Logical and Probabilistic Models of Belief Change

Eric Pacuit

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

July 11, 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/

Plan for today

- Introduction to belief revision
- AGM
- Possible worlds model
- Bayesian models (time permitting)

Belief Change

- Computer science:
  - updating databases (Doyle 1979 and Fagin et al. 1983)
- Philosophy (epistemology/philosophy of science):
  - scientific theory change and revisions of probability assignments;
  - belief change (Levi 1977, 1980, Harper 1977) and its rationality.

AGM

Carlos **A**lchourrón, Peter **G**ärdenfors, and David **M**akinson.

C. Alchourrón, P. Gärdenfors and D. Makinson. *On the logic of theory change: Partial meet contraction and revision functions*. Journal of Symbolic Logic, 50, 510 - 530, 1985.

Classics

I Levi. *Subjunctives, Dispositions and Chances*. Synthese 34:423-455, 1977.

W. Spohn. *Ordinal conditional functions: A dynamic theory of epistemic states*. in W.L. Harper and B. Skyrms, eds., Causation in Decision, Belief Change and Statistics, vol 2, pp. 105-134, 1988.

W. Harper. *Rational Conceptual Change*. PSA 1976, pp. 462494.

Belief Change

Consider the following beliefs of a rational agent:

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

Thus, the agent believes:

- q The bird caught in the trap is white.

Belief Change

Consider the following beliefs of a rational agent:

p_1 All Europeans swans are white.

p_2 The bird caught in the trap is a swan.

p_3 The bird caught in the trap comes from Sweden.

p_4 Sweden is part of Europe.

Thus, the agent believes:

q The bird caught in the trap is white.

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

Belief Change

Consider the following beliefs of a rational agent:

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

Thus, the agent believes:

- q The bird caught in the trap is white.

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

Belief Change

Consider the following beliefs of a rational agent:

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

Thus, the agent believes:

- q The bird caught in the trap is white.

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

Belief Change

Consider the following beliefs of a rational agent:

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

Thus, the agent believes:

- q The bird caught in the trap is white.

Problem: Logical considerations alone are insufficient to answer this question! Why?? **There are several logically consistent ways to incorporate $\neg q$!**

Belief Change

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

Belief Change

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

Thus, the agent believes:

- q The bird caught in the trap is white.

New belief:

- $\neg q$ The bird caught in the trap is not white.

Belief Change

p_1 All Europeans swans are white.

p_2 The bird caught in the trap is a swan.

p_3 The bird caught in the trap comes from Sweden.

p_4 Sweden is part of Europe.

Thus, the agent believes:

~~q The bird caught in the trap is white.~~

New belief:

$\neg q$ The bird caught in the trap is not white.

Belief Change

~~p_1 All Europeans swans are white.~~

~~p_2 The bird caught in the trap is a swan.~~

~~p_3 The bird caught in the trap comes from Sweden.~~

~~p_4 Sweden is part of Europe.~~

Thus, the agent believes:

~~q The bird caught in the trap is white.~~

New belief:

$\neg q$ The bird caught in the trap is not white.

Belief Change

p_1 All Europeans swans are white.

~~p_2 The bird caught in the trap is a swan.~~

~~p_3 The bird caught in the trap comes from Sweden.~~

~~p_4 Sweden is part of Europe.~~

Thus, the agent believes:

~~q The bird caught in the trap is white.~~

New belief:

$\neg q$ The bird caught in the trap is not white.

Belief Change

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

Thus, the agent believes:

- ~~q The bird caught in the trap is white.~~

New belief:

- $\neg q$ The bird caught in the trap is not white.

Belief Change

~~p_1~~ ~~All Europeans swans are white.~~

p_2 The bird caught in the trap is a swan.

p_3 The bird caught in the trap comes from Sweden.

p_4 Sweden is part of Europe.

Thus, the agent believes:

~~q~~ ~~The bird caught in the trap is white.~~

New belief:

$\neg q$ The bird caught in the trap is not white.

Belief Change

p_1 All Europeans swans are white.

~~p_2 The bird caught in the trap is a swan.~~

p_3 The bird caught in the trap comes from Sweden.

p_4 Sweden is part of Europe.

Thus, the agent believes:

~~q The bird caught in the trap is white.~~

New belief:

$\neg q$ The bird caught in the trap is not white.

Belief Change

p_1 All Europeans swans are white.

p_2 The bird caught in the trap is a swan.

~~p_3 The bird caught in the trap comes from Sweden.~~

p_4 Sweden is part of Europe.

Thus, the agent believes:

~~q The bird caught in the trap is white.~~

New belief:

$\neg q$ The bird caught in the trap is not white.

Belief Change

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

Thus, the agent believes:

- ~~q The bird caught in the trap is white.~~

New belief:

- $\neg q$ The bird caught in the trap is not white.

Belief Change

p_1 All Europeans swans are white.

p_2 The bird caught in the trap is a swan.

p_3 The bird caught in the trap comes from Sweden.

p_4 Sweden is part of Europe.

Thus, the agent believes:

~~q The bird caught in the trap is white.~~

New belief:

$\neg q$ The bird caught in the trap is not white.

Belief Change - Guiding Principles

1. When accepting a new piece of information, an agent should aim at a minimal change of his old beliefs.

2. If there are different ways to effect a belief change, the agent should give up those beliefs which are least entrenched.

Textbooks

S. O. Hansson. *A Textbook of Belief Dynamics. Theory Change and Database Updating. Dordrecht*. Kluwer Academic Publishers, 1999.

P. Gärdenfors. *Knowledge in Flux. Modeling the Dynamics of Epistemic States*. The MIT Press, 1988.

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

Epistemic States

- Belief sets
- (Ellis's belief systems)
- Possible worlds models
- (Doyle's truth maintenance systems)
- Spohn's generalized possible worlds model
- Bayesian models
- Generalized Bayesian models
- (Johnson-Laird mental models)
- ...

Epistemic States

- Belief sets
- (Ellis's belief systems)
- Possible worlds models
- (Doyle's truth maintenance systems)
- Spohn's generalized possible worlds model
- Bayesian models
- Generalized Bayesian models
- (Johnson-Laird mental models)
- ...

J. Halpern. *Reasoning about uncertainty*. The MIT Press, 2003.

AGM

Language of Beliefs in AGM:

propositional logic: atomic propositions p, q, r, ...
connectives: negation ($\neg$), conjunction ($\wedge$), disjunction ($\vee$), implication ($\rightarrow$), and equivalence ($\leftrightarrow$).

AGM

Language of Beliefs in AGM:

propositional logic: atomic propositions p, q, r, ...
connectives: negation ($\neg$), conjunction ($\wedge$), disjunction ($\vee$), implication ($\rightarrow$), and equivalence ($\leftrightarrow$).

Rationality constraints::

1. Belief sets should be consistent
2. Belief sets should be closed under logical consequence

Classical Consequence

For any set A of sentences, $Cn(A)$ is the set of **logical consequences** of A.

$Cn : \wp(\mathcal{L}) \rightarrow \wp(\mathcal{L})$ satisfying the following three conditions:

- $A \subseteq Cn(A)$ (inclusion);
- If $A \subseteq B$, then $Cn(A) \subseteq Cn(B)$ (monotony);
- $Cn(A) = Cn(Cn(A))$ (idempotence)

If p can be derived from A by classical propositional logic, then $p \in Cn(A)$.

Write $A \vdash p$ when $p \in Cn(A)$.

K is a belief set just in case $K = Cn(K)$.

K is a belief set just in case $K = Cn(K)$.

- Logical omniscience; explicit vs. implicit beliefs; "A belief set is not what you actually believe, but what you are committed to believe" (Levi 1991).

K is a belief set just in case $K = Cn(K)$.

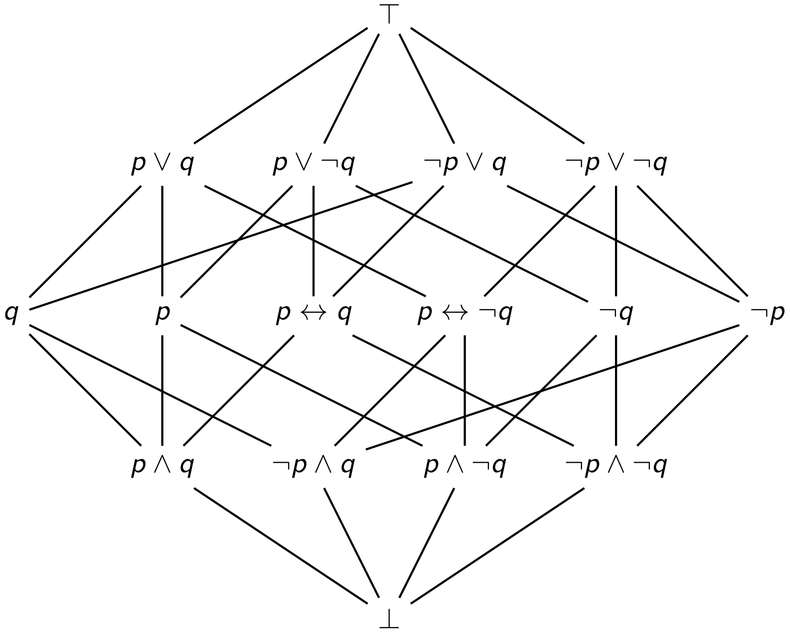
- Logical omniscience; explicit vs. implicit beliefs; "A belief set is not what you actually believe, but what you are committed to believe" (Levi 1991).
- Belief *bases* vs. belief *sets*. $B_1 = \{p, p \leftrightarrow q\}$, $B_2 = \{p, q\}$.

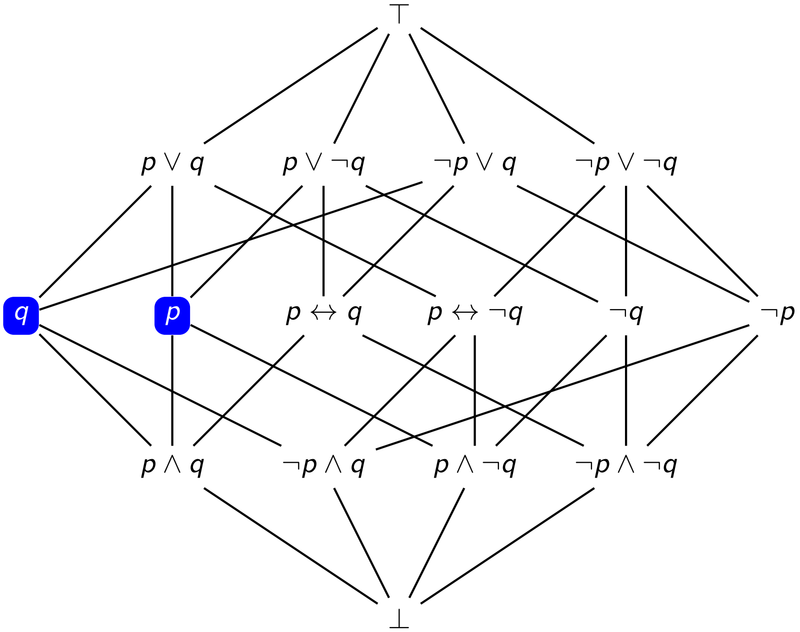
K is a belief set just in case $K = Cn(K)$.

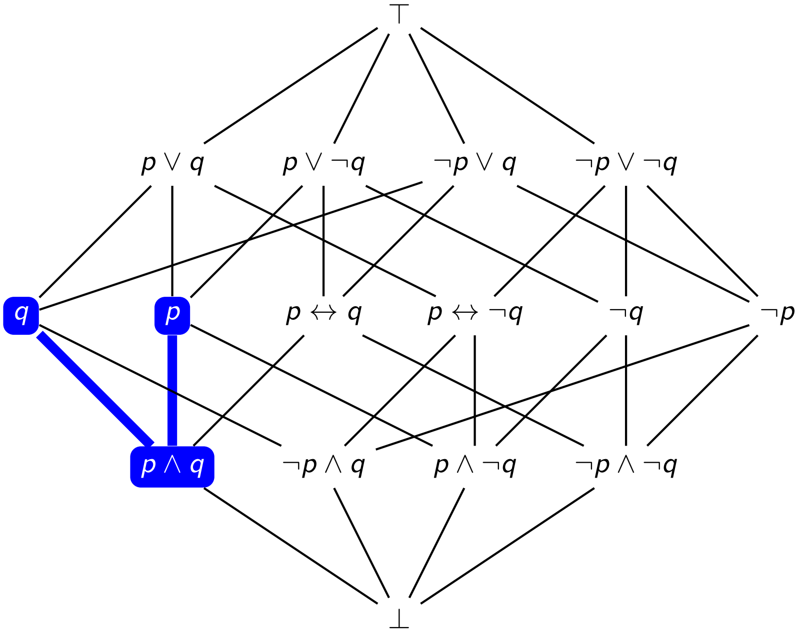
- Logical omniscience; explicit vs. implicit beliefs; "A belief set is not what you actually believe, but what you are committed to believe" (Levi 1991).
- Belief *bases* vs. belief *sets*. $B_1 = \{p, p \leftrightarrow q\}$, $B_2 = \{p, q\}$. $Cn(B_1) = Cn(B_2)$.

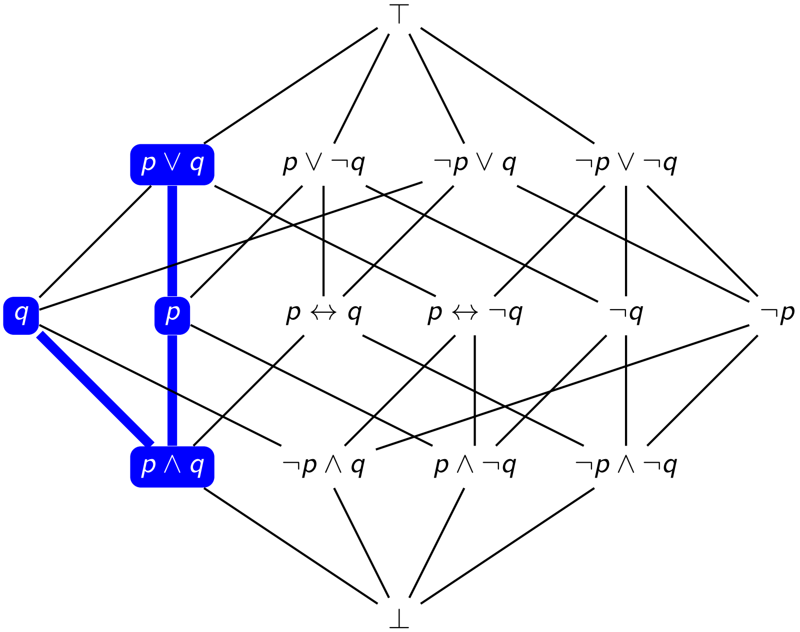
K is a belief set just in case $K = Cn(K)$.

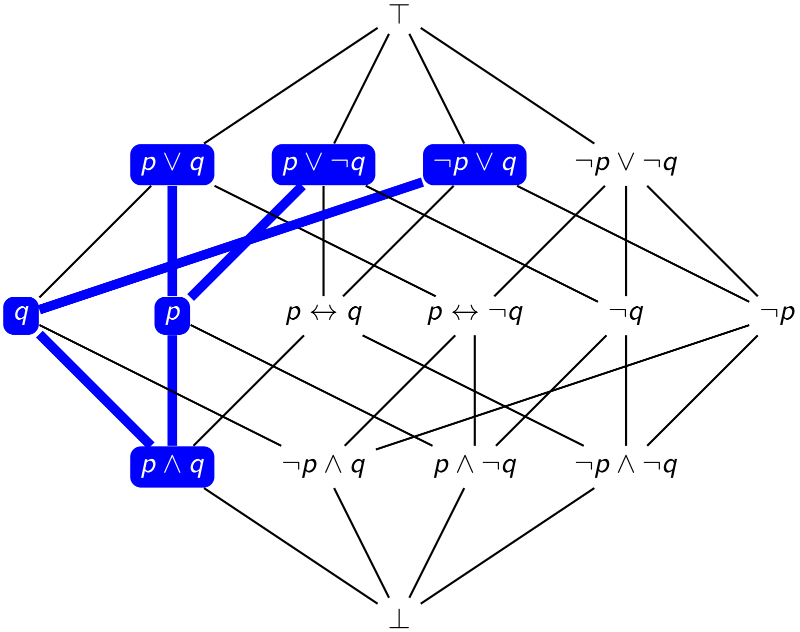
- Logical omniscience; explicit vs. implicit beliefs; "A belief set is not what you actually believe, but what you are committed to believe" (Levi 1991).
- Belief *bases* vs. belief *sets*. $B_1 = \{p, p \leftrightarrow q\}$, $B_2 = \{p, q\}$. $Cn(B_1) = Cn(B_2)$. What happens when we receive the evidence that $\neg p$?

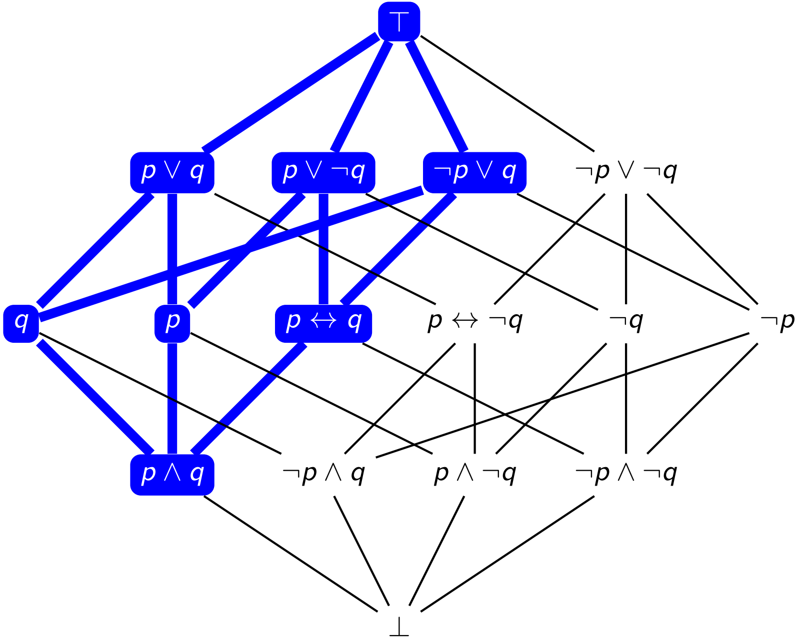

$\top$
$p \vee q$
$p \vee \neg q$
$\neg p \vee q$
$\neg p \vee \neg q$
q
p
$p \leftrightarrow q$
$p \leftrightarrow \neg q$
$\neg q$
$\neg p$
$p \wedge q$
$\neg p \wedge q$
$p \wedge \neg q$
$\neg p \wedge \neg q$
$\bot$


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

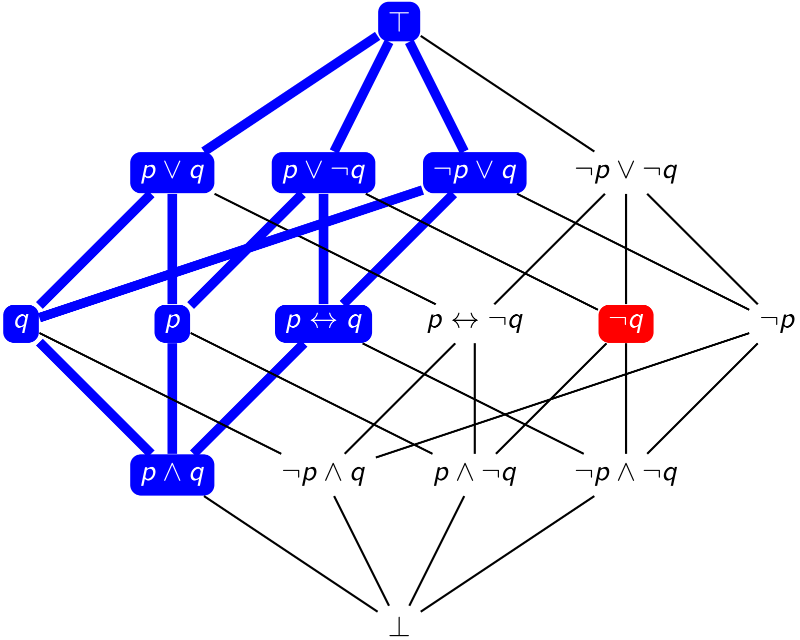

⊤
$p \vee q$
$p \vee \neg q$
$\neg p \vee q$
$\neg p \vee \neg q$
q
p
$p \leftrightarrow q$
$p \leftrightarrow \neg q$
$\neg q$
$\neg p$
$p \wedge q$
$\neg p \wedge q$
$p \wedge \neg q$
$\neg p \wedge \neg q$
⊥


⊤
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
$p \leftrightarrow q$
$p \leftrightarrow \neg q$
$\neg q$
$\neg p$
$p \wedge q$
$\neg p \wedge q$
$p \wedge \neg q$
$\neg p \wedge \neg q$
$\bot$


⊤
$p \vee q$
$p \vee \neg q$
$\neg p \vee q$
$\neg p \vee \neg q$
q
p
$p \leftrightarrow q$
$p \leftrightarrow \neg q$
$\neg q$
$\neg p$
$p \wedge q$
$\neg p \wedge q$
$p \wedge \neg q$
$\neg p \wedge \neg q$
⊥

$$K \circ \varphi$$

$$K \circ \varphi$$

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

Belief Change

Minimal change: "The criterion of informational economy demands that as few beliefs as possible be given up so that the change is in some sense a minimal change of K to accommodate for A." (Gärdenfors 1988, p. 53).

Keep the most entrenched beliefs: "...beliefs are only given up when there are no less entrenched candidates.... If one of two beliefs must be retracted in order to accommodate some new fact, the less entrenched belief will be relinquished, while the more entrenched persists." (Boutilier 1996, pp. 264-265).

Belief Change

Suppose that K is the current beliefs.

If you give priority to the new information φ, then there are three belief change operations:

Belief Change

Suppose that K is the current beliefs.

If you give priority to the new information φ, then there are three belief change operations:

1. **Expansion**: $K + \varphi$; φ is added to K.

Belief Change

Suppose that K is the current beliefs.

If you give priority to the new information φ, then there are three belief change operations:

1. **Expansion**: $K + \varphi$; φ is added to K.

2. **Contraction**: $K \dot{-} \varphi$; φ is removed from K.

Belief Change

Suppose that K is the current beliefs.

If you give priority to the new information φ, then there are three belief change operations:

1. **Expansion**: $K + \varphi$; φ is added to K.
2. **Contraction**: $K \dot{-} \varphi$; φ is removed from K.
3. **Revision**: $K * \varphi$; φ is added and other things are removed.

Expansion Postulates

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

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

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

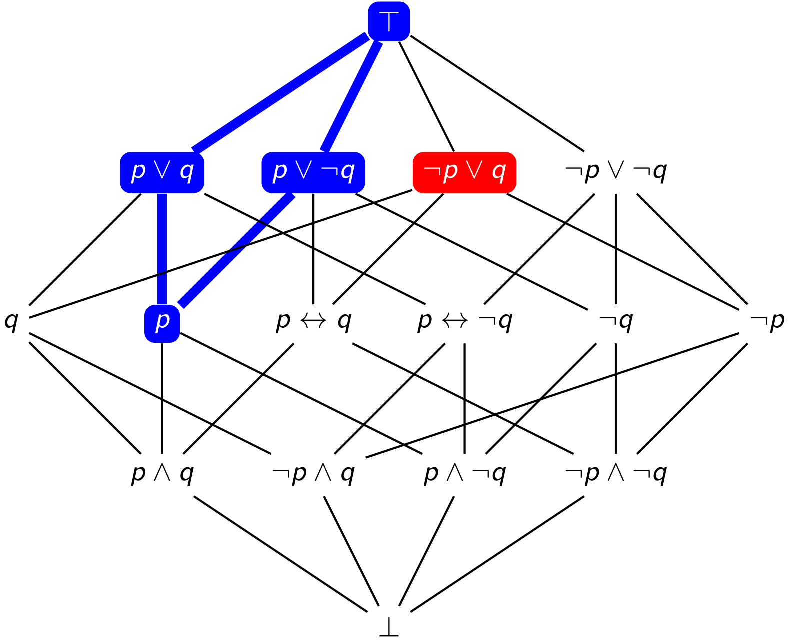
($E4$) If $\alpha \in K$, then $K + \alpha = K$

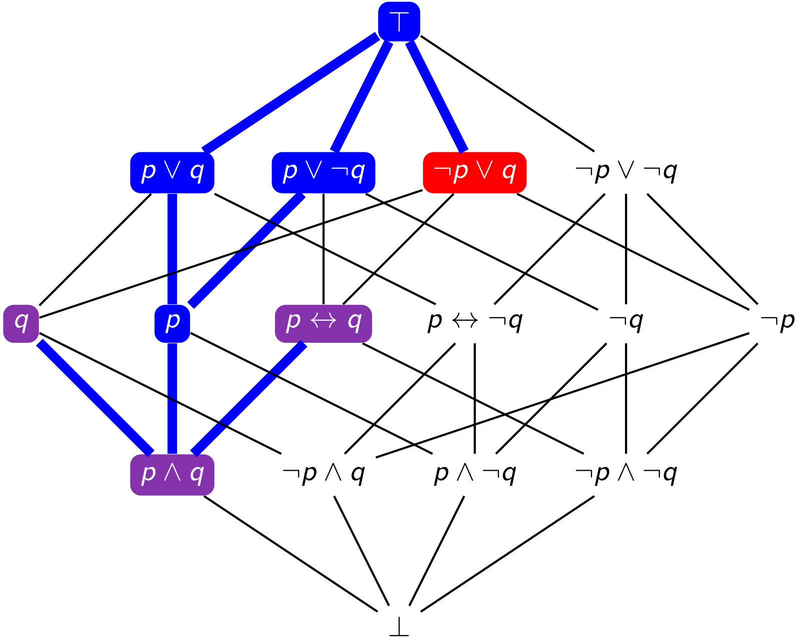
($E5$) If $K \subseteq K'$, then $K + \alpha \subseteq K' + \alpha$

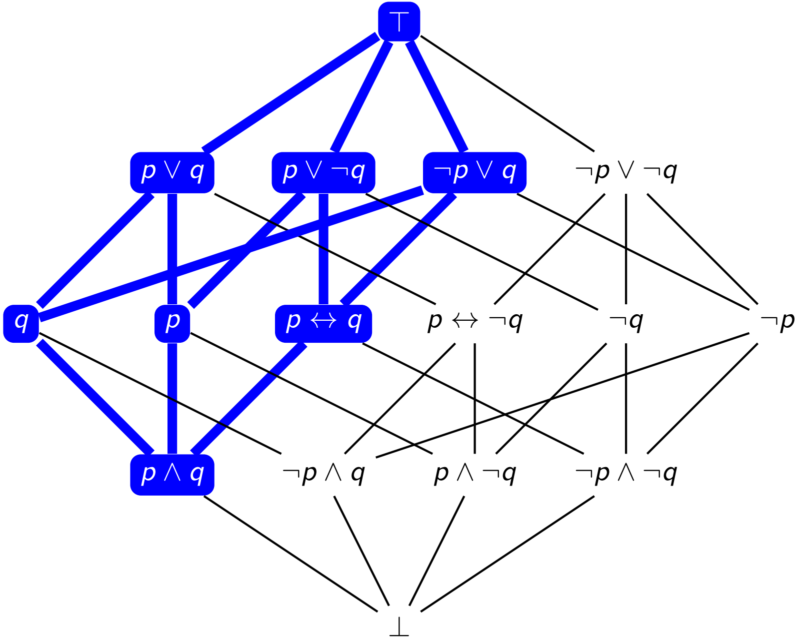
(Minimality) For all belief sets K and all sentences α, $K + \alpha$ is the smallest belief set that satisfies (E1), (E2), and (E3).

Expansion

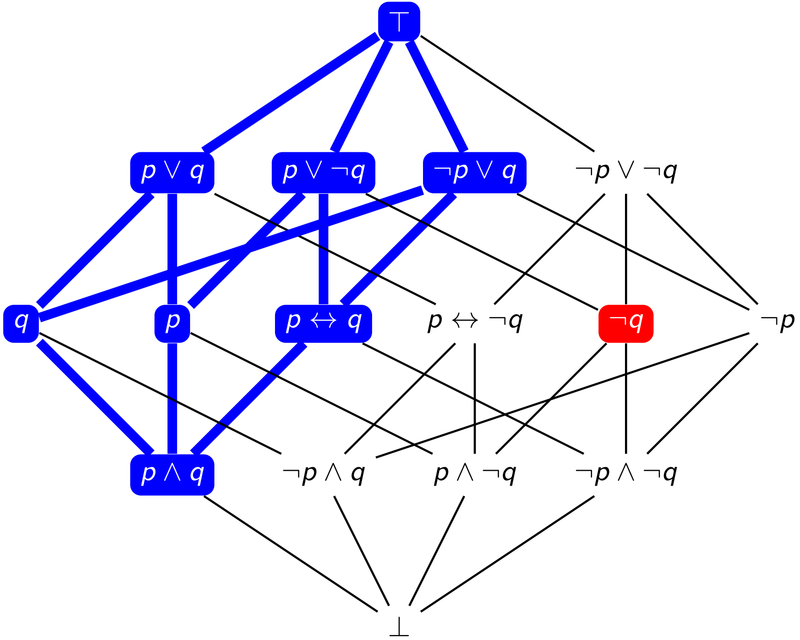
Theorem Let $+$ be a function on belief sets and formulas. Then, $+$ satisfies minimality if and only if $K + \alpha = Cn(K \cup \{\alpha\})$.

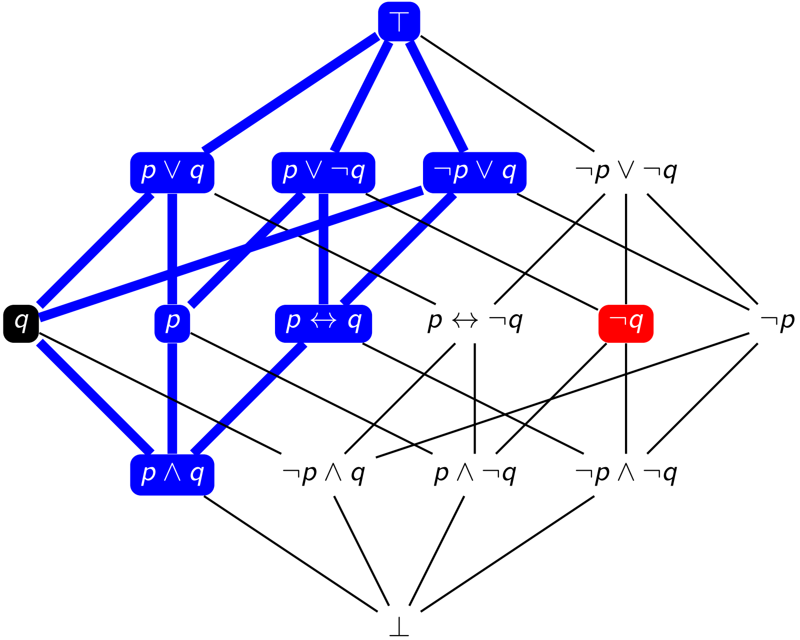

⊤
p ∨ q
p ∨ ¬q
¬p ∨ q
¬p ∨ ¬q
q
p
p ↔ q
p ↔ ¬q
¬q
¬p
p ∧ q
¬p ∧ q
p ∧ ¬q
¬p ∧ ¬q
⊥


⊤
$p \vee q$
$p \vee \neg q$
$\neg p \vee q$
$\neg p \vee \neg q$
q
p
$p \leftrightarrow q$
$p \leftrightarrow \neg q$
$\neg q$
$\neg p$
$p \wedge q$
$\neg p \wedge q$
$p \wedge \neg q$
$\neg p \wedge \neg q$
⊥



⊤
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
⊥



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

Contraction Postulates

(*C*1) $K \dot{-} \alpha$ is deductively closed

(*C*2) $K \dot{-} \alpha \subseteq K$

(*C*3) If $\alpha \notin K$ or $\vdash \alpha$ then $K \dot{-} \alpha = K$

(*C*4) If $\not\vdash \alpha$, then $\alpha \notin K \dot{-} \alpha$

(*C*5) If $\vdash \alpha \leftrightarrow \beta$, then $K \dot{-} \alpha = K \dot{-} \beta$

(*C*6) $K \subseteq Cn((K \dot{-} \alpha) \cup \{\alpha\})$

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

Let K be a belief set and φ a formula.

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

$A \in K \bot \varphi$ 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 $\gamma(K \perp \alpha) = \{K\}$

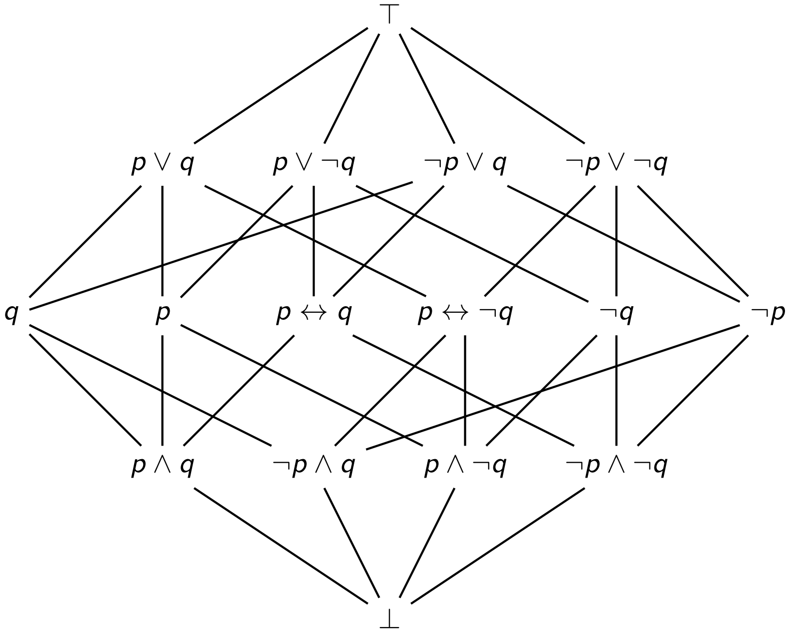
Let K be a set of formulas. A function $\dot{-}$ is a **partial meet contraction** for K if there is a selection function γ for K such that for all formula α:

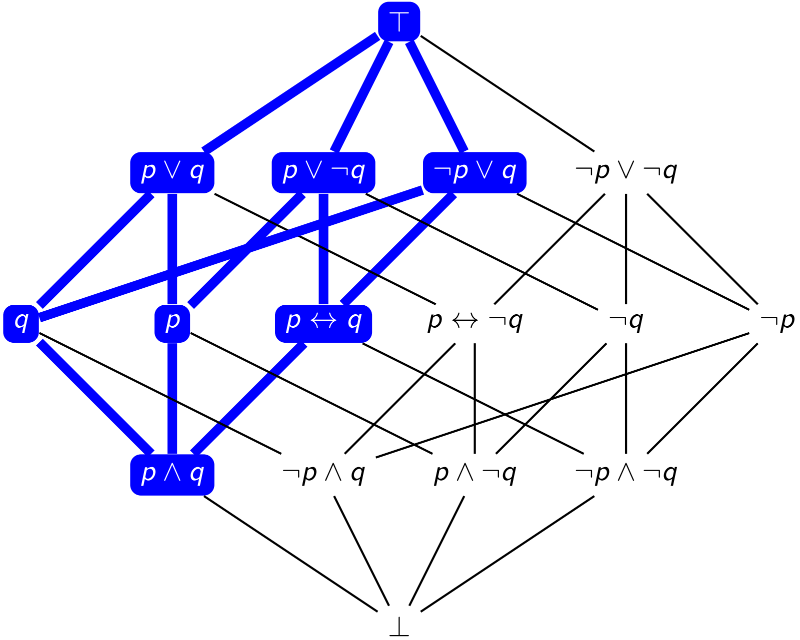
$$K \dot{-} \alpha = \bigcap \gamma(K \bot \alpha)$$

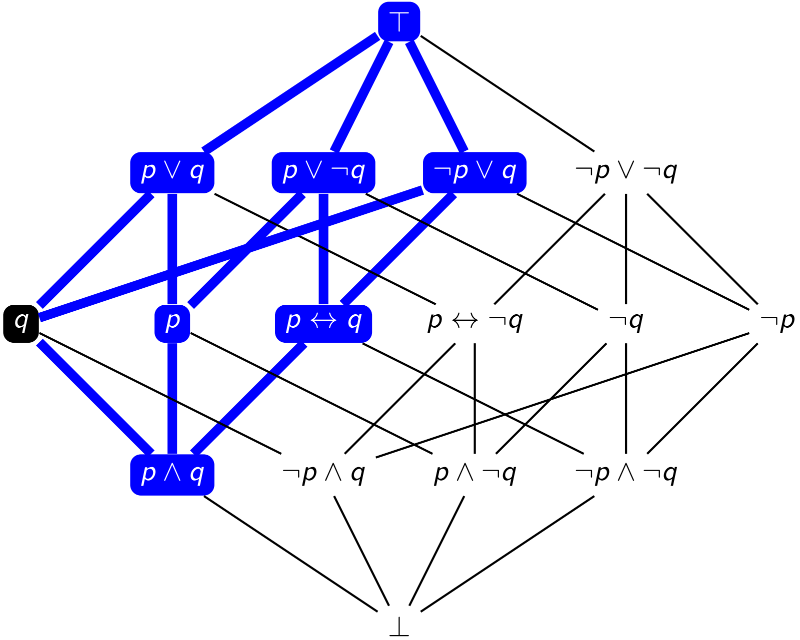
Let K be a set of formulas. A function $\dot{-}$ is a **partial meet contraction** for K if there is a selection function γ for K such that for all formula α:

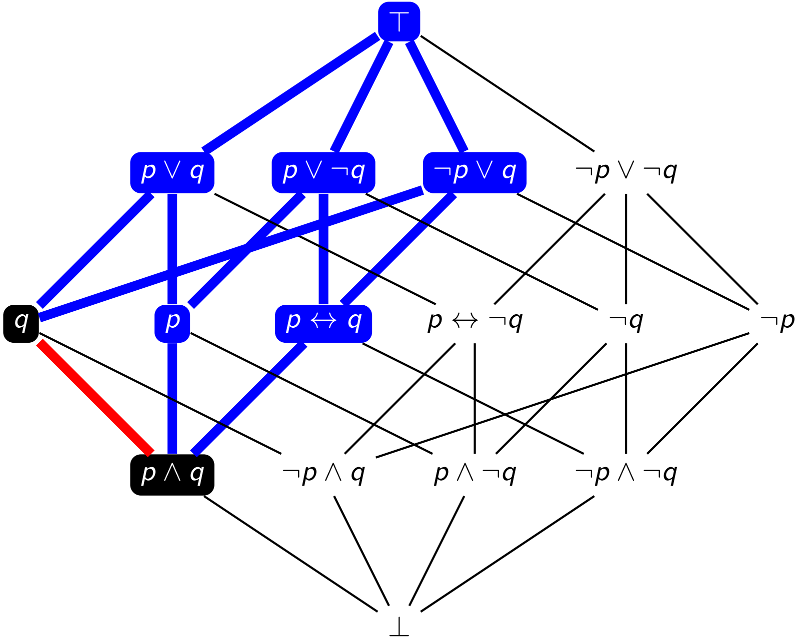
$$K \dot{-} \alpha = \bigcap \gamma(K \bot \alpha)$$

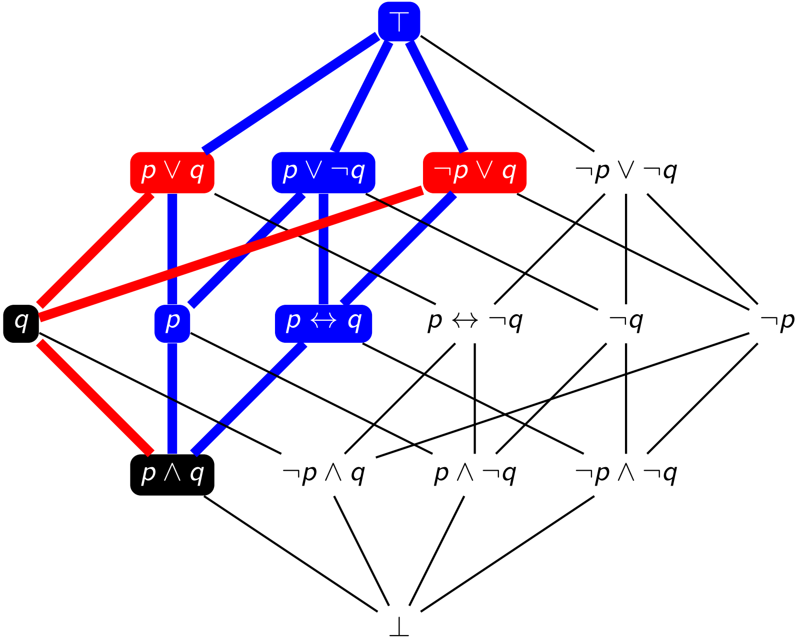
- γ selects exactly one element of $K \bot \alpha$ (maxichoice contraction)
- γ selects the entire set $K \bot \alpha$ (full meet contraction)

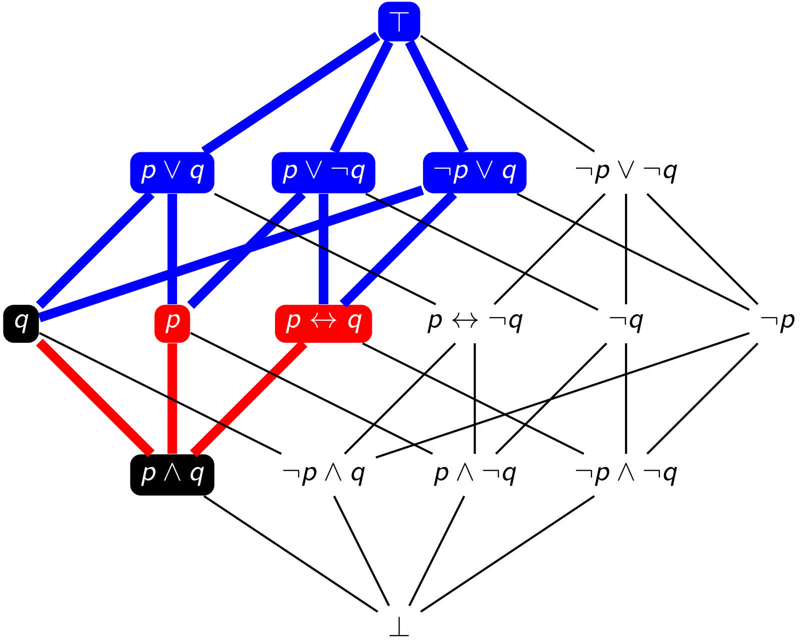

⊤
p ∨ q
p ∨ ¬q
¬p ∨ q
¬p ∨ ¬q
q
p
p ↔ q
p ↔ ¬q
¬q
¬p
p ∧ q
¬p ∧ q
p ∧ ¬q
¬p ∧ ¬q
⊥

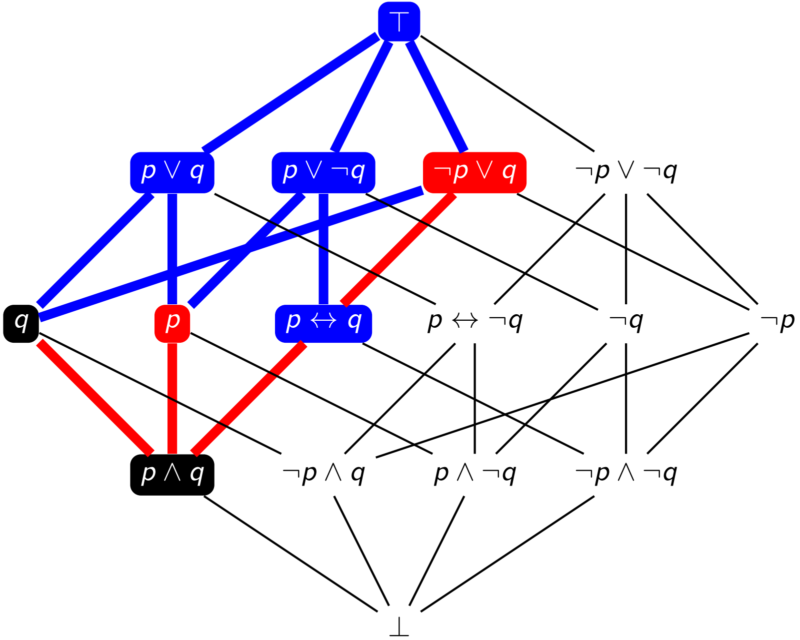

⊤
$p \vee q$
$p \vee \neg q$
$\neg p \vee q$
$\neg p \vee \neg q$
q
p
$p \leftrightarrow q$
$p \leftrightarrow \neg q$
$\neg q$
$\neg p$
$p \wedge q$
$\neg p \wedge q$
$p \wedge \neg q$
$\neg p \wedge \neg q$
⊥


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

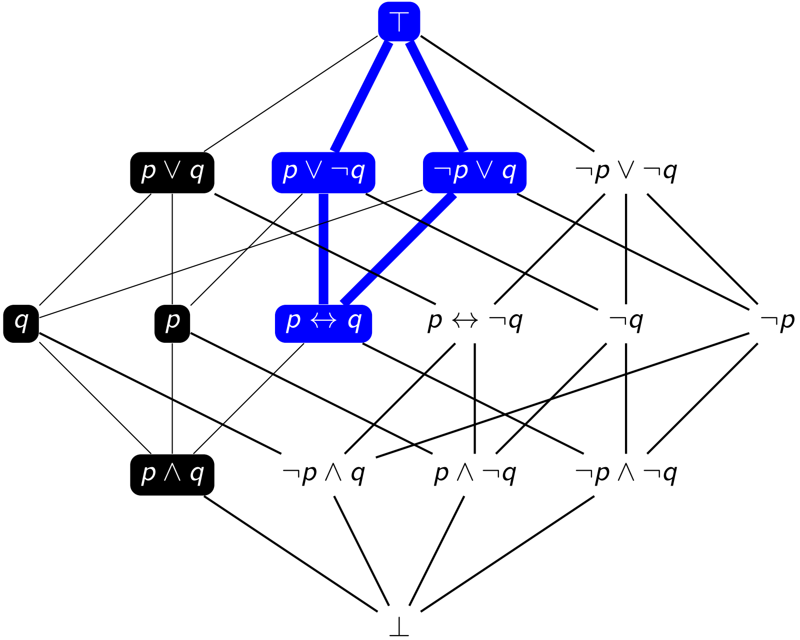

⊤
$p \vee q$
$p \vee \neg q$
$\neg p \vee q$
$\neg p \vee \neg q$
q
p
$p \leftrightarrow q$
$p \leftrightarrow \neg q$
$\neg q$
$\neg p$
$p \wedge q$
$\neg p \wedge q$
$p \wedge \neg q$
$\neg p \wedge \neg q$
⊥


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

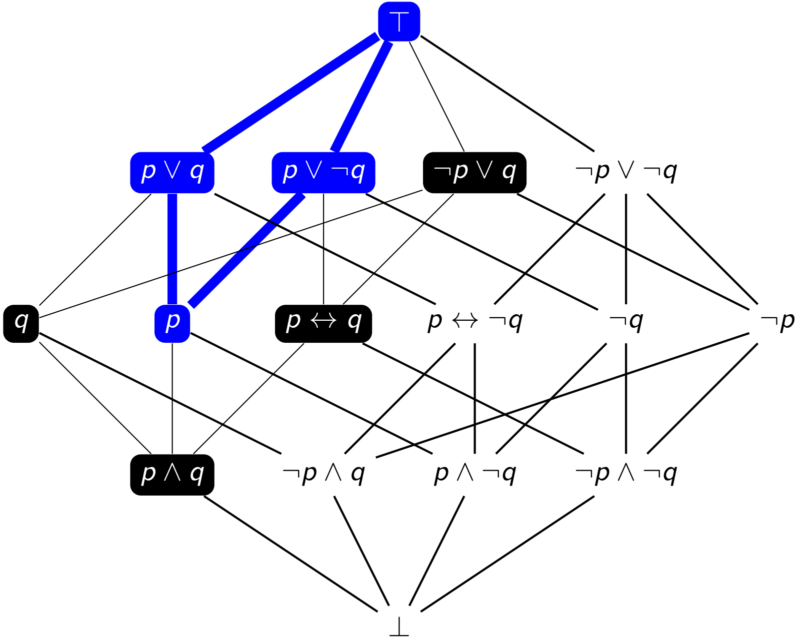

⊤
p ∨ q
p ∨ ¬q
¬p ∨ q
¬p ∨ ¬q
q
p
p ↔ q
p ↔ ¬q
¬q
¬p
p ∧ q
¬p ∧ q
p ∧ ¬q
¬p ∧ ¬q
⊥

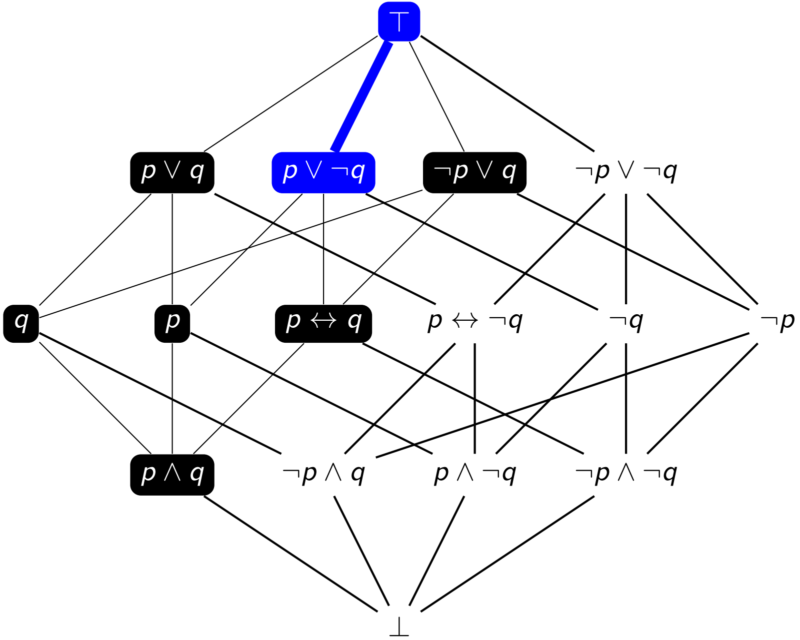


⊤
p ∨ q
p ∨ ¬q
¬p ∨ q
¬p ∨ ¬q
q
p
p ↔ q
p ↔ ¬q
¬q
¬p
p ∧ q
¬p ∧ q
p ∧ ¬q
¬p ∧ ¬q
⊥



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


⊤
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
⊥

Levi Identity

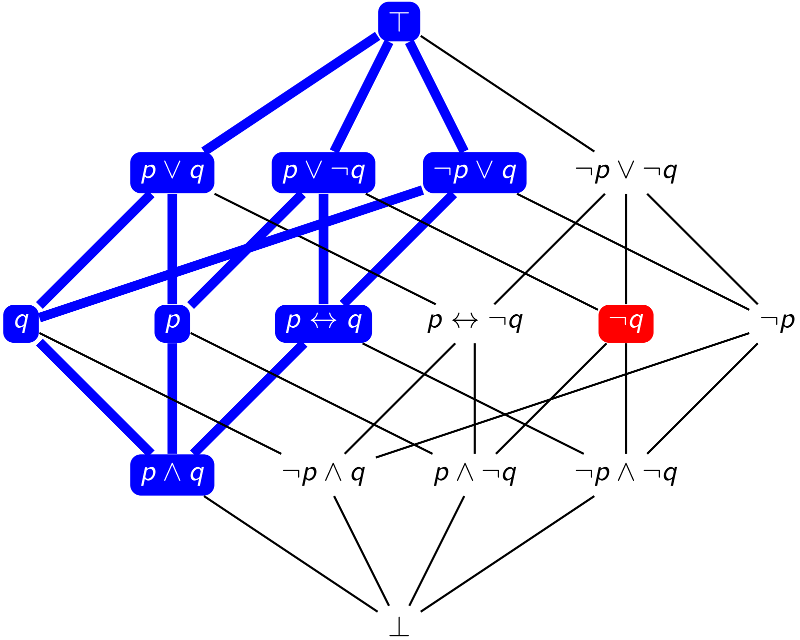
$$K * \varphi = (K \dot{-} \neg\varphi) + \varphi$$

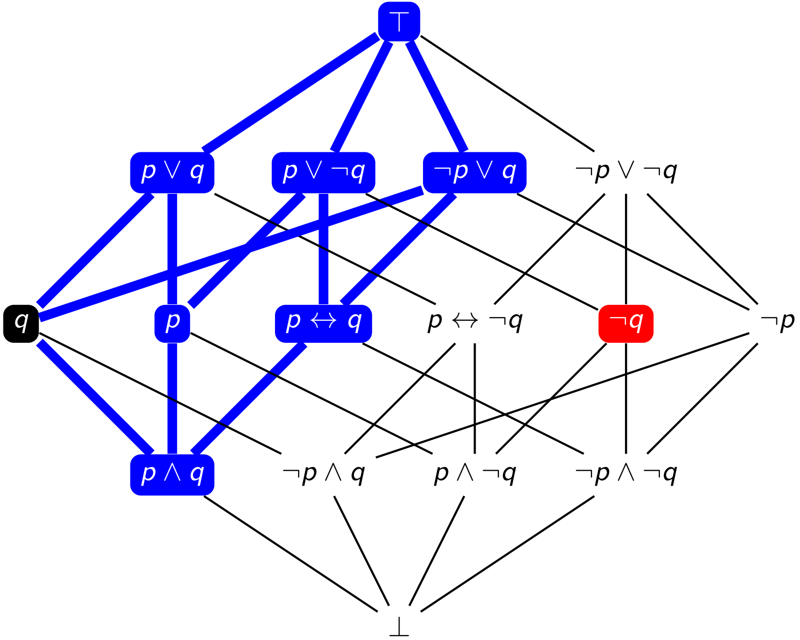
Let K be a set of formulas. A function $\dot{-}$ is a **partial meet contraction** for K if there is a selection function γ for K such that for all formula α:

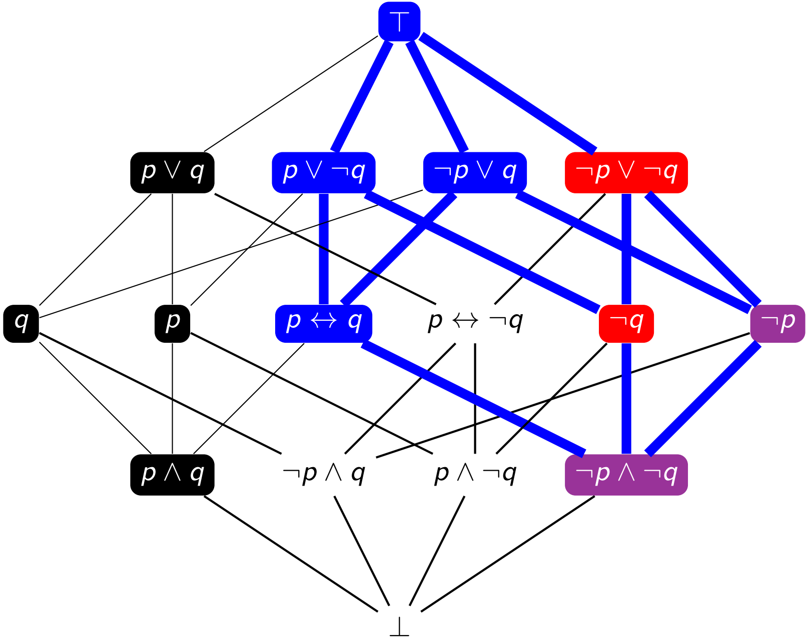
$$K \dot{-} \alpha = \bigcap \gamma(K \perp \alpha)$$

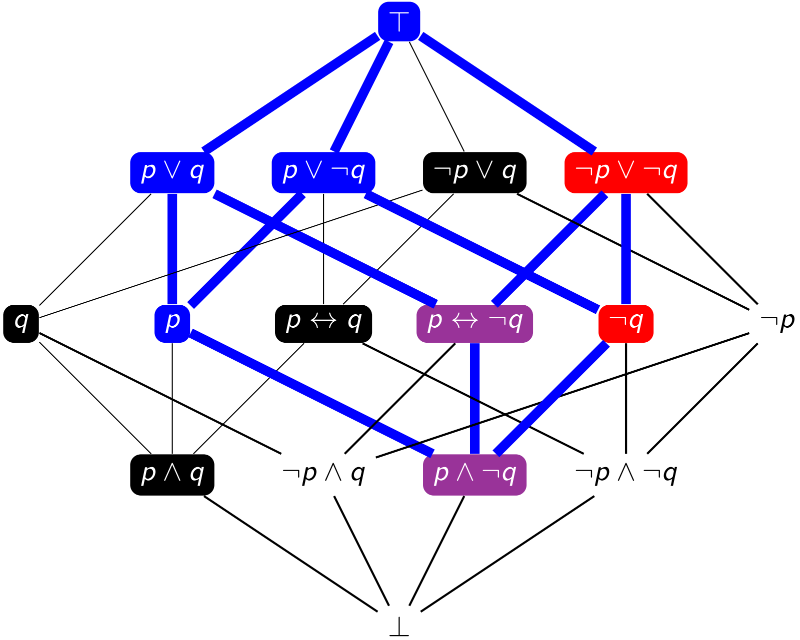
- γ selects exactly one element of $K \perp \alpha$ (maxichoice contraction)
- γ selects the entire set $K \perp \alpha$ (full meet contraction)

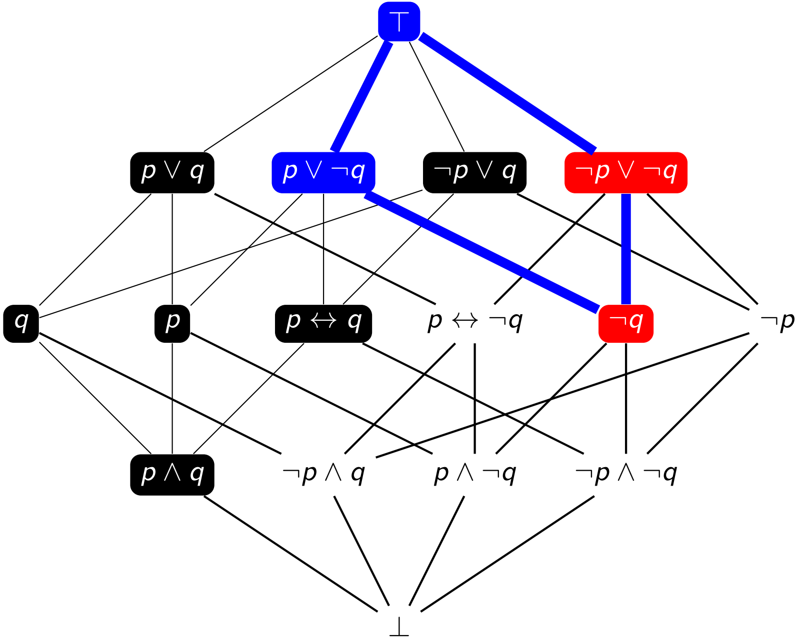
Then, $K * \alpha = Cn(\bigcap \gamma(K \perp \neg\alpha) \cup \{\alpha\})$


⊤
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
⊥


⊤
$p \vee q$
$p \vee \neg q$
$\neg p \vee q$
$\neg p \vee \neg q$
q
p
$p \leftrightarrow q$
$p \leftrightarrow \neg q$
$\neg q$
$\neg p$
$p \wedge q$
$\neg p \wedge q$
$p \wedge \neg q$
$\neg p \wedge \neg q$
⊥


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

AGM Postulates

$(AGM1)$ $K * \varphi$ is deductively closed

$(AGM2)$ $\varphi \in K * \varphi$

$(AGM3)$ $K * \varphi \subseteq Cn(K \cup \{\varphi\})$

$(AGM4)$ If $\neg\varphi \notin K$ then $K * \varphi = Cn(K \cup \{\varphi\})$

$(AGM5)$ $K * \varphi$ is inconsistent only if φ is inconsistent

$(AGM6)$ If φ and ψ are logically equivalent then $K * \varphi = K * \psi$

$(AGM7)$ $K * (\varphi \wedge \psi) \subseteq Cn(K * \varphi \cup \{\psi\})$

$(AGM8)$ If $\neg\psi \notin K * \varphi$ then $Cn(K * \varphi \cup \{\psi\}) \subseteq K * (\varphi \wedge \psi)$

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 $AGM1$ - $AGM8$.

There is a **transitive relation** $\preceq$ on $K \perp \alpha$ such that

$\gamma(K \perp \alpha) = \{K' \in K \perp \alpha \mid K'' \preceq K' \text{ for all } K'' \in K \perp \alpha\}$

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

Evaluating the AGM postulates

Counterexample to AGM 2 (Success)

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

Counterexample to AGM 2 (Success)

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

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

Counterexample to AGM 2 (Success)

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

You are walking down a street and see someone holding a sign reading "The World will End Tomorrow", but you don't add this add this to your beliefs. Is this a counterexample to AGM 2?

Counterexample to AGM 2 (Success)

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

You are walking down a street and see someone holding a sign reading "The World will End Tomorrow", but you don't add this add this to your beliefs. Is this a counterexample to AGM 2? No

Counterexample to AGM 2 (Success)

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

You are walking down a street and see someone holding a sign reading "The World will End Tomorrow", but you don't add this add this to your beliefs. Is this a counterexample to AGM 2? No

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

Counterexample to AGM 2 (Success)

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

You are walking down a street and see someone holding a sign reading "The World will End Tomorrow", but you don't add this add this to your beliefs. Is this a counterexample to AGM 2? No

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

Counterexample to AGM 2 (Success)

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

You are walking down a street and see someone holding a sign reading "The World will End Tomorrow", but you don't add this add this to your beliefs. Is this a counterexample to AGM 2? No

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

Counterexample to AGM 2 (Success)

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

You are walking down a street and see someone holding a sign reading "The World will End Tomorrow", but you don't add this add this to your beliefs. Is this a counterexample to AGM 2? No

Two people, Ann and Bob, are reliable sources of information on whether The Netherlands will win the world cup. They are equally reliable. AGM assumes that the most recent evidence that you received takes precedent. Ann says "yes" and a little bit later, Bob says "no". Why should the, possibly arbitrary, order in which you receive the information give more weight to Bob's announcement? Is this a counterexample to AGM 2?

Counterexample to AGM 2 (Success)

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

You are walking down a street and see someone holding a sign reading "The World will End Tomorrow", but you don't add this add this to your beliefs. Is this a counterexample to AGM 2? No

Two people, Ann and Bob, are reliable sources of information on whether The Netherlands will win the world cup. They are equally reliable. AGM assumes that the most recent evidence that you received takes precedent. Ann says "yes" and a little bit later, Bob says "no". Why should the, possibly arbitrary, order in which you receive the information give more weight to Bob's announcement? Is this a counterexample to AGM 2? No

Rott's Counterexample

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

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

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

Rott's Counterexample

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

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

Rott's Counterexample

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

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

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

$$A, \neg B, \neg C \in K * (A \vee B)$$

Rott's Counterexample

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

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

Rott's Counterexample

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

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

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

$$\neg A, B, \neg C \in K * (A \vee B \vee C)$$

Rott's Counterexample

Data:

- $A, \neg B, \neg C \in K * (A \vee B)$
- $\neg A, B, \neg C \in K * (A \vee B \vee C)$

Theory:

- $(AGM7)$ $K * (\varphi \wedge \psi) \subseteq Cn(K * \varphi \cup \{\psi\})$
- $(AGM8)$ If $\neg\psi \not\in K * \varphi$ then $Cn(K * \varphi \cup \{\psi\}) \subseteq K * (\varphi \wedge \psi)$
- So, if $\psi \in K * \varphi$, then $K * \varphi \subseteq K * (\varphi \wedge \psi)$

Rott's Counterexample

Problem:

- $\neg A, B \in K * (A \vee B \vee C)$
- $A \vee B \in K * (A \vee B \vee C)$
- $K * (A \vee B \vee C) \subseteq K * ((A \vee B \vee C) \wedge (A \vee B)) = K * (A \vee B)$
- $A, \neg B \in K * (A \vee B)$

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

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

"...Rott seems to take the point about meta-information to explain why the example conflicts with the theoretical principles, whereas I want to conclude that it shows why the example does not conflict with the theoretical principles, since I take the relevance of the meta-information to show that the conditions for applying the principles in question are not met by the example.... I think proper attention to the relation between concrete examples and the abstract models will allow us to reconcile some of the beautiful properties [of the abstract theory of belief revision] with the complexity of concrete reasoning." (Stalnaker, 204)

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

Counterexamples to Recovery (C6)

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

While reading a book about Cleopatra I learned that she had both a son and a daughter. I therefore believe both that Cleopatra had a son (s) and Cleopatra had a daughter (d).

Counterexamples to Recovery (C6)

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

While reading a book about Cleopatra I learned that she had both a son and a daughter. I therefore believe both that Cleopatra had a son (s) and Cleopatra had a daughter (d). Later I learn from a well-informed friend that the book in question is just a historical novel. I accordingly contract my belief that Cleopatra had a child ($s \vee d$).

Counterexamples to Recovery (C6)

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

While reading a book about Cleopatra I learned that she had both a son and a daughter. I therefore believe both that Cleopatra had a son (s) and Cleopatra had a daughter (d). Later I learn from a well-informed friend that the book in question is just a historical novel. I accordingly contract my belief that Cleopatra had a child ($s \vee d$). However, shortly thereafter I learn from a reliable source that in fact Cleopatra had a child. I thereby reintroduce $s \vee d$ to my collection of beliefs without also returning either s or d.

S. Hansson. *A Textbook of Belief Dynamics, Theory Change and Database Updating*. Kluwer Academic Publishers, 1999.

Evaluating counterexamples

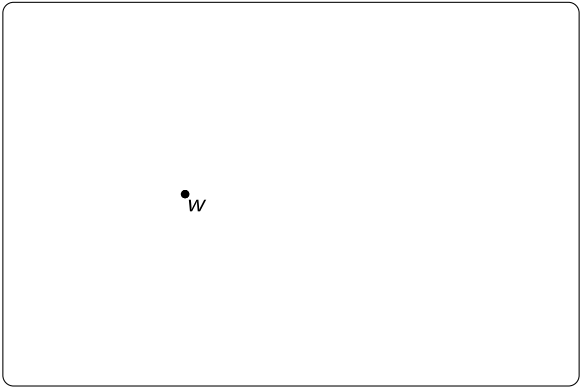
. . . information about how I learn some of the things I learn, about the sources of my information, or about what I believe about what I believe and don't believe. If the story we tell in an example makes certain information about any of these things relevant, then it needs to be included in a proper model of the story, if it is to play the right role in the evaluation of the abstract principles of the model.

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

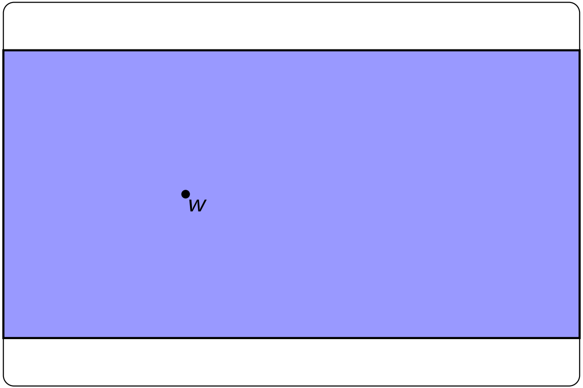
Belief Revision: The Semantic View

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

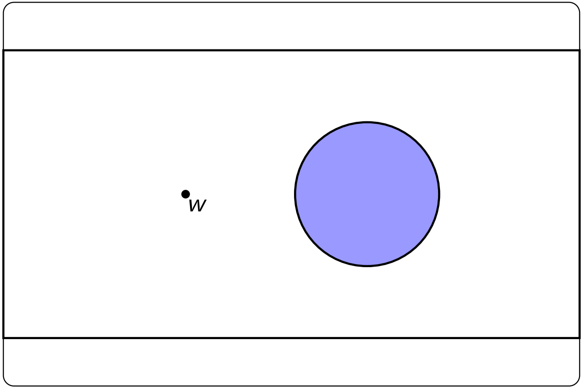
EP. *Dynamic Epistemic Logic II: Logics of information change*. Philosophy Compass, Vol. 8, Iss. 9, pgs. 815 - 833, 2013.



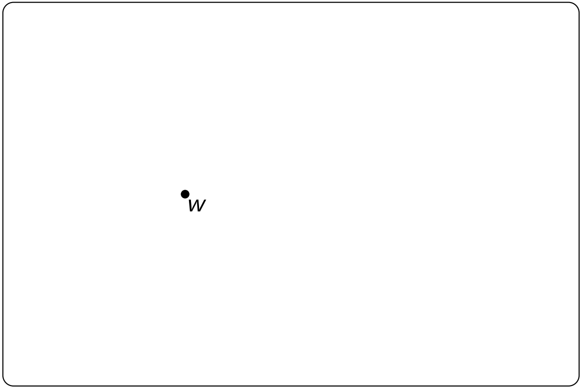
- The set of states, with a distinguished state denoting the "actual world"



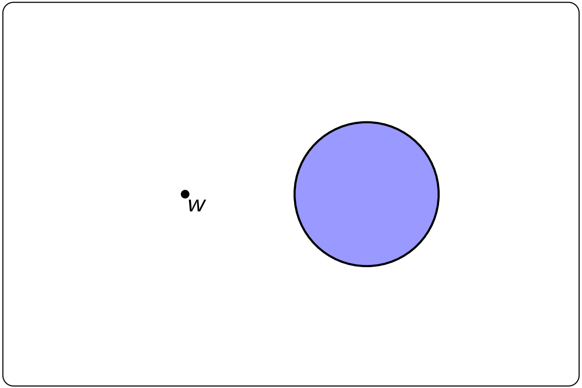
- The set of states, with a distinguished state denoting the “actual world”.
- The agent’s (hard) information (i.e., the states consistent with what the agent knows)



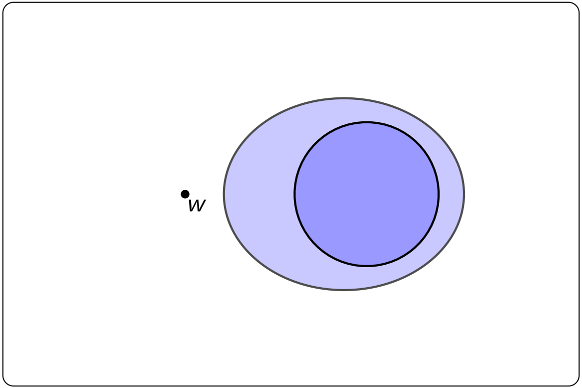
- The agent's (hard) information (i.e., the states consistent with what the agent knows)
- The agent's beliefs (soft information—the states consistent with what the agent believes)



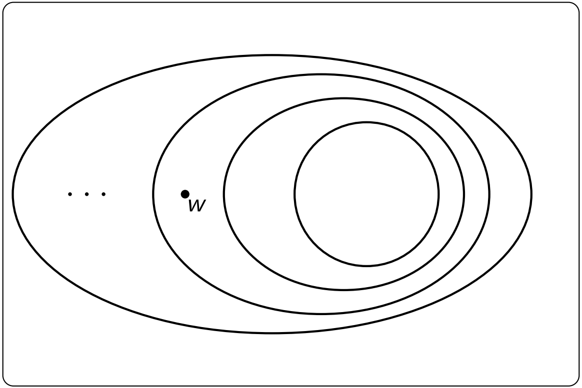
- The states consistent with what the agent knows with a distinguished state (the "actual world")
- Each state is associated with a propositional valuation for the underlying propositional language



- 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"



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

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

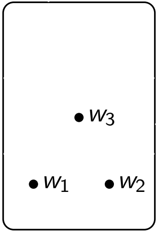
$$w \preceq_{\mathcal{F}} v \text{ iff for all } S \in \mathcal{F}, \text{ if } v \in S \text{ then } w \in S$$

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.

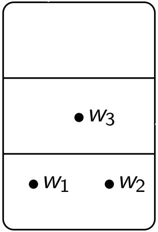
Belief Revision via Plausibility

- $W = \{w_1, w_2, w_3\}$



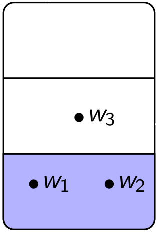
Belief Revision via Plausibility

- $W = \{w_1, w_2, w_3\}$
- $w_1 \preceq w_2$ and $w_2 \preceq w_1$ (w_1 and w_2 are equi-plausbile)
- $w_1 \prec w_3$ ($w_1 \preceq w_3$ and $w_3 \not\preceq w_1$)
- $w_2 \prec w_3$ ($w_2 \preceq w_3$ and $w_3 \not\preceq w_2$)

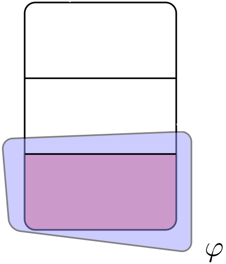


Belief Revision via Plausibility

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



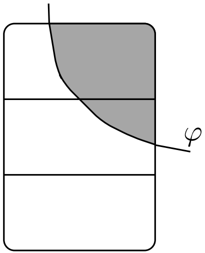
Belief Revision via Plausibility



Belief: $B\varphi$

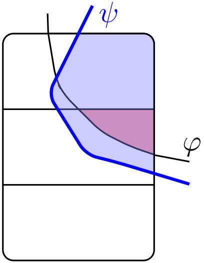
$$Min_{\preceq}(W) \subseteq [\![\varphi]\!]_{\mathcal{M}}$$

Belief Revision via Plausibility



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

Belief Revision via Plausibility



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

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

Plausibility Models

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

Truth: $\mathcal{M}, w \models \varphi$ is defined as follows:

- $\mathcal{M}, w \models p$ iff $w \in V(p)$ (with $p \in \text{At}$)
- $\mathcal{M}, w \models \neg\varphi$ if $\mathcal{M}, w \not\models \varphi$
- $\mathcal{M}, w \models \varphi \wedge \psi$ if $\mathcal{M}, w \models \varphi$ and $\mathcal{M}, w \models \psi$
- $\mathcal{M}, w \models K_i\varphi$ if for each $v \in W$, if $w \sim_i v$, then $\mathcal{M}, v \models \varphi$

Plausibility Models

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

Truth: $\mathcal{M}, w \models \varphi$ is defined as follows:

- $\mathcal{M}, w \models p$ iff $w \in V(p)$ (with $p \in \mathrm{At}$)
- $\mathcal{M}, w \models \neg\varphi$ if $\mathcal{M}, w \not\models \varphi$
- $\mathcal{M}, w \models \varphi \wedge \psi$ if $\mathcal{M}, w \models \varphi$ and $\mathcal{M}, w \models \psi$
- $\mathcal{M}, w \models K_i\varphi$ if for each $v \in W$, if $w \sim_i v$, then $\mathcal{M}, v \models \varphi$

Plausibility Models

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

Plausibility Relation: $\preceq_i \subseteq W \times W$ where $w \preceq_i v$ means "w is at least as plausible as v."

Assumptions:

1. $\preceq_i$ is reflexive and transitive (and well-founded)
2. *plausibility implies possibility*: if $w \preceq_i v$ then $w \sim_i v$.
3. *locally-connected*: if $w \sim_i v$ then either $w \preceq_i v$ or $v \preceq_i w$.

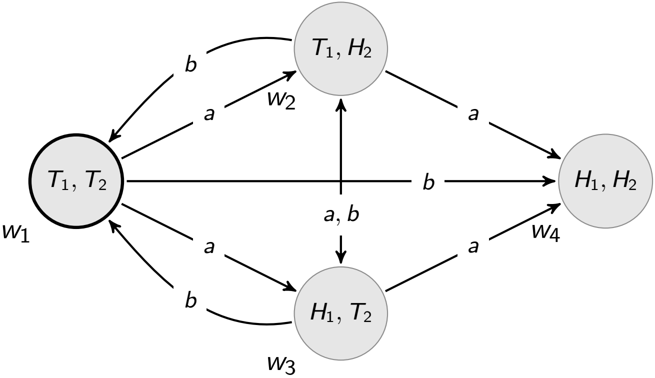
Plausibility Models

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

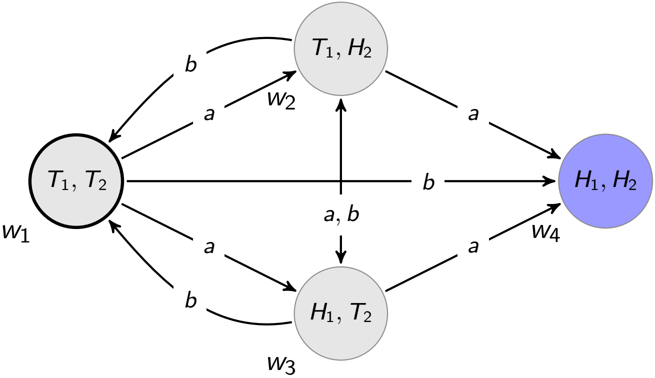
Truth: $\mathcal{M}, w \models \varphi$ is defined as follows:

- $\mathcal{M}, w \models p$ iff $w \in V(p)$ (with $p \in \mathsf{At}$)
- $\mathcal{M}, w \models \neg\varphi$ if $\mathcal{M}, w \not\models \varphi$
- $\mathcal{M}, w \models \varphi \wedge \psi$ if $\mathcal{M}, w \models \varphi$ and $\mathcal{M}, w \models \psi$
- $\mathcal{M}, w \models K_i\varphi$ if for each $v \in W$, if $w \sim_i v$, then $\mathcal{M}, v \models \varphi$
- $\mathcal{M}, w \models B_i\varphi$ if for each $v \in Min_{\preceq_i}([w]_i)$, $\mathcal{M}, v \models \varphi$
  $[w]_i = \{v \mid w \sim_i v\}$ is the agent's **information cell**.

Example

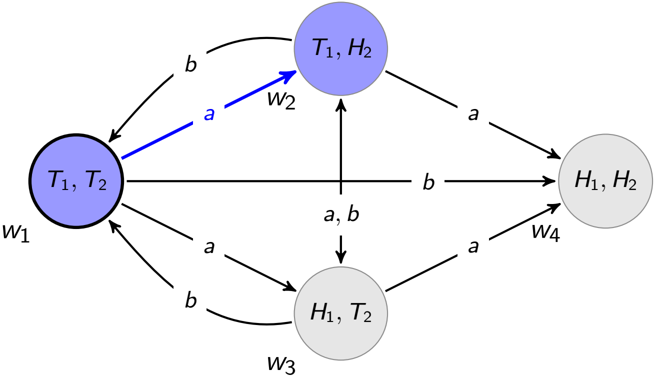


Example



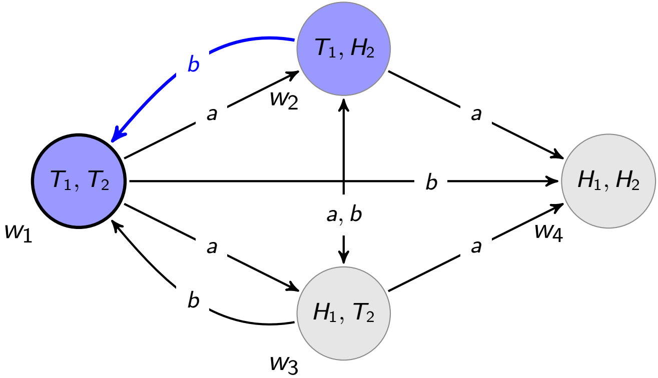
- $w_1 \models B_a(H_1 \wedge H_2) \wedge B_b(H_1 \wedge H_2)$

Example



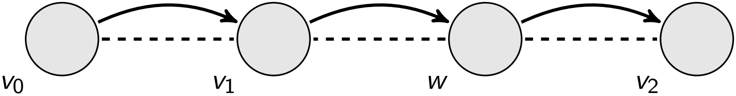
- $w_1 \models B_a(H_1 \wedge H_2) \wedge B_b(H_1 \wedge H_2)$
- $w_1 \models B_a^{T_1} H_2$

Example

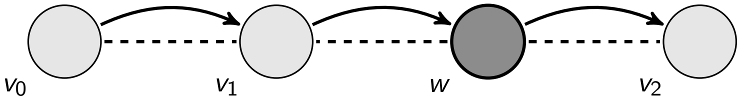


- $w_1 \models B_a(H_1 \wedge H_2) \wedge B_b(H_1 \wedge H_2)$
- $w_1 \models B_a^{T_1} H_2$
- $w_1 \models B_b^{T_1} T_2$

Grades of Doxastic Strength

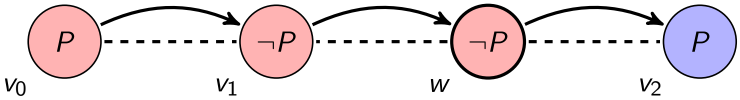


Grades of Doxastic Strength



Suppose that w is the current state.

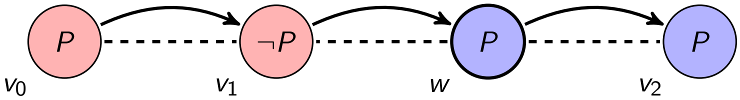
Grades of Doxastic Strength



Suppose that w is the current state.

- **Belief** (BP)

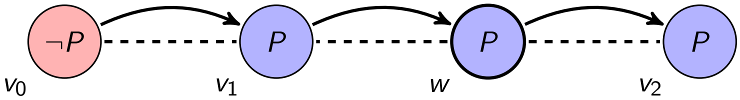
Grades of Doxastic Strength



Suppose that w is the current state.

- **Belief** (BP)
- **Robust Belief** ($[\preceq]P$)

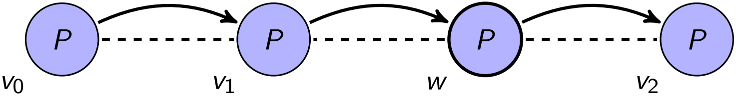
Grades of Doxastic Strength



Suppose that w is the current state.

- **Belief** (BP)
- **Robust Belief** ($[\preceq]P$)
- **Strong Belief** ($B^s P$)

Grades of Doxastic Strength



Suppose that w is the current state.

- **Belief** (BP)
- **Robust Belief** ($[\preceq]P$)
- **Strong Belief** (B^sP)
- **Knowledge** (KP)

Is $B\varphi \rightarrow B^{\psi}\varphi$ valid?

Is $B\varphi \to B^{\psi}\varphi$ valid?

Is $B^{\alpha}\varphi \to B^{\alpha\wedge\beta}\varphi$ valid?

Is $B\varphi \to B^{\psi}\varphi$ valid?

Is $B^{\alpha}\varphi \to B^{\alpha\wedge\beta}\varphi$ valid?

Is $B\varphi \to B^{\psi}\varphi \vee B^{\neg\psi}\varphi$ valid?

Is $B\varphi \to B^{\psi}\varphi$ valid?

Is $B^{\alpha}\varphi \to B^{\alpha\wedge\beta}\varphi$ valid?

Is $B\varphi \to B^{\psi}\varphi \vee B^{\neg\psi}\varphi$ valid?

Exercise: Prove that B, B^{φ} and B^{s} are definable in the language with K and $[\preceq]$ modalities.

$\mathcal{M}, w \models B^{\varphi}\psi$ if for each $v \in Min_{\preceq}([w] \cap [\![\varphi]\!])$, $\mathcal{M}, v \models \varphi$
where $[\![\varphi]\!] = \{w \mid \mathcal{M}, w \models \varphi\}$ and $[w] = \{v \mid w \sim v\}$

$\mathcal{M}, w \models B^{\varphi}\psi$ if for each $v \in Min_{\preceq}([w] \cap [\![\varphi]\!])$, $\mathcal{M}, v \models \varphi$
where $[\![\varphi]\!] = \{w \mid \mathcal{M}, w \models \varphi\}$ and $[w] = \{v \mid w \sim v\}$

Core Logical Principles:

1. $B^{\varphi}\varphi$
2. $B^{\varphi}\psi \rightarrow B^{\varphi}(\psi \vee \chi)$
3. $(B^{\varphi}\psi_1 \wedge B^{\varphi}\psi_2) \rightarrow B^{\varphi}(\psi_1 \wedge \psi_2)$
4. $(B^{\varphi_1}\psi \wedge B^{\varphi_2}\psi) \rightarrow B^{\varphi_1 \vee \varphi_2}\psi$
5. $(B^{\varphi}\psi \wedge B^{\psi}\varphi) \rightarrow (B^{\varphi}\chi \leftrightarrow B^{\psi}\chi)$

J. Burgess. *Quick completeness proofs for some logics of conditionals*. Notre Dame Journal of Formal Logic 22, 76 – 84, 1981.

Types of Beliefs: Logical Characterizations

- $\mathcal{M}, w \models K_i\varphi$ iff $\mathcal{M}, w \models B_i^{\psi}\varphi$ for all ψ
  i knows φ iff *i* continues to believe φ given any new information

Types of Beliefs: Logical Characterizations

- $\mathcal{M}, w \models K_i\varphi$ iff $\mathcal{M}, w \models B_i^{\psi}\varphi$ for all ψ
  i knows φ iff *i* continues to believe φ given any new information

- $\mathcal{M}, w \models [\preceq_i]\varphi$ iff $\mathcal{M}, w \models B_i^{\psi}\varphi$ for all ψ with $\mathcal{M}, w \models \psi$.
  i robustly believes φ iff *i* continues to believe φ given any true formula.

Types of Beliefs: Logical Characterizations

- $\mathcal{M}, w \models K_i\varphi$ iff $\mathcal{M}, w \models B_i^{\psi}\varphi$ for all ψ
  i knows φ iff *i* continues to believe φ given any new information

- $\mathcal{M}, w \models [\preceq_i]\varphi$ iff $\mathcal{M}, w \models B_i^{\psi}\varphi$ for all ψ with $\mathcal{M}, w \models \psi$.
  i robustly believes φ iff *i* continues to believe φ given any true formula.

- $\mathcal{M}, w \models B_i^s\varphi$ iff $\mathcal{M}, w \models B_i\varphi$ and $\mathcal{M}, w \models B_i^{\psi}\varphi$ for all ψ with $\mathcal{M}, w \models \neg K_i(\psi \to \neg\varphi)$.
  i strongly believes φ iff *i* believes φ and continues to believe φ given any evidence (truthful or not) that is not known to contradict φ.

Dynamic Epistemic Logic

The key idea of dynamic epistemic logic is that we can represent changes in agents' epistemic states by *transforming models*.
In the simplest case, we model an agent's acquisition of knowledge by the elimination of possibilities from an initial epistemic model.

Finding out that φ

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

$$\Big\Downarrow$$

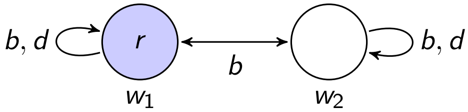
Find out that φ

$$\Big\Downarrow$$

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

Example: College Park and Amsterdam

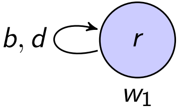
Recall the College Park agent who doesn't know whether it's raining in Amsterdam, whose epistemic state is represented by the model:



What happens when the Amsterdam agent calls the College Park agent on the phone and says, "It's raining in Amsterdam"?
We model the change in b's epistemic state by eliminating all epistemic possibilities in which it's *not* raining in Amsterdam.

Example: College Park and Amsterdam

Recall the College Park agent who doesn't know whether it's raining in Amsterdam, whose epistemic state is represented by the model:



What happens when the Amsterdam agent calls the College Park agent on the phone and says, "It's raining in Amsterdam"? We model the change in *b*'s epistemic state by eliminating all epistemic possibilities in which it's *not* raining in Amsterdam.

Model Update

We can easily give a formal definition that captures the idea of knowledge acquisition as the elimination of possibilities.
Given $\mathcal{M} = \langle W, \{R_a \mid a \in \text{Agt}\}, V\rangle$, the *updated model* $\mathcal{M}_{|\varphi}$ is obtained by deleting from $\mathcal{M}$ all worlds in which φ was false.
Formally, $\mathcal{M}_{|\varphi} = \langle W_{|\varphi}, \{R_{a_{|\varphi}} \mid a \in \text{Agt}\}, V_{|\varphi}\rangle$ is the model s.th.:

$W_{|\varphi} = \{v \in W \mid \mathcal{M}, v \models \varphi\}$;
$R_{a_{|\varphi}}$ is the restriction of R_a to $W_{|\varphi}$;
$V_{|\varphi}(p)$ is the intersection of $V(p)$ and $W_{|\varphi}$.

In the single-agent case, this models the agent learning φ. In the multi-agent case, this models all agents *publicly* learning φ.

Public Announcement Logic

One of the **big ideas** of dynamic epistemic logic is to add to our formal language operators that can describe the kinds of model updates that we just saw for the College Park and Amsterdam example.

The language of Public Announcement Logic (PAL) is given by:

$$\varphi ::= p \mid \neg\varphi \mid (\varphi \wedge \varphi) \mid K_a\varphi \mid [!\varphi]\varphi$$

Read $[!\varphi]\psi$ as "after (every) true announcement of φ, ψ."

Read $\langle !\varphi\rangle\psi := \neg[!\varphi]\neg\psi$ as "after a true announcement of φ, ψ."

Public Announcement Logic

The truth clause for the dynamic operator $[!\varphi]$ is:

- $\mathcal{M}, w \models [!\varphi]\psi$ iff $\mathcal{M}, w \models \varphi$ *implies* $\mathcal{M}_{|\varphi}, w \models \psi$.

  So if φ is false, $[!\varphi]\psi$ is vacuously true. Here is the $\langle !\varphi \rangle$ clause:

- $\mathcal{M}, w \models \langle !\varphi \rangle\psi$ iff $\mathcal{M}, w \models \varphi$ *and* $\mathcal{M}_{|\varphi}, w \models \psi$.

Main Idea: we evaluate $[!\varphi]\psi$ and $\langle !\varphi \rangle\psi$ not by looking at *other worlds in the same model*, but rather by looking at a new model.

Public Announcement Logic

Suppose $\mathcal{M} = \langle W, \{\sim_i\}_{i\in\mathcal{A}}, \{\preceq_i\}_{i\in\mathcal{A}}, V\rangle$ is a multi-agent Kripke Model

$$\mathcal{M}, w \models [!\psi]\varphi \text{ iff } \mathcal{M}, w \models \psi \text{ implies } \mathcal{M}|_\psi, w \models \varphi$$

where $\mathcal{M}|_\psi = \langle W', \{\sim_i'\}_{i\in\mathcal{A}}, \{\preceq_i'\}_{i\in\mathcal{A}}, V'\rangle$ with

- $W' = W \cap \{w \mid \mathcal{M}, w \models \psi\}$
- For each i, $\sim_i' = \sim_i \cap (W' \times W')$
- For each i, $\preceq_i' = \preceq_i \cap (W' \times W')$
- for all $p \in \mathsf{At}$, $V'(p) = V(p) \cap W'$

Public Announcement Logic

$$[!\psi]p \quad \leftrightarrow \quad (\psi \to p)$$

Public Announcement Logic

$$\begin{aligned} [!\psi]p &\leftrightarrow (\psi \to p) \\ [!\psi]\neg\varphi &\leftrightarrow (\psi \to \neg[!\psi]\varphi) \end{aligned}$$

Public Announcement Logic

$$\begin{aligned}
[!\psi]p &\leftrightarrow (\psi \rightarrow p) \\
[!\psi]\neg\varphi &\leftrightarrow (\psi \rightarrow \neg[!\psi]\varphi) \\
[!\psi](\varphi \wedge \chi) &\leftrightarrow ([!\psi]\varphi \wedge [!\psi]\chi)
\end{aligned}$$

Public Announcement Logic

$$
\begin{aligned}
[!\psi]p &\leftrightarrow (\psi \rightarrow p) \\
[!\psi]\neg\varphi &\leftrightarrow (\psi \rightarrow \neg[!\psi]\varphi) \\
[!\psi](\varphi \wedge \chi) &\leftrightarrow ([!\psi]\varphi \wedge [!\psi]\chi) \\
[!\psi][!\varphi]\chi &\leftrightarrow [!(\psi \wedge [!\psi]\varphi)]\chi
\end{aligned}
$$

Public Announcement Logic

$$
\begin{aligned}
[!\psi]p &\leftrightarrow (\psi \rightarrow p) \\
[!\psi]\neg\varphi &\leftrightarrow (\psi \rightarrow \neg[!\psi]\varphi) \\
[!\psi](\varphi \wedge \chi) &\leftrightarrow ([!\psi]\varphi \wedge [!\psi]\chi) \\
[!\psi][!\varphi]\chi &\leftrightarrow [!(\psi \wedge [!\psi]\varphi)]\chi \\
[!\psi]K_i\varphi &\leftrightarrow (\psi \rightarrow K_i(\psi \rightarrow [!\psi]\varphi))
\end{aligned}
$$

Public Announcement Logic

$$
\begin{aligned}
[!\psi]p &\leftrightarrow (\psi \rightarrow p) \\
[!\psi]\neg\varphi &\leftrightarrow (\psi \rightarrow \neg[!\psi]\varphi) \\
[!\psi](\varphi \wedge \chi) &\leftrightarrow ([!\psi]\varphi \wedge [!\psi]\chi) \\
[!\psi][!\varphi]\chi &\leftrightarrow [!(\psi \wedge [!\psi]\varphi)]\chi \\
[!\psi]K_i\varphi &\leftrightarrow (\psi \rightarrow K_i(\psi \rightarrow [!\psi]\varphi))
\end{aligned}
$$

Theorem Every formula of Public Announcement Logic is equivalent to a formula of Epistemic Logic.

Public Announcement vs. Conditional Belief

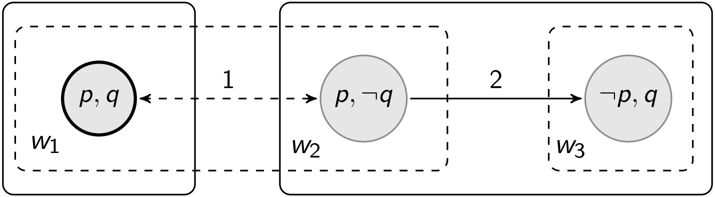
Are $[!\varphi]B\psi$ and $B^{\varphi}\psi$ different?

Public Announcement vs. Conditional Belief

Are $[!\varphi]B\psi$ and $B^{\varphi}\psi$ different? Yes!

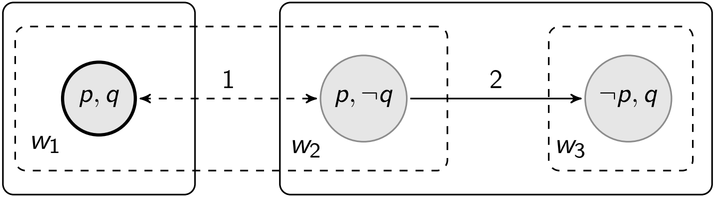
Public Announcement vs. Conditional Belief

Are $[!\varphi]B\psi$ and $B^{\varphi}\psi$ different? Yes!



Public Announcement vs. Conditional Belief

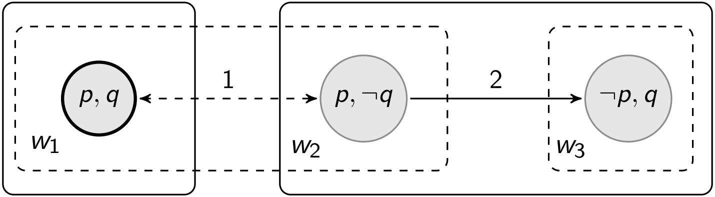
Are $[!\varphi]B\psi$ and $B^{\varphi}\psi$ different? Yes!



- $w_1 \models B_1 B_2 q$

Public Announcement vs. Conditional Belief

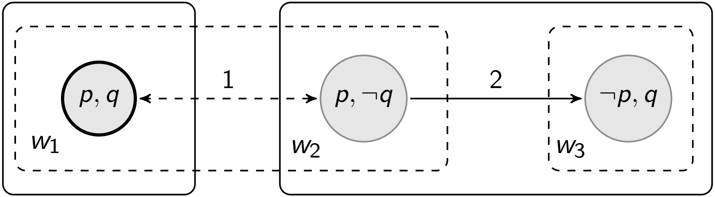
Are $[!\varphi]B\psi$ and $B^{\varphi}\psi$ different? Yes!



- $w_1 \models B_1 B_2 q$
- $w_1 \models B_1^p B_2 q$

Public Announcement vs. Conditional Belief

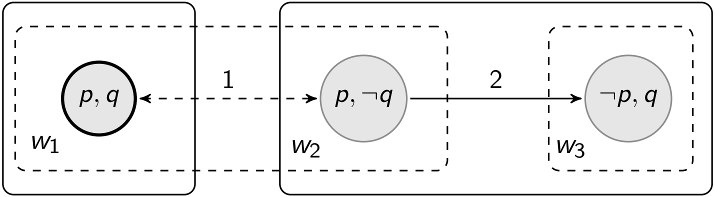
Are $[!\varphi]B\psi$ and $B^{\varphi}\psi$ different? Yes!



- $w_1 \models B_1 B_2 q$
- $w_1 \models B_1^p B_2 q$
- $w_1 \models [!p]\neg B_1 B_2 q$

Public Announcement vs. Conditional Belief

Are $[!\varphi]B\psi$ and $B^{\varphi}\psi$ different? Yes!



- $w_1 \models B_1 B_2 q$
- $w_1 \models B_1^p B_2 q$
- $w_1 \models [!p]\neg B_1 B_2 q$
- More generally, $B_i^p(p \wedge \neg K_i p)$ is satisfiable but $[!p]B_i(p \wedge \neg K_i p)$ is not.

The Logic of Public Observation

- $[!\varphi]K\psi \leftrightarrow (\varphi \rightarrow K(\varphi \rightarrow [!\varphi]\psi))$

The Logic of Public Observation

- $[!\varphi]K\psi \leftrightarrow (\varphi \rightarrow K(\varphi \rightarrow [!\varphi]\psi))$

- $[!\varphi][\preceq]\psi \leftrightarrow (\varphi \rightarrow [\preceq](\varphi \rightarrow [!\varphi]\psi))$

The Logic of Public Observation

- $[!\varphi]K\psi \leftrightarrow (\varphi \rightarrow K(\varphi \rightarrow [!\varphi]\psi))$

- $[!\varphi][\preceq]\psi \leftrightarrow (\varphi \rightarrow [\preceq](\varphi \rightarrow [!\varphi]\psi))$

- **Belief**: $[!\varphi]B\psi \not\leftrightarrow (\varphi \rightarrow B(\varphi \rightarrow [!\varphi]\psi))$

The Logic of Public Observation

- $[!\varphi]K\psi \leftrightarrow (\varphi \rightarrow K(\varphi \rightarrow [!\varphi]\psi))$

- $[!\varphi][\preceq]\psi \leftrightarrow (\varphi \rightarrow [\preceq](\varphi \rightarrow [!\varphi]\psi))$

- **Belief**: $[!\varphi]B\psi \not\leftrightarrow (\varphi \rightarrow B(\varphi \rightarrow [!\varphi]\psi))$

  $[!\varphi]B\psi \leftrightarrow (\varphi \rightarrow B^{\varphi}[!\varphi]\psi)$

The Logic of Public Observation

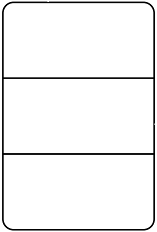
- $[!\varphi]K\psi \leftrightarrow (\varphi \rightarrow K(\varphi \rightarrow [!\varphi]\psi))$

- $[!\varphi][\preceq]\psi \leftrightarrow (\varphi \rightarrow [\preceq](\varphi \rightarrow [!\varphi]\psi))$

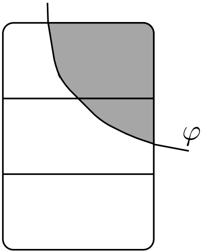
- **Belief**: $[!\varphi]B\psi \not\leftrightarrow (\varphi \rightarrow B(\varphi \rightarrow [!\varphi]\psi))$

  $[!\varphi]B\psi \leftrightarrow (\varphi \rightarrow B^{\varphi}[!\varphi]\psi)$
  $[!\varphi]B^{\alpha}\psi \leftrightarrow (\varphi \rightarrow B^{\varphi \wedge [!\varphi]\alpha}[!\varphi]\psi)$

Belief Revision via Plausibility

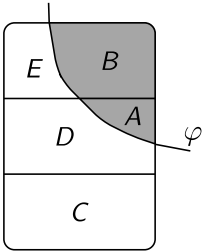


Belief Revision via Plausibility



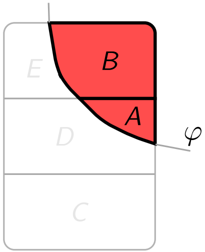
Incorporate the new information φ

Belief Revision via Plausibility



Incorporate the new information φ

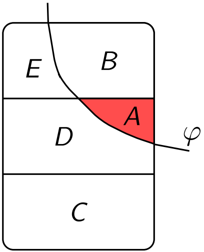
Belief Revision via Plausibility



Public Announcement: Information from an infallible source

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

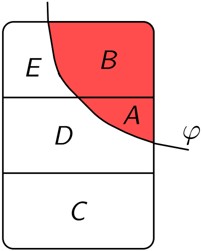
Belief Revision via Plausibility



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

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

Belief Revision via Plausibility



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