Reasoning about Knowledge and Beliefs

Lecture 6

Eric Pacuit

University of Maryland, College Park

pacuit.org
epacuit@umd.edu

October 2, 2013

Summary

(Multi-agent) **S5** is a logic of “knowledge”

(Multi-agent) **KD45** is a logic of “belief”

Summary

(Multi-agent) **S5** is a logic of "knowledge"

(Multi-agent) **KD45** is a logic of "belief"

Two issues:

- Modeling awareness/unawareness
- Logics with both knowledge and belief operators

Unawareness

Why would an agent not know some fact φ? (i.e., why would $\neg K_i\varphi$ be true?)

Unawareness

Why would an agent not know some fact φ? (i.e., why would $\neg K_i\varphi$ be true?)

- The agent may or may not believe φ, but has not ruled out all the $\neg\varphi$-worlds

Unawareness

Why would an agent not know some fact φ? (i.e., why would $\neg K_i\varphi$ be true?)

- The agent may or may not believe φ, but has not ruled out all the $\neg\varphi$-worlds
- The agent may believe φ and ruled-out the $\neg\varphi$-worlds, but this was based on "bad" evidence, or was not justified, or the agent was "epistemically lucky" (e.g., Gettier cases),...

Unawareness

Why would an agent not know some fact φ? (i.e., why would $\neg K_i\varphi$ be true?)

- The agent may or may not believe φ, but has not ruled out all the $\neg\varphi$-worlds
- The agent may believe φ and ruled-out the $\neg\varphi$-worlds, but this was based on “bad” evidence, or was not justified, or the agent was “epistemically lucky” (e.g., Gettier cases),...
- The agent has not yet entertained possibilities relevant to the truth of φ (the agent is unaware of φ).

Can we model unawareness in state-space models?

Can we model unawareness in state-space models?

E. Dekel, B. Lipman and A. Rustichini. *Standard State-Space Models Preclude Unawareness*. Econometrica, 55:1, pp. 159 - 173 (1998).

Properties of Unawareness

1. $U\varphi \rightarrow (\neg K\varphi \wedge \neg K \neg K\varphi)$

Properties of Unawareness

1. $U\varphi \to (\neg K\varphi \wedge \neg K\neg K\varphi)$
2. $\neg KU\varphi$

Properties of Unawareness

1. $U\varphi \to (\neg K\varphi \land \neg K \neg K\varphi)$
2. $\neg KU\varphi$
3. $U\varphi \to UU\varphi$

Properties of Unawareness

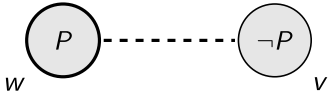
1. $U\varphi \to (\neg K\varphi \wedge \neg K\neg K\varphi)$
2. $\neg KU\varphi$
3. $U\varphi \to UU\varphi$

Theorem. In any logic where U satisfies the above axiom schemes, we have

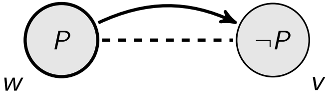
1. If K satisfies Necessitation (from φ infer $K\varphi$), then for all formulas φ, $\neg U\varphi$ is derivable (the agent is aware of everything); and
2. If K satisfies Monotonicity (from $\varphi \to \psi$ infer $K\varphi \to K\psi$), then for all φ and ψ, $U\varphi \to \neg K\psi$ is derivable (if the agent is unaware of something then the agent does not know anything).

B. Schipper. *Online Bibliography on Models of Unawareness*. http://www.econ.ucdavis.edu/faculty/schipper/unaw.htm.

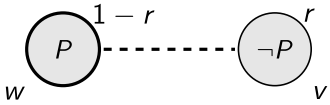
J. Halpern. *Alternative semantics for unawareness*. Games and Economic Behavior, 37, 321-339, 2001.



Ann does not know that P



Ann does not know that P, but she believes that $\neg P$



Ann does not know that P, but she believes that $\neg P$ is true to degree r.

Combining Logics of Knowledge and Belief

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

- $W \neq \emptyset$ is a set of states;
- each $\sim_i$ is an equivalence relation on W;
- each R_i is a serial, transitive, Euclidean relation on W; and
- V is a valuation function.

Combining Logics of Knowledge and Belief

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

- $W \neq \emptyset$ is a set of states;
- each $\sim_i$ is an equivalence relation on W;
- each R_i is a serial, transitive, Euclidean relation on W; and
- V is a valuation function.

What is the relationship between knowledge (K_i) and believe (B_i)?

- Each K_i is **S5**

Combining Logics of Knowledge and Belief

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

- $W \neq \emptyset$ is a set of states;
- each $\sim_i$ is an equivalence relation on W;
- each R_i is a serial, transitive, Euclidean relation on W; and
- V is a valuation function.

What is the relationship between knowledge (K_i) and believe (B_i)?

- Each K_i is **S5**
- Each B_i is **KD45**

Combining Logics of Knowledge and Belief

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

- $W \neq \emptyset$ is a set of states;
- each $\sim_i$ is an equivalence relation on W;
- each R_i is a serial, transitive, Euclidean relation on W; and
- V is a valuation function.

What is the relationship between knowledge (K_i) and believe (B_i)?

- Each K_i is **S5**
- Each B_i is **KD45**
- $K_i\varphi \to B_i\varphi$? "knowledge implies belief"

Combining Logics of Knowledge and Belief

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

- $W \neq \emptyset$ is a set of states;
- each $\sim_i$ is an equivalence relation on W;
- each R_i is a serial, transitive, Euclidean relation on W; and
- V is a valuation function.

What is the relationship between knowledge (K_i) and believe (B_i)?

- Each K_i is **S5**
- Each B_i is **KD45**
- $K_i\varphi \to B_i\varphi$? "knowledge implies belief"
- $B_i\varphi \to B_iK_i\varphi$? "positive certainty"

Combining Logics of Knowledge and Belief

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

- $W \neq \emptyset$ is a set of states;
- each $\sim_i$ is an equivalence relation on W;
- each R_i is a serial, transitive, Euclidean relation on W; and
- V is a valuation function.

What is the relationship between knowledge (K_i) and believe (B_i)?

- Each K_i is **S5**
- Each B_i is **KD45**
- $K_i\varphi \rightarrow B_i\varphi$? "knowledge implies belief"
- $B_i\varphi \rightarrow B_iK_i\varphi$? "positive certainty"
- $B_i\varphi \rightarrow K_iB_i\varphi$? "strong introspection"

An Issue

- Suppose that p is something you are certain of (you *believe* it with probability one), but is false: $\neg p \wedge Bp$

An Issue

- Suppose that p is something you are certain of (you *believe* it with probability one), but is false: $\neg p \wedge Bp$
- Assuming 1. B satisfies **KD45**, 2. K satisfies **S5**, 3. knowledge implies believe and 4. positive certainty leads to a contradiction.

An Issue

- Suppose that p is something you are certain of (you *believe* it with probability one), but is false: $\neg p \wedge Bp$
- Assuming 1. B satisfies **KD45**, 2. K satisfies **S5**, 3. knowledge implies believe and 4. positive certainty leads to a contradiction.
- $Bp \rightarrow BKp$

An Issue

- Suppose that p is something you are certain of (you *believe* it with probability one), but is false: $\neg p \wedge Bp$
- Assuming 1. B satisfies **KD45**, 2. K satisfies **S5**, 3. knowledge implies believe and 4. positive certainty leads to a contradiction.
- $Bp \rightarrow BKp$
- $\neg p \rightarrow \neg Kp$

An Issue

- Suppose that p is something you are certain of (you *believe* it with probability one), but is false: $\neg p \wedge Bp$
- Assuming 1. B satisfies **KD45**, 2. K satisfies **S5**, 3. knowledge implies believe and 4. positive certainty leads to a contradiction.
- $Bp \rightarrow BKp$
- $\neg p \rightarrow \neg Kp \rightarrow K\neg Kp$

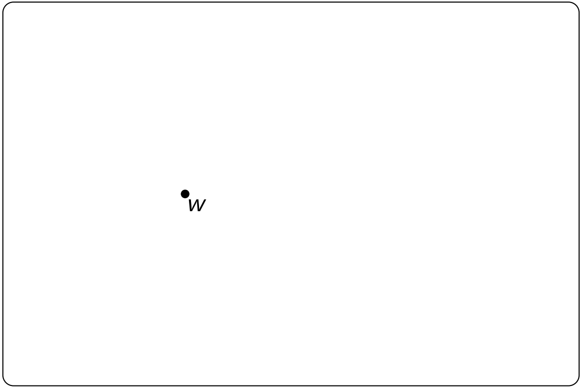
An Issue

- Suppose that p is something you are certain of (you *believe* it with probability one), but is false: $\neg p \wedge Bp$
- Assuming 1. B satisfies **KD45**, 2. K satisfies **S5**, 3. knowledge implies believe and 4. positive certainty leads to a contradiction.
- $Bp \rightarrow BKp$
- $\neg p \rightarrow \neg Kp \rightarrow K\neg Kp \rightarrow B\neg Kp$

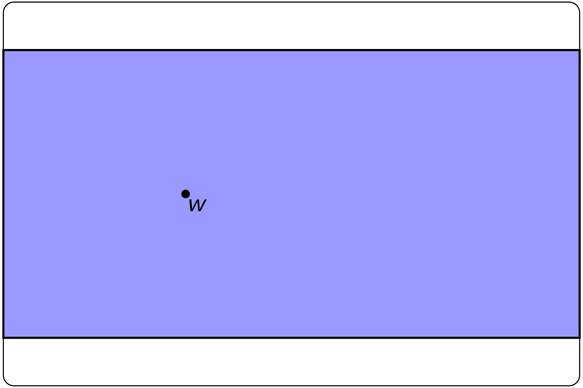
An Issue

- Suppose that p is something you are certain of (you *believe* it with probability one), but is false: $\neg p \wedge Bp$
- Assuming 1. B satisfies **KD45**, 2. K satisfies **S5**, 3. knowledge implies believe and 4. positive certainty leads to a contradiction.
- $Bp \rightarrow BKp$
- $\neg p \rightarrow \neg Kp \rightarrow K\neg Kp \rightarrow B\neg Kp$
- So, $BKp \wedge B\neg Kp$ also holds, but this contradictions $B\varphi \rightarrow \neg B\neg\varphi$.

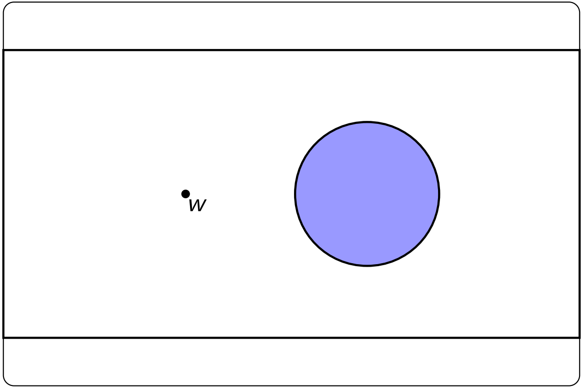
J. Halpern. *Should Knowledge Entail Belief?*. Journal of Philosophical Logic, 25:5, 1996, pp. 483-494.



- 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)

Digression on Belief Change, I

Digression on Belief Change, I

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.

Digression on Belief Change, I

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 ($\neg q$).

Digression on Belief Change, I

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?

Digression on Belief Change, I

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

Digression on Belief Change, I

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?

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

There are several logically consistent ways to incorporate $\neg q$!

Digression on Belief Change, II

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

Digression on Belief Change, III

Belief revision is a matter of choice, and the choices are to be made in such a way that:

1. The resulting theory squares with the experience;
2. It is simple; and
3. The choices disturb the original theory as little as possible.

Digression on Belief Change, III

Belief revision is a matter of choice, and the choices are to be made in such a way that:

1. The resulting theory squares with the experience;
2. It is simple; and
3. The choices disturb the original theory as little as possible.

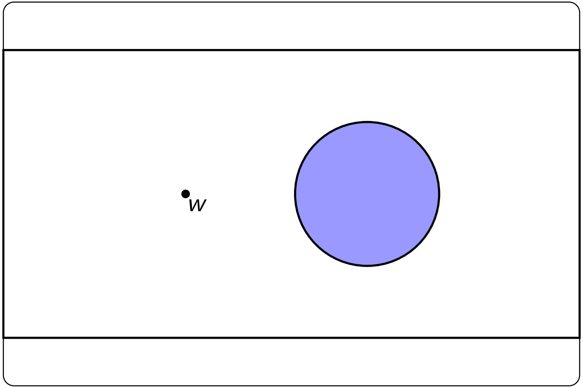
Research has relied on the following related guiding ideas:

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.

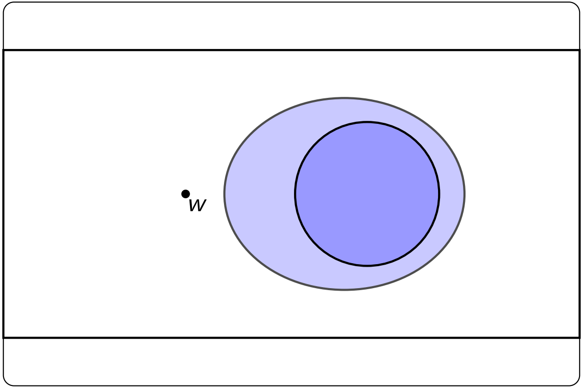
Digression: Belief Revision

A.P. Pedersen and H. Arló-Costa. *"Belief Revision."*. In Continuum Companion to Philosophical Logic. Continuum Press, 2011.

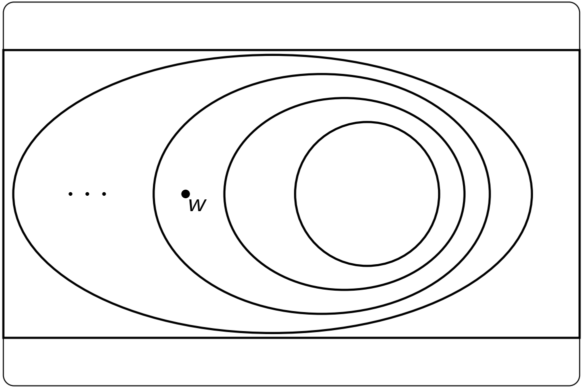
Hans Rott. *Change, Choice and Inference: A Study of Belief Revision and Nonmonotonic Reasoning*. Oxford University Press, 2001.



- 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 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

Sphere Models

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

- 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 $\bigcup \mathcal{F} = W$ (or $[w] = \{v \mid w \sim v\}$)

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.

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 $p \in \mathsf{At}$)
- $\mathcal{M}, w \models \neg\varphi$ 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 $\mathcal{M}, v \models \varphi$

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 $\mathcal{M}, v \models \varphi$

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: $\preceq_i \subseteq W \times W$. $w \preceq_i v$ means

"w is at least as plausible as v."

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: $\preceq_i \subseteq W \times W$. $w \preceq_i v$ means

"w is at least as plausible as v."

Properties of $\preceq_i$: reflexive, transitive, and *well-founded*.

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: $\preceq_i \subseteq W \times W$. $w \preceq_i v$ means

"w is at least as plausible as v."

Properties of $\preceq_i$: reflexive, transitive, and *well-founded*.

Most Plausible: For $X \subseteq W$, let

$$Min_{\preceq_i}(X) = \{v \in W \mid v \preceq_i w \text{ for all } w \in X\}$$

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: $\preceq_i \subseteq W \times W$. $w \preceq_i v$ means

"w is at least as plausible as v."

Properties of $\preceq_i$: reflexive, transitive, and *well-founded*.

Most Plausible: For $X \subseteq W$, let

$$Min_{\preceq_i}(X) = \{v \in W \mid v \preceq_i w \text{ for all } w \in X\}$$

Assumptions:

1. *plausibility implies possibility*: if $w \preceq_i v$ then $w \sim_i v$.
2. *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 $p \in \text{At}$)
- $\mathcal{M}, w \models \neg\varphi$ 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 $\mathcal{M}, v \models \varphi$
- $\mathcal{M}, w \models B_i\varphi$ 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**.