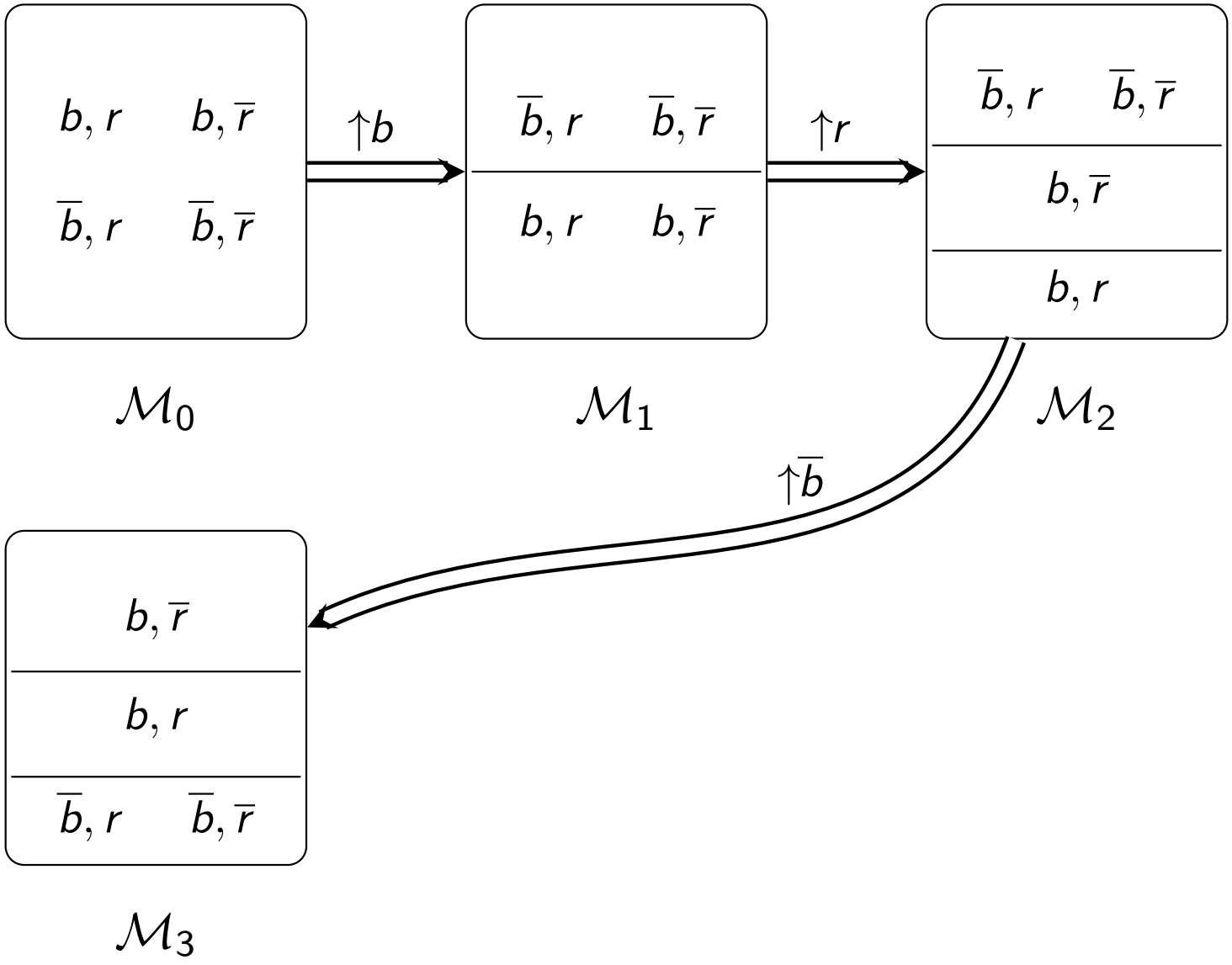
Suppose that you are in the forest and happen to a see strange-looking animal.

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.

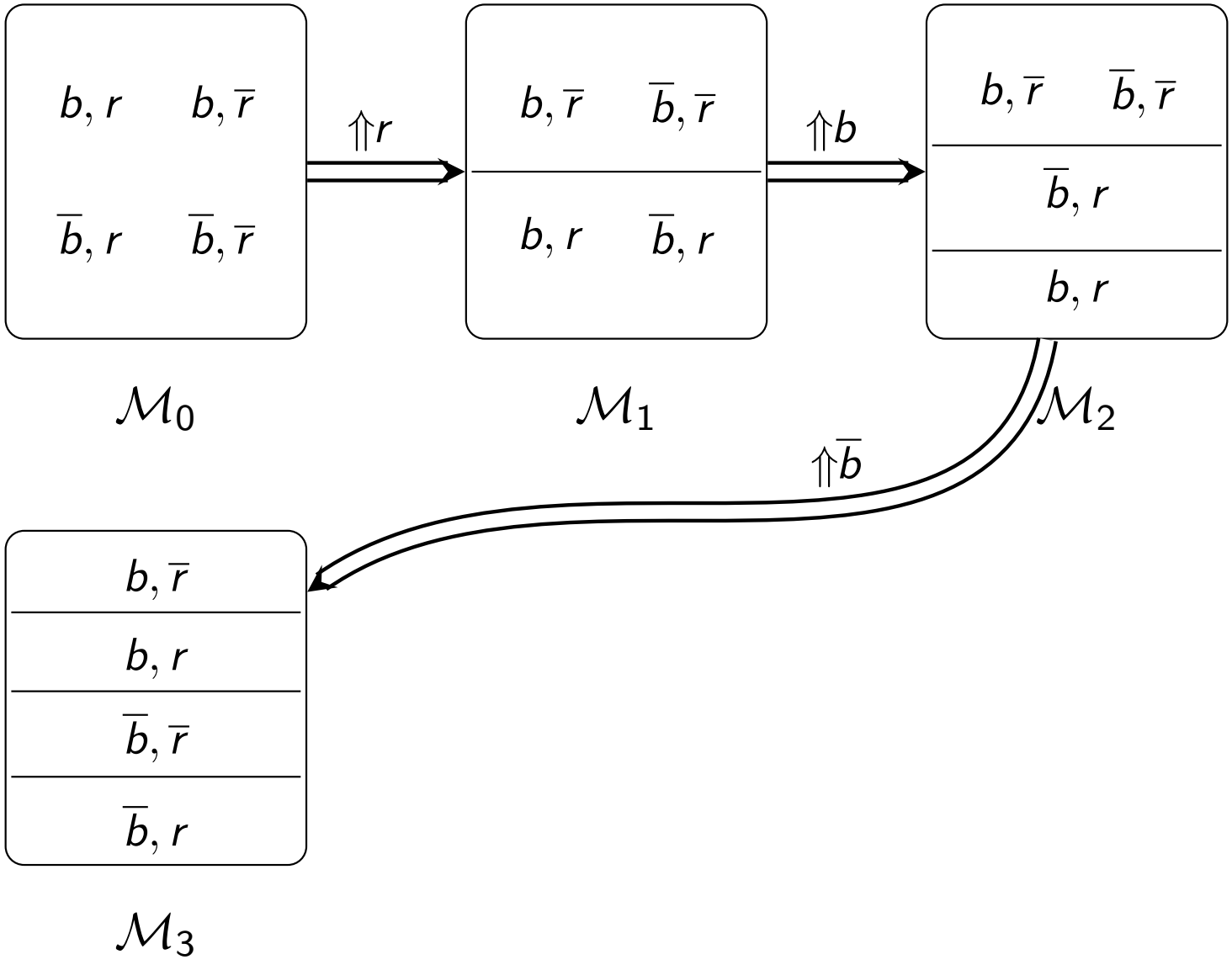
Suppose that you are in the forest and happen to a see strange-looking animal. You consult your animal guidebook and find a picture that seems to match the animal you see. The guidebook says that the animal is a type of bird, so that is what you conclude: The animal before you is a bird. After looking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

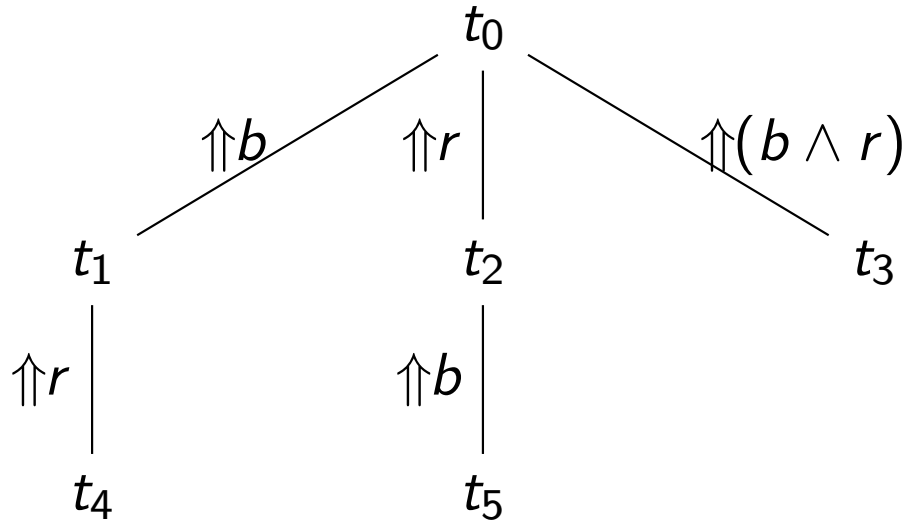

b, r $b, \overline{r}$
$\overline{b}, r$ $\overline{b}, \overline{r}$
$\uparrow b$
$\overline{b}, r$ $\overline{b}, \overline{r}$
b, r $b, \overline{r}$
$\uparrow r$
$\overline{b}, r$ $\overline{b}, \overline{r}$
$b, \overline{r}$
b, r
$\mathcal{M}_0$
$\mathcal{M}_1$
$\mathcal{M}_2$
$\uparrow \overline{b}$
$b, \overline{r}$
b, r
$\overline{b}, r$ $\overline{b}, \overline{r}$
$\mathcal{M}_3$

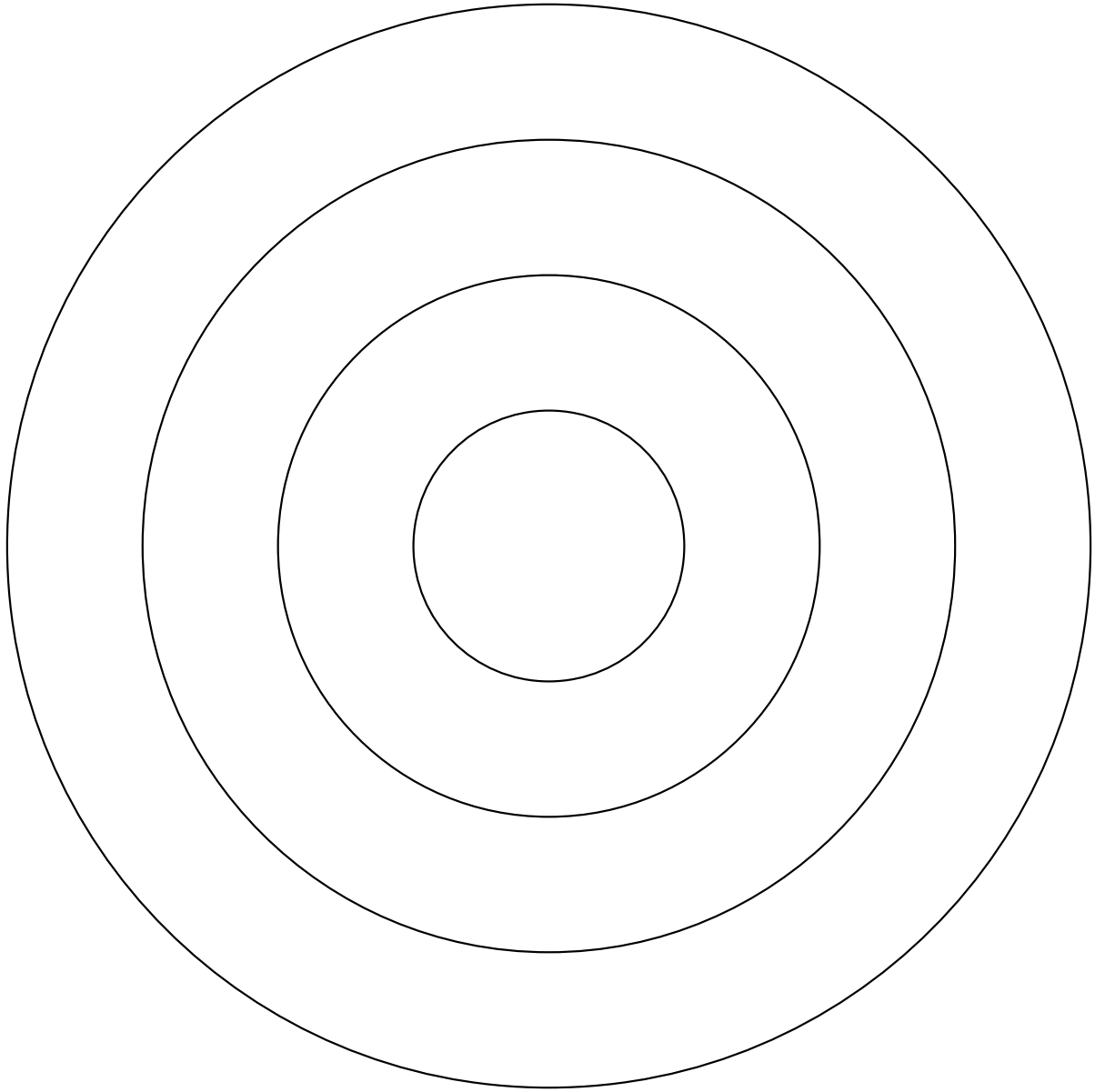
Note that in the last model, $\mathcal{M}_3$, the agent does not believe that the bird is red.

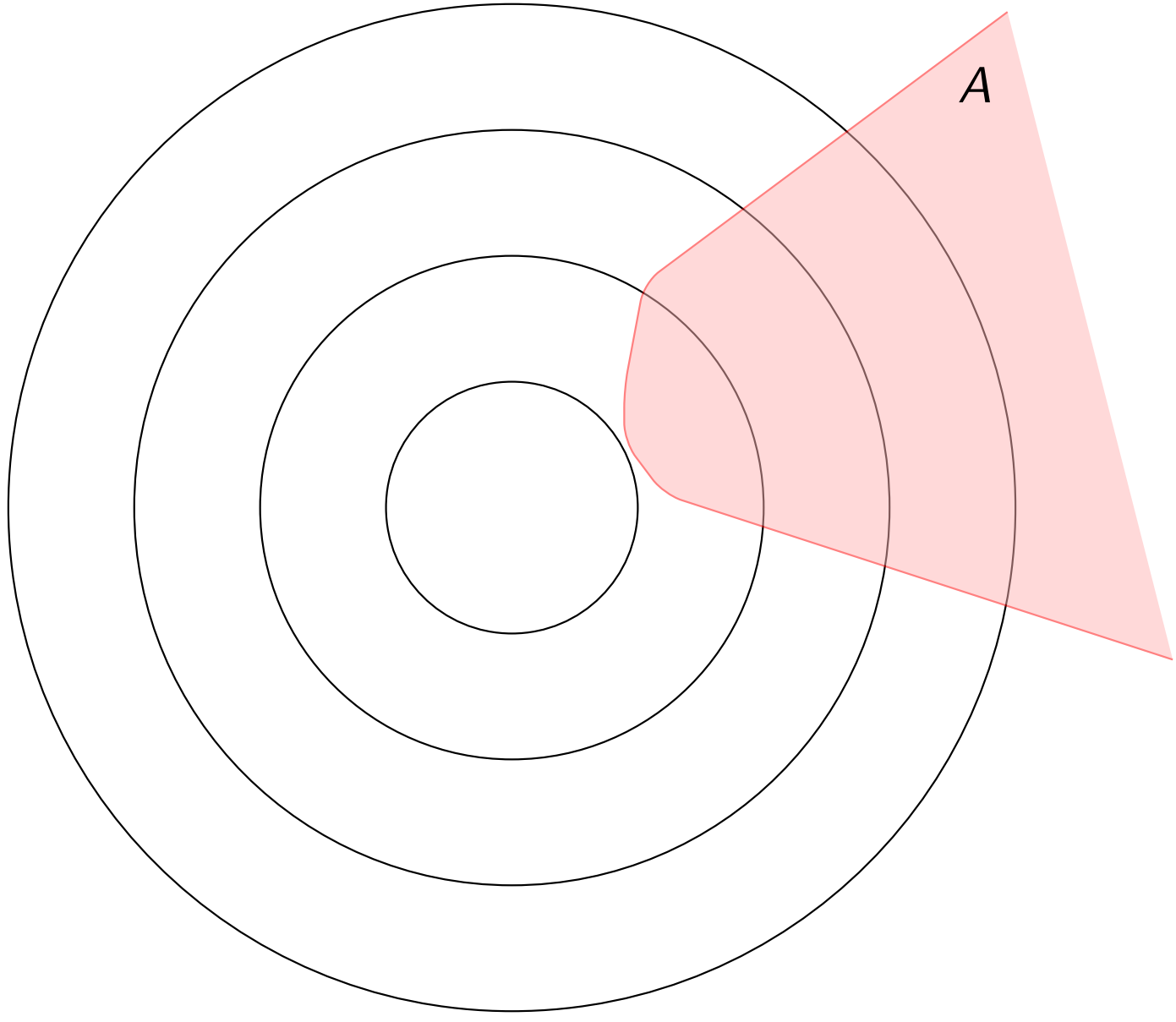
Note that in the last model, $\mathcal{M}_3$, the agent does not believe that the bird is red. The problem is that there does not seem to be any justification for why the agent drops her belief that the bird is red. This seems to result from the accidental fact that the agent started by updating with 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:

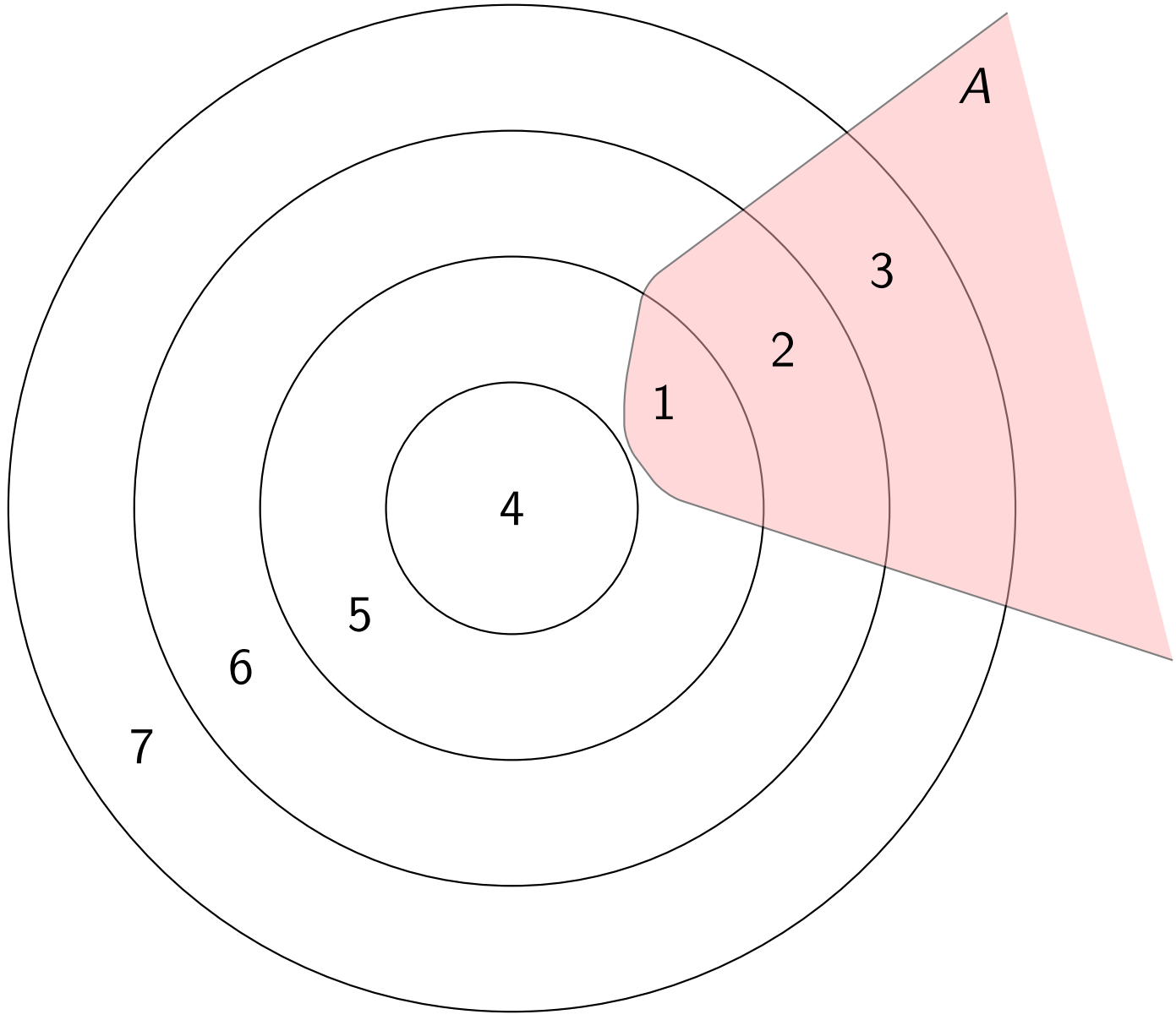

b, r $b, \overline{r}$
$\overline{b}, r$ $\overline{b}, \overline{r}$
$\Uparrow r$
$b, \overline{r}$ $\overline{b}, \overline{r}$
b, r $\overline{b}, r$
$\Uparrow b$
$b, \overline{r}$ $\overline{b}, \overline{r}$
$\overline{b}, r$
b, r
$\mathcal{M}_0$
$\mathcal{M}_1$
$\mathcal{M}_2$
$\Uparrow \overline{b}$
$b, \overline{r}$
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

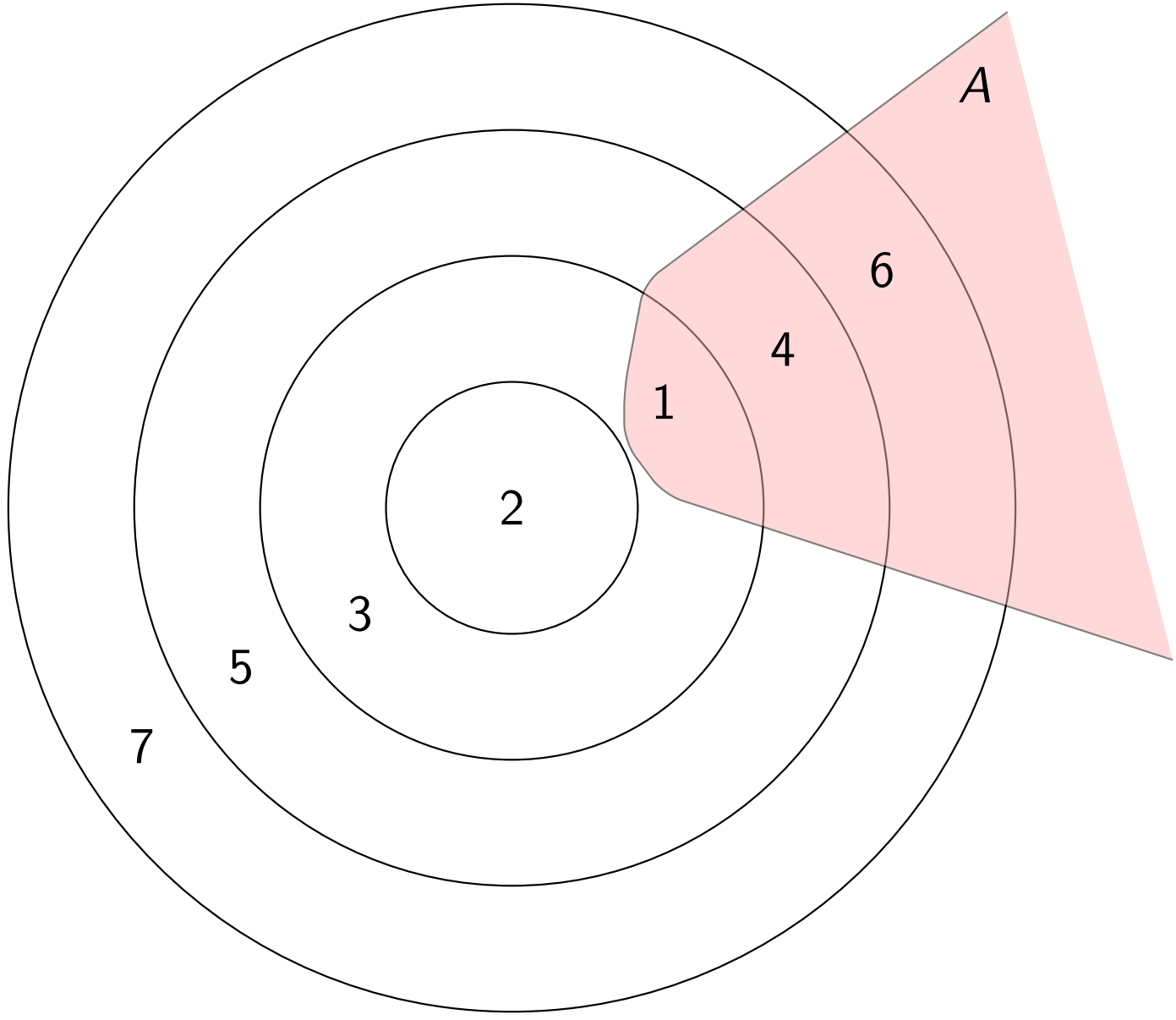



A

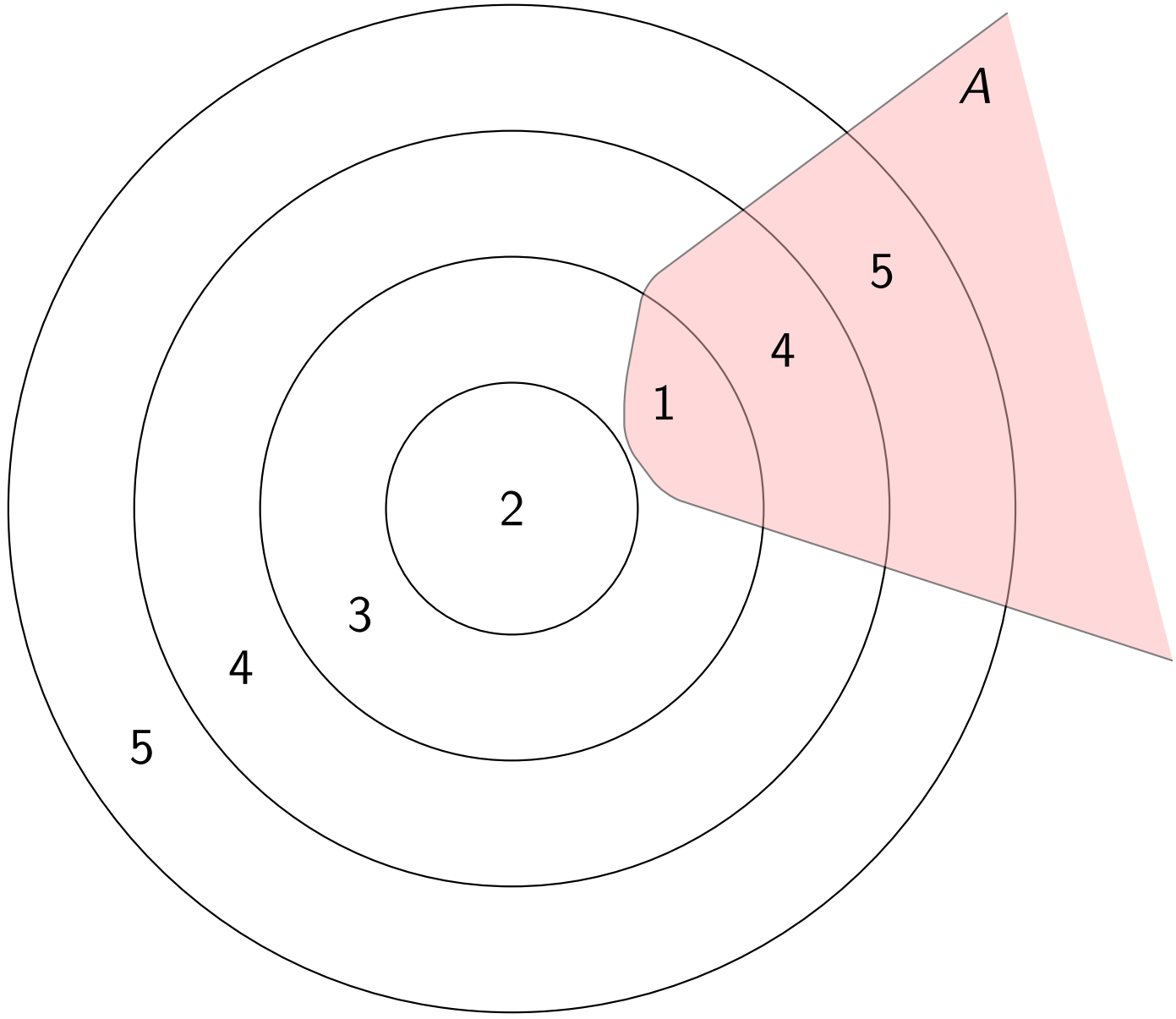
Lexicographic Revision



Restrained Revision



Natural Revision



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 $(K * A) * B = K * B$

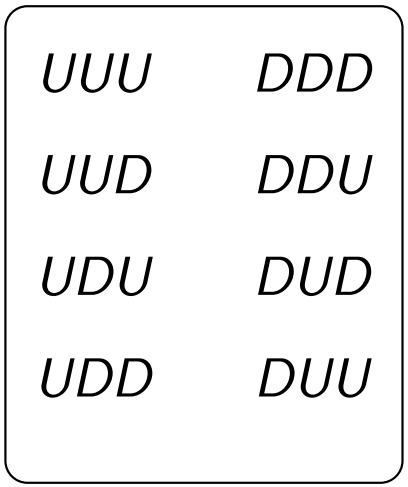
- Postulate I1 demands if $A \rightarrow B$ is a theorem (with respect to the background theory), then first learning B followed by the more specific information A is equivalent to directly learning the more specific information A.

- 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 $B \in (K * B) * A$.

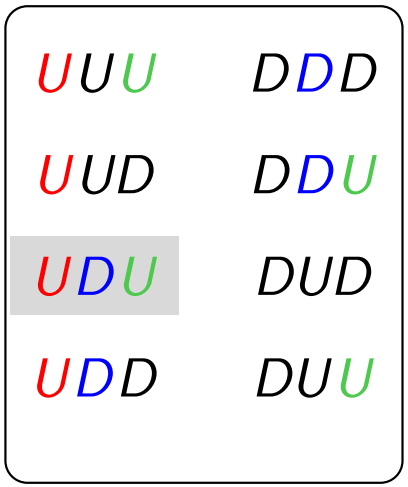
I4 If $\neg B \notin K * A$ then $\neg B \notin (K * B) * A$.

Stalnaker's Counterexample to I1



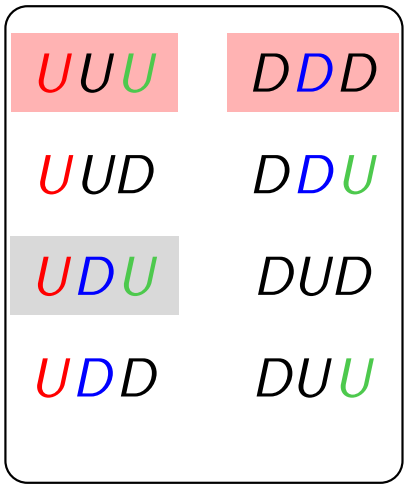
- Three switches wired such that a light is on iff all three switches are up or all three are down.

Stalnaker's Counterexample to I1



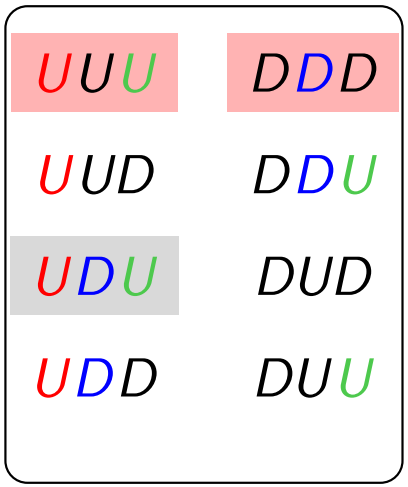
- Three switches wired such that a light is on iff all three switches are up or all three are down.
- Three independent (reliable) observers report on the switches: Alice says switch 1 is U, Bob says switch 2 is D and Carla says switch 3 is U.

Stalnaker's Counterexample to I1



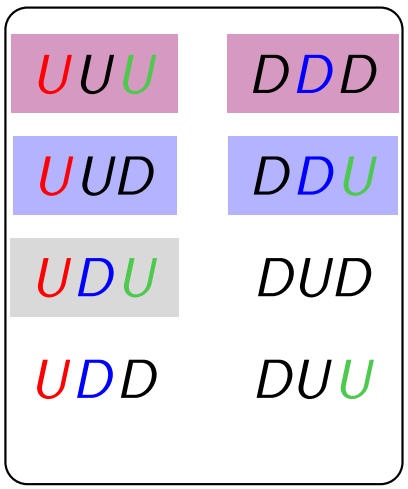
- Three switches wired such that a light is on iff all three switches are up or all three are down.
- Three independent (reliable) observers report on the switches: Alice says switch 1 is U, Bob says switch 2 is D and Carla says switch 3 is U.
- I receive the information that the light is on. What should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

Stalnaker's Counterexample to I1



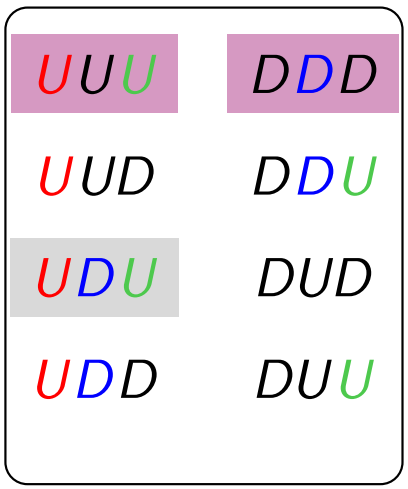
- Three switches wired such that a light is on iff all three switches are up or all three are down.
- Three independent (reliable) observers report on the switches: Alice says switch 1 is U, Bob says switch 2 is D and Carla says switch 3 is U.
- I receive the information that the light is on. What should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

Stalnaker's Counterexample to I1



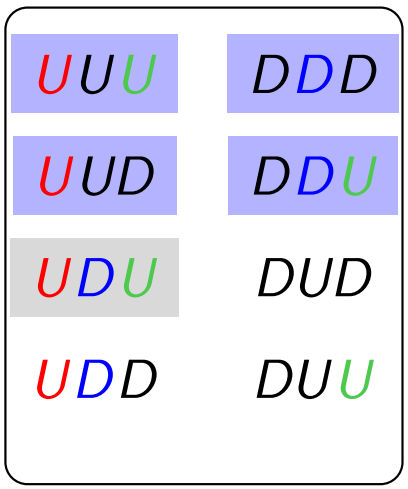
- Suppose there are two switches: L_1 is the main switch and L_2 is a secondary switch controlled by the first two lights. (So $L_1 \rightarrow L_2$, but not the converse)

Stalnaker's Counterexample to I1



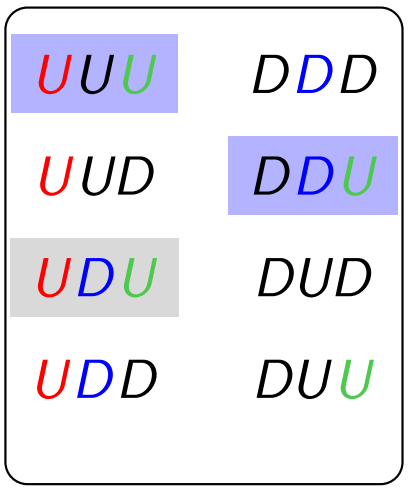
- Suppose there are two switches: L_1 is the main switch and L_2 is a secondary switch controlled by the first two lights. (So $L_1 \rightarrow L_2$, but not the converse)
- Suppose I receive $L_1 \wedge L_2$, this does not change the story.

Stalnaker's Counterexample to I1



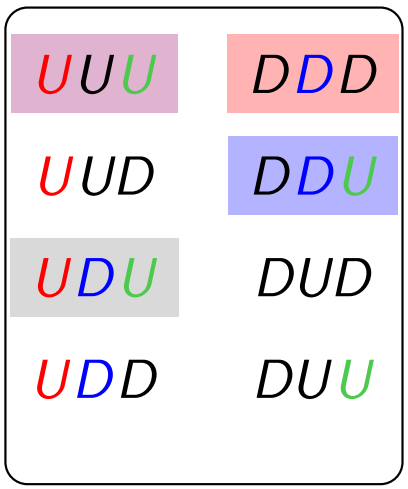
- Suppose there are two switches: L_1 is the main switch and L_2 is a secondary switch controlled by the first two lights. (So $L_1 \rightarrow L_2$, but not the converse)
- Suppose I receive $L_1 \wedge L_2$, this does not change the story.
- Suppose I learn that L_2. This is irrelevant to Carla's report,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

Stalnaker's Counterexample to I1



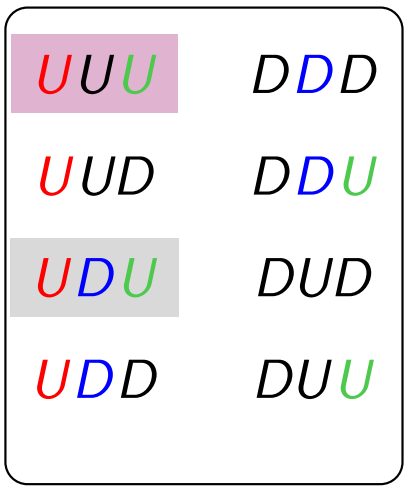
- Suppose there are two switches: L_1 is the main switch and L_2 is a secondary switch controlled by the first two lights. (So $L_1 \rightarrow L_2$, but not the converse)
- Suppose I receive $L_1 \wedge L_2$, this does not change the story.
- Suppose I learn that L_2. This is irrelevant to Carla's report,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

Stalnaker's Counterexample to I1



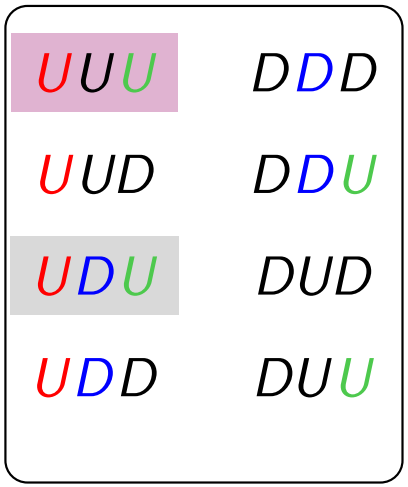
- Suppose there are two switches: L_1 is the main switch and L_2 is a secondary switch controlled by the first two lights. (So $L_1 \rightarrow L_2$, but not the converse)
- Suppose I receive $L_1 \wedge L_2$, this does not change the story.
- Suppose I learn that L_2. This is irrelevant to Carla's report,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



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

Stalnaker's Counterexample to I1



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

Stalnaker's Counterexample to I2

- Two fair coins are flipped and placed in two boxes and two independent and reliable observers deliver reports about the status (heads up or tails up) of the coins in the opaque boxes.

Stalnaker's Counterexample to I2

- Two fair coins are flipped and placed in two boxes and two independent and reliable observers deliver reports about the status (heads up or tails up) of the coins in the opaque boxes.

- Alice reports that the coin in box 1 is lying heads up, Bert reports that the coin in box 2 is lying heads up.

Stalnaker's Counterexample to I2

- Two fair coins are flipped and placed in two boxes and two independent and reliable observers deliver reports about the status (heads up or tails up) of the coins in the opaque boxes.
- Alice reports that the coin in box 1 is lying heads up, Bert reports that the coin in box 2 is lying heads up.
- Two new independent witnesses, whose reliability trumps that of Alice's and Bert's, provide additional reports on the status of the coins. Carla reports that the coin in box 1 is lying tails up, and Dora reports that the coin in box 2 is lying tails up.

Stalnaker's Counterexample to I2

- Two fair coins are flipped and placed in two boxes and two independent and reliable observers deliver reports about the status (heads up or tails up) of the coins in the opaque boxes.
- Alice reports that the coin in box 1 is lying heads up, Bert reports that the coin in box 2 is lying heads up.
- Two new independent witnesses, whose reliability trumps that of Alice's and Bert's, provide additional reports on the status of the coins. Carla reports that the coin in box 1 is lying tails up, and Dora reports that the coin in box 2 is lying tails up.
- Finally, Elmer, a third witness considered the most reliable overall, reports that the coin in box 1 is lying heads up.

H_i (T_i): the coin in box i facing heads (tails) up.

H_i (T_i): the coin in box i facing heads (tails) up.

- The first revision results in the belief set $K' = K * (H_1 \wedge H_2)$, where K is the agents original set of beliefs.

H_i (T_i): the coin in box i facing heads (tails) up.

- The first revision results in the belief set $K' = K * (H_1 \wedge H_2)$, where K is the agents original set of beliefs.
- After receiving the reports, the belief set is $K' * (T_1 \wedge T_2) * H_1$.

H_i (T_i): the coin in box i facing heads (tails) up.

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

H_i (T_i): the coin in box i facing heads (tails) up.

- The first revision results in the belief set $K' = K * (H_1 \wedge H_2)$, where K is the agents original set of beliefs.
- After receiving the reports, the belief set is $K' * (T_1 \wedge T_2) * H_1$.
- Since Elmers report is irrelevant to the status of the coin in box 2, it seems natural to assume that $H_1 \wedge T_2 \in K' * (T_1 \wedge T_2) * H_1$.
- The problem: Since $(T_1 \wedge T_2) \rightarrow \neg H_1$ is a theorem (given the background theory), by I2 it follows that $K' * (T_1 \wedge T_2) * H_1 = K' * H_1$.

H_i (T_i): the coin in box i facing heads (tails) up.

- The first revision results in the belief set $K' = K * (H_1 \wedge H_2)$, where K is the agents original set of beliefs.
- After receiving the reports, the belief set is $K' * (T_1 \wedge T_2) * H_1$.
- Since Elmers report is irrelevant to the status of the coin in box 2, it seems natural to assume that $H_1 \wedge T_2 \in K' * (T_1 \wedge T_2) * H_1$.
- The problem: Since $(T_1 \wedge T_2) \rightarrow \neg H_1$ is a theorem (given the background theory), by I2 it follows that $K' * (T_1 \wedge T_2) * H_1 = K' * H_1$.

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

...[Postulate I2] directs us to take back the totality of any information that is overturned. Specifically, if we first receive information α, and then receive information that conflicts with α, we should return to the belief state we were previously in, before learning α. But this directive is too strong. Even if the new information conflicts with the information just received, it need not necessarily cast doubt on all of that information.

(Stalnaker, pg. 207–208)

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?

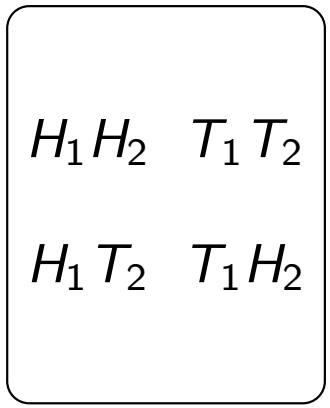
1. There is no suitable way to formalize the scenario in such a way that the AGM postulates (possibly including postulates of iterated belief revision) can be saved;

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

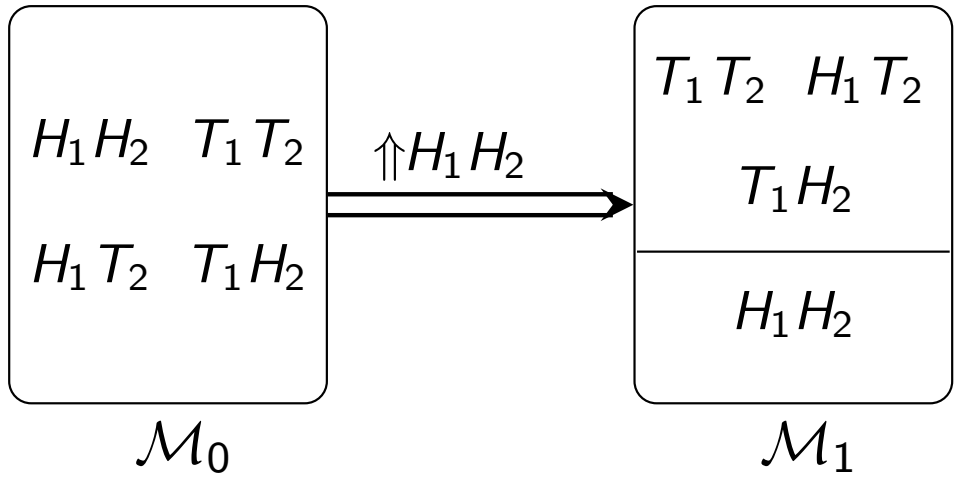
3. There is a suitable and systematic way to make the meta-information explicit, but this is something that the AGM framework cannot properly accommodate.

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

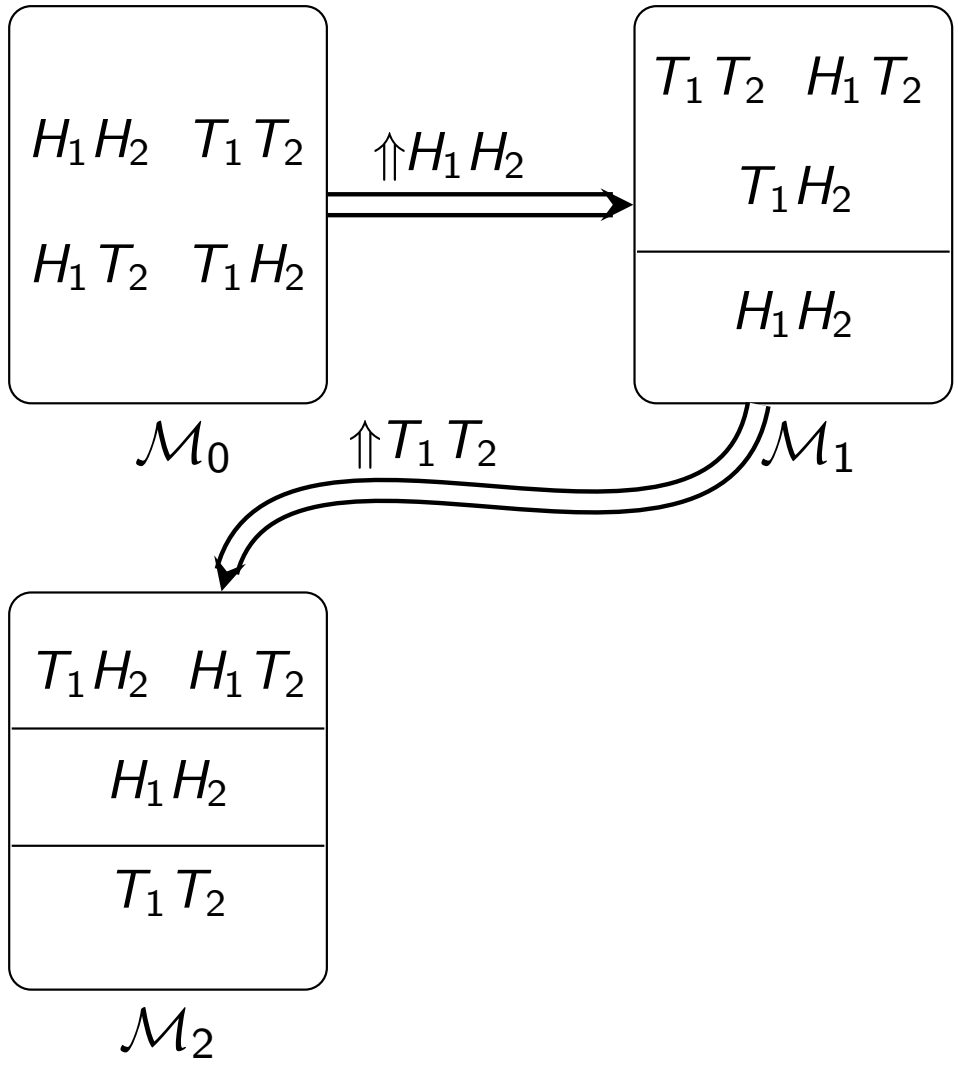
Heuristic Diagnosis of Stalnaker's Example



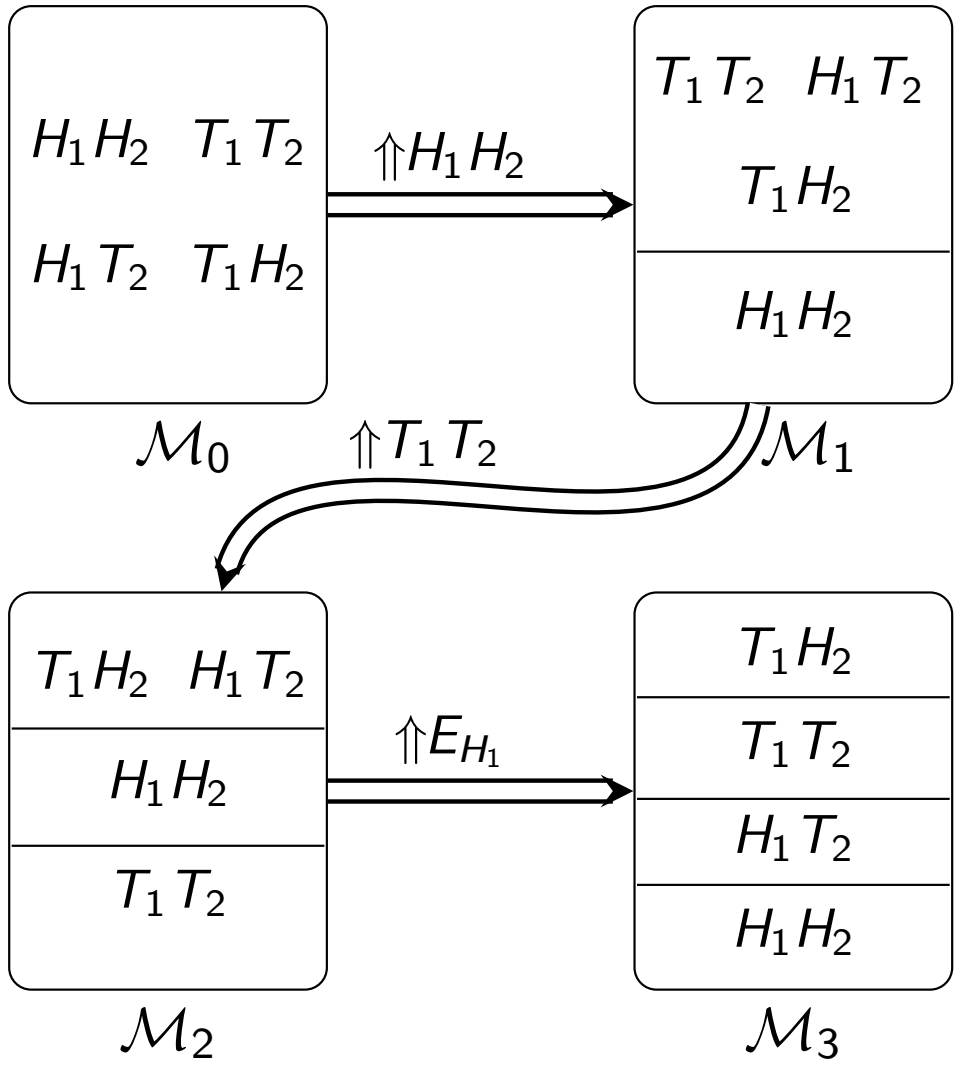
Heuristic Diagnosis of Stalnaker's Example



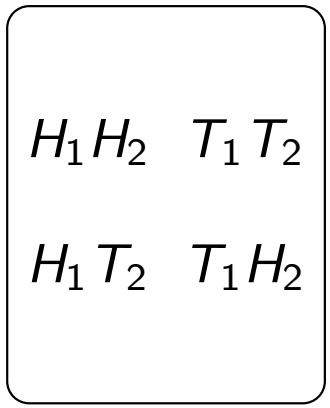
Heuristic Diagnosis of Stalnaker's Example



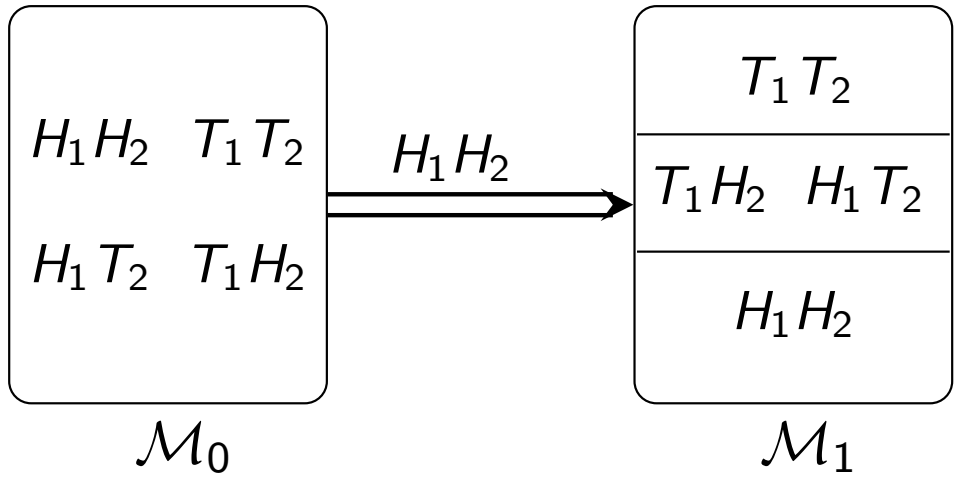
Heuristic Diagnosis of Stalnaker's Example



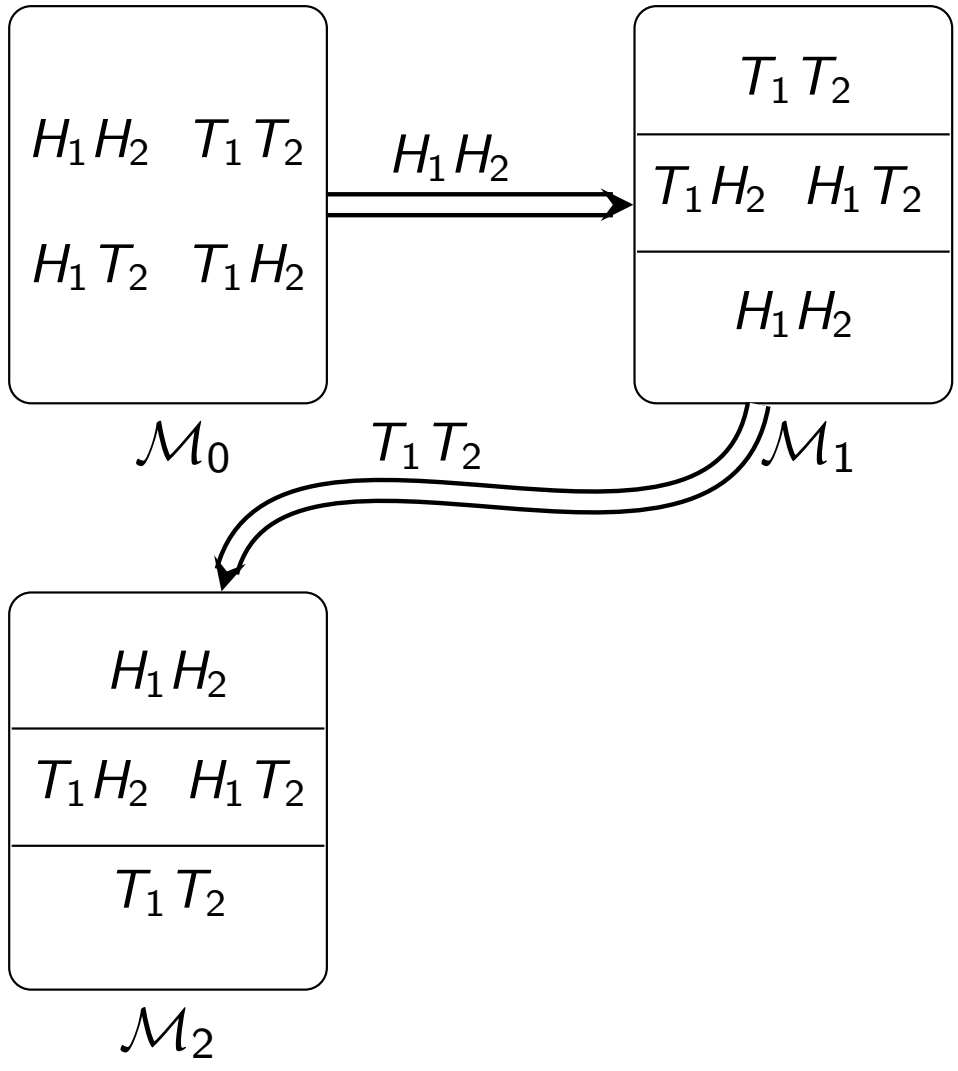
Heuristic Diagnosis of Stalnaker's Example



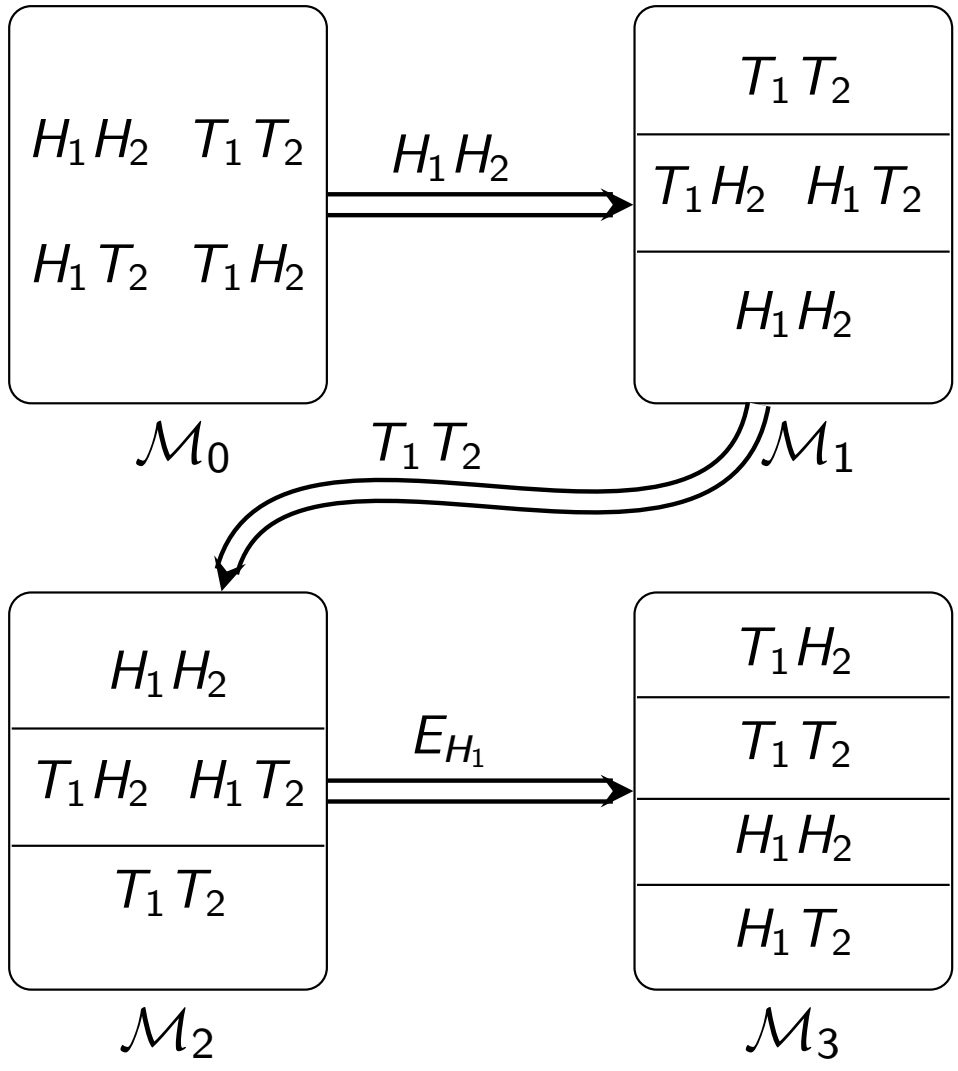
Heuristic Diagnosis of Stalnaker's Example



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

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

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.

Irreducibility to Single Revision

R. Booth and J. Chandler. *The Irreducibility of Iterated to Single Revision*. Journal of Philosophical Logic, forthcoming.

Knowledge and Issues

J. van Benthem and S. Minica. *Toward a Dynamic Logic of Questions*. Journal of Philosophical Logic, 41(4), pp 633 - 669, 2012.

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

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

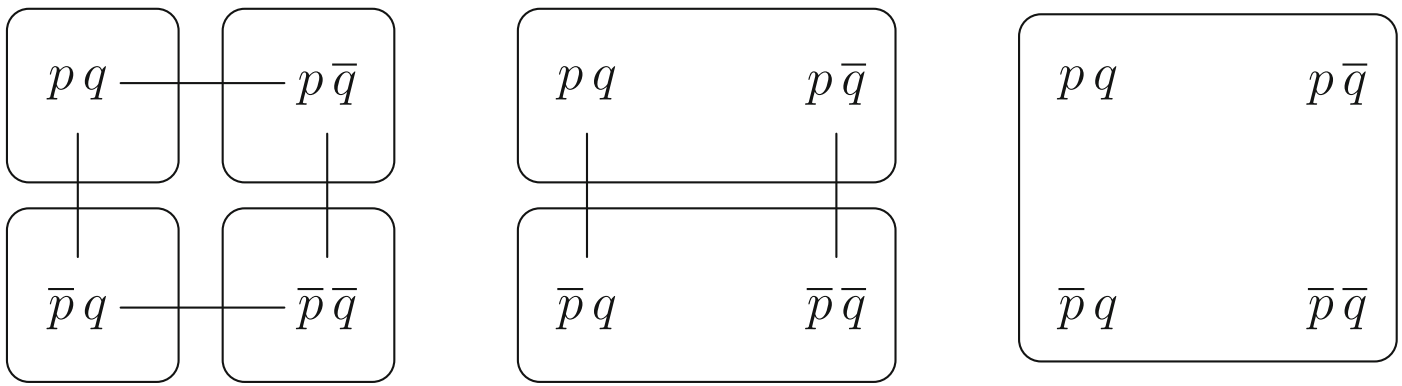


Fig. 1 Examples of epistemic issue models

- $\mathcal{M}, w \models K\varphi$ iff for all v, if $w \sim v$, then $\mathcal{M}, v \models \varphi$
- $\mathcal{M}, w \models Q\varphi$ iff for all v, if $w \approx v$, then $\mathcal{M}, v \models \varphi$
- $\mathcal{M}, w \models R\varphi$ iff for all v, if $w \ (\sim \cap \approx) \ v$, then $\mathcal{M}, v \models \varphi$

- $\mathcal{M}, w \models K\varphi$ iff for all v, if $w \sim v$, then $\mathcal{M}, v \models \varphi$
- $\mathcal{M}, w \models Q\varphi$ iff for all v, if $w \approx v$, then $\mathcal{M}, v \models \varphi$
- $\mathcal{M}, w \models R\varphi$ iff for all v, if $w \ (\sim \cap \approx) \ v$, then $\mathcal{M}, v \models \varphi$

Knowledge dynamics, issue dynamics

Coarsening at Random

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

Three Prisoner's Problem

Three prisoners A, B and C have been tried for murder and their verdicts will told to them tomorrow morning. They know only that one of them will be declared guilty and will be executed while the others will be set free. The identity of the condemned prisoner is revealed to the very reliable prison guard, but not to the prisoners themselves. Prisoner A asks the guard "Please give this letter to one of my friends — to the one who is to be released. We both know that at least one of them will be released".

Three Prisoner's Problem

An hour later, A asks the guard "Can you tell me which of my friends you gave the letter to? It should give me no clue regarding my own status because, regardless of my fate, each of my friends had an equal chance of receiving my letter." The guard told him that B received his letter.

Prisoner A then concluded that the probability that he will be released is $1/2$ (since the only people without a verdict are A and C).

Three Prisoner's Problem

But, A thinks to himself:

Three Prisoner's Problem

But, *A* thinks to himself:

> *Before I talked to the guard my chance of being executed was 1 in 3. Now that he told me B has been released, only C and I remain, so my chances of being executed have gone from 33.33% to 50%. What happened? I made certain not to ask for any information relevant to my own fate...*

Three Prisoner's Problem

But, A thinks to himself:

> *Before I talked to the guard my chance of being executed was 1 in 3. Now that he told me B has been released, only C and I remain, so my chances of being executed have gone from 33.33% to 50%. What happened? I made certain not to ask for any information relevant to my own fate...*

Explain what is wrong with A's reasoning.

A pair (w, l), where $w \in W$ is the actual world, and l is the agent's local state, which essentially characterizes her information.

W is the "naive space".

For the purposes of this paper, it is assumed that l has the form $\langle o_1, \ldots, o_n \rangle$, where o_j is the observation that the agent makes at time j, $j = 1, \ldots, n$.

A pair $(w, \langle o_1, \ldots, o_n \rangle)$ is called a run.

Consider the three-prisoners puzzle in more detail:

- The naive space is $W = \{w_a, w_b, w_c\}$, where w_x is the world where x is not executed.
- We are only interested in runs of length 1, so $n = 1$. The set O of observations (what agent can be told) is $\{\{w_a, w_b\}, \{w_a, w_c\}\}$. Here "$\{w_a, w_b\}$" corresponds to the observation that either a or b will not be executed (i.e., the jailer saying "c will be executed"); similarly, $\{w_a, w_c\}$ corresponds to the jailer saying "b will be executed".

The sophisticated space consists of the four runs

$$(w_a, \langle\{w_a, w_b\}\rangle), (w_a, \langle\{w_b, w_c\}\rangle), (w_b, \langle\{w_a, w_b\}\rangle), (w_c, \langle\{w_a, w_c\}\rangle)$$

Note that there is no run with observation $\{w_b, w_c\}$, since the jailer will not tell a that he will be executed.

According to the story, the prior Pr_W in the naive space has $Pr_W(w) = 1/3$ for $w \in W$. The full distribution Pr on the runs is not completely specified by the story. In particular, we are not told the probability with which the jailer will say b and c if a will not be executed.

All runs are of the form $r = (w, \langle U \rangle)$, where $w \in U$.

All runs are of the form $r = (w, \langle U \rangle)$, where $w \in U$.

Question: After observing U, the agent can compute her posterior on W by conditioning on U. Roughly speaking, this amounts to asking whether observing U is the same as discovering that U is true.

$$\mathcal{R} = \{(w, \langle U \rangle) \mid U \in O, w \in U\}$$

For $r = (w, \langle U \rangle)$, let $X_W(r) = w$ and $X_O(r) = U$

Pr is a probability measure on $\mathcal{R}$.

Pr_W is the marginal distribution of X_W

Pr_O is the marginal distribution of X_O.

Let Pr be a prior on $\mathcal{R}$. Let $Pr' = Pr(\cdot \mid X_O = U)$ be the posterior after observing U.

Main question: under what conditions we have

$$Pr'_W(V) = Pr_W(V \mid U) \qquad \text{for all } V \subseteq W$$

$$Pr(X_W = w \mid X_O = U) = Pr(X_W = w \mid X_W \in U) \quad \text{for all } w \in U$$

Theorem. Fix a probability Pr on $\mathcal{R}$ and a set $U \subseteq W$. The following are equivalent:

1. If $Pr(X_O = U) > 0$, then $Pr(X_W = w \mid X_O = U) = Pr(X_W = w \mid X_W \in U)$ for all $w \in U$
2. The event $X_W = w$ is independent of the event $X_O = U$ given $X_W \in U$ for all $w \in U$.
3. $Pr(X_O = U \mid X_W = w) = Pr(X_O = U \mid X_W \in U)$ for all $w \in U$ such that $Pr(X_W = w) > 0$
4. $Pr(X_O = U \mid X_W = w) = Pr(X_O = U \mid X_W = w')$ for all $w, w' \in U$ such that $Pr(X_W = w) > 0$ and $Pr(X_W = w') > 0$

In the three-prisoner's puzzle, what is a's prior distribution Pr on $\mathcal{R}$? We assumed that the marginal distribution Pr_W on W is uniform. Apart from this, Pr is unspecified.

Now suppose that a observes $\{w_a, w_c\}$ ("the jailer says b"). Naive conditioning would lead a to adopt the distribution $Pr_W(\cdot \mid \{w_a, w_c\})$. This satisfies $Pr_W(w_a \mid \{w_a, w_c\}) = 1/2$.

Sophisticated conditioning leads a to adopt the distribution $Pr' = Pr(\cdot \mid X_O = \{w_a, w_c\})$

By part (4) of the Theorem, naive conditioning is appropriate (i.e., $Pr'_W = Pr_W(\cdot \mid \{w_a, w_c\})$ only if the jailer is equally likely to say b in both worlds w_a and w_c. Since the jailer must say that b will be executed in world w_c, it follows that
$Pr(X_O = \{w_a, w_c\} | X_W = w_c) = 1$.

Thus, conditioning is appropriate only if the jailer's protocol is such that he definitely says b in w_a, i.e., even if both b and c are executed.

But if this is the case, when the jailer says c, conditioning Pr_W on $\{w_a, w_b\}$ is not appropriate, since then a knows that he will be executed.

The world cannot be w_a, for then the jailer would have said b. Therefore, *no matter what the jailer's protocol is*, conditioning in the naive space cannot coincide with conditioning in the sophisticated space for both of his responses.

Suppose that $O = \{U_1, U_2\}$, and both U_1 and U_2 are observed with positive probability. (This is the case for both Monty Hall and the three-prisoners puzzle.) Then the CAR condition cannot hold for both U_1 and U_2 unless $Pr(X_W \in U_1 \cap U_2)$ is either 0 or 1.

Proposition The CAR condition holds for all distributions Pr on $\mathcal{R}$ if and only if O consists of pairwise disjoint subsets of W.