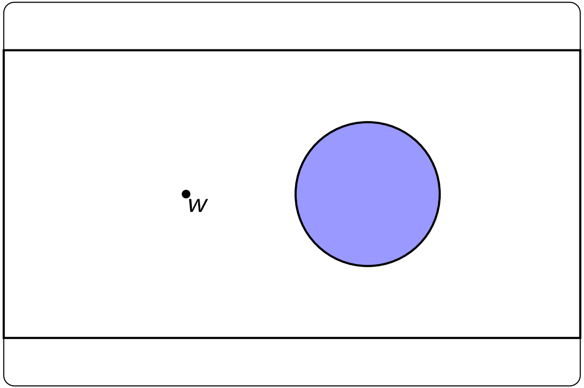
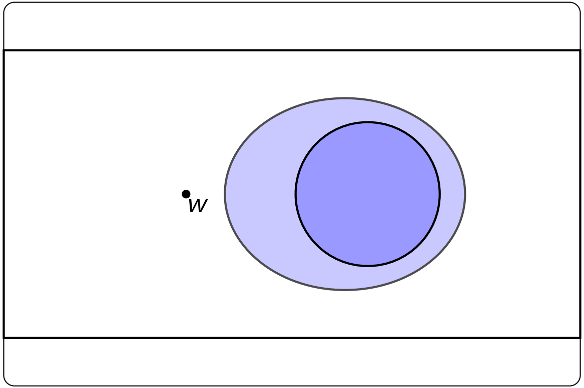
Modal Logic: Logics of Knowledge and Belief

Eric Pacuit, University of Maryland

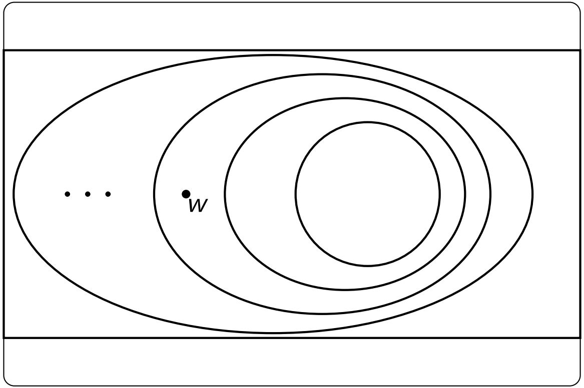
November 15, 2023



- 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 her beliefs and w is still epistemically possible, then w is also consistent with her beliefs.

Plausibility Models

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

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

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

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

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

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

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

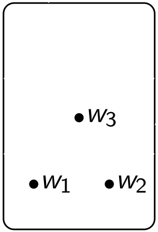
Epistemic-Plausibility Models: $\mathcal{M} = \langle W, \{\sim_i\}_{i \in \mathrm{Agt}}, \{\preceq_i\}_{i \in \mathrm{Agt}}, V \rangle$

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 $\mathcal{M}, v \models \varphi$
- $\mathcal{M}, w \models B_i\varphi$ 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**.

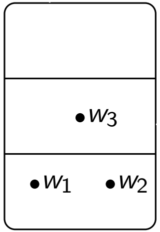
Beliefs via Plausibility

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



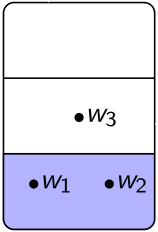
Beliefs via Plausibility

- $W = \{w_1, w_2, w_3\}$
- $w_1 \preceq w_2$ 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 $w_3 \npreceq w_1$)
- $w_2 \prec w_3$ ($w_2 \preceq w_3$ and $w_3 \npreceq w_2$)

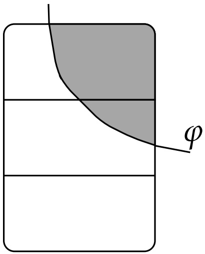


Beliefs via Plausibility

- $W = \{w_1, w_2, w_3\}$
- $w_1 \preceq w_2$ 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 $w_3 \npreceq w_1$)
- $w_2 \prec w_3$ ($w_2 \preceq w_3$ and $w_3 \npreceq w_2$)
- $\{w_1, w_2\} \subseteq Min_{\preceq}([w_i])$

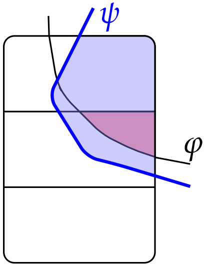


Beliefs via Plausibility



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

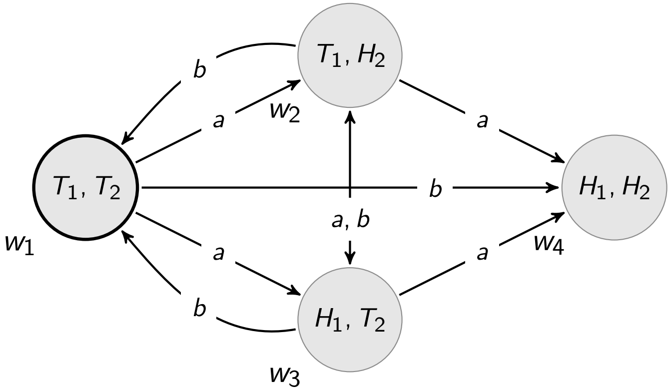
Beliefs via Plausibility



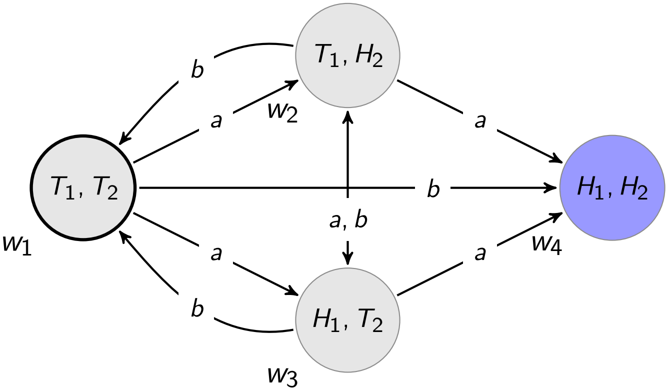
Conditional Belief: $B^{\varphi}\psi$

$$Min_{\preceq}(\llbracket\varphi\rrbracket_{\mathcal{M}}) \subseteq \llbracket\psi\rrbracket_{\mathcal{M}}$$

Example

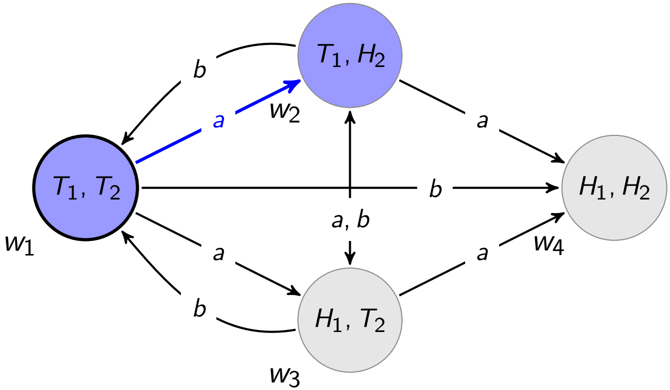


Example



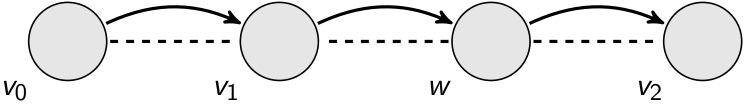
- $w_1 \models B_a(H_1 \wedge H_2) \wedge B_b(H_1 \wedge H_2)$

Example

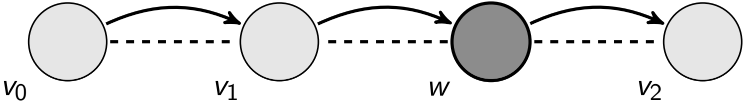


- $w_1 \models B_a(H_1 \wedge H_2) \wedge B_b(H_1 \wedge H_2)$
- $w_1 \models B_a^{T_1} H_2$
- $w_1 \models B_b^{T_1} T_2$

Grades of Doxastic Strength

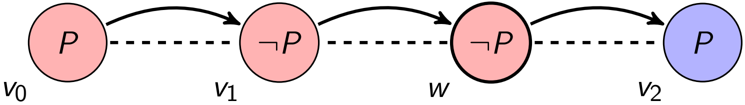


Grades of Doxastic Strength



Suppose that w is the current state.

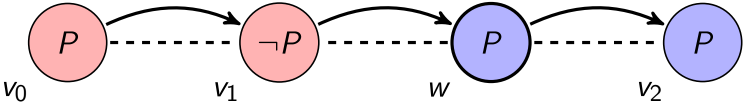
Grades of Doxastic Strength



Suppose that w is the current state.

- **Belief** (BP)

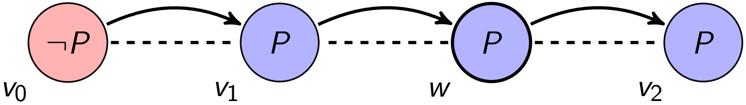
Grades of Doxastic Strength



Suppose that w is the current state.

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

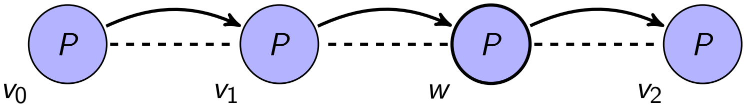
Grades of Doxastic Strength



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** $(B^s P)$
- **Knowledge** (KP)

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

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

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

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

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

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

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 $\mathcal{M}, w \models \neg K_i(\psi \rightarrow \neg\varphi)$.
  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.

Additional Axioms

Success: $B_i^{\varphi}\varphi$

Additional Axioms

Success:	$B_i^{\varphi}\varphi$
Knowledge entails belief	$K_i\varphi \to B_i^{\psi}\varphi$

Additional Axioms

Success:	$B_i^{\varphi}\varphi$
Knowledge entails belief	$K_i\varphi \to B_i^{\psi}\varphi$
Full introspection:	$B_i^{\varphi}\psi \to K_iB_i^{\varphi}\psi$ and $\neg B_i^{\varphi}\psi \to K_i\neg B_i^{\varphi}\psi$

Additional Axioms

Success:	$B_i^{\varphi}\varphi$
Knowledge entails belief	$K_i\varphi \to B_i^{\psi}\varphi$
Full introspection:	$B_i^{\varphi}\psi \to K_i B_i^{\varphi}\psi$ and $\neg B_i^{\varphi}\psi \to K_i \neg B_i^{\varphi}\psi$
Cautious Monotonicity:	$(B_i^{\varphi}\alpha \wedge B_i^{\varphi}\beta) \to B_i^{\varphi\wedge\beta}\alpha$

Additional Axioms

Success:	$B_i^{\varphi}\varphi$
Knowledge entails belief	$K_i\varphi \to B_i^{\psi}\varphi$
Full introspection:	$B_i^{\varphi}\psi \to K_iB_i^{\varphi}\psi$ and $\neg B_i^{\varphi}\psi \to K_i\neg B_i^{\varphi}\psi$
Cautious Monotonicity:	$(B_i^{\varphi}\alpha \wedge B_i^{\varphi}\beta) \to B_i^{\varphi\wedge\beta}\alpha$
Rational Monotonicity:	$(B_i^{\varphi}\alpha \wedge \neg B_i^{\varphi}\neg\beta) \to B_i^{\varphi\wedge\beta}\alpha$

Additional Axioms

Success:	$B_i^{\varphi}\varphi$
Knowledge entails belief	$K_i\varphi \to B_i^{\psi}\varphi$
Full introspection:	$B_i^{\varphi}\psi \to K_iB_i^{\varphi}\psi$ and $\neg B_i^{\varphi}\psi \to K_i\neg B_i^{\varphi}\psi$
Cautious Monotonicity:	$(B_i^{\varphi}\alpha \wedge B_i^{\varphi}\beta) \to B_i^{\varphi\wedge\beta}\alpha$
Rational Monotonicity:	$(B_i^{\varphi}\alpha \wedge \neg B_i^{\varphi}\neg\beta) \to B_i^{\varphi\wedge\beta}\alpha$