Knowledge, belief and (un)awareness
Adding Beliefs

**Epistemic Models:** \( M = \langle W, \{\sim i \}_{i \in A}, V \rangle \)

**Truth:** \( M, w \models \varphi \) is defined as follows:

- \( M, w \models p \) iff \( w \in V(p) \) (with \( p \in At \))
- \( M, w \models \neg \varphi \) if \( M, w \not\models \varphi \)
- \( M, w \models \varphi \land \psi \) if \( M, w \models \varphi \) and \( M, w \models \psi \)
- \( M, w \models K_i \varphi \) if for each \( v \in W \), if \( w \sim_i v \), then \( M, v \models \varphi \)
Adding Beliefs

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

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 $\mathcal{M}, w \not\models \varphi$
- $\mathcal{M}, w \models \varphi \land \psi$ 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 $\mathcal{M}, v \models \varphi$
Adding Beliefs

**Epistemic-Doxastic Models:** \( \mathcal{M} = \langle W, \{\sim_i\}_{i \in A}, \{\preceq_i\}_{i \in A}, V \rangle \)

**Plausibility Relation:** \( \preceq_i \subseteq W \times W \). \( w \preceq_i v \) means

“\( v \) is at least as plausible as \( w \).”
Adding Beliefs

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

Plausibility Relation: \( \preceq_i \subseteq W \times W. \) \( w \preceq_i v \) means

“\( v \) is at least as plausible as \( w \).”

Properties of \( \preceq_i \): reflexive, transitive, and well-founded.
Adding Beliefs

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

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

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

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

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

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

Plausibility Relation: \( \preceq_i \subseteq W \times W \). \( w \preceq_i v \) means

"\( v \) is at least as plausibile as \( w \)."

Properties of \( \preceq_i \): reflexive, transitive, and well-founded.

Most Plausible: For \( X \subseteq W \), let

\[ \text{Min}_{\preceq_i}(X) = \{ v \in W \mid v \preceq_i w \text{ for all } w \in X \} \]

Assumptions:

1. \textit{plausibility implies possibility}: if \( w \preceq_i v \) then \( w \sim_i v \).
2. \textit{locally-connected}: if \( w \sim_i v \) then either \( w \preceq_i v \) or \( v \preceq_i w \).
Adding Beliefs

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

**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 \( \mathcal{M}, w \not\models \varphi \)
- \( \mathcal{M}, w \models \varphi \land \psi \) 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 \( v \in \text{Min}_{\preceq_i}([w]_i) \), \( \mathcal{M}, v \models \varphi \)

\([w]_i = \{ v \mid w \sim_i v \}\) is the agent’s information cell.
Suppose that \( w \) is the current state. Knowledge (\( KP \)), Belief (\( BP \)), Safe Belief (\( □P \)), Strong Belief (\( BS_P \)).
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** ($\Box P$)
Suppose that $w$ is the current state.

- **Belief** ($BP$)
- **Robust Belief** ($\Box P$)
- **Strong Belief** ($B^s P$)
Suppose that $w$ is the current state.

- **Belief** ($BP$)
- **Robust Belief** ($\square P$)
- **Strong Belief** ($B^s P$)
- **Knowledge** ($KP$)
Conditional Beliefs

- \( w_1 \sim w_2 \sim w_3 \)

Diagram:

```
  W_3
  W_1   W_2
```

\( \phi \): Agent \( i \) believes \( \psi \), given that \( \phi \) is true.

\( M, w \mid = \phi \) if for each \( v \in [w_i] \cap [\phi] \),

\( M, v \mid = \phi \) where \( [\phi] = \{ w \mid M, w \mid = \phi \} \).
Conditional Beliefs

- $w_1 \sim w_2 \sim w_3$
- $w_1 \preceq w_2$ and $w_2 \preceq w_1$ ($w_1$ and $w_2$ are equi-plausible)
- $w_1 \prec w_3$ ($w_1 \preceq w_3$ and $w_3 \not\preceq w_1$)
- $w_2 \prec w_3$ ($w_2 \preceq w_3$ and $w_3 \not\preceq w_2$)

$\{w_1, w_2\} \subseteq \text{Min} \preceq (\{w_i\})$
Conditional Beliefs

- $w_1 \sim w_2 \sim w_3$
- $w_1 \preceq w_2$ and $w_2 \preceq w_1$ ($w_1$ and $w_2$ are equi-plausible)
- $w_1 \prec w_3$ ($w_1 \preceq w_3$ and $w_3 \npreceq w_1$)
- $w_2 \prec w_3$ ($w_2 \preceq w_3$ and $w_3 \npreceq w_2$)
- $\{w_1, w_2\} \subseteq \text{Min}_\preceq([w_i])$
Conditional Beliefs

$B_i^{\varphi} \psi$: Agent $i$ believes $\psi$, given that $\varphi$ is true.
Conditional Beliefs

$B_i^{\varphi} \psi$: Agent $i$ believes $\psi$, given that $\varphi$ is true.
Conditional Beliefs

$B_i^{\varphi} \psi$: Agent $i$ believes $\psi$, given that $\varphi$ is true.

$\mathcal{M}, w \models B_i^{\varphi} \psi$ if for each $v \in Min_{\preceq i}([w]_i \cap [\varphi])$, $\mathcal{M}, v \models \varphi$

where $[\varphi] = \{w \mid \mathcal{M}, w \models \varphi\}$
Conditional Beliefs, continued

\( B^\varphi_i \psi: \) Agent \( i \) believes \( \psi \), given that \( \varphi \) is true.

\( \mathcal{M}, w \models B^\varphi_i \psi \) if for each \( v \in Min_{\preceq i}([w]_i \cap [\varphi]) \), \( \mathcal{M}, v \models \varphi \)

where \( [\varphi] = \{ w \mid \mathcal{M}, w \models \varphi \} \)
Conditional Beliefs, continued

\( B_i^\varphi \psi \): Agent \( i \) believes \( \psi \), given that \( \varphi \) is true.

\[ \mathcal{M}, w \models B_i^\varphi \psi \text{ if for each } v \in \text{Min}_{\preceq_i}([w]_i \cap \llbracket \varphi \rrbracket), \mathcal{M}, v \models \varphi \]

where \( \llbracket \varphi \rrbracket = \{ w \mid \mathcal{M}, w \models \varphi \} \)

Is \( B\varphi \rightarrow B\psi \varphi \) valid?
Conditional Beliefs, continued

\( B_i^\varphi \psi \): Agent \( i \) believes \( \psi \), given that \( \varphi \) is true.

\[ M, w \models B_i^\varphi \psi \text{ if for each } v \in \text{Min}_{\leq i}(\{w_i \cap [\varphi]\}), M, v \models \varphi. \]

where \([\varphi]\) = \{w \mid M, w \models \varphi\}

Is \( B\varphi \rightarrow B^\psi \varphi \) valid? \( \text{No} \)
Conditional Beliefs, continued

\( B_i^{\varphi} \psi \): Agent \( i \) believes \( \psi \), given that \( \varphi \) is true.

\( \mathcal{M}, w \models B_i^{\varphi} \psi \) if for each \( v \in \text{Min}_{\leq i}(\llbracket w \rrbracket_i \cap \llbracket \varphi \rrbracket) \), \( \mathcal{M}, v \models \varphi \)

where \( \llbracket \varphi \rrbracket = \{ w | \mathcal{M}, w \models \varphi \} \)

Is \( B\varphi \rightarrow B\psi \varphi \) valid? No

What about \( B\varphi \rightarrow B\psi \varphi \lor B\neg\psi \varphi \)?
Conditional Beliefs, continued

\( B_i^{\varphi} \psi \): Agent \( i \) believes \( \psi \), given that \( \varphi \) is true.

\( M, w \models B_i^{\varphi} \psi \) if for each \( v \in \text{Min}_{\preceq i}(\{w\}_i \cap [\varphi]) \), \( M, v \models \varphi \)

where \([\varphi] = \{w \mid M, w \models \varphi\}\)

Is \( B\varphi \rightarrow B^\psi \varphi \) valid? No

What about \( B\varphi \rightarrow B^\psi \varphi \lor B^{\neg \psi} \varphi \)? Yes (but need connectedness...)

Logic and Artificial Intelligence 6/19
Conditional Beliefs, continued

\( B^\varphi \psi \): Agent \( i \) believes \( \psi \), given that \( \varphi \) is true.

\( \mathcal{M}, w \models B^\varphi \psi \) if for each \( v \in \text{Min}_{\leq i}([w]_i \cap [] \varphi) \), \( \mathcal{M}, v \models \varphi \)
where \( [] \varphi = \{ w \mid \mathcal{M}, w \models \varphi \} \)

Is \( B\varphi \rightarrow B^\psi \varphi \) valid? No

What about \( B\varphi \rightarrow B^\psi \varphi \lor B^\neg \psi \varphi \)? Yes (but need connectedness...)

What does it mean if \( B^\neg \varphi \perp \) is true at a state?
Conditional Beliefs, continued

\(B^\varphi_i \psi\): Agent \(i\) believes \(\psi\), given that \(\varphi\) is true.

\(\mathcal{M}, w \models B^\varphi_i \psi\) if for each \(v \in Min_{\leq_i}([w]_i \cap \llbracket \varphi \rrbracket)\), \(\mathcal{M}, v \models \varphi\)

where \(\llbracket \varphi \rrbracket = \{w \mid \mathcal{M}, w \models \varphi\}\)

Is \(B\varphi \rightarrow B^\psi \varphi\) valid? No

What about \(B\varphi \rightarrow B^\psi \varphi \lor B^{\neg \psi} \varphi\)? Yes (but need connectedness...)

What does it mean if \(B^{\neg \varphi} \bot\) is true at a state? The agent knows \(\varphi\)
Conditional Beliefs, continued

\[ M, w \models B_{i}^{\varphi} \psi \text{ if for each } \psi \in Min_{\leq i}(\lceil w \rceil i \cap \lceil \varphi \rceil), M, w \models \varphi \]
where \( \lceil \varphi \rceil = \{ w \mid M, w \models \varphi \} \)

Core Logical Principles:

1. \( B\varphi \varphi \)
2. \( B\varphi \psi \rightarrow B\varphi (\psi \lor \chi) \)
3. \( (B\varphi \psi_1 \land B\varphi \psi_2) \rightarrow B\varphi (\psi_1 \land \psi_2) \)
4. \( (B\varphi_1 \psi \land B\varphi_2 \psi) \rightarrow B\varphi_1 \lor \varphi_2 \psi \)
5. \( (B\varphi \psi \land B\psi \varphi) \rightarrow (B\varphi \chi \leftrightarrow B\psi \chi) \)

Types of Beliefs: Logical Characterizations

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

- $\mathcal{M}, w \models \Box_i \varphi$ iff $\mathcal{M}, w \models B_i^\psi \varphi$ for all $\psi$ with $\mathcal{M}, w \models \psi$.
  
i robustly believes $\varphi$ iff $i$ continues to believe $\varphi$ 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 $\psi$
  
i knows $\varphi$ iff $i$ continues to believe $\varphi$ given any new information

- ▶ $\mathcal{M}, w \models \Box_i\varphi$ iff $\mathcal{M}, w \models B_i^\psi \varphi$ for all $\psi$ with $\mathcal{M}, w \models \psi$.
  
i robustly believes $\varphi$ iff $i$ continues to believe $\varphi$ 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 $\psi$
  
i strongly believes $\varphi$ iff $i$ believes $\varphi$ and continues to believe $\varphi$ given any evidence (truthful or not) that is not known to contradict $\varphi$. 
Unawareness

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

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

- The agent many or may not believe \( \varphi \), but has not ruled out all the \( \neg \varphi \)-worlds
Unawareness

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

- The agent may or may not believe $\varphi$, but has not ruled out all the $\neg \varphi$-worlds.
- The agent may believe $\varphi$ 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),...
Why would an agent not know some fact $\varphi$? (i.e., why would $\neg K_i \varphi$ be true?)

- The agent many or may not believe $\varphi$, but has not ruled out all the $\neg \varphi$-worlds.
- The agent may believe $\varphi$ 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 $\varphi$ (the agent is unaware of $\varphi$).
Can we model unawareness in state-space models?
Can we model unawareness in state-space models?

While Watson never reports it, Sherlock Holmes once noted an even more curious incident, that of the dog that barked and the cat that howled in the night.
While Watson never reports it, Sherlock Holmes once noted an even more curious incident, that of the dog that barked and the cat that howled in the night. When Watson objected that the dog did not bark and the cat did not howl, Holmes replied “that is the curious incident to which I refer.”
While Watson never reports it, Sherlock Holmes once noted an even more curious incident, that of the dog that barked and the cat that howled in the night. When Watson objected that the dog did not bark and the cat did not howl, Holmes replied “that is the curious incident to which I refer.” Holmes knew that this meant that no one, neither man nor dog, had intruded on the premises the previous night.
While Watson never reports it, Sherlock Holmes once noted an even more curious incident, that of the dog that barked and the cat that howled in the night. When Watson objected that the dog did not bark and the cat did not howl, Holmes replied “that is the curious incident to which I refer.” Holmes knew that this meant that no one, neither man nor dog, had intruded on the premises the previous night. For had a man intruded, the dog would have barked. Had a dog intruded, the cat would have howled. Hence the lack of either of these two signals means that there could not have been a human or canine intruder.
Modeling Watson’s Unawareness
Let \( E = \{w_2\} \) be the event that there was an intruder.
Modeling Watson’s Unawareness

- Let $E = \{w_2\}$ be the event that there was an intruder.
- $K(E) = \{w_2\}$ (at $w_2$, Watson knows there is a human intruder) and $-K(E) = \{w_1, w_3\}$
Modeling Watson’s Unawareness

- Let $E = \{w_2\}$ be the event that there was an intruder.
- $K(E) = \{w_2\}$ (at $w_2$, Watson knows there is a human intruder) and $-K(E) = \{w_1, w_3\}$
- $K(-K(E)) = \{w_3\}$ (at $w_3$, Watson knows that she does not know $E$), and $-K(-K(E)) = \{w_1, w_2\}$
Let $E = \{w_2\}$ be the event that there was an intruder.

$K(E) = \{w_2\}$ (at $w_2$, Watson knows there is a human intruder) and $-K(E) = \{w_1, w_3\}$

$K(-K(E)) = \{w_3\}$ (at $w_3$, Watson knows that she does not know $E$), and $-K(-K(E)) = \{w_1, w_2\}$

$-K(E) \cap -K(-K(E)) = \{w_1\}$ and, in fact, $\bigcap_{i=1}^{\infty} (-K)^i(E) = \{w_1\}$
Modeling Watson’s Unawareness

- $E = \{ w_2 \}$
- $K(E) = \{ w_2 \}$, $-K(E) = \{ w_1, w_3 \}$
- $K(-K(E)) = \{ w_3 \}$, $-K(-K(E)) = \{ w_1, w_2 \}$
- $-K(E) \cap -K(-K(E)) = \{ w_1 \}$,
  $\bigcap_{i=1}^{\infty} (-K)^i(E) = \{ w_1 \}$

Let $U(F) = \bigcap_{i=1}^{\infty} (-K)^i(F)$. Then,
- $U(\emptyset) = U(W) = U(\{w_1\}) = U(\{w_2, w_3\}) = \emptyset$
- $U(E) = U(\{w_3\}) = U(\{w_1, w_3\}) = U(\{w_1, w_2\}) = \{ w_1 \}$
Modeling Watson’s Unawareness

- \( E = \{ w_2 \} \)
- \( K(E) = \{ w_2 \}, \quad -K(E) = \{ w_1, w_3 \} \)
- \( K(-K(E)) = \{ w_3 \}, \quad -K(-K(E)) = \{ w_1, w_2 \} \)
- \( -K(E) \cap -K(-K(E)) = \{ w_1 \}, \quad \bigcap_{i=1}^{\infty} (-K)^i(E) = \{ w_1 \} \)

Let \( U(F) = \bigcap_{i=1}^{\infty} (-K)^i(F) \). Then,

- \( U(\emptyset) = U(W) = U(\{ w_1 \}) = U(\{ w_2, w_3 \}) = \emptyset \)
- \( U(E) = U(\{ w_3 \}) = U(\{ w_1, w_3 \}) = U(\{ w_1, w_2 \}) = \{ w_1 \} \)

Then, \( U(E) = \{ w_1 \} \) and \( U(U(E)) = U(\{ w_1 \}) = \emptyset \)
Properties of Unawareness

1. \[ U\varphi \rightarrow (\neg K\varphi \land \neg K\neg K\varphi) \]
Properties of Unawareness

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

2. $\neg KU\varphi$
Properties of Unawareness

1. $U\varphi \rightarrow (\neg K\varphi \land \neg K\neg K\varphi)$
2. $\neg KU\varphi$
3. $U\varphi \rightarrow UU\varphi$
Properties of Unawareness

1. \( U\varphi \rightarrow (\neg K\varphi \land \neg K\neg K\varphi) \)

2. \( \neg KU\varphi \)

3. \( U\varphi \rightarrow UU\varphi \)

**Theorem.** In any logic where \( U \) satisfies the above axiom schemes, we have

1. If \( K \) satisfies Necessitation (from \( \varphi \) infer \( K\varphi \)), then for all formulas \( \varphi \), \( \neg U\varphi \) is derivable (the agent is aware of everything); and

2. If \( K \) satisfies Monotonicity (from \( \varphi \rightarrow \psi \) infer \( K\varphi \rightarrow \psi \)), then for all \( \varphi \) and \( \psi \), \( U\varphi \rightarrow \neg K\psi \) is derivable (if the agent is unaware of something then the agent does not know anything).

Epistemic-Probability Models
Adding Probabilities

**Epistemic-Probability Model:** \( \mathcal{M} = \langle W, \{\sim_i\}_{i \in A}, \{P_i\}_{i \in A}, V \rangle \)

where each \( \sim_i \) is an equivalence relation on \( W \) is an epistemic model and \( P_i : W \rightarrow \Delta(W) \) assigns to each state a probability measure over \( W \), and \( V \) is a valuation function.

\[ \Delta(W) = \{ p : W \rightarrow [0, 1] \mid p \text{ is a probability measure } \} \]
Adding Probabilities

**Epistemic-Probability Model:** \( \mathcal{M} = \langle W, \{\sim_i\}_{i \in A}, \{P_i\}_{i \in A}, V \rangle \) where each \( \sim_i \) is an equivalence relation on \( W \) is an epistemic model and \( P_i : W \rightarrow \Delta(W) \) assigns to each state a probability measure over \( W \), and \( V \) is a valuation function.

\( \Delta(W) = \{ p : W \rightarrow [0,1] \mid p \text{ is a probability measure} \} \)

Write \( p_i^w \) for the \( i \)'s probability measure at state \( w \). We make two natural assumptions:

1. For all \( v \in W \), if \( p_i^w(v) > 0 \) then \( p_i^w = p_i^v \); and
2. For all \( v \not\in [w]_i \), \( p_i^w(v) = 0 \).
Common Prior

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

Common Prior: $p : W \rightarrow [0, 1]$ is a probability measure (assume $W$ finite)

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 $\mathcal{M}, w \not\models \varphi$
- $\mathcal{M}, w \models \varphi \land \psi$ 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 $\mathcal{M}, v \models \varphi$
- $\mathcal{M}, w \models B^r \varphi$ iff $p([\varphi] | [w]_i) = \frac{p([\varphi] \cap [w]_i)}{p([w]_i)} \geq r$
An Example

Suppose Ann chooses H and Bob chooses M. Are these choices rational?

Yes.

Ann (Bob) knows that Bob (Ann) is rational.

\[ P_A(L) \geq 0 \cdot P_A(L) + 2 \cdot P_A(R) \]

\[ 3,3 \]

\[ 0,0 \]

\[ 0,0 \]

\[ 1,1 \]
An Example

<table>
<thead>
<tr>
<th>Ann</th>
<th>Bob</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>(U)</td>
<td>3,3</td>
</tr>
<tr>
<td>(D)</td>
<td>0,0</td>
</tr>
</tbody>
</table>

A set of information states

\[
1 \cdot P_A(L) + 0 \cdot P_A(R) \geq 0 \cdot P_A(L) + 2 \cdot P_A(R)
\]
An Example

A set of information states

```
<table>
<thead>
<tr>
<th></th>
<th>L</th>
<th>R</th>
</tr>
</thead>
<tbody>
<tr>
<td>U</td>
<td>3,3</td>
<td>0,0</td>
</tr>
<tr>
<td>D</td>
<td>0,0</td>
<td>1,1</td>
</tr>
</tbody>
</table>
```
An Example

A set of information states

Bob

<table>
<thead>
<tr>
<th></th>
<th>L</th>
<th>R</th>
</tr>
</thead>
<tbody>
<tr>
<td>U</td>
<td>3,3</td>
<td>0,0</td>
</tr>
<tr>
<td>D</td>
<td>0,0</td>
<td>1,1</td>
</tr>
</tbody>
</table>

Ann

\[1 \cdot P_A(L) + 0 \cdot P_A(R) \geq 0 \cdot P_A(L) + 2 \cdot P_A(R)\]
An Example

<table>
<thead>
<tr>
<th></th>
<th>Bob</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>L</strong></td>
<td>3,3</td>
</tr>
<tr>
<td><strong>R</strong></td>
<td>0,0</td>
</tr>
</tbody>
</table>

Ann (Bob) knows that Bob (Ann) is rational.

$$1 \cdot P_A(L) + 0 \cdot P_A(R) \geq 0 \cdot P_A(L) + 2 \cdot P_A(R)$$

A set of information states
An Example

A common prior

<table>
<thead>
<tr>
<th></th>
<th>Bob</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>L</td>
</tr>
<tr>
<td><strong>U</strong></td>
<td>3,3</td>
</tr>
<tr>
<td><strong>D</strong></td>
<td>0,0</td>
</tr>
</tbody>
</table>

$$A \text{ common prior}$$

<table>
<thead>
<tr>
<th></th>
<th>L</th>
<th>R</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>U</strong></td>
<td>$\frac{1}{8}$</td>
<td>$\frac{1}{16}$</td>
</tr>
<tr>
<td><strong>D</strong></td>
<td>$\frac{1}{8}$</td>
<td>$\frac{1}{8}$</td>
</tr>
</tbody>
</table>
An Example

<table>
<thead>
<tr>
<th></th>
<th>Bob</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ann U</td>
<td>3,3</td>
</tr>
<tr>
<td>Ann D</td>
<td>0,0</td>
</tr>
</tbody>
</table>

Suppose Ann chooses $U$ and Bob chooses $R$.

Are these choices rational?

Yes.

Ann (Bob) knows that Bob (Ann) is rational.

$$1 \cdot P_A(L) + 0 \cdot P_A(R) \geq 0 \cdot P_A(L) + 2 \cdot P_A(R)$$
An Example

Suppose Ann chooses $U$ and Bob chooses $R$

Are these choices rational?
An Example

Suppose Ann chooses $U$ and Bob chooses $R$

Are these choices rational?

<table>
<thead>
<tr>
<th></th>
<th>Bob</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Ann</td>
<td>$L$</td>
<td>$R$</td>
</tr>
<tr>
<td>$U$</td>
<td>3,3</td>
<td>0,0</td>
</tr>
<tr>
<td>$D$</td>
<td>0,0</td>
<td>1,1</td>
</tr>
</tbody>
</table>

Suppose Ann chooses $U$ and Bob chooses $R$

Are these choices rational?
An Example

Suppose Ann chooses $U$ and Bob chooses $R$

Are these choices rational?
An Example

Suppose Ann chooses $U$ and Bob chooses $R$

Are these choices rational?

$$3 \cdot P_A(L) + 0 \cdot P_A(R) \geq 0 \cdot P_A(L) + 1 \cdot P_A(R)$$
An Example

Suppose Ann chooses $U$ and Bob chooses $R$

Are these choices rational?

\[ 3 \cdot P_A(L) + 0 \cdot P_A(R) \geq 0 \cdot P_A(L) + 1 \cdot P_A(R) \]
Suppose Ann chooses $U$ and Bob chooses $R$.

Are these choices rational?

$$3 \cdot \frac{1}{2} + 0 \cdot P_A(R) \geq 0 \cdot \frac{1}{2} + 1 \cdot P_A(R)$$
An Example

Suppose Ann chooses $U$ and Bob chooses $R$.

Are these choices rational?

\[
\begin{array}{c|cc}
  & L & R \\ 
 U & 3,3 & 0,0 \\ 
 D & 0,0 & 1,1 \\ 
\end{array}
\]

\[
3 \cdot \frac{1}{2} + 0 \cdot \frac{1}{2} \geq 0 \cdot \frac{1}{2} + 1 \cdot \frac{1}{2}
\]
Suppose Ann chooses $U$ and Bob chooses $R$.

Are these choices rational?
An Example

Suppose Ann chooses $U$ and Bob chooses $R$

Are these choices *rational*?

\[
\begin{array}{c|cc}
 & L & R \\
\hline
U & 3,3 & 0,0 \\
D & 0,0 & 1,1 \\
\end{array}
\]

\[
0 \cdot \frac{2}{12} + 1 \cdot \frac{10}{12} \geq 3 \cdot \frac{2}{12} + 0 \cdot \frac{10}{12}
\]
An Example

Suppose Ann chooses $U$ and Bob chooses $R$.

*Are these choices rational?*

Yes.

![Game Matrix](image)

<table>
<thead>
<tr>
<th></th>
<th>Bob</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Ann</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$U$</td>
<td></td>
<td>3,3</td>
</tr>
<tr>
<td>$D$</td>
<td></td>
<td>0,0</td>
</tr>
</tbody>
</table>

Logic and Artificial Intelligence
An Example

Suppose Ann chooses $U$ and Bob chooses $R$.

Are these choices rational?

Yes.

Bob (Ann) knows that Ann (Bob) is rational.

$0 \cdot \frac{1}{6} + 1 \cdot \frac{5}{6} \geq 3 \cdot \frac{1}{6} + 0 \cdot \frac{5}{6}$
Two Issues

Zero probability $\neq$ "impossible"

Different "types" of players can make the same choice.

Are Ann and Bob rational? Yes.

Do they know that each other is rational? No.

$\text{Pr}_\text{Bob}(\text{Irrat}_\text{Ann}) = 0$
Two Issues

1. Zero probability $\neq$ “impossible”

<table>
<thead>
<tr>
<th></th>
<th>L</th>
<th>R</th>
</tr>
</thead>
<tbody>
<tr>
<td>U</td>
<td>3,3</td>
<td>0,0</td>
</tr>
<tr>
<td>D</td>
<td>0,0</td>
<td>1,1</td>
</tr>
</tbody>
</table>

Bob

Ann

$\text{Pr}_\text{Bob}(\text{Irrat}_\text{Ann}) = 0$
Two Issues

1. Zero probability ≠ “impossible”

2. Different “types” of players can make the same choice

<table>
<thead>
<tr>
<th></th>
<th>Bob</th>
</tr>
</thead>
<tbody>
<tr>
<td>L</td>
<td>3,3</td>
</tr>
<tr>
<td>U</td>
<td>0,0</td>
</tr>
</tbody>
</table>

Ann and Bob are rational, but they do not know that each other is rational (though $\Pr_{Bob}(\text{Irrat}_{Ann}) = 0$).
Two Issues

1. Zero probability ≠ “impossible”

2. Different “types” of players can make the same choice

- Are Ann and Bob rational?
Two Issues

<table>
<thead>
<tr>
<th></th>
<th>U</th>
<th>D</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>L</strong></td>
<td>3,3</td>
<td>0,0</td>
</tr>
<tr>
<td><strong>R</strong></td>
<td>0,0</td>
<td>1,1</td>
</tr>
</tbody>
</table>

1. Zero probability $\neq$ “impossible”
2. Different “types” of players can make the same choice
   - Are Ann and Bob rational? **Yes.**
Two Issues

1. Zero probability \( \neq \) “impossible”

2. Different “types” of players can make the same choice

- Are Ann and Bob rational? Yes.
- Do they know that each other is rational? No.
Two Issues

1. Zero probability \( \neq \) “impossible”

2. Different “types” of players can make the same choice

- Are Ann and Bob rational? Yes.
- Do they know that each other is rational? No. (though \( Pr_{Bob}(Irrat(Ann)) = 0 \))
Next: Common Knowledge