Logic and Probabilistic Models of Belief Change

Eric Pacuit

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

February 4, 2016

$$K \circ \varphi$$

$K \circ \varphi$

Initial set of beliefs

$K \circ \varphi$

Initial set of beliefs

New evidence φ

Belief change operator: $\circ : \mathcal{B} \times \mathcal{L} \to \mathcal{B}$

$$K \circ \varphi$$

Initial set of beliefs

New evidence φ

Conservatism: "an ordinary rational person continues to believe something that he or she starts out believing in the absence of a special reason to doubt it"

G. Harman. *Rationality, in An Invitation to Cognitive Science*. The MIT Press, 1995, pgs. 175 - 212.

Belief Change

Suppose that K is the current beliefs.

If you give priority to the new information φ:

Belief Change

Suppose that K is the current beliefs.

If you give priority to the new information φ:

1. **Expansion**: $K + \varphi$; φ is added to K giving a new belief set K'.

Belief Change

Suppose that K is the current beliefs.

If you give priority to the new information φ:

1. **Expansion**: $K + \varphi$; φ is added to K giving a new belief set K'.
2. **Contraction**: $K \dot{-} \varphi$; φ is removed from K giving a new belief set K'

Belief Change

Suppose that K is the current beliefs.

If you give priority to the new information φ:

1. **Expansion**: $K + \varphi$; φ is added to K giving a new belief set K'.

2. **Contraction**: $K \dot{-} \varphi$; φ is removed from K giving a new belief set K'

3. **Revision**: $K * \varphi$; φ is added and other formulas are removed, so that the resulting new belief set K' is consistent.

Expansion Postulates

$(E1)$ $K + \alpha$ is deductively closed

$(E2)$ $\alpha \in K + \alpha$

$(E3)$ $K \subseteq K + \alpha$

$(E4)$ If $\alpha \in K$, then $K + \alpha = K$

$(E5)$ If $K \subseteq K'$, then $K + \alpha \subseteq K' + \alpha$

(Minimality) For all belief sets K and all sentences α, $K + \alpha$ is the smallest belief set that satisfies (E1), (E2), and (E3).

Expansion

Theorem Let $+$ be a function on belief sets and formulas. Then, $+$ satisfies minimality of and only if $K + \alpha = Cn(K \cup \{\alpha\})$.

Contraction Postulates

$(C1)$ $K \dot{-} \alpha$ is deductively closed

$(C2)$ $K \dot{-} \alpha \subseteq K$

$(C3)$ If $\alpha \notin K$ or $\vdash \alpha$ then $K \dot{-} \alpha = K$

$(C4)$ If $\nvdash \alpha$, then $\alpha \notin K \dot{-} \alpha$

$(C5)$ If $\vdash \alpha \leftrightarrow \beta$, then $K \dot{-} \alpha = K \dot{-} \beta$

$(C6)$ $K \subseteq Cn((K \dot{-} \alpha) \cup \{\alpha\})$

Definition. An operator $-$ is a **withdrawal** if and only if it satisfies (C1-C5).

Definition. An operator $-$ is a **withdrawal** if and only if it satisfies (C1-C5).

The following is a withdrawal:

$$K - \alpha = \begin{cases} K & \text{if } \alpha \notin K \\ Cn(\emptyset) & \text{if } \alpha \in K \end{cases}$$

Definition. An operator $-$ is a **withdrawal** if and only if it satisfies (C1-C5).

The following is a withdrawal:

$$K - \alpha = \begin{cases} K & \text{if } \alpha \notin K \\ Cn(\emptyset) & \text{if } \alpha \in K \end{cases}$$

Minimal Information Loss (Recovery): $K \subseteq Cn((K \dot{-} \alpha) \cup \{\alpha\})$

Levi Identity

$$K * \varphi = (K \dot{-} \neg\varphi) + \varphi$$

AGM Postulates

AGM 1: $K * \varphi$ is deductively closed

AGM 2: $\varphi \in K * \varphi$

AGM 3: $K * \varphi \subseteq Cn(K \cup \{\varphi\})$

AGM 4: If $\neg\varphi \notin K$ then $K * \varphi = Cn(K \cup \{\varphi\})$

AGM 5: $K * \varphi$ is inconsistent only if φ is inconsistent

AGM 6: If φ and ψ are logically equivalent then $K * \varphi = K * \psi$

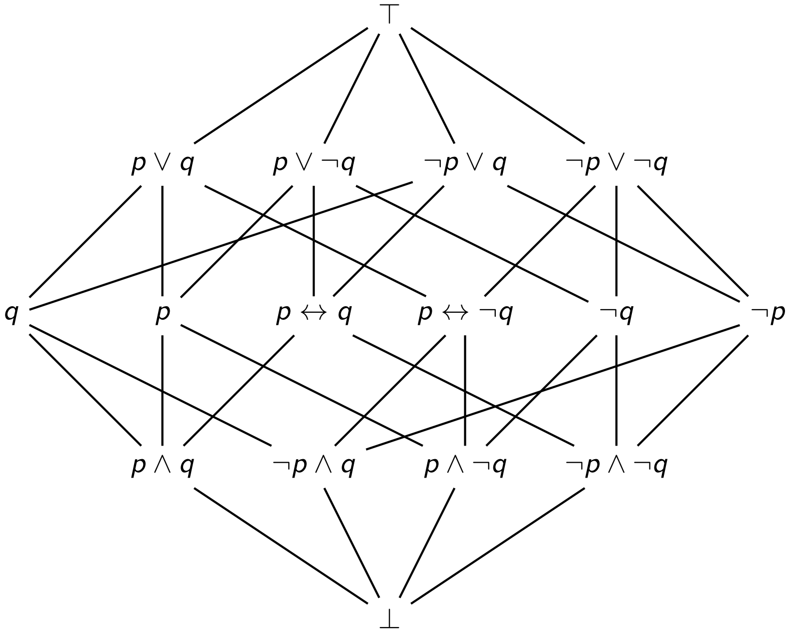
AGM 7: $K * (\varphi \wedge \psi) \subseteq Cn(K * \varphi \cup \{\psi\})$

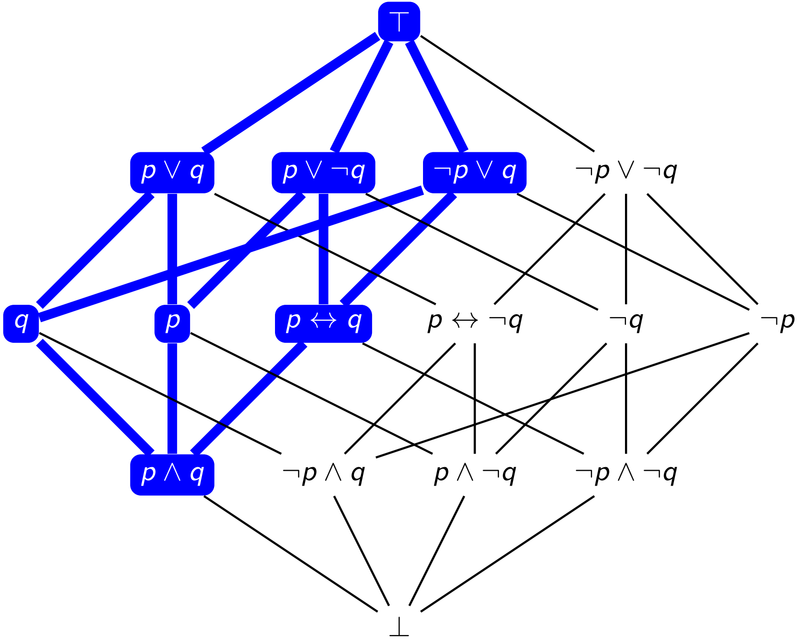
AGM 8: if $\neg\psi \notin K * \varphi$ then $Cn(K * \varphi \cup \{\psi\}) \subseteq K * (\varphi \wedge \psi)$

Recovery?

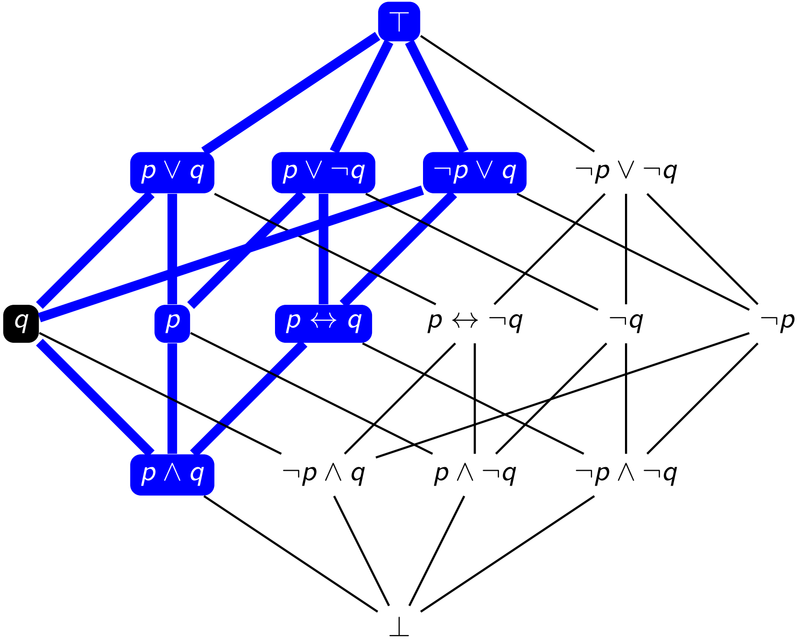
Theorem. Let K be any theory. Then for each withdrawal operations $-$ on K, there is a unique contraction operation $\dot{-}$ that is **revision equivalent** to $-$, and this $\dot{-}$ is the greatest element of $[-]$ (the equivalence class of withdrawal operations revision equivalent to $-$).

D. Makinson. *On the status of Recovery*. Journal of Philosophical Logic, 16, pp. 383 - 394, 1987.


$\top$
$p \vee q$
$p \vee \neg q$
$\neg p \vee q$
$\neg p \vee \neg q$
q
p
$p \leftrightarrow q$
$p \leftrightarrow \neg q$
$\neg q$
$\neg p$
$p \wedge q$
$\neg p \wedge q$
$p \wedge \neg q$
$\neg p \wedge \neg q$
$\bot$



$\top$
$p \vee q$
$p \vee \neg q$
$\neg p \vee q$
$\neg p \vee \neg q$
q
p
$p \leftrightarrow q$
$p \leftrightarrow \neg q$
$\neg q$
$\neg p$
$p \wedge q$
$\neg p \wedge q$
$p \wedge \neg q$
$\neg p \wedge \neg q$
$\bot$


$\top$
$p \vee q$
$p \vee \neg q$
$\neg p \vee q$
$\neg p \vee \neg q$
q
p
$p \leftrightarrow q$
$p \leftrightarrow \neg q$
$\neg q$
$\neg p$
$p \wedge q$
$\neg p \wedge q$
$p \wedge \neg q$
$\neg p \wedge \neg q$
$\perp$

Let K be a belief set and φ a formula.

$K \perp \varphi$ is the **remainder set** of K.

$A \in K \perp \varphi$ iff

1. $A \subseteq K$
2. $\varphi \notin Cn(A)$
3. There is no B such that $A \subset B \subseteq K$ and $\varphi \notin Cn(B)$.

- $K \perp \alpha = \{K\}$ iff $\neg\alpha \notin Cn(K)$
- $K \perp \alpha = \emptyset$ iff $\alpha \in Cn(\emptyset)$
- If $K' \subseteq K$ and $\alpha \notin Cn(K')$ then there is some T such that $K' \subseteq T \in K \perp \alpha$.

A **selection function** γ for K is a function on $K \perp \alpha$ such that:

- If $K \perp \alpha \neq \emptyset$, then $\gamma(K \perp \alpha) \subseteq K \perp \alpha$ and $\gamma(K \perp \alpha) \neq \emptyset$
- If $K \perp \alpha = \emptyset$, then $\gamma(K \perp \alpha) = \{K\}$

Let K be a set of formulas. A function $\dot{-}$ is a **partial meet contraction** for K if there is a selection function γ for K such that for all formula α:

$$K \dot{-} \alpha = \bigcap \gamma(K \perp \alpha)$$

Let K be a set of formulas. A function $\dot{-}$ is a **partial meet contraction** for K if there is a selection function γ for K such that for all formula α:

$$K \dot{-} \alpha = \bigcap \gamma(K \perp \alpha)$$

Then $K * \alpha = Cn(\bigcap \gamma(K \perp \neg\alpha) \cup \{\alpha\})$

Let K be a set of formulas. A function $\dot{-}$ is a **partial meet contraction** for K if there is a selection function γ for K such that for all formula α:

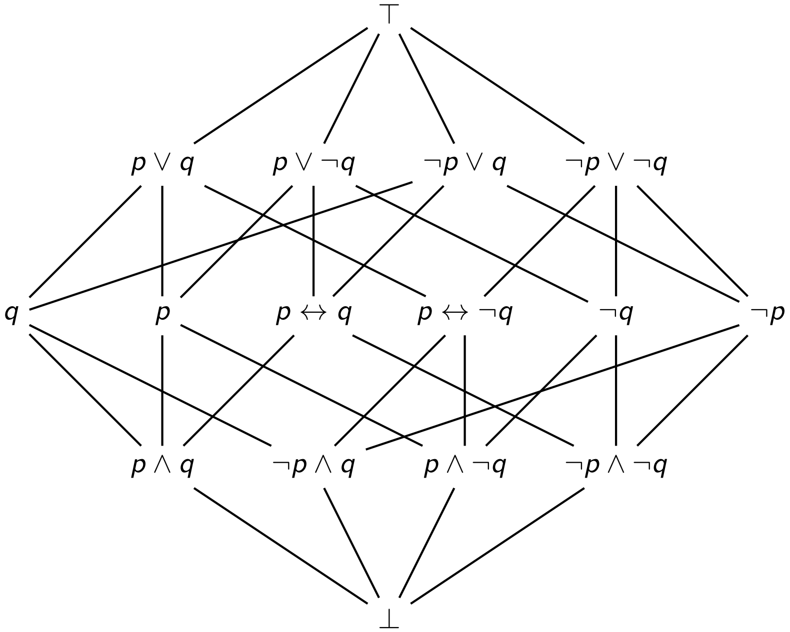
$$K \dot{-} \alpha = \bigcap \gamma(K \perp \alpha)$$

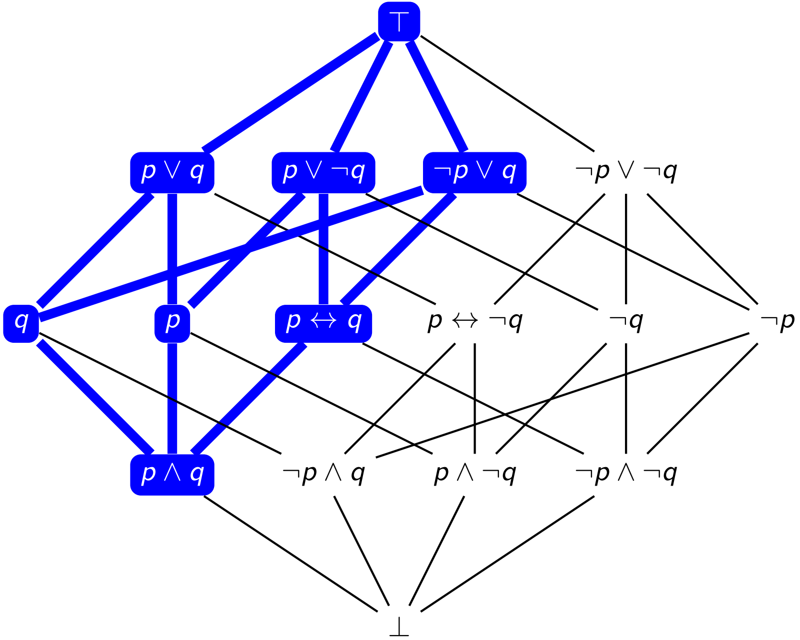
Then $K * \alpha = Cn(\bigcap \gamma(K \perp \neg\alpha) \cup \{\alpha\})$

- γ selects exactly one element of $K \perp \alpha$ (maxichoice contraction)
- γ selects the entire set $K \perp \alpha$ (full meet contraction)

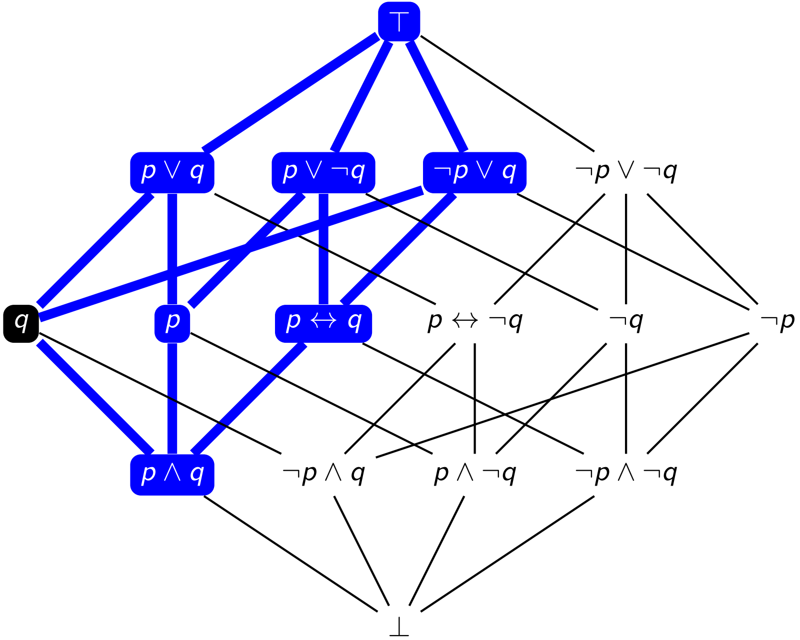
Theorem (AGM 1985). Let K be a belief set and let $*$ be a function on $\mathcal{L}$. Then

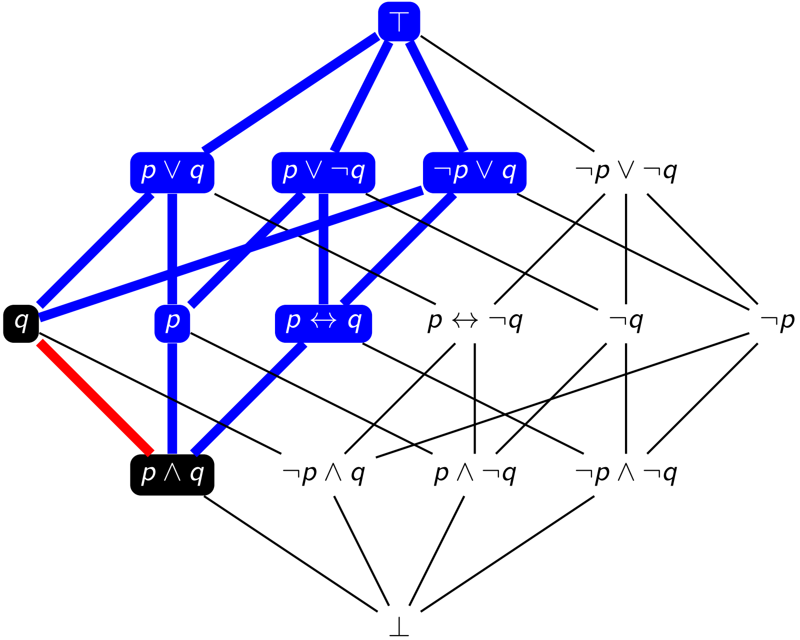
- The function $*$ is a partial meet revision for K if and only if it satisfies the postulates $AGM1$ - $AGM6$
- The function $*$ is a transitively relational partial meet revision for K if and only if it satisfies $AGM1$ - $AGM8$.

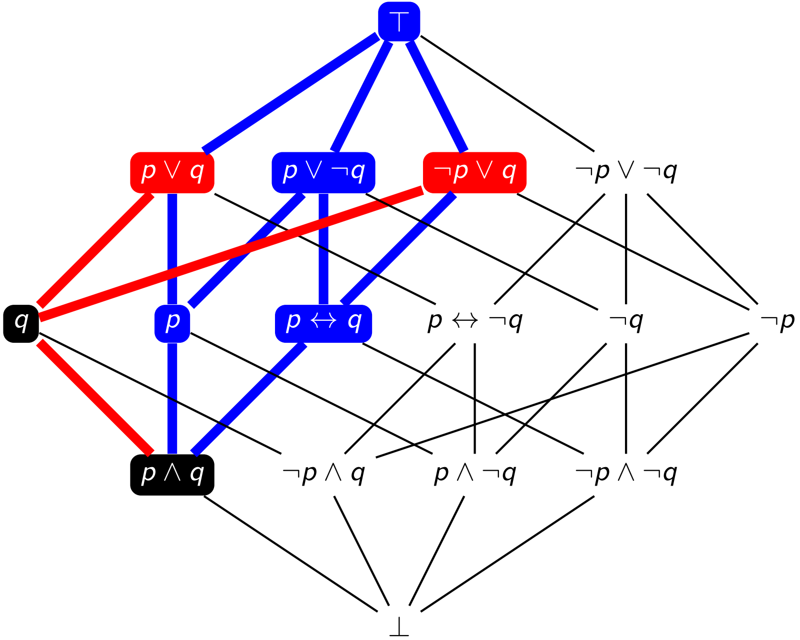

$\top$
$p \vee q$
$p \vee \neg q$
$\neg p \vee q$
$\neg p \vee \neg q$
q
p
$p \leftrightarrow q$
$p \leftrightarrow \neg q$
$\neg q$
$\neg p$
$p \wedge q$
$\neg p \wedge q$
$p \wedge \neg q$
$\neg p \wedge \neg q$
$\bot$

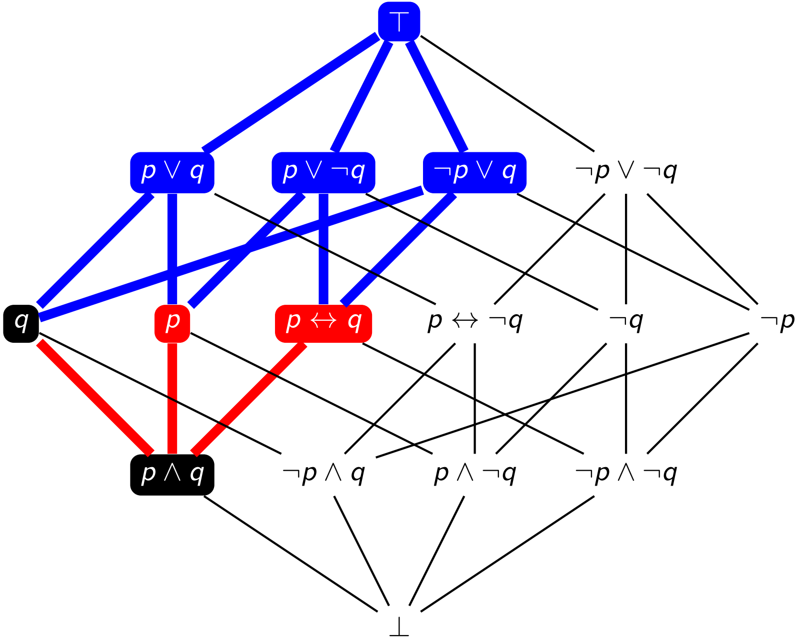


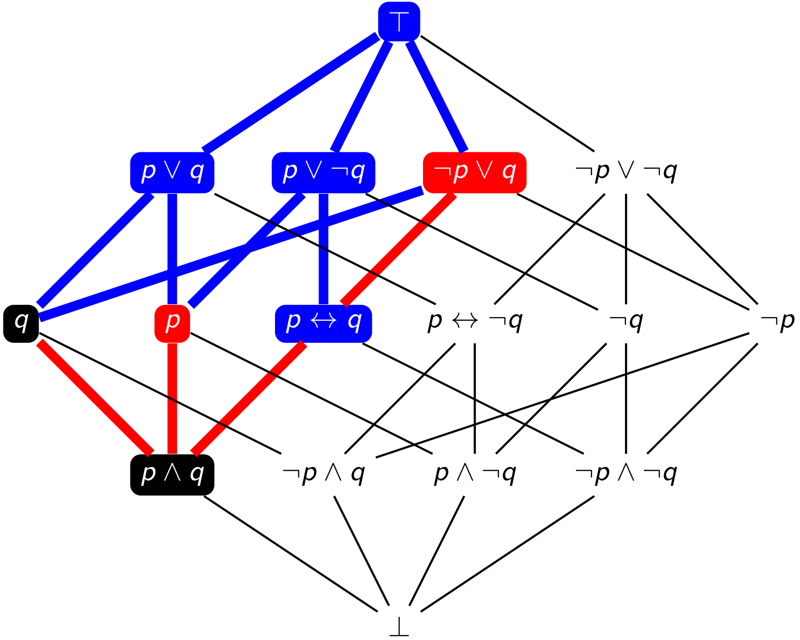
$\top$
$p \vee q$
$p \vee \neg q$
$\neg p \vee q$
$\neg p \vee \neg q$
q
p
$p \leftrightarrow q$
$p \leftrightarrow \neg q$
$\neg q$
$\neg p$
$p \wedge q$
$\neg p \wedge q$
$p \wedge \neg q$
$\neg p \wedge \neg q$
$\perp$

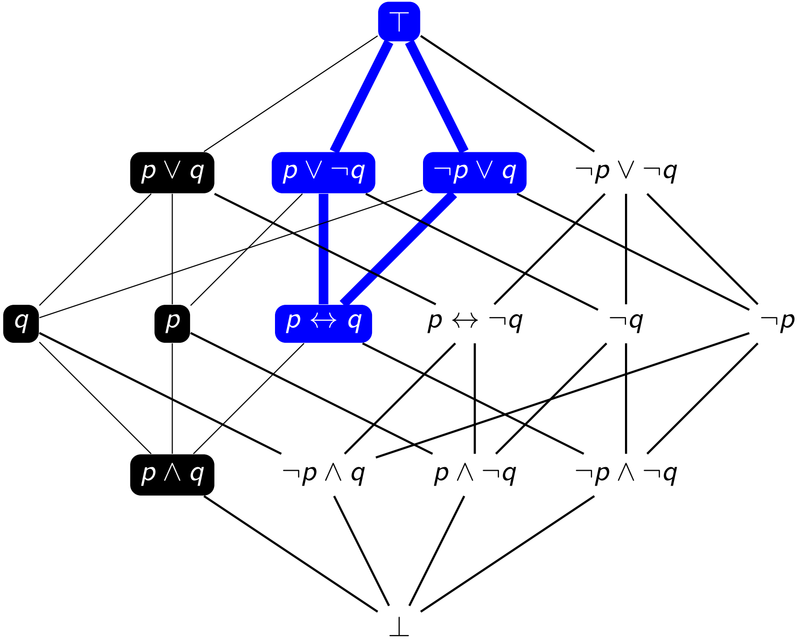

⊤
p ∨ q
p ∨ ¬q
¬p ∨ q
¬p ∨ ¬q
q
p
p ↔ q
p ↔ ¬q
¬q
¬p
p ∧ q
¬p ∧ q
p ∧ ¬q
¬p ∧ ¬q
⊥

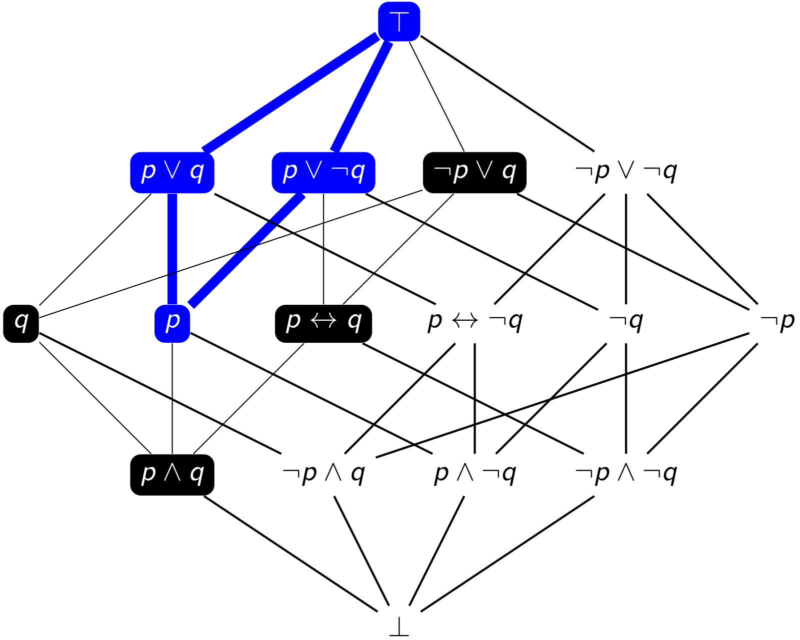

$\top$
$p \vee q$
$p \vee \neg q$
$\neg p \vee q$
$\neg p \vee \neg q$
q
p
$p \leftrightarrow q$
$p \leftrightarrow \neg q$
$\neg q$
$\neg p$
$p \wedge q$
$\neg p \wedge q$
$p \wedge \neg q$
$\neg p \wedge \neg q$
$\perp$

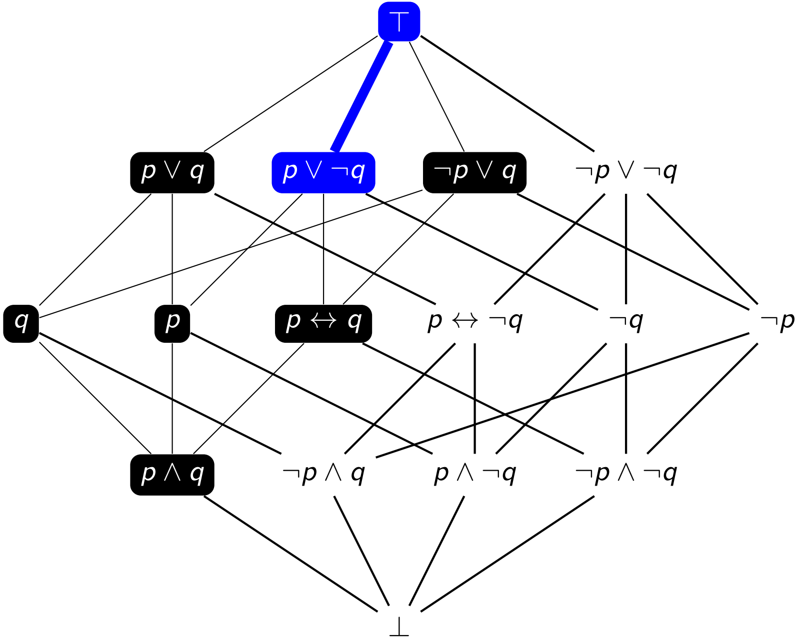

⊤
p ∨ q
p ∨ ¬q
¬p ∨ q
¬p ∨ ¬q
q
p
p ↔ q
p ↔ ¬q
¬q
¬p
p ∧ q
¬p ∧ q
p ∧ ¬q
¬p ∧ ¬q
⊥


$\top$
$p \vee q$
$p \vee \neg q$
$\neg p \vee q$
$\neg p \vee \neg q$
q
p
$p \leftrightarrow q$
$p \leftrightarrow \neg q$
$\neg q$
$\neg p$
$p \wedge q$
$\neg p \wedge q$
$p \wedge \neg q$
$\neg p \wedge \neg q$
$\bot$


$\top$
$p \vee q$
$p \vee \neg q$
$\neg p \vee q$
$\neg p \vee \neg q$
q
p
$p \leftrightarrow q$
$p \leftrightarrow \neg q$
$\neg q$
$\neg p$
$p \wedge q$
$\neg p \wedge q$
$p \wedge \neg q$
$\neg p \wedge \neg q$
$\perp$


$\top$
$p \vee q$
$p \vee \neg q$
$\neg p \vee q$
$\neg p \vee \neg q$
q
p
$p \leftrightarrow q$
$p \leftrightarrow \neg q$
$\neg q$
$\neg p$
$p \wedge q$
$\neg p \wedge q$
$p \wedge \neg q$
$\neg p \wedge \neg q$
$\bot$


⊤
p ∨ q
p ∨ ¬q
¬p ∨ q
¬p ∨ ¬q
q
p
p ↔ q
p ↔ ¬q
¬q
¬p
p ∧ q
¬p ∧ q
p ∧ ¬q
¬p ∧ ¬q
⊥


⊤
p ∨ q
p ∨ ¬q
¬p ∨ q
¬p ∨ ¬q
q
p
p ↔ q
p ↔ ¬q
¬q
¬p
p ∧ q
¬p ∧ q
p ∧ ¬q
¬p ∧ ¬q
⊥

There is a **transitive relation** $\preceq$ on $K \perp \alpha$ such that
$\gamma(K\perp\alpha) = \{K' \in K\perp\alpha K'' \preceq K'$ for all $K'' \in K\perp\alpha\}$

$$K \dot{-} \alpha = \begin{cases} \bigcap\{K' \in K\perp\alpha \mid K' \text{ is } \preceq\text{-maximal}\} & \text{if } \alpha \notin Cn(\emptyset) \\ K & \text{otherwise} \end{cases}$$

Epistemic Entrenchment

Even if all sentences in a knowledge set are accepted or considered as facts (so that they are assigned maximal probability), this does not mean that all sentences are are of equal value for planning or problem-solving purposes. Certain pieces of our knowledge and beliefs about the world are more important than others when planning future actions, conducting scientific investigations, or reasoning in general. We will say that some sentences in a knowledge system have a higher degree of epistemic entrenchment than others. This degree of entrenchment will, intuitively, have a bearing on what is abandoned from a knowledge set, and what is retained, when a contraction or a revision is carried out.

P Gärdenfors and D. Makinson. *Revisions of Knowledge Systems Using Epistemic Entrenchment*. Proceedings of the Second Conference on Theoretical Aspects of Reasoning about Knowledge, 1988, 83-95.

Epistemic Entrenchment

$$\leq_K \subseteq \mathcal{L} \times \mathcal{L}$$

Note that the relation $\leq$ is only defined in relation to a given K — different knowledge sets may be associated with different orderings of epistemic entrenchment.

Epistemic Entrenchment

$$\leq_K \subseteq \mathcal{L} \times \mathcal{L}$$

Note that the relation $\leq$ is only defined in relation to a given K — different knowledge sets may be associated with different orderings of epistemic entrenchment.

- $\varphi \leq \psi$: φ is at most as entrenched as ψ
- $\varphi < \psi$: φ is less entrenched than ψ
  $((\varphi \leq \psi)$ and $\mathrm{not}(\psi \leq \varphi))$
- $\varphi \equiv \psi$: φ and ψ are equally entrenched
  $((\varphi \leq \psi)$ and $(\psi \leq \varphi))$

Epistemic Entrenchment

Transitivity: If $\varphi \leq \psi$ and $\psi \leq \chi$, then $\varphi \leq \chi$

Dominance: If $\varphi \vdash \psi$, then $\varphi \leq \psi$

Conjunctiveness: Either $\varphi \leq (\varphi \wedge \psi)$ or $\psi \leq (\varphi \wedge \psi)$.

Minimality: If the belief set K is consistent, then $\varphi \notin K$ if and only if $\varphi \leq \psi$ for all ψ.

Maximality: If $\psi \leq \varphi$ for all ψ, then $\varphi \in Cn(\emptyset)$.

$$K \dot{-} \alpha = \begin{cases} K \cap \{\beta \mid \alpha < \alpha \vee \beta\} & \text{if } \alpha \notin Cn(\emptyset) \\ K & \text{otherwise} \end{cases}$$

$\varphi \leq \psi$ if and only if $\varphi \notin K \dot{-} (\varphi \wedge \psi)$ or $\vdash \varphi \wedge q$

Epistemic Entrenchment

$q \in K \dot{-} p$ if and only if $q \in K$ and either $p < (p \vee q)$ or $p \in Cn(\emptyset)$

$q \in K \dot{-} p$ if and only if $q \in K$ and either $q < p \vee q$ or $\vdash p$.

$\varphi < \psi$ iff $\psi \in K \dot{-} \varphi \wedge \psi$

Severe Withdrawal

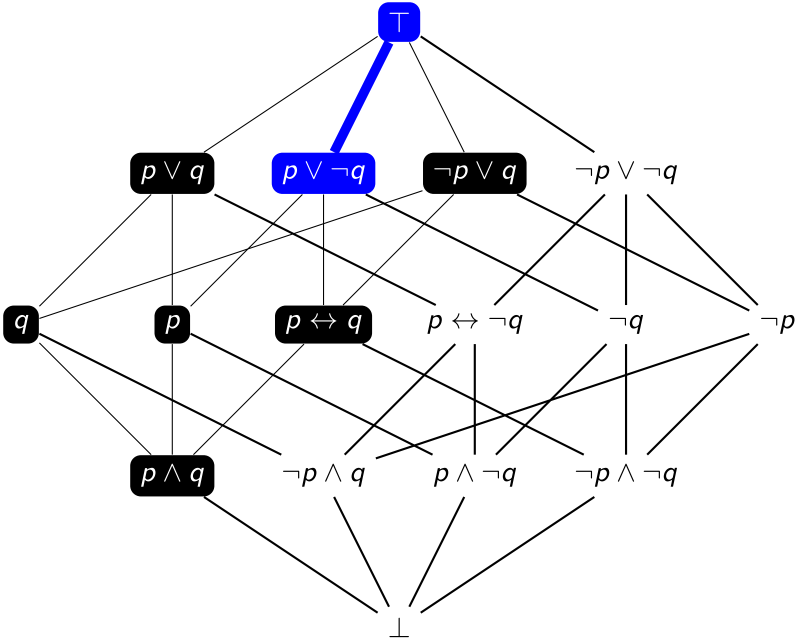
$$K \dot{-} \alpha = \begin{cases} K \cap \{\beta \mid \alpha < \beta\} & \text{if } \alpha \notin Cn(\emptyset) \\ K & \text{otherwise} \end{cases}$$

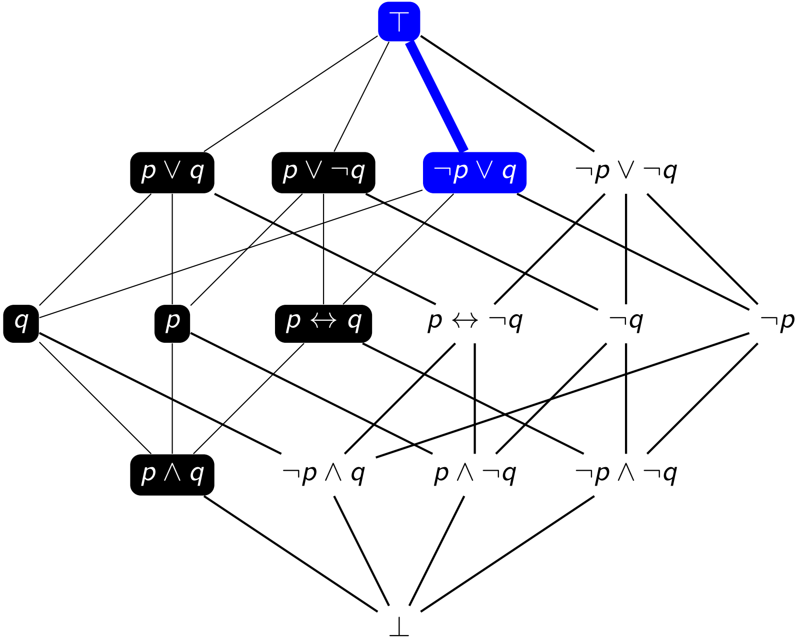
Preference. Objects held in higher regard should be afforded a more favorable treatment.

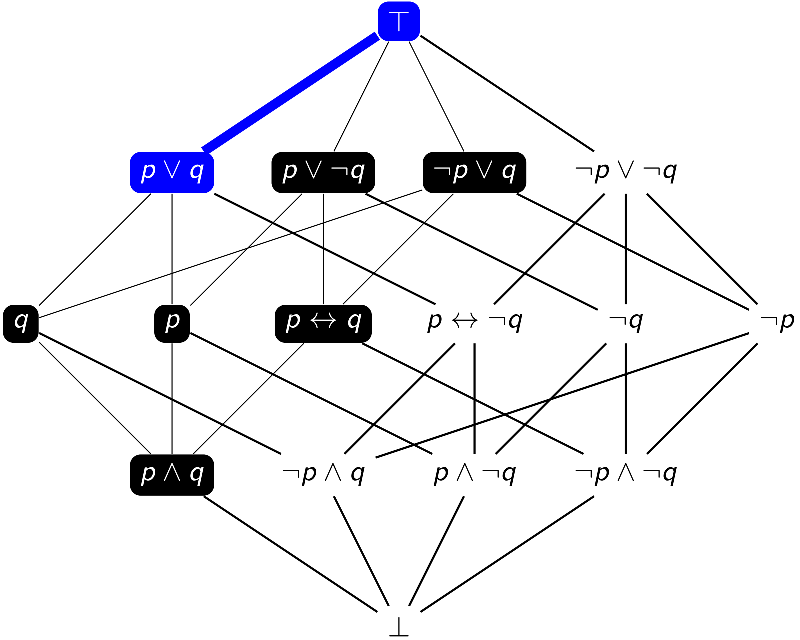
Theorem. Severe withdrawal satisfies every postulate except recovery.

H. Rott and M. Pagnucco. *Sever withdrawal (and recover)*. Journal of Philsophical Logic, 28, pp. 501 - 547, 1999.

Levi Contraction


$\top$
$p \vee q$
$p \vee \neg q$
$\neg p \vee q$
$\neg p \vee \neg q$
q
p
$p \leftrightarrow q$
$p \leftrightarrow \neg q$
$\neg q$
$\neg p$
$p \wedge q$
$\neg p \wedge q$
$p \wedge \neg q$
$\neg p \wedge \neg q$
$\perp$


⊤
p ∨ q
p ∨ ¬q
¬p ∨ q
¬p ∨ ¬q
q
p
p ↔ q
p ↔ ¬q
¬q
¬p
p ∧ q
¬p ∧ q
p ∧ ¬q
¬p ∧ ¬q
⊥


$\top$
$p \vee q$
$p \vee \neg q$
$\neg p \vee q$
$\neg p \vee \neg q$
q
p
$p \leftrightarrow q$
$p \leftrightarrow \neg q$
$\neg q$
$\neg p$
$p \wedge q$
$\neg p \wedge q$
$p \wedge \neg q$
$\neg p \wedge \neg q$
$\bot$

Informational Value

$$V : \mathbb{K} \rightarrow \mathbb{R}$$

For theories, K_1, K_2, if $K_1 \subseteq K_2$, then $V(K_1) \leq V(K_2)$

$$K \dot{-} \alpha = \begin{cases} \bigcap \delta(S(K, \alpha)) & \text{if } \alpha \in Cn(\emptyset) \\ K & \text{otherwise} \end{cases}$$

$\delta(S(K, \alpha)) = \{K' \in S(K, \alpha) \mid V(K'') \leq V(K')$ for all $K'' \in S(K, \alpha)\}$

Relevance

R. Parikh. *Relevance sensitive belief structures*. Annals of Mathematics and Artificial Intelligence, 28, pp. 259 - 285, 2000.

G. Kourousias and D. Makinson. *Parallel Interpolation, Splitting, and Relevance in Belief Change*. Journal of Symbolic Logic, 72(3), pp. 994-1002, 2007 .

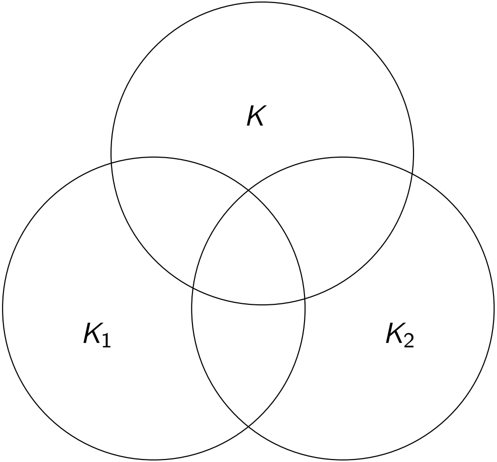
Hans Rott. *Two Dogmas of Belief Revision*. The Journal of Philosophy, 97:9, pp. 503-522 (2000).

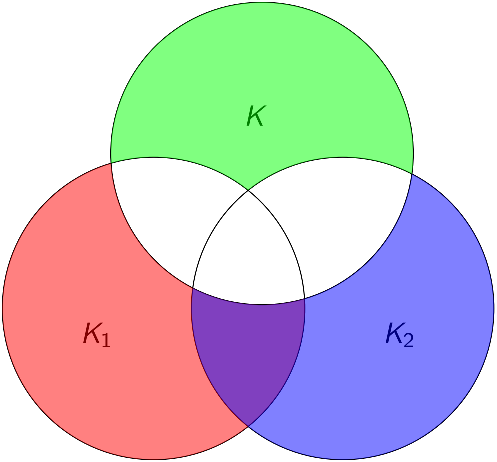
Two Dogmas of Belief Revision

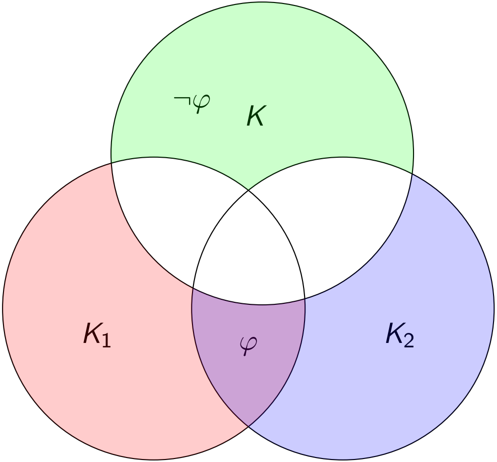
1. When accepting a new piece of information, an agent should aim at a minimal change of his old beliefs.
2. If there are different ways to effect a belief change, the agent should give up those beliefs that are least entrenched.

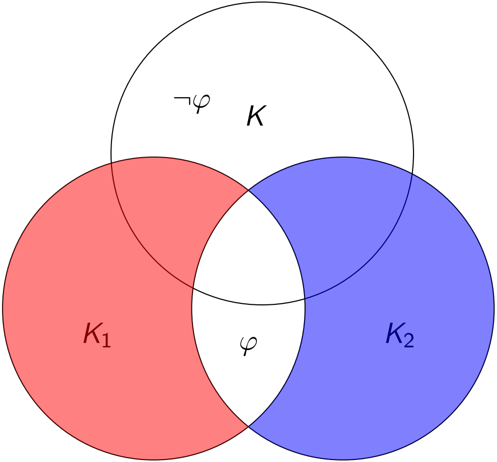
Observation 1

No two distinct belief-contravening candidate revisions of a consistent and logically closed belief set by a sentence can be set-theoretically compared in terms of the sets of beliefs on which they differ with the original belief set.

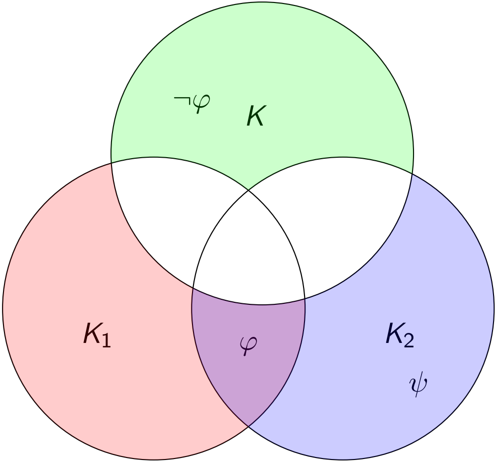

K
K_1
K_2

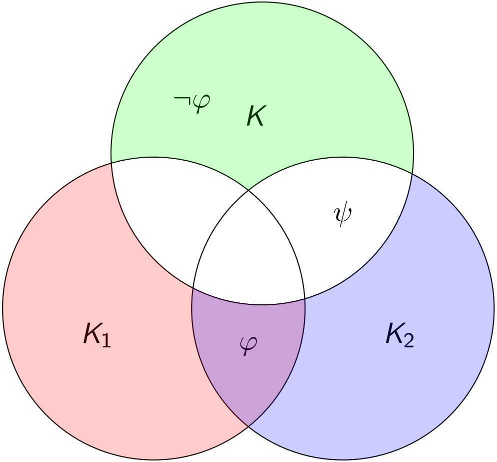

K
K_1
K_2


$\neg\varphi$
K
K_1
φ
K_2

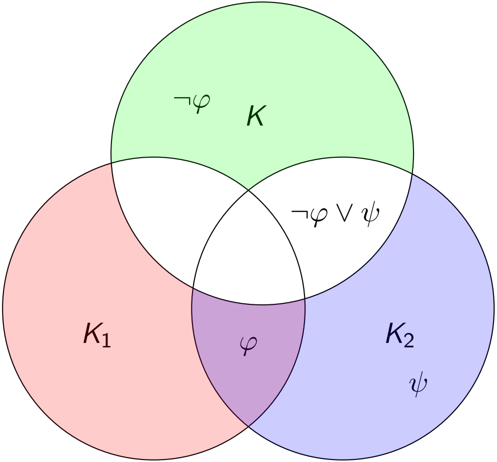


$\neg\varphi$
K
K_1
φ
K_2

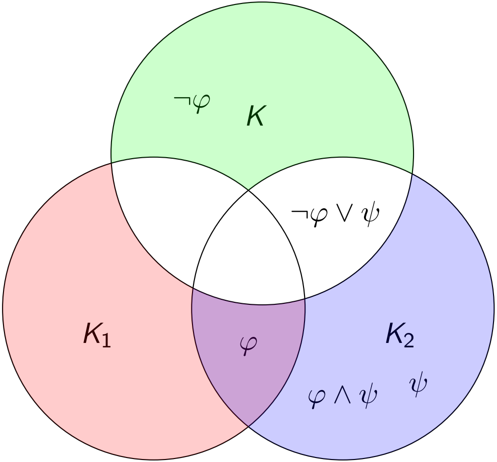

$\neg\varphi$
K
K_1
φ
K_2
ψ

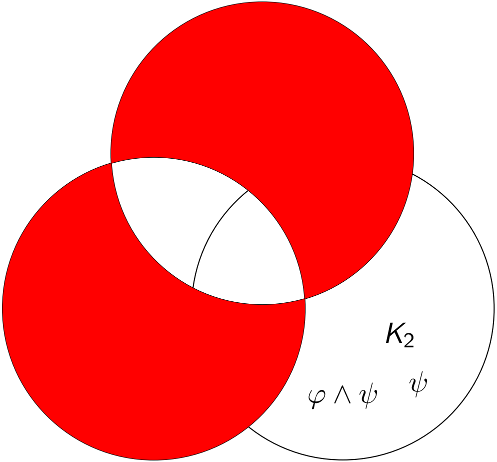


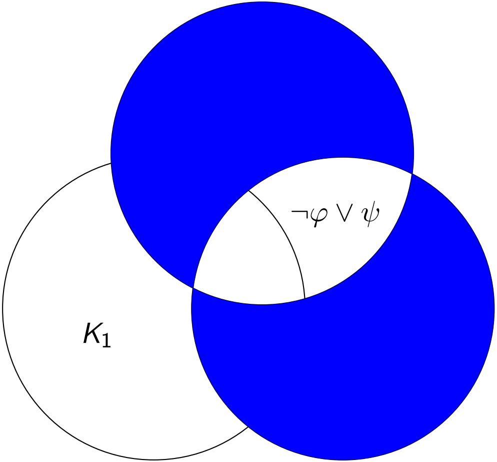
$\neg\varphi$
K
ψ
K_1
φ
K_2

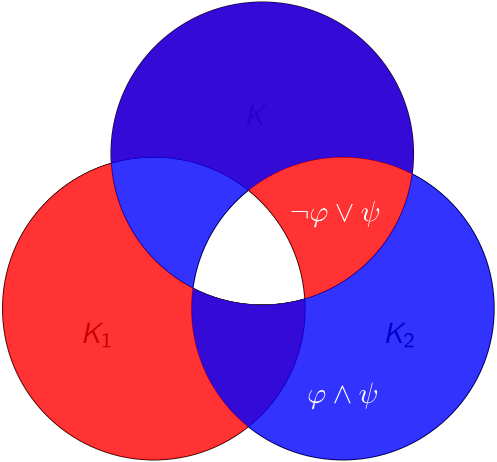


$\neg\varphi$
K
$\neg\varphi \vee \psi$
K_1
φ
K_2
ψ


$\neg\varphi$
K
$\neg\varphi \vee \psi$
K_1
φ
K_2
$\varphi \wedge \psi$
ψ

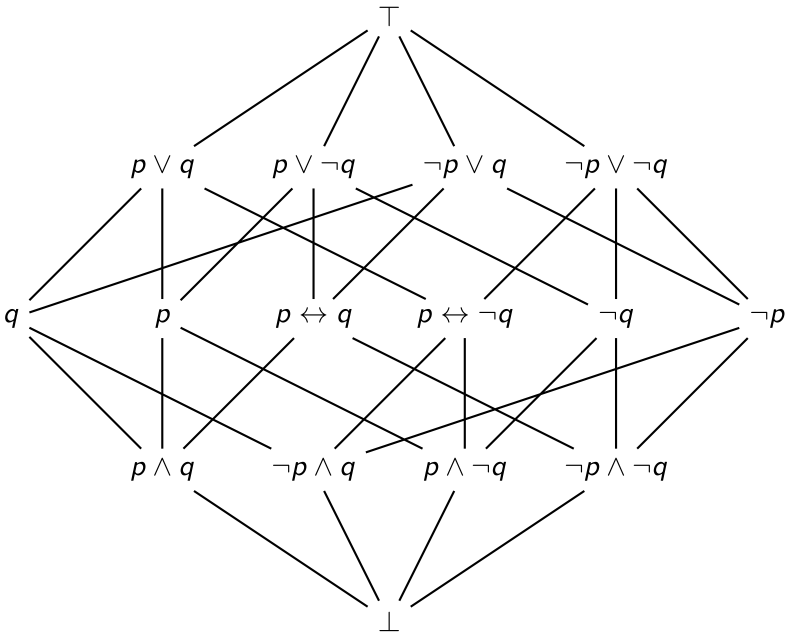

K_2
$\varphi \wedge \psi$
ψ

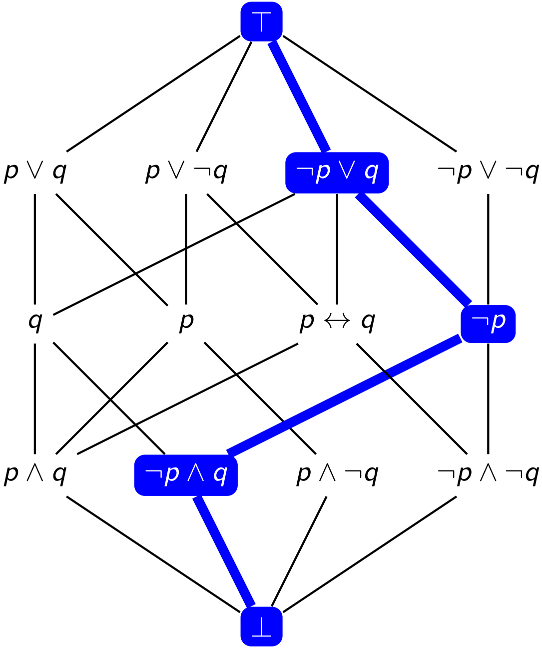

$\neg\varphi \vee \psi$
K_1


K
$\neg\varphi \vee \psi$
K_1
K_2
$\varphi \wedge \psi$

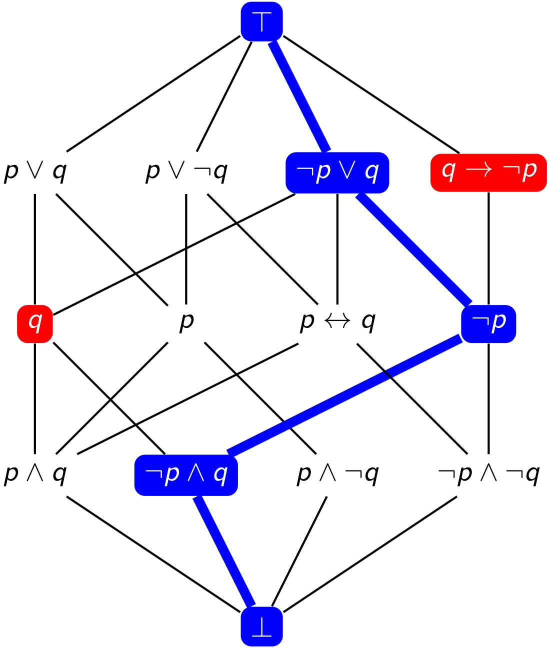
Observation 2

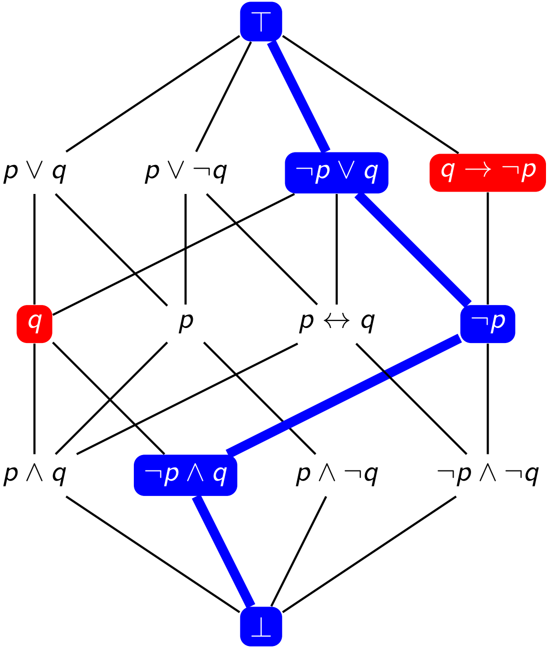
Suppose we want to revise a belief set by a sentence φ and find two elements of the belief set that non-redundantly entail the negation of φ. Then it may well be rational, according to the standard belief revision constructions, to restore consistency by removing the more entrenched and retain the less entrenched belief. In fact, such a situation can always be identified in an anamnestic revision by a consistent and moderately surprising sentence.


$\top$
$p \vee q$
$p \vee \neg q$
$\neg p \vee q$
$\neg p \vee \neg q$
q
p
$p \leftrightarrow q$
$p \leftrightarrow \neg q$
$\neg q$
$\neg p$
$p \wedge q$
$\neg p \wedge q$
$p \wedge \neg q$
$\neg p \wedge \neg q$
$\bot$

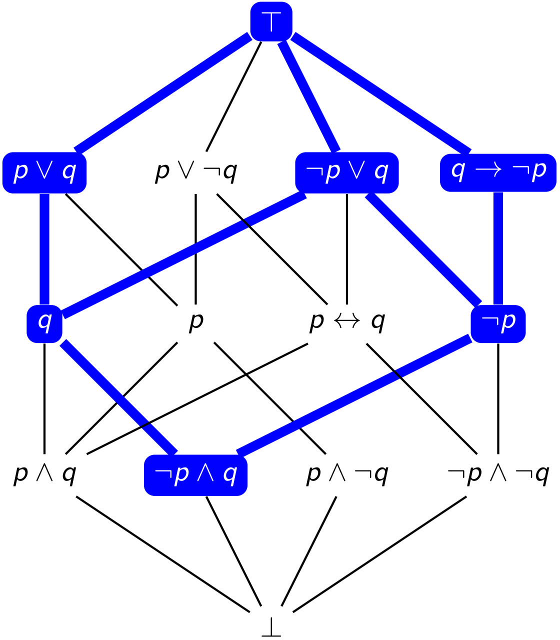


$\top$
$p \vee q$
$p \vee \neg q$
$\neg p \vee q$
$\neg p \vee \neg q$
q
p
$p \leftrightarrow q$
$\neg p$
$p \wedge q$
$\neg p \wedge q$
$p \wedge \neg q$
$\neg p \wedge \neg q$
$\bot$

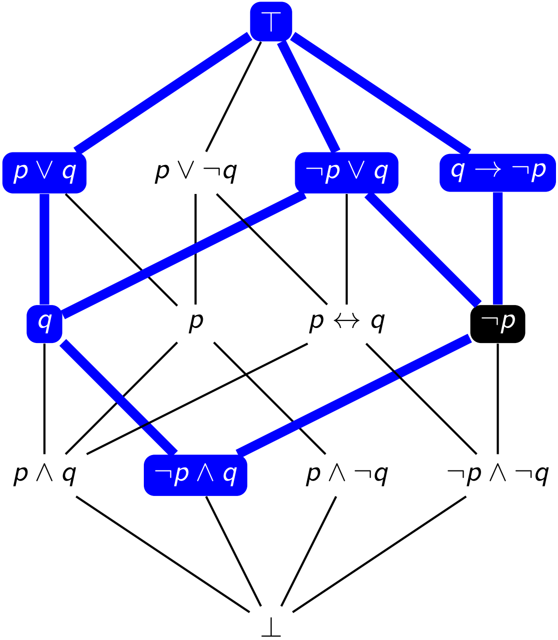

$\top$
$p \vee q$
$p \vee \neg q$
$\neg p \vee q$
$q \to \neg p$
q
p
$p \leftrightarrow q$
$\neg p$
$p \wedge q$
$\neg p \wedge q$
$p \wedge \neg q$
$\neg p \wedge \neg q$
$\bot$



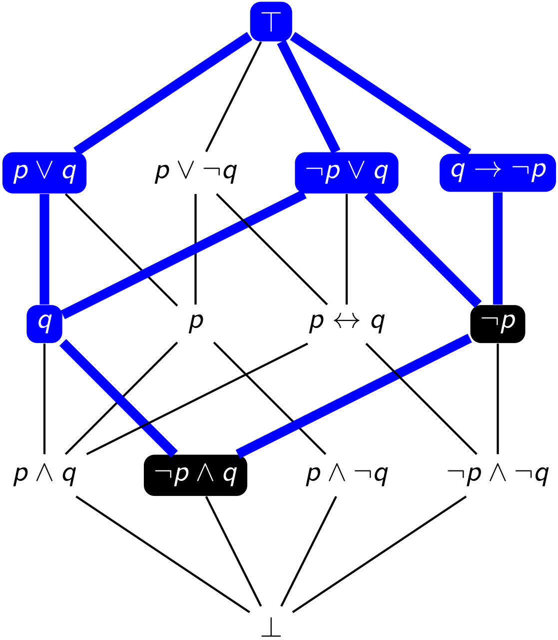
$$q < q \to \neg p$$



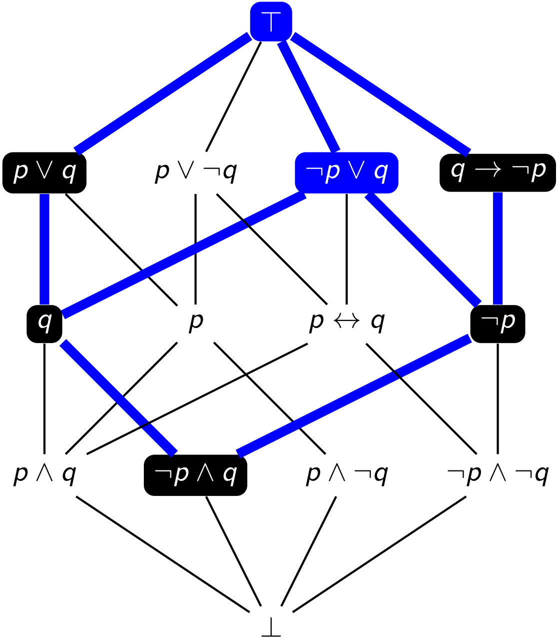
$K = Cn(\neg p \wedge q)$



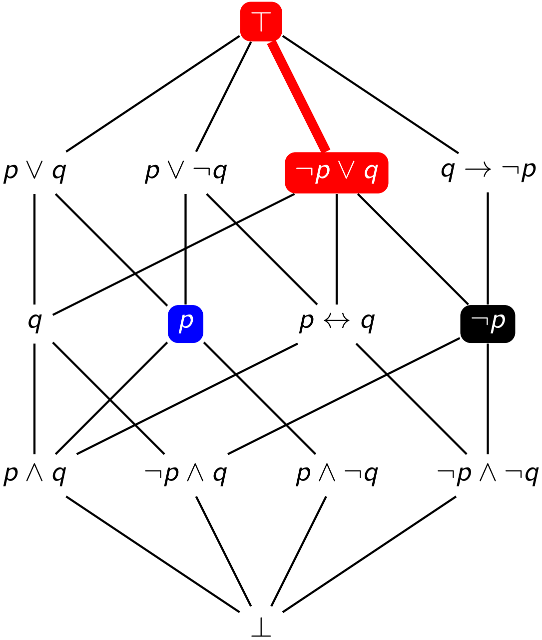
$K = Cn(\neg p \wedge q)$
$K * p =$
$Cn(\{\varphi \in K \mid \neg p < \varphi\} \cup \{p\})$



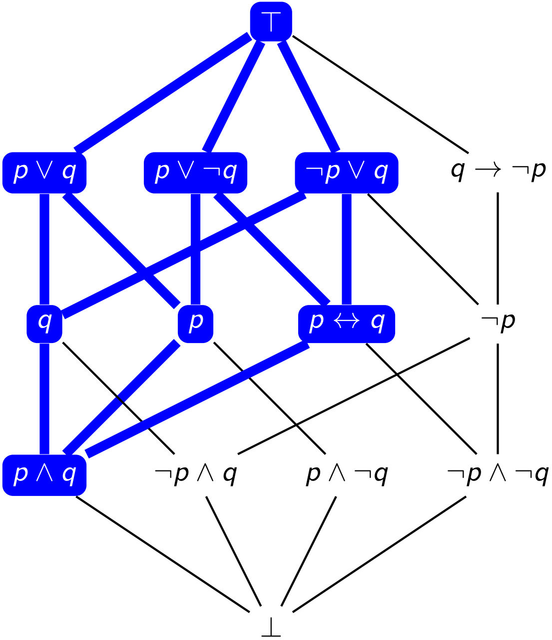
$K = Cn(\neg p \wedge q)$
$K * p =$
$Cn(\{\varphi \in K \mid \neg p < \varphi\} \cup \{p\})$



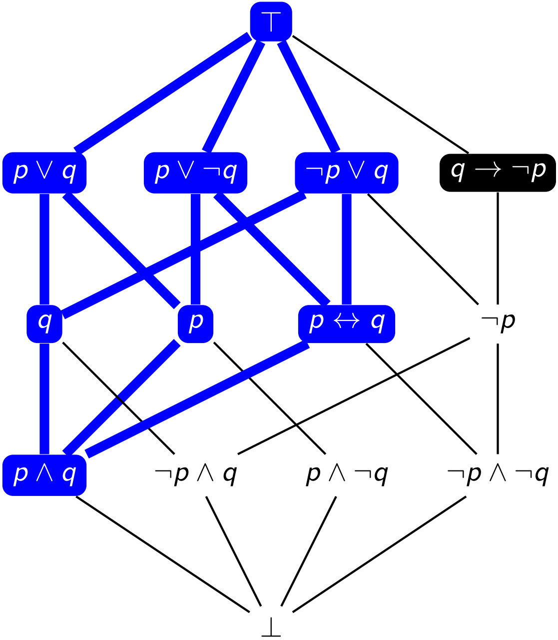
$K = Cn(\neg p \wedge q)$
$K * p =$
$Cn(\{\varphi \in K \mid \neg p < \varphi\} \cup \{p\})$



$K = Cn(\neg p \wedge q)$
$K * p =$
$Cn(\{\varphi \in K \mid \neg p < \varphi\} \cup \{p\})$



$K = Cn(\neg p \wedge q)$
$K * p =$
$Cn(\{\varphi \in K \mid \neg p < \varphi\} \cup \{p\})$
$= Cn(p \wedge q)$



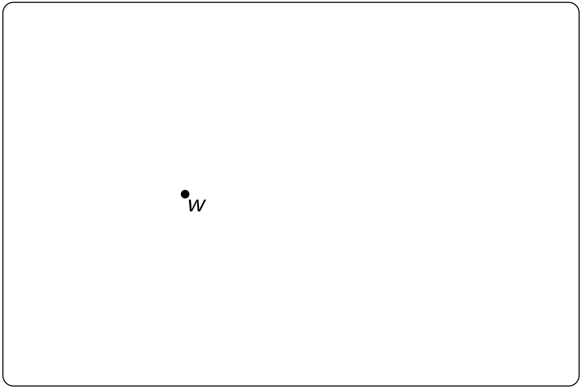
$K = Cn(\neg p \wedge q)$
$K * p =$
$= Cn(p \wedge q)$
$q < q \to \neg p$

Sphere models

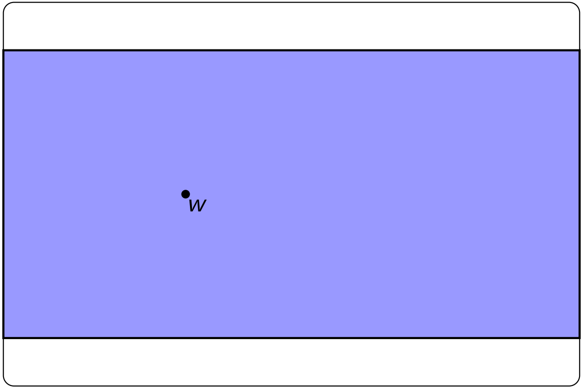
Belief Revision: The Semantic View

A. Grove. *Two modelings for theory change*. Journal of Philosophical Logic, 17, pgs. 157 - 170, 1988.

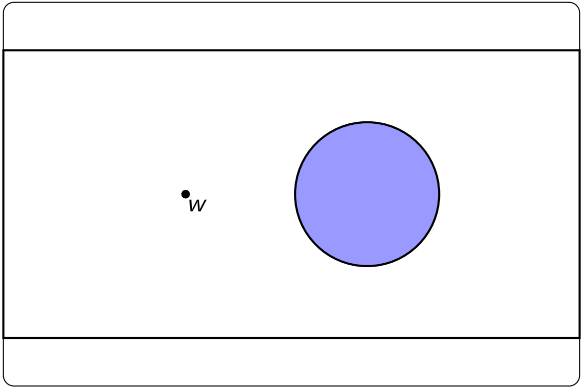
EP. *Dynamic Epistemic Logic II: Logics of information change*. Philosophy Compass, Vol. 8, Iss. 9, pgs. 815 - 833, 2013.



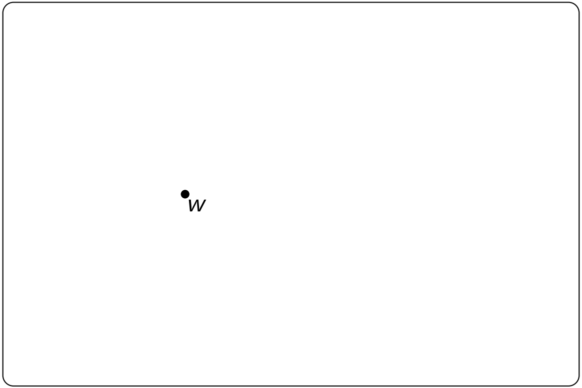
- The set of states, with a distinguished state denoted the "actual world"



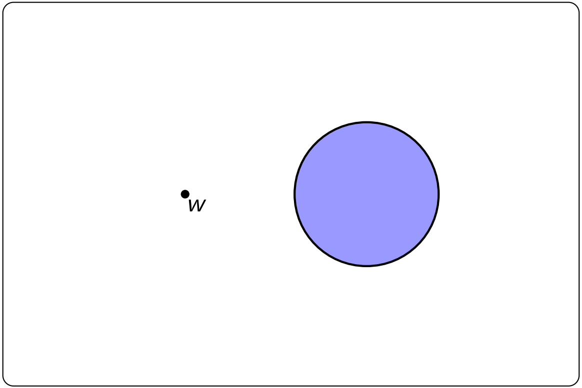
- The set of states, with a distinguished state denoted the "actual world"
- The agent's (hard) information (i.e., the states consistent with what the agent knows)



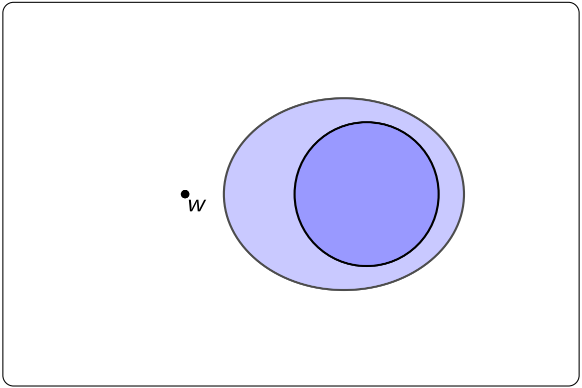
- The agent's (hard) information (i.e., the states consistent with what the agent knows)
- The agent's beliefs (soft information—-the states consistent with what the agent believes)



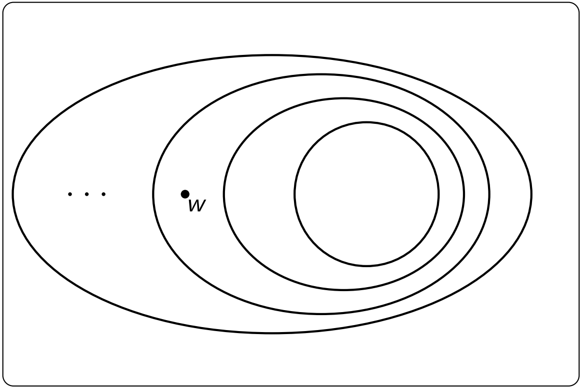
- The states consistent with what the agent knows with a distinguished state (the “actual world”)
- Each state is associated with a propositional valuation for the underlying propositional language



- The agent's beliefs (soft information—-the states consistent with what the agent believes)



- The agent's beliefs (soft information—the states consistent with what the agent believes)
- The agent's "contingency plan": when the stronger beliefs fail, go with the weaker ones.



- The agent's beliefs (soft information—the states consistent with what the agent believes)
- The agent's "contingency plan": when the stronger beliefs fail, go with the weaker ones.

Sphere Models

Let W be a set of states, A set $\mathcal{F} \subseteq \wp(W)$ is called a system of spheres provided:

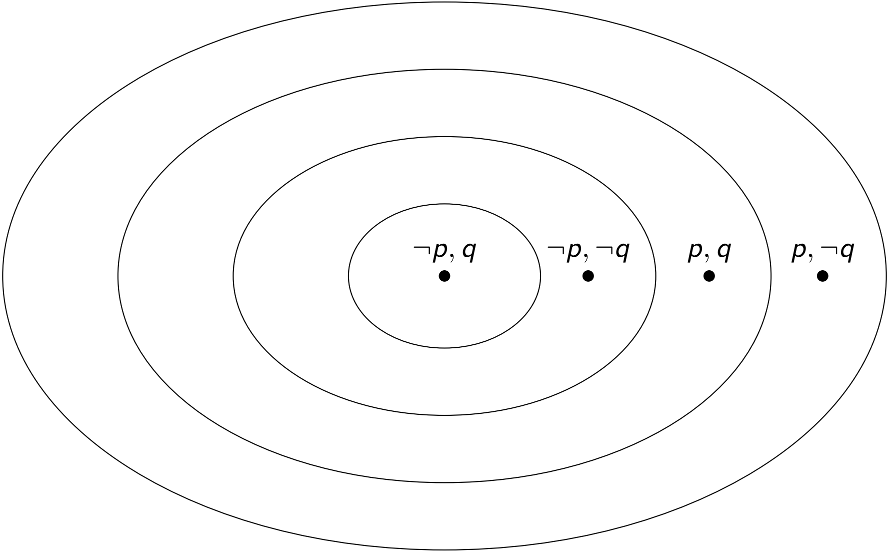
- For each $S, S' \in \mathcal{F}$, either $S \subseteq S'$ or $S' \subseteq S$
- For any $P \subseteq W$ there is a smallest $S \in \mathcal{F}$ (according to the subset relation) such that $P \cap S \neq \emptyset$
- The spheres are non-empty $\bigcap \mathcal{F} \neq \emptyset$ and cover the entire information cell $\bigcup \mathcal{F} = W$

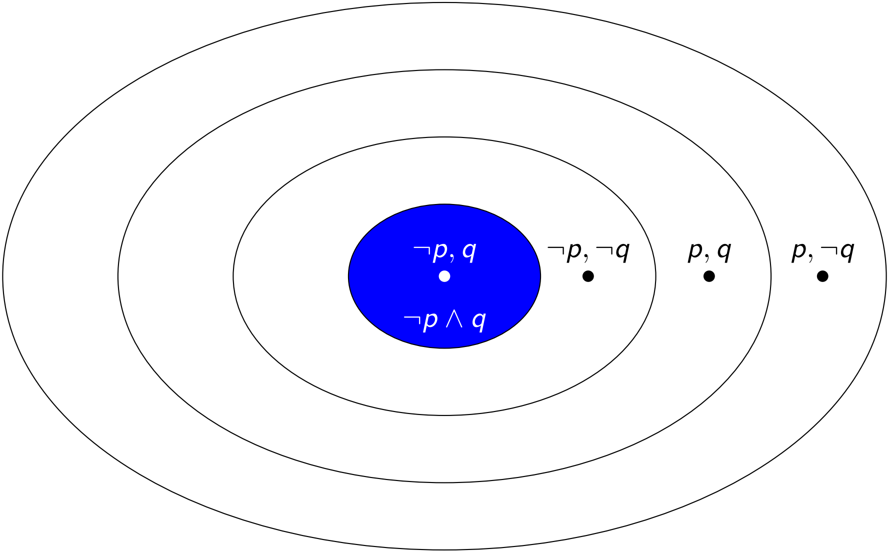
Let $\mathcal{F}$ be a system of spheres on W: for $w, v \in W$, let

$$w \preceq_{\mathcal{F}} v \text{ iff for all } S \in \mathcal{F}, \text{ if } v \in S \text{ then } w \in S$$

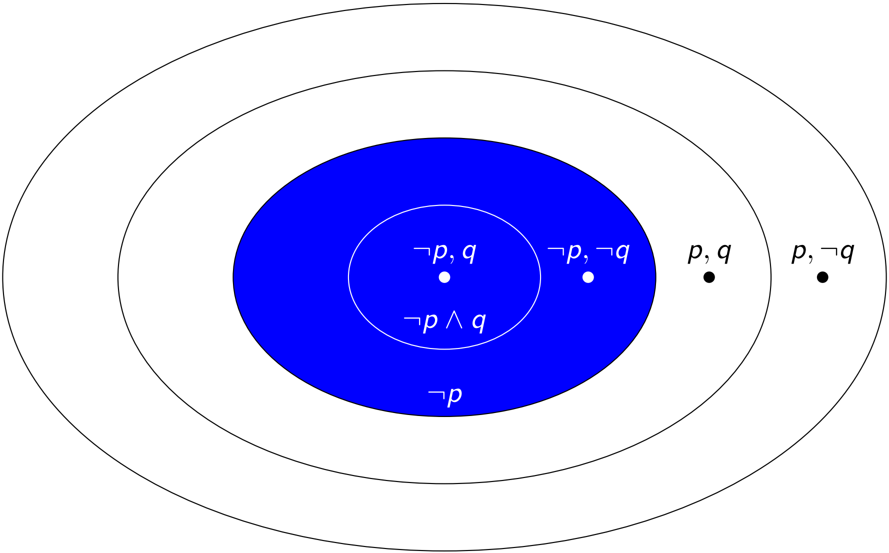
Then, $\preceq_{\mathcal{F}}$ is reflexive, transitive, and well-founded.

$w \preceq_{\mathcal{F}} v$ means that no matter what the agent learns in the future, as long as world v is still consistent with his beliefs and w is still epistemically possible, then w is also consistent with his beliefs.

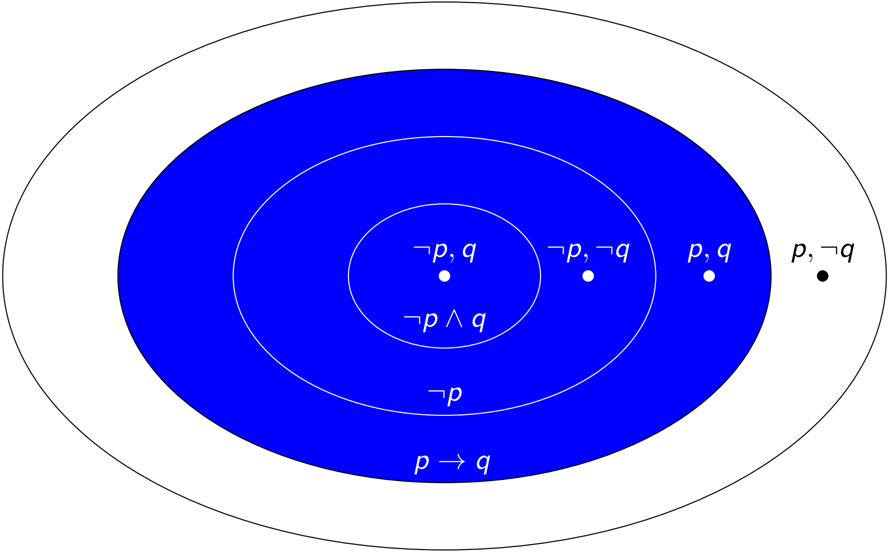

$\neg p, q$
$\neg p, \neg q$
p, q
$p, \neg q$



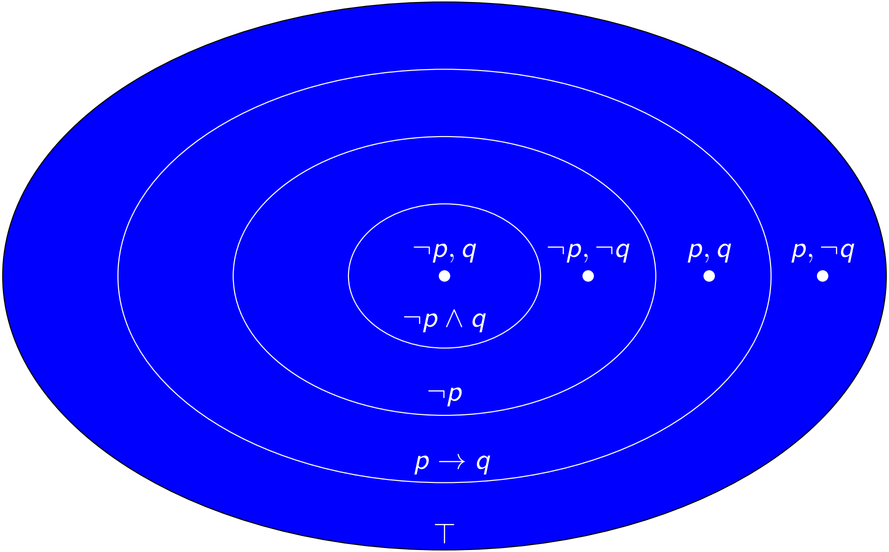
$$\bot \prec \neg p \wedge q \prec \neg p \prec p \to q \prec \top$$



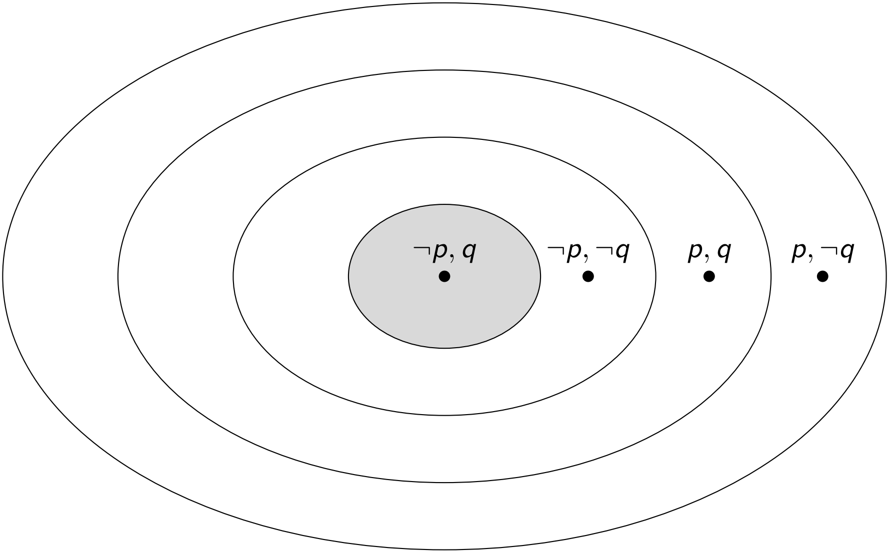
$$\bot \prec \neg p \wedge q \prec \neg p \prec p \to q \prec \top$$



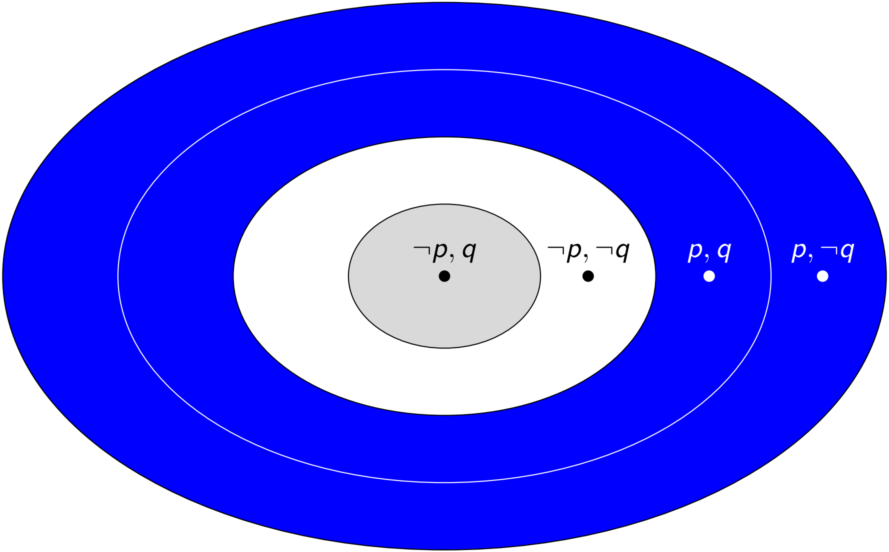
$$\bot \prec \neg p \wedge q \prec \neg p \prec p \to q \prec \top$$



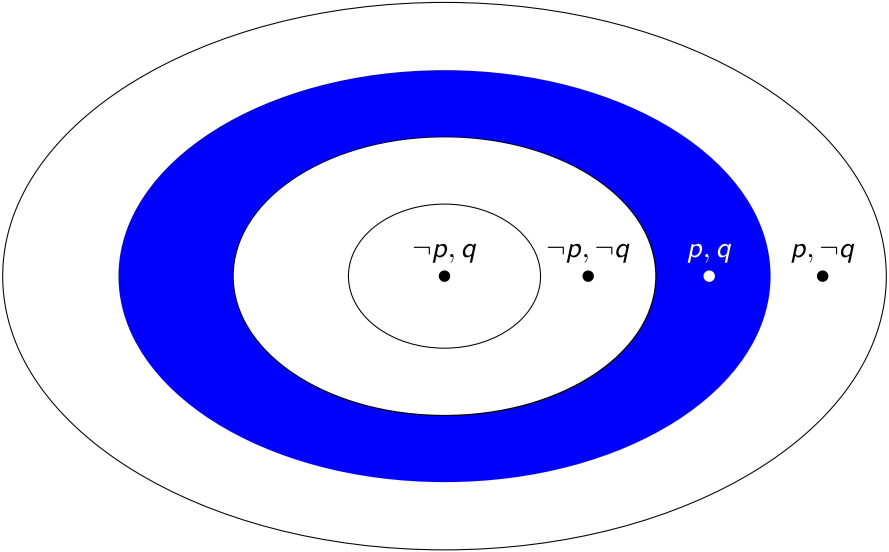
$$\perp \; \prec \; \neg p \wedge q \; \prec \; \neg p \; \prec \; p \rightarrow q \; \prec \; \top$$



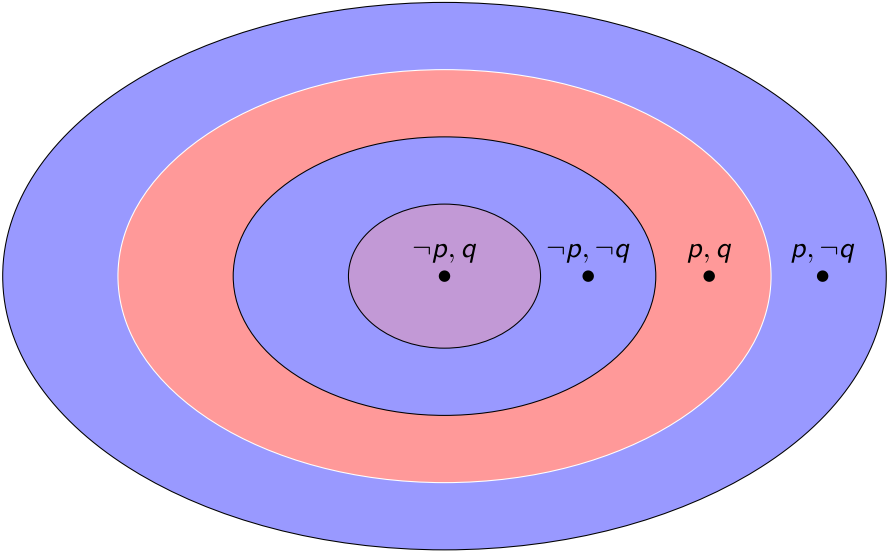
$K = Cn(\{\neg p, q\})$ $\quad K * p = Cn(\{p, q\})$ $\quad q < q \to \neg p$



$K = Cn(\{\neg p, q\}) \qquad K * p = Cn(\{p, q\}) \qquad q < q \to \neg p$



$K = Cn(\{\neg p, q\})$ $\quad K * p = Cn(\{p, q\})$ $\quad q < q \to \neg p$



$K = Cn(\{\neg p, q\}) \qquad K * p = Cn(\{p, q\}) \qquad q < q \to \neg p$

Bayesian Models

Rational Beliefs

Beliefs can represent the world more or less accurately....the more accurate the better.

Rational Beliefs

Beliefs can represent the world more or less accurately....the more accurate the better.

But we can also judge some beliefs as being more *rational* than others.

Rational Beliefs

Beliefs can represent the world more or less accurately....the more accurate the better.

But we can also judge some beliefs as being more *rational* than others.

Accuracy and rationality are linked, they are not the same:

Rational Beliefs

Beliefs can represent the world more or less accurately....the more accurate the better.

But we can also judge some beliefs as being more *rational* than others.

Accuracy and rationality are linked, they are not the same: a fool may hold a belief irrationally — as a result of a lucky guess or wishful thinking — yet it might happen to be correct.

Rational Beliefs

Beliefs can represent the world more or less accurately....the more accurate the better.

But we can also judge some beliefs as being more *rational* than others.

Accuracy and rationality are linked, they are not the same: a fool may hold a belief irrationally — as a result of a lucky guess or wishful thinking — yet it might happen to be correct. Conversely, a detective might hold a belief on the basis of a careful and exhaustive examination of all the evidence and yet the evidence may be misleading, and the belief may turn out to be wrong.

Conceptions of Belief

Binary: "all-out" belief. For any statement p, the agent either does or does not believe p. It is natural to take an *unqualified* assertion as a statement of belief of the speaker.

Conceptions of Belief

Binary: "all-out" belief. For any statement p, the agent either does or does not believe p. It is natural to take an *unqualified* assertion as a statement of belief of the speaker.

Graded: beliefs come in degrees. We are *more confident* in some of our beliefs than in others.

Conceptions of Belief

Binary: "all-out" belief. For any statement p, the agent either does or does not believe p. It is natural to take an *unqualified* assertion as a statement of belief of the speaker.

Graded: beliefs come in degrees. We are *more confident* in some of our beliefs than in others.

Eric Schwitzgebel. *Belief*. In The Stanford Encyclopedia of Philosophy.

Franz Huber. *Formal Theories of Belief*. In The Stanford Encyclopedia of Philosophy.

Conceptions of Beliefs: Questions

What are the *formal constraints* on rational belief?

- rational graded beliefs should obey the laws of probability

Conceptions of Beliefs: Questions

What are the *formal constraints* on rational belief?

- rational graded beliefs should obey the laws of probability
- rational all-out beliefs should be consistent/deductively closed

Conceptions of Beliefs: Questions

What are the *formal constraints* on rational belief?

- rational graded beliefs should obey the laws of probability
- rational all-out beliefs should be consistent/deductively closed
- how should we justify these constraints?

D. Christensen. *Putting Logic in its Place*. Oxford University Press.

Probability

Kolmogorov Axioms:

1. For each E, $0 \leq p(E) \leq 1$
2. $p(W) = 1$, $p(\emptyset) = 0$
3. If $E_1, \ldots, E_n, \ldots$ are pairwise disjoint ($E_i \cap E_j = \emptyset$ for $i \neq j$), then $p(\bigcup_i E_i) = \sum_i p(E_i)$

Probability

Kolmogorov Axioms:

1. For each E, $0 \leq p(E) \leq 1$
2. $p(W) = 1$, $p(\emptyset) = 0$
3. If $E_1, \ldots, E_n, \ldots$ are pairwise disjoint ($E_i \cap E_j = \emptyset$ for $i \neq j$), then $p(\bigcup_i E_i) = \sum_i p(E_i)$

- $p(\overline{E}) = 1 - p(E)$ ($\overline{E}$ is the complement of E)
- If $E \subseteq F$ then $p(E) \leq p(F)$
- $p(E \cup F) = p(E) + p(F) + p(E \cap F)$

Conditional Probability

The probability of E *given* F, dented $p(E|F)$, is defined to be

$$p(E|F) = \frac{p(E \cap F)}{p(F)}.$$

Conditional Probability

The probability of E *given* F, dented $p(E|F)$, is defined to be

$$p(E|F) = \frac{p(E \cap F)}{p(F)}.$$

Bayes Theorem: $p(E|F) = p(F|E)\frac{p(E)}{p(F)}$

Bayes theorem is important because it expresses the quantity $p(E|F)$ (the probability of a hypothesis E given the evidence F) — which is something people often find hard to assess — in terms of quantities that can be drawn directly from experiential knowledge.

Bayes theorem is important because it expresses the quantity $p(E|F)$ (the probability of a hypothesis E given the evidence F) — which is something people often find hard to assess — in terms of quantities that can be drawn directly from experiential knowledge.

Example: Suppose you are in a casino and you hear a person at the next gambling table announce "Twelve". We want to know wether he was rolling a pair of dice or a roulette wheel.

Bayes theorem is important because it expresses the quantity $p(E|F)$ (the probability of a hypothesis E given the evidence F) — which is something people often find hard to assess — in terms of quantities that can be drawn directly from experiential knowledge.

Example: Suppose you are in a casino and you hear a person at the next gambling table announce "Twelve". We want to know wether he was rolling a pair of dice or a roulette wheel. That is, compare $p(Dice \mid Twelve)$ with $p(Roulette \mid Twelve)$.

Bayes theorem is important because it expresses the quantity $p(E|F)$ (the probability of a hypothesis E given the evidence F) — which is something people often find hard to assess — in terms of quantities that can be drawn directly from experiential knowledge.

Example: Suppose you are in a casino and you hear a person at the next gambling table announce "Twelve". We want to know wether he was rolling a pair of dice or a roulette wheel. That is, compare $p(Dice \mid Twelve)$ with $p(Roulette \mid Twelve)$. Based on our background knowledge of gambling we have $p(Twelve \mid Dice) = 1/36$ and $p(Twelve \mid Roulette) = 1/38$.

Bayes theorem is important because it expresses the quantity $p(E|F)$ (the probability of a hypothesis E given the evidence F) — which is something people often find hard to assess — in terms of quantities that can be drawn directly from experiential knowledge.

Example: Suppose you are in a casino and you hear a person at the next gambling table announce "Twelve". We want to know wether he was rolling a pair of dice or a roulette wheel. That is, compare $p(\mathit{Dice} \mid \mathit{Twelve})$ with $p(\mathit{Roulette} \mid \mathit{Twelve})$. Based on our background knowledge of gambling we have $p(\mathit{Twelve} \mid \mathit{Dice}) = 1/36$ and $p(\mathit{Twelve} \mid \mathit{Roulette}) = 1/38$. Based on our observations about the casino, we can judge the prior probabilities $p(\mathit{Dice})$ and $p(\mathit{Roulette})$.

Bayes theorem is important because it expresses the quantity $p(E|F)$ (the probability of a hypothesis E given the evidence F) — which is something people often find hard to assess — in terms of quantities that can be drawn directly from experiential knowledge.

Example: Suppose you are in a casino and you hear a person at the next gambling table announce "Twelve". We want to know wether he was rolling a pair of dice or a roulette wheel. That is, compare $p(Dice \mid Twelve)$ with $p(Roulette \mid Twelve)$. Based on our background knowledge of gambling we have $p(Twelve \mid Dice) = 1/36$ and $p(Twelve \mid Roulette) = 1/38$. Based on our observations about the casino, we can judge the prior probabilities $p(Dice)$ and $p(Roulette)$. But this is now enough to *calculate* the required probabilities.

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's reasoning

Consider the following events:

G_A: "Prisoner A will be declared guilty" (we have $p(G_A) = 1/3$)

I_B: "Prisoner B will be declared innocent" (we have $p(I_B) = 2/3$)

A's reasoning

Consider the following events:

G_A: "Prisoner A will be declared guilty" (we have $p(G_A) = 1/3$)

I_B: "Prisoner B will be declared innocent" (we have $p(I_B) = 2/3$)

We have $p(I_B \mid G_A) = 1$: "If A is declared guilty then B will be declared innocent."

A's reasoning

Consider the following events:

G_A: "Prisoner A will be declared guilty" (we have $p(G_A) = 1/3$)

I_B: "Prisoner B will be declared innocent" (we have $p(I_B) = 2/3$)

We have $p(I_B \mid G_A) = 1$: "If A is declared guilty then B will be declared innocent."

Bayes Theorem:

$$p(G_A \mid I_B) =$$

A's reasoning

Consider the following events:

G_A: "Prisoner A will be declared guilty" (we have $p(G_A) = 1/3$)

I_B: "Prisoner B will be declared innocent" (we have $p(I_B) = 2/3$)

We have $p(I_B \mid G_A) = 1$: "If A is declared guilty then B will be declared innocent."

Bayes Theorem:

$$p(G_A \mid I_B) = p(I_B \mid G_A)\frac{p(G_A)}{p(I_B)} =$$

A's reasoning

Consider the following events:

G_A: "Prisoner A will be declared guilty" (we have $p(G_A) = 1/3$)

I_B: "Prisoner B will be declared innocent" (we have $p(I_B) = 2/3$)

We have $p(I_B \mid G_A) = 1$: "If A is declared guilty then B will be declared innocent."

Bayes Theorem:

$$p(G_A \mid I_B) = p(I_B \mid G_A)\frac{p(G_A)}{p(I_B)} = 1 \cdot \frac{1/3}{2/3} = 1/2$$

A's reasoning, corrected

But, A did not receive the information that B will be declared innocent, but rather that "the guard said that B will be declared innocent." So, A should have conditioned on the event:

I'_B: "The guard said that B will be declared innocent"

A's reasoning, corrected

But, A did not receive the information that B will be declared innocent, but rather that "the guard said that B will be declared innocent." So, A should have conditioned on the event:

I'_B: "The guard said that B will be declared innocent"

Given that $p(I'_B \mid G_A)$ is $1/2$ (given that A is guilty, there is a 50-50 chance that the guard could have given the letter to B or C).

A's reasoning, corrected

But, A did not receive the information that B will be declared innocent, but rather that "the guard said that B will be declared innocent." So, A should have conditioned on the event:

I'_B: "The guard said that B will be declared innocent"

Given that $p(I'_B \mid G_A)$ is $1/2$ (given that A is guilty, there is a 50-50 chance that the guard could have given the letter to B or C). This gives us the following correct calculation:

$$p(G_A \mid I'_B) = p(I'_B \mid G_A)\frac{p(G_A)}{p(I'_B)} = 1/2 \cdot \frac{1/3}{1/2} = 1/3$$

When does conditioning on the "naive" space give the same results as conditioning on the "sophisticated" space?

Revising probabilities

- *Generalized conditioning*: Jeffrey conditioning, Adams conditioning, MRE (Maximum relative entropy)
- *Weakening probability measures*: finite-additive measures, plausibility measures, Dempster-Shafer belief functions, Sphon measures
- *Generalizations of probability*: Lexicographic probability measures, conditional probability measures, non-standard probability measures

CPS (Popper Space)

A **conditional probability space** (CPS) over $(W, \mathfrak{A})$ is a tuple $(W, \mathfrak{A}, \mathfrak{B}, \mu)$ such that $\mathfrak{A}$ is an algebra over W, $\mathfrak{B}$ is a set of subsets of W (not necessarily an algebra) that does not contain $\emptyset$ and $\mu : \mathfrak{A} \times \mathfrak{B} \to [0, 1]$ satisfying the following conditions:

1. $\mu(U \mid U) = 1$ if $U \in \mathfrak{B}$
2. $\mu(E_1 \cup E_1 \mid U) = \mu(E_1 \mid U) + \mu(E_2 \mid U)$ if $E_1 \cap E_2 = \emptyset$, $U \in \mathfrak{B}$ and $E_1, E_2 \in \mathfrak{A}$
3. $\mu(E \mid U) = \mu(E \mid X) * \mu(X \mid U)$ if $E \subseteq X \subseteq U$, $U, X \in \mathfrak{B}$ and $E \in \mathfrak{A}$.

LPS (Lexicographic Probability Space)

A **lexicographic probability space** (LPS) (of length α) is a tuple $(W, \mathcal{F}, \vec{\mu})$ where W is a set of possible worlds, $\mathcal{F}$ is an algebra over W and $\vec{\mu}$ is a sequence of (finitely/countable additive) probability measures on $(W, \mathcal{F})$ indexed by ordinals $< \alpha$.

Fix an *LPS* $\vec{\mu} = (\mu_0, \ldots, \mu_n)$

- E is certain: $\mu_0(E) = 1$

Fix an *LPS* $\vec{\mu} = (\mu_0, \ldots, \mu_n)$

- E is certain: $\mu_0(E) = 1$
- E is absolutely certain: $\mu_i(E) = 1$ for all $i = 1, \ldots, n$

Fix an *LPS* $\vec{\mu} = (\mu_0, \ldots, \mu_n)$

- E is certain: $\mu_0(E) = 1$
- E is absolutely certain: $\mu_i(E) = 1$ for all $i = 1, \ldots, n$
- E is *assumed*: there exists k such that $\mu_i(E) = 1$ for all $i \leq k$ and $\mu_i(E) = 0$ for all $k < i < n$.

NPS (non-standard probability measures)

$\mathbb{R}^*$ is a *non-Archimedean* field that includes the real numbers as a subfield but also has *infinitesimals*.

For all $b \in \mathbb{R}^*$ such that $-r < b < r$ for some $r \in \mathbb{R}$, there is a unique closest real number a such that $|a - b|$ is an infinitesimal. Let $st(b)$ denote the closest standard real to b.

NPS (non-standard probability measures)

$\mathbb{R}^*$ is a *non-Archimedean* field that includes the real numbers as a subfield but also has *infinitesimals*.

For all $b \in \mathbb{R}^*$ such that $-r < b < r$ for some $r \in \mathbb{R}$, there is a unique closest real number a such that $|a - b|$ is an infinitesimal. Let $st(b)$ denote the closest standard real to b.

A **nonstandard probability space** (NPS) is a tuple $(W, \mathcal{F}, \mu)$ where W is a set of possible worlds, $\mathcal{F}$ is an algebra over W and μ assigns to elements of $\mathcal{F}$, nonnegative elements of $\mathbb{R}^*$ such that $\mu(W) = 1$, $\mu(E \cup F) = \mu(E) + \mu(F)$ if E and F are disjoint.

J. Halpern. *Lexicographic probability, conditional probability, and nonstandard probability*. Games and Economic Behavior, 68:1, pgs. 155 - 179, 2010.

Indeterminate Probability

- Allow probability functions to take on sets of values instead of a single value
- Work with sets of probabilities rather than a single probability

Precisification Given a function $\sigma : \mathcal{F} \rightarrow \wp([0,1])$, a probability function $p : \mathcal{F} \rightarrow [0,1]$ of σ if and only if $p(A) \in \sigma(A)$ for each $A \in \mathcal{F}$.

Indeterminate Probability A function $\sigma : \mathcal{F} \rightarrow \wp([0,1])$ such that whenever $x \in \sigma(A)$ there is some precisifcation of σ, p for which $p(A) = x$.

Ambiguation If Π is a set of probability functions, the *ambiguation* of Π is the indeterminate probability function that assigns to each A

$$\sigma(A) = \{x \mid p(A) = x \text{ for some } p \in \Pi\}$$

Ambiguation If Π is a set of probability functions, the *ambiguation* of Π is the indeterminate probability function that assigns to each A

$$\sigma(A) = \{x \mid p(A) = x \text{ for some } p \in \Pi\}$$

Observation. The map that takes and indeterminate probability function to the class of its precisfications is clearly 1-1. However, the ambiguation of a set of probability functions can have precisfications not in the ambiguated set.

Convexity A class of probability functions Π is **convex** if and only if whenever $p, q \in \Pi$, every mixture of p and q is in Π as well. I.e., $\alpha p + (1-\alpha)q \in \Pi$ for all $\alpha \in (0,1)$.

Proposition. If P is convex with σ it ambiguation, then $\sigma(A)$ is an interval for each A.

Upper and Lower Probabilities

If σ is an indeterminate probability function, define

- **Lower probability**: $\sigma_*(A) = \inf\{x \mid x \in \sigma(A)\}$
- **Upper probability**: $\sigma^*(A) = \sup\{x \mid x \in \sigma(A)\}$

How do qualitative and quantitative belief relate to each other?

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

In view of the fact that we have a reasonably clear picture of what the logics of qualitative and quantitative belief are like, what conclusions can we draw form this on how qualitative and quantitative belief ought to relate to each other, assuming that they satisfy their respective logics? How do they relate to each other in the case of an agent who is a perfect reasoner?

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."

Topics

- Classic papers (Makinson, Diaconis & Zabel, KLM, ...)
- Beliefs, credences and probability (Leitgeb's stability theory of belief, Pettigrew, Fitelson & Shear)
- Revising probabilities (List, Dietrich & Bradley, Halpern)
- Conditioning vs. learning (Osherson et al., Curpi et al.)
- Context shifts (Halpern & Grünwald, Romeijn, Pettigrew)
- Lottery, Preface and Review paradox (Leitgeb, Easerwen & Fitelson)
- Iterated belief change, long-term dynamics, convergence results (Huttegger, EP)
- Bayesian reasoning, reasoning to the best explanation, case-base reasoning (Gilboa et al., Douven and Shubach)

Preface Paradox

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

Preface Paradox

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

Preface Paradox

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

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

Preface Paradox

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

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

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

Preface Paradox

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

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

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

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

Preface Paradox

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

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

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

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

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

Preface Paradox

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

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

Preface Paradox

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

1\. does not necessarily imply 2.

Preface Paradox: The Problem

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

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

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

Bel_t is the set of propositions believed at time t

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

$t' > t$

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

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

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

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

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

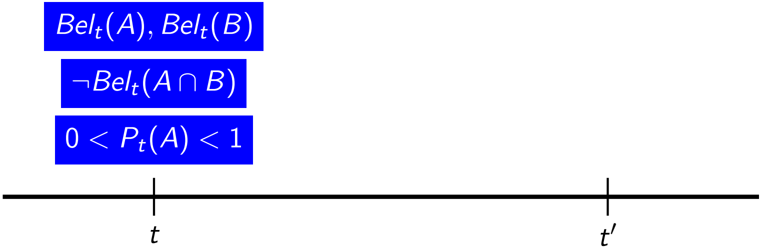
P3 WHen the agent learns, this is captured probabilistically by conditionalization.

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

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

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

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



Assumption

$Bel_t(A), Bel_t(B)$

$\neg Bel_t(A \cap B)$

$0 < P_t(A) < 1$

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

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

t — Receives evidence A — t'

By P3

$Bel_t(A), Bel_t(B)$

$\neg Bel_t(A \cap B)$

$0 < P_t(A) < 1$

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

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

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

t — Receives evidence A — t'

By P1

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

$\neg Bel_t(A \cap B)$

$0 < P_t(A) < 1$

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

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

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

t Receives evidence A t'

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

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

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

By P2

$Bel_t(A)$

$\neg Bel_t(A \cap B)$

$0 < P_t(A) < 1$

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

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

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

t Receives evidence A t'

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

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

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

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

Lottery Paradox

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

I. Douven and T. Williamson. *Generalizing the Lottery Paradox*. British Journal of the Philosophy of Science, 57, 755 - 779, 2006.

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

Lottery Paradox

Lottery Paradox

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

Lottery Paradox

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

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

Lottery Paradox

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

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

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

Lottery Paradox

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

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

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

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

Lottery Paradox

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

Lottery Paradox

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

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

Lottery Paradox

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

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

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

The Lottery Paradox

Kyburg: The following are inconsistent,

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

Subjective Probabilities

Should a rational agent's graded beliefs satisfy the laws of probability?

J. Joyce. *Bayesianism*. in Handbook of Rationality.

Ann: "the probability it will rain tomorrow is 0.9" means "Ann's degree of belief is fairly high (0.9) that it will rain tomorrow. Of course whether it will actually rain, depends on objective events taking place in the external worlds."

Ramsey, de Finetti and Savage (1)

How do we *measure* a (rational) agent's subjective probabilities?

Ramsey, de Finetti and Savage (1)

How do we *measure* a (rational) agent's subjective probabilities?

Suppose we are wondering about Ann's degree of belief about whether a coin will land heads (H) or tails (T).

Ramsey, de Finetti and Savage (1)

How do we *measure* a (rational) agent's subjective probabilities?

Suppose we are wondering about Ann's degree of belief about whether a coin will land heads (H) or tails (T).

Why don't we just *ask* her?

Ramsey, de Finetti and Savage (1)

How do we *measure* a (rational) agent's subjective probabilities?

Suppose we are wondering about Ann's degree of belief about whether a coin will land heads (H) or tails (T).

Why don't we just *ask* her? reported vs. "actual" degrees of belief.

Ramsey, de Finetti and Savage (1)

How do we *measure* a (rational) agent's subjective probabilities?

Suppose we are wondering about Ann's degree of belief about whether a coin will land heads (H) or tails (T).

Why don't we just *ask* her? reported vs. "actual" degrees of belief.

What we need: systematic procedures for linking the probability calculus (graded beliefs) to claims about objectively observable behavior, such as preferences revealed by choice behavior.

Ramsey, de Finetti and Savage (2)

Suppose we are wondering about Ann's degree of belief about whether a coin will land heads (H) or tails (T).

Offer Ann two bets:

- L_1 If the coin lands heads, you win a sports car; otherwise you win nothing
- L_2 If the coin does not land heads, you win a sports car; otherwise you win nothing.

Ramsey, de Finetti and Savage (2)

Suppose we are wondering about Ann's degree of belief about whether a coin will land heads (H) or tails (T).

Offer Ann two bets:

- L_1 If the coin lands heads, you win a sports car; otherwise you win nothing
- L_2 If the coin does not land heads, you win a sports car; otherwise you win nothing.

If Ann chooses L_1, she believes H is more probable than T
If Ann chooses L_2, she believes T is more probable than H
If Ann is indifferent, she believes H and T are equally probable (i.e., $p_A(H) = p_A(T) = 1/2$)

The Dutch Book Argument

But, why *should* a rational agent's graded beliefs satisfy the Kolmogorov axioms?

The Dutch Book Argument

But, why *should* a rational agent's graded beliefs satisfy the Kolmogorov axioms?

Anyone whose beliefs violate the laws of probability is *practically irrational*.

F. P. Ramsey. *Truth and Probability*. 1931.

B. de Finetti. *La prévision: Ses lois logiques, ses sources subjectives*. 1937.

Alan Hájek. *Dutch Book Arguments*. Oxford Handbook of Rational and Social Choice, 2008.

The EU-Thesis

Expected Money/Value/Utility: Given an agent's beliefs and desires, the **expected utility** of an **action** leading to a set of outcomes *Out* is:

$$\sum_{o \in Out} \text{[how likely the act will lead to } o\text{]} \times \text{[how much the agent desires } o\text{]}$$

The EU-Thesis

Expected Money/Value/Utility: Given an agent's beliefs and desires, the **expected utility** of an **action** leading to a set of outcomes *Out* is:

$$\sum_{o \in Out} \text{[how likely the act will lead to } o] \times \text{[how much the agent desires } o]$$

1. principle of maximizing expected monetary value
2. principle of maximizing expected value
3. principle of maximizing expected utility

Betting Behavior

The *EU*-thesis entails that a person satisfying 1-3 will reveal the strengths of her beliefs in her betting behavior.

Betting Behavior

The *EU*-thesis entails that a person satisfying 1-3 will reveal the strengths of her beliefs in her betting behavior.

A **wager**: $W_X = [\ a \text{ if } X,\ b \text{ otherwise}]$: "you get a EUR if X is true and b EUR otherwise.
(X's truth does not depend causally on W)

Betting Behavior

The *EU*-thesis entails that a person satisfying 1-3 will reveal the strengths of her beliefs in her betting behavior.

A **wager**: $W_X = [\ a \text{ if } X,\ b \text{ otherwise}]$: "you get a EUR if X is true and b EUR otherwise.
(X's truth does not depend causally on W)

The EU-thesis entails that the agent's level of confidence in X will be revealed by the monetary value she puts on W_X.

Betting Behavior

fair price f **for** W_X: the sum of money at which she is indifferent between receiving a payment of f EUR or having W_X go into effect.

Betting Behavior

fair price f **for** W_X: the sum of money at which she is indifferent between receiving a payment of f EUR or having W_X go into effect.

$f = ExpVal(W_X) = C(X) \cdot a + (1 - C(X)) \cdot b$ implies $C(X) = \frac{f-b}{a-b}$

Betting Behavior

fair price f **for** W_X: the sum of money at which she is indifferent between receiving a payment of f EUR or having W_X go into effect.

$$f = ExpVal(W_X) = C(X) \cdot a + (1 - C(X)) \cdot b \text{ implies } C(X) = \frac{f-b}{a-b}$$

If she is indifferent between $63,81$ EUR and [100 EUR if it rains, 0 EUR otherwise], then she believes to degree 0.6381 that it will rain.

Dutch Book

An agent will swap an (set of) wagers with the (sum of) their fair prices.

Dutch Book

Suppose that X and Y are logically incompatible ($X \cap Y = \emptyset$)

Dutch Book

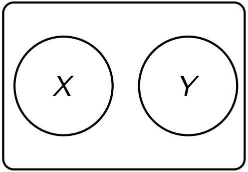
Suppose that X and Y are logically incompatible ($X \cap Y = \emptyset$)

$$W_X = [1 \text{ if } X,\ 0 \text{ else}]$$
$$W_Y = [1 \text{ if } Y,\ 0 \text{ else}]$$
$$W_{X \vee Y} = [1 \text{ if } X \vee Y,\ 0 \text{ else}]$$

Dutch Book

Suppose that X and Y are logically incompatible ($X \cap Y = \emptyset$)

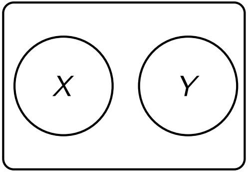
$$W_X = [1 \text{ if } X, \text{ 0 else}]$$
$$W_Y = [1 \text{ if } Y, \text{ 0 else}]$$
$$W_{X \vee Y} = [1 \text{ if } X \vee Y, \text{ 0 else}]$$



Dutch Book

Suppose that X and Y are logically incompatible ($X \cap Y = \emptyset$)

Fair price: $f = 0.25$ for $W_X = [1 \text{ if } X, 0 \text{ else}]$
Fair price: $f = 0.25$ for $W_Y = [1 \text{ if } Y, 0 \text{ else}]$
Fair price: $f = 0.6$ for $W_{X \vee Y} = [1 \text{ if } X \vee Y, 0 \text{ else}]$



Dutch Book

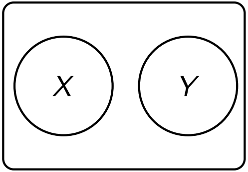
Suppose that X and Y are logically incompatible ($X \cap Y = \emptyset$)

Fair price: $f = 0.25$ for $W_X = [1 \text{ if } X, 0 \text{ else}]$

Fair price: $f = 0.25$ for $W_Y = [1 \text{ if } Y, 0 \text{ else}]$

Fair price: $f = 0.6$ for $W_{X \vee Y} = [1 \text{ if } X \vee Y, 0 \text{ else}]$

Consider $\mathcal{W}_1 = \{0.6, W_X, W_Y\}$ and $\mathcal{W}_2 = \{0.5, W_{X \vee Y}\}$



Dutch Book

Suppose that X and Y are logically incompatible ($X \cap Y = \emptyset$)

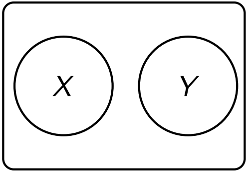
Fair price: $f = 0.25$ for $W_X = [1 \text{ if } X, 0 \text{ else}]$

Fair price: $f = 0.25$ for $W_Y = [1 \text{ if } Y, 0 \text{ else}]$

Fair price: $f = 0.6$ for $W_{X \vee Y} = [1 \text{ if } X \vee Y, 0 \text{ else}]$

Consider $\mathcal{W}_1 = \{0.6, W_X, W_Y\}$ and $\mathcal{W}_2 = \{0.5, W_{X \vee Y}\}$

- indifferent between $\mathcal{W}_1$ and $\mathcal{W}_2$



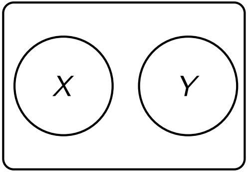
Dutch Book

Suppose that X and Y are logically incompatible ($X \cap Y = \emptyset$)

Fair price: $f = 0.25$ for $W_X = [1 \text{ if } X, 0 \text{ else}]$
Fair price: $f = 0.25$ for $W_Y = [1 \text{ if } Y, 0 \text{ else}]$
Fair price: $f = 0.6$ for $W_{X \vee Y} = [1 \text{ if } X \vee Y, 0 \text{ else}]$

Consider $\mathcal{W}_1 = \{0.6, W_X, W_Y\}$ and $\mathcal{W}_2 = \{0.5, W_{X \vee Y}\}$

- indifferent between $\mathcal{W}_1$ and $\mathcal{W}_2$
- swap $\mathcal{W}_1$ for $\mathcal{W}_2$

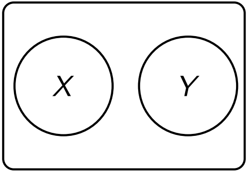


Dutch Book

Suppose that X and Y are logically incompatible ($X \cap Y = \emptyset$)

Fair price: $f = 0.25$ for $W_X = [1 \text{ if } X, 0 \text{ else}]$
Fair price: $f = 0.25$ for $W_Y = [1 \text{ if } Y, 0 \text{ else}]$
Fair price: $f = 0.6$ for $W_{X \vee Y} = [1 \text{ if } X \vee Y, 0 \text{ else}]$

Consider $\mathcal{W}_1 = \{0.6, W_X, W_Y\}$ and $\mathcal{W}_2 = \{0.5, W_{X \vee Y}\}$



- indifferent between $\mathcal{W}_1$ and $\mathcal{W}_2$
- swap $\mathcal{W}_1$ for $\mathcal{W}_2$
- But $\mathcal{W}_1$ is always better:

Dutch Book

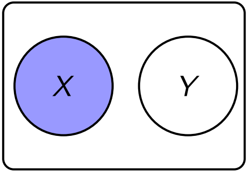
Suppose that X and Y are logically incompatible ($X \cap Y = \emptyset$)

Fair price: $f = 0.25$ for $W_X = [1 \text{ if } X, 0 \text{ else}]$

Fair price: $f = 0.25$ for $W_Y = [1 \text{ if } Y, 0 \text{ else}]$

Fair price: $f = 0.6$ for $W_{X \vee Y} = [1 \text{ if } X \vee Y, 0 \text{ else}]$

Consider $\mathcal{W}_1 = \{0.6, W_X, W_Y\}$ and $\mathcal{W}_2 = \{0.5, W_{X \vee Y}\}$



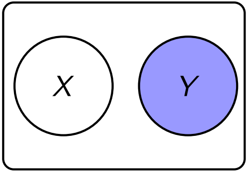
- indifferent between $\mathcal{W}_1$ and $\mathcal{W}_2$
- swap $\mathcal{W}_1$ for $\mathcal{W}_2$
- But $\mathcal{W}_1$ is always better:
  - If X is true
    payoff($\mathcal{W}_1$) $= 1.6 >$ payoff($\mathcal{W}_2$)$=1.5$

Dutch Book

Suppose that X and Y are logically incompatible ($X \cap Y = \emptyset$)

Fair price: $f = 0.25$ for $W_X = [1 \text{ if } X, 0 \text{ else}]$
Fair price: $f = 0.25$ for $W_Y = [1 \text{ if } Y, 0 \text{ else}]$
Fair price: $f = 0.6$ for $W_{X \vee Y} = [1 \text{ if } X \vee Y, 0 \text{ else}]$

Consider $\mathcal{W}_1 = \{0.6, W_X, W_Y\}$ and $\mathcal{W}_2 = \{0.5, W_{X \vee Y}\}$



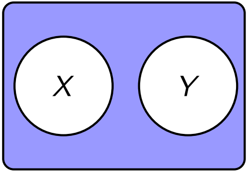
- indifferent between $\mathcal{W}_1$ and $\mathcal{W}_2$
- swap $\mathcal{W}_1$ for $\mathcal{W}_2$
- But $\mathcal{W}_1$ is always better:
  - If X is true
    payoff($\mathcal{W}_1$) $= 1.6 >$ payoff($\mathcal{W}_2$)$=1.5$
  - If Y is true
    payoff($\mathcal{W}_1$) $= 1.6 >$ payoff($\mathcal{W}_2$)$= 1.5$

Dutch Book

Suppose that X and Y are logically incompatible ($X \cap Y = \emptyset$)

Fair price: $f = 0.25$ for $W_X = [1 \text{ if } X, 0 \text{ else}]$
Fair price: $f = 0.25$ for $W_Y = [\ 1 \text{ if } Y, 0 \text{ else}]$
Fair price: $f = 0.6$ for $W_{X \vee Y} = [1 \text{ if } X \vee Y, 0 \text{ else}]$

Consider $\mathcal{W}_1 = \{0.6, W_X, W_Y\}$ and $\mathcal{W}_2 = \{0.5, W_{X \vee Y}\}$



- indifferent between $\mathcal{W}_1$ and $\mathcal{W}_2$
- swap $\mathcal{W}_1$ for $\mathcal{W}_2$
- But $\mathcal{W}_1$ is always better:
    - If X is true
      payoff($\mathcal{W}_1$) $= 1.6 >$ payoff($\mathcal{W}_2$)$=1.5$
    - If Y is true
      payoff($\mathcal{W}_1$) $= 1.6 >$ payoff($\mathcal{W}_2$)$= 1.5$
    - If neither X nor Y is true
      payoff($\mathcal{W}_1$) $= 0.6 >$ payoff($\mathcal{W}_2$)$=0.5$

Dutch Book Theorem

Theorem. Imagine and EU-maximizer who satisfies 1-3 and has a precise degree of belief for every proposition she considers. If these beliefs violate the laws of probability, then she will make Dutch Book against herself.

Dutch Book Theorem

Theorem. Imagine and EU-maximizer who satisfies 1-3 and has a precise degree of belief for every proposition she considers. If these beliefs violate the laws of probability, then she will make Dutch Book against herself.

This assumes there is an agent who

1. Meets conditions 1-3

Dutch Book Theorem

Theorem. Imagine and EU-maximizer who satisfies 1-3 and has a precise degree of belief for every proposition she considers. If these beliefs violate the laws of probability, then she will make Dutch Book against herself.

This assumes there is an agent who

1. Meets conditions 1-3
2. sets a fair price for every wager she considers

Dutch Book Theorem

Theorem. Imagine and EU-maximizer who satisfies 1-3 and has a precise degree of belief for every proposition she considers. If these beliefs violate the laws of probability, then she will make Dutch Book against herself.

This assumes there is an agent who

1. Meets conditions 1-3
2. sets a fair price for every wager she considers
3. maximizes expected utility

Dutch Book Theorem

Theorem. Imagine and EU-maximizer who satisfies 1-3 and has a precise degree of belief for every proposition she considers. If these beliefs violate the laws of probability, then she will make Dutch Book against herself.

This assumes there is an agent who

1. Meets conditions 1-3
2. sets a fair price for every wager she considers
3. maximizes expected utility

allow agents to have incomplete or imprecise preferences

Dutch Book Theorem

Theorem. Imagine and EU-maximizer who satisfies 1-3 and has a precise degree of belief for every proposition she considers. If these beliefs violate the laws of probability, then she will make Dutch Book against herself.

This assumes there is an agent who

1. Meets conditions 1-3
2. sets a fair price for every wager she considers
3. maximizes expected utility

justify probabilistic coherence and EU simultaneously: Savage's Representation Theorem (discussed later in the semester)

Thank you!