Ten Puzzles and Paradoxes about Knowledge and Belief

ESSLLI 2013, Düsseldorf

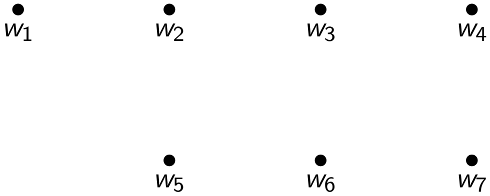
Wes Holliday **Eric Pacuit**

August 15, 2013

Robert Aumann. *Agreeing to Disagree*. Annals of Statistics **4** (1976).

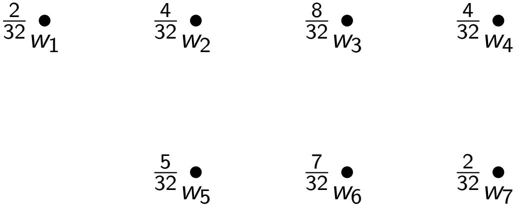
Theorem. Suppose that n agents share a common prior and have different private information. If there is common knowledge in the group of the posterior probabilities, then the posteriors must be equal.

2 Scientists Perform an Experiment



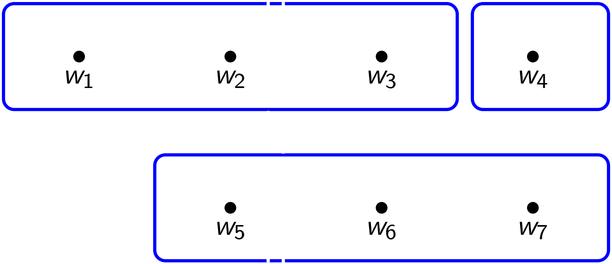
They agree the true state is one of seven different states.

2 Scientists Perform an Experiment



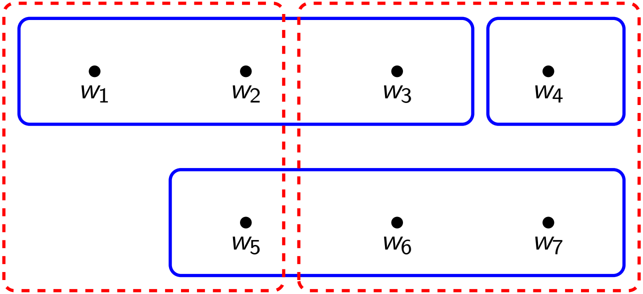
They agree on a common prior.

2 Scientists Perform an Experiment



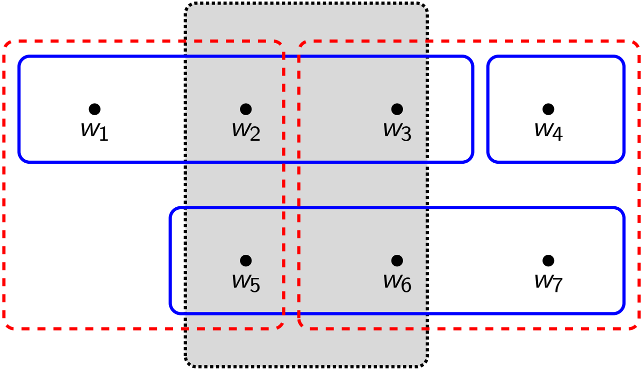
They agree that Experiment 1 would produce the blue partition.

2 Scientists Perform an Experiment



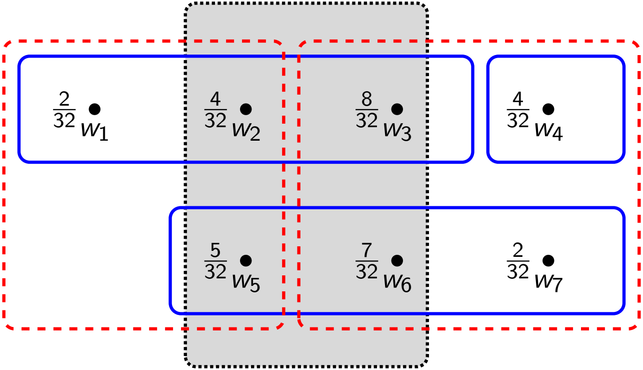
They agree that Experiment 1 would produce the blue partition and Experiment 2 the red partition.

2 Scientists Perform an Experiment



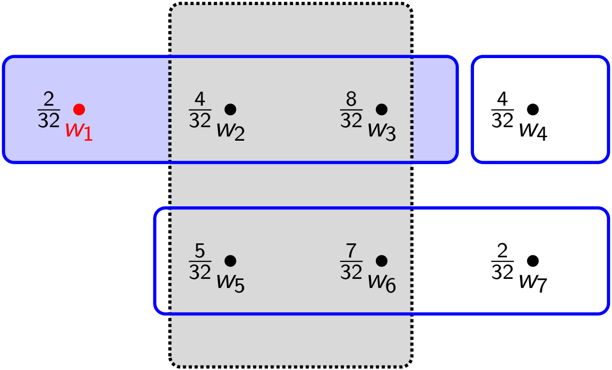
They are interested in the truth of $E = \{w_2, w_3, w_5, w_6\}$.

2 Scientists Perform an Experiment



So, they agree that $P(E) = \frac{24}{32}$.

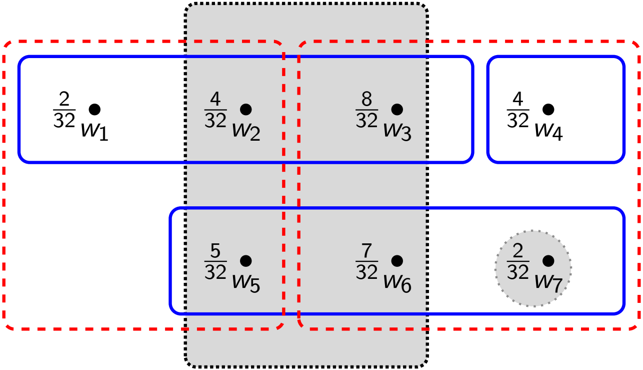
2 Scientists Perform an Experiment



Also, that *if the true state is* w_1, then Experiment 1 will yield

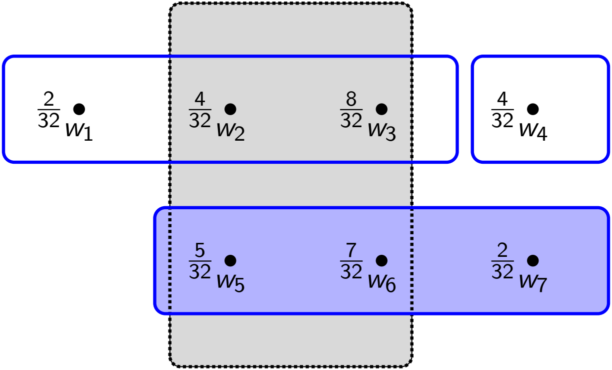
$$P(E|I) = \frac{P(E \cap I)}{P(I)} = \frac{12}{14}$$

2 Scientists Perform an Experiment



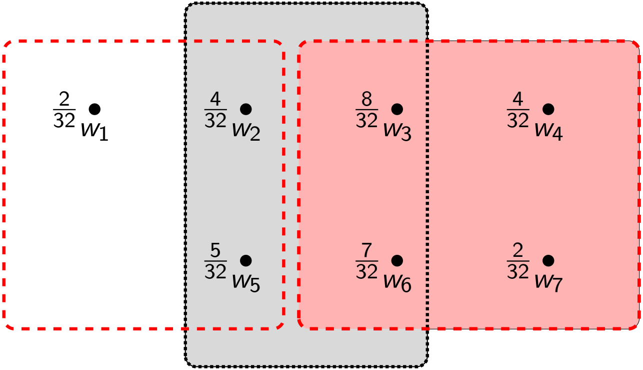
Suppose the true state is w_7 and the agents preform the experiments.

2 Scientists Perform an Experiment



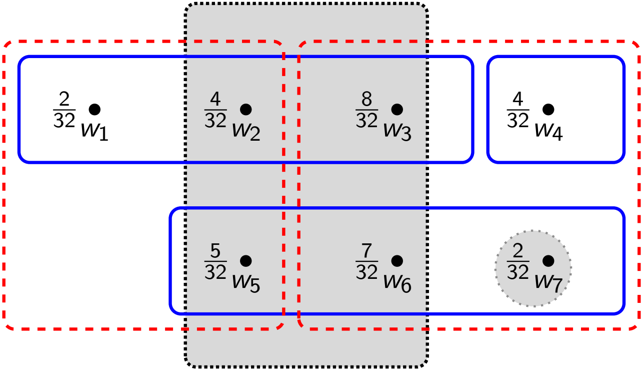
Suppose the true state is w_7, then $Pr_1(E) = \frac{12}{14}$

2 Scientists Perform an Experiment



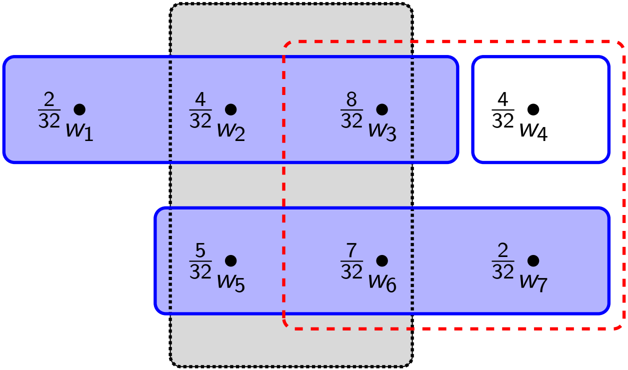
Then $Pr_1(E) = \frac{12}{14}$ and $Pr_2(E) = \frac{15}{21}$

2 Scientists Perform an Experiment



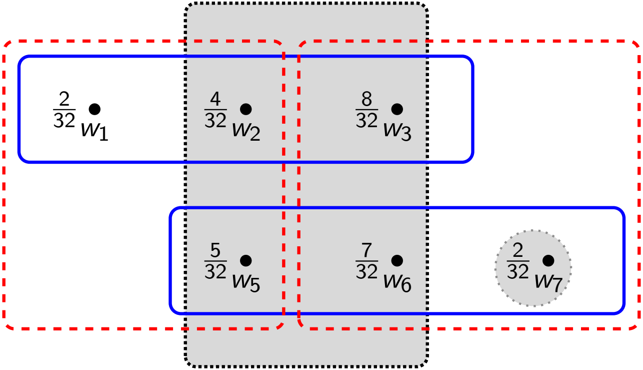
Suppose they exchange emails with the new subjective probabilities: $Pr_1(E) = \frac{12}{14}$ and $Pr_2(E) = \frac{15}{21}$

2 Scientists Perform an Experiment



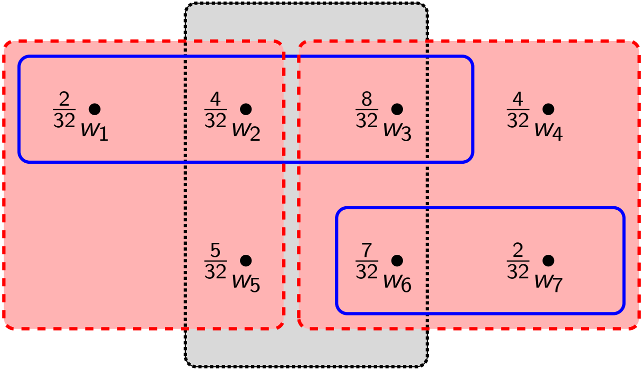
Agent 2 learns that w_4 is **NOT** the true state (same for Agent 1).

2 Scientists Perform an Experiment



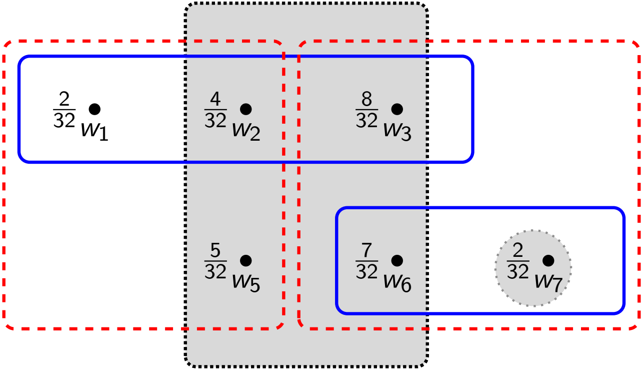
Agent 2 learns that w_4 is **NOT** the true state (same for Agent 1).

2 Scientists Perform an Experiment



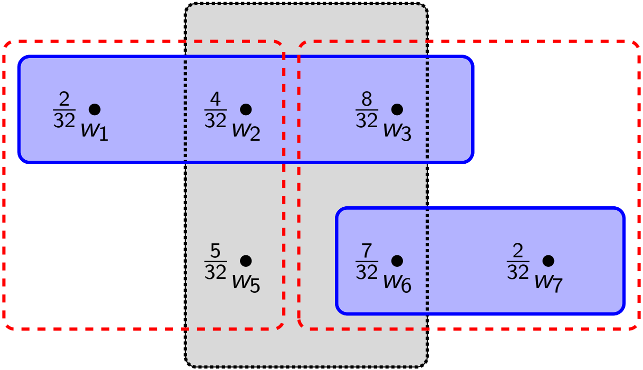
Agent 1 learns that w_5 is **NOT** the true state (same for Agent 1).

2 Scientists Perform an Experiment



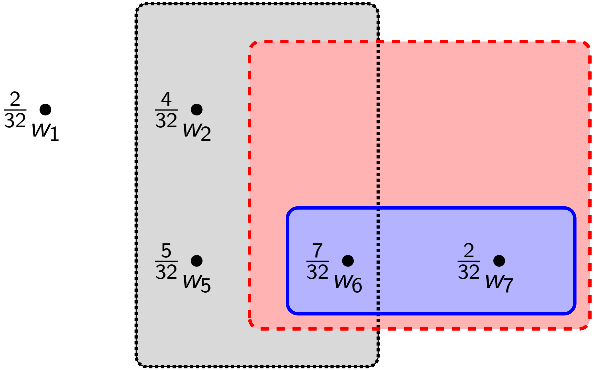
The new probabilities are $Pr_1(E|I') = \frac{7}{9}$ and $Pr_2(E|I') = \frac{15}{17}$

2 Scientists Perform an Experiment



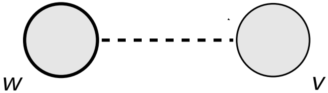
After exchanging this information ($Pr_1(E|I') = \frac{7}{9}$ and $Pr_2(E|I') = \frac{15}{17}$), Agent 2 learns that w_3 is **NOT** the true state.

2 Scientists Perform an Experiment



No more revisions are possible and the agents agree on the posterior probabilities.

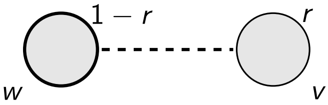
Adding Probabilities



$\mathcal{M} = \langle W, \{\Pi_i\}_{i \in \mathcal{A}} \rangle$
Π_i is agent i's partition with $\Pi_i(w)$ the partition cell containing w.

$K_i(E) = \{w \mid \Pi_i(w) \subseteq E\}$

Adding Probabilities



$\mathcal{M} = \langle W, \{\Pi_i\}_{i\in\mathcal{A}}, \{p_i\}_{i\in\mathcal{A}}\rangle$
for each i, $p_i : W \to [0,1]$ is a probability measure

$B_i^r(E) = \{w \mid p_i(E \mid \Pi_i(w)) = \frac{\pi_i(E\ \cap\ \Pi_i(w))}{p_i(\Pi_i(w))} \geq r\}$

1. $B_i^r(B_i^r(E)) = B_i^r(E)$

2. If $E \subseteq F$ then $B_i^r(E) \subseteq B_i^r(F)$

3. $\pi(E \mid B_i^r(E)) \geq r$

What is common belief in a probabilistic setting?

Fact. For all $i \in \mathcal{A}$ and $E \subseteq W$, $K_i C(E) = C(E)$.

Fact. For all $i \in \mathcal{A}$ and $E \subseteq W$, $K_i C(E) = C(E)$.

> Suppose you are told "Ann and Bob are going together,"' and respond "sure, that's common knowledge." What you mean is not only that everyone knows this, but also that the announcement is pointless, occasions no surprise, reveals nothing new; in effect, that the situation after the announcement does not differ from that before. ...the event "Ann and Bob are going together" — call it E — is common knowledge if and only if some event — call it F — happened that entails E and also entails all players' knowing F (like all players met Ann and Bob at an intimate party). *(Aumann, pg. 271, footnote 8)*

Fact. For all $i \in \mathcal{A}$ and $E \subseteq W$, $K_i C(E) = C(E)$.

An event F is **self-evident** if $K_i(F) = F$ for all $i \in \mathcal{A}$.

Fact. An event E is commonly known iff some self-evident event that entails E obtains.

Fact. For all $i \in \mathcal{A}$ and $E \subseteq W$, $K_i C(E) = C(E)$.

An event F is **self-evident** if $K_i(F) = F$ for all $i \in \mathcal{A}$.

Fact. An event E is commonly known iff some self-evident event that entails E obtains.

Common r-belief

The typical example of an event that creates common knowledge is a **public announcement**.

Common r-belief

The typical example of an event that creates common knowledge is a **public announcement**.

Shouldn't one always allow for some small probability that a participant was absentminded, not listening, sending a text, checking facebook, proving a theorem, asleep, ...

Common r-belief

The typical example of an event that creates common knowledge is a **public announcement**.

Shouldn't one always allow for some small probability that a participant was absentminded, not listening, sending a text, checking facebook, proving a theorem, asleep, ...

"We show that the weaker concept of "common belief" can function successfully as a substitute for common knowledge in the theory of equilibrium of Bayesian games."

D. Monderer and D. Samet. *Approximating Common Knowledge with Common Beliefs*. Games and Economic Behavior (1989).

Common r-belief: definition

$$B_i^r(E) = \{w \mid p_i(E \mid \Pi_i(w)) \geq r\}$$

Common r-belief: definition

$$B_i^r(E) = \{w \mid p_i(E \mid \Pi_i(w)) \geq r\}$$

An event E is **evident r-belief** if for each $i \in \mathcal{A}$, $E \subseteq B_i^r(E)$

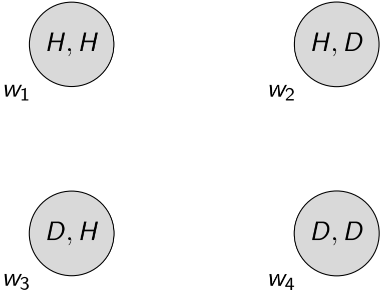
Common r-belief: definition

$$B_i^r(E) = \{w \mid p_i(E \mid \Pi_i(w)) \geq r\}$$

An event E is **evident** r**-belief** if for each $i \in \mathcal{A}$, $E \subseteq B_i^r(E)$

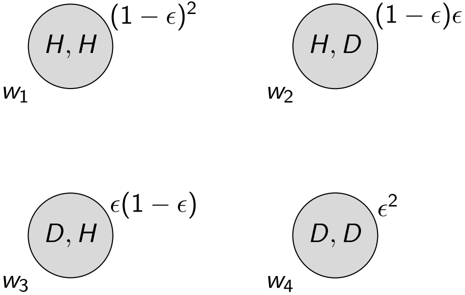
An event F is **common** r**-belief** at w if there exists an evident r-belief event E such that $w \in E$ and for all $i \in \mathcal{A}$, $E \subseteq B_i^r(F)$

Common r-belief: example



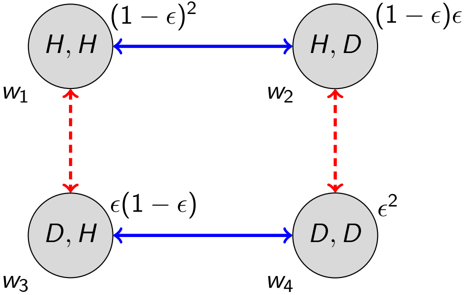
Two agents either hear (H) or don't hear (D) the announcement.

Common r-belief: example



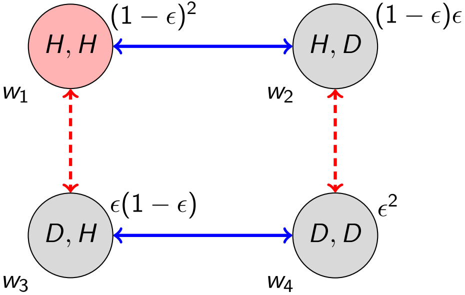
The probability that an agent hears is $1 - \epsilon$.

Common *r*-belief: example



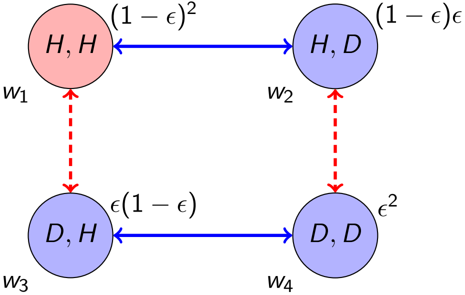
The agents *know* their "type".

Common r-belief: example



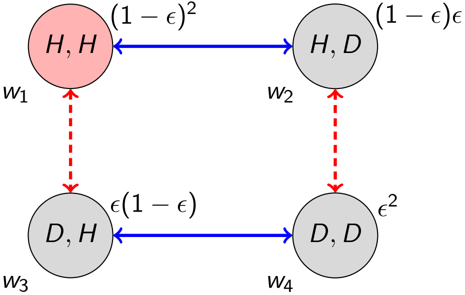
The event "everyone hears" ($E = \{w_1\}$)

Common r-belief: example



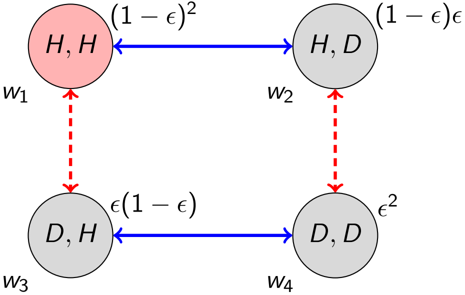
The event "everyone hears" ($E = \{w_1\}$) is **not** common knowledge

Common r-belief: example



The event "everyone hears" ($E = \{w_1\}$) is **not** common knowledge, but it is common $(1-\epsilon)$-belief

Common r-belief: example



The event "everyone hears" ($E = \{w_1\}$) is **not** common knowledge, but it is common $(1-\epsilon)$-belief:
$B_i^{(1-\epsilon)}(E) = \{w \mid p_i(E \mid \Pi_i(w)) \geq 1-\epsilon\} = \{w_1\} = E$, for $i = 1, 2$

Common r-belief

Theorem. If the posteriors of an event X are common r-belief at some state w, then any two posteriors can differ by at most $2(1-r)$.

D. Samet and D. Monderer. *Approximating Common Knowledge with Common Beliefs*. Games and Economic Behavior, Vol. 1, No. 2, 1989.

Recap

Assuming common prior...

- there cannot be common knowledge that the posterior probabilities are different.
- like-minded individuals cannot agree to make different decisions.
- common belief to a “high degree” implies that the posterior probabilities are very close.

Recap

Assuming common prior...

- there cannot be common knowledge that the posterior probabilities are different.
- like-minded individuals cannot agree to make different decisions.
- common belief to a "high degree" implies that the posterior probabilities are very close.

Assumptions

- The truth axiom and $p_i(E \mid B_i^r(E)) \geq r$.
- The (interpersonal) sure-thing principle

Sure-Thing Principle

Should I study or have a beer?

Sure-Thing Principle

Should I study or have a beer? Either I pass or I won't pass the exam.

Sure-Thing Principle

Should I study or have a beer? Either I pass or I won't pass the exam. If I pass, it is better to drink and pass, so I should drink. If I fail, it is better to drink and fail, so I should drink.

Sure-Thing Principle

Should I study or have a beer? Either I pass or I won't pass the exam. If I pass, it is better to drink and pass, so I should drink. If I fail, it is better to drink and fail, so I should drink. I should drink in either case, so I should have a drink.

Sure-Thing Principle

It is not the logical principle $\varphi \to \chi$ and $\psi \to \chi$ then $\varphi \vee \psi \to \chi$.

Sure-Thing Principle

It is not the logical principle $\varphi \to \chi$ and $\psi \to \chi$ then $\varphi \lor \psi \to \chi$.
There is a book I want to read which was written by one of two authors.

Sure-Thing Principle

It is not the logical principle $\varphi \to \chi$ and $\psi \to \chi$ then $\varphi \vee \psi \to \chi$. There is a book I want to read which was written by one of two authors. If I know it is written by author A then I will read it. If I know it is written by author B then I will read it.

Sure-Thing Principle

It is not the logical principle $\varphi \to \chi$ and $\psi \to \chi$ then $\varphi \vee \psi \to \chi$. There is a book I want to read which was written by one of two authors. If I know it is written by author A then I will read it. If I know it is written by author B then I will read it. If I know it is written by either author A or author B then I may not choose to read the book.

Sure-Thing Principle

There are three candidates, republican, independent and democrat.

Sure-Thing Principle

There are three candidates, republican, independent and democrat. I will buy stock if the democrat looses and I will buy stock if the republican looses.

Sure-Thing Principle

There are three candidates, republican, independent and democrat. I will buy stock if the democrat looses and I will buy stock if the republican looses. Either the republican or democrat will loose. So, I should buy the stock.

R. Aumann, S. Hart and M. Perry. *Conditioning and the Sure-Thing Principle*. manuscript, 2005.

The Nixon Diamond

You're told (from a reliable source) that Nixon is a republican, which suggests that he is a Hawk. You're also told (from a reliable source) that Nixon is a Quaker, which suggests that he is a Dove.

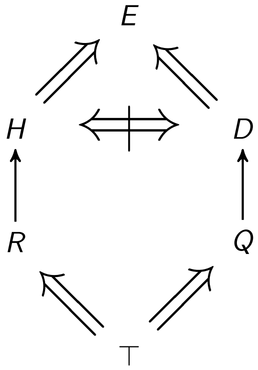
The Nixon Diamond

You're told (from a reliable source) that Nixon is a republican, which suggests that he is a Hawk. You're also told (from a reliable source) that Nixon is a Quaker, which suggests that he is a Dove. Either being a Hawk or a Dove implies having extreme political views.

The Nixon Diamond

You're told (from a reliable source) that Nixon is a republican, which suggests that he is a Hawk. You're also told (from a reliable source) that Nixon is a Quaker, which suggests that he is a Dove. Either being a Hawk or a Dove implies having extreme political views. Should you conclude that Nixon has extreme political views?

Floating Conclusions



J. Horty. *Skepticism and floating conclusions*. Artificial Intelligence, 135, pp. 55 - 72, 2002.

Your parents have 1M inheritance which will is split between you mother and father (each may give you 0.5M).

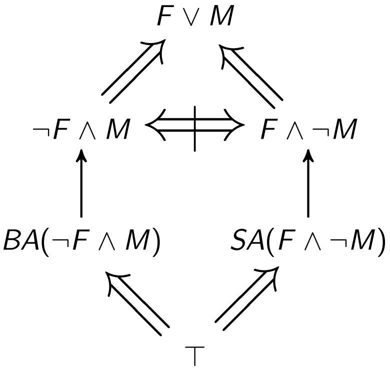
Your parents have 1M inheritance which will is split between you mother and father (each may give you 0.5M). Your brother (a reliable source) says that you will receive the money from your Mother (but not your Father).

Your parents have 1M inheritance which will is split between you mother and father (each may give you 0.5M). Your brother (a reliable source) says that you will receive the money from your Mother (but not your Father). Your sister (a reliable source) says that you will receive the money from your Father (but not your Mother).

Your parents have 1M inheritance which will is split between you mother and father (each may give you 0.5M). Your brother (a reliable source) says that you will receive the money from your Mother (but not your Father). Your sister (a reliable source) says that you will receive the money from your Father (but not your Mother). You want to buy a yacht which requires a large deposit and you can only afford it provided you inherit the money.

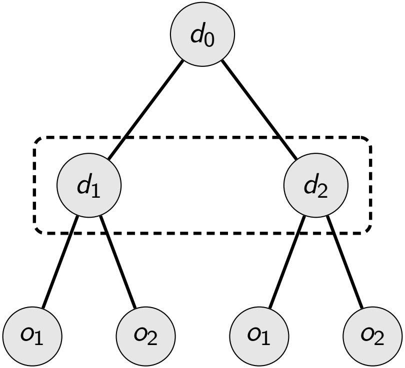
Your parents have 1M inheritance which will is split between you mother and father (each may give you 0.5M). Your brother (a reliable source) says that you will receive the money from your Mother (but not your Father). Your sister (a reliable source) says that you will receive the money from your Father (but not your Mother). You want to buy a yacht which requires a large deposit and you can only afford it provided you inherit the money. Should you make a deposit on the yacht?

Floating Conclusions, II



The Absent-Minded Driver

Games of Imperfect Information



The Absent-Minded Driver

An individual is sitting late at night in a bar planning his midnight trip home. In order to get home he has to take the highway and get off at the second exit.

The Absent-Minded Driver

An individual is sitting late at night in a bar planning his midnight trip home. In order to get home he has to take the highway and get off at the second exit. Turning at the first exit leads into a disastrous area (payoff 0). Turning at the second exit yields the highest reward (payoff 4).

The Absent-Minded Driver

An individual is sitting late at night in a bar planning his midnight trip home. In order to get home he has to take the highway and get off at the second exit. Turning at the first exit leads into a disastrous area (payoff 0). Turning at the second exit yields the highest reward (payoff 4). If he continues beyond the second exit, he cannot go back and at the end of the highway he will find a motel where he can spend the night (payoff 1).

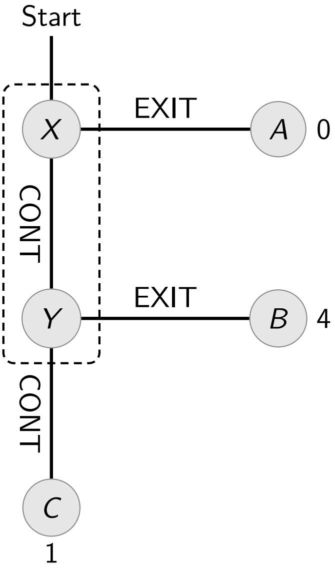
The Absent-Minded Driver

The driver is absentminded and is aware of this fact. At an intersection, he cannot tell whether it is the first or the second intersection and he cannot remember how many he has passed (one can make the situation more realistic by referring to the 17th intersection).

The Absent-Minded Driver

The driver is absentminded and is aware of this fact. At an intersection, he cannot tell whether it is the first or the second intersection and he cannot remember how many he has passed (one can make the situation more realistic by referring to the 17th intersection). While sitting at the bar, all he can do is to decide whether or not to exit at an intersection. (pg. 7)

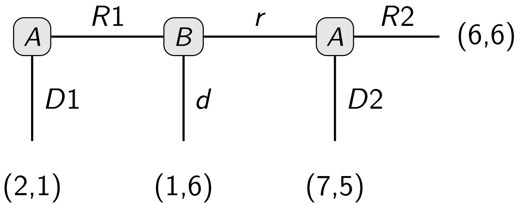
M. Piccione and A. Rubinstein. *On the Interpretation of Decision Problems with Imperfect Recall*. Games and Econ Behavior, 20, pgs. 3- 24, 1997.


Start
EXIT
X
A
0
CONT
EXIT
Y
B
4
CONT
C
1

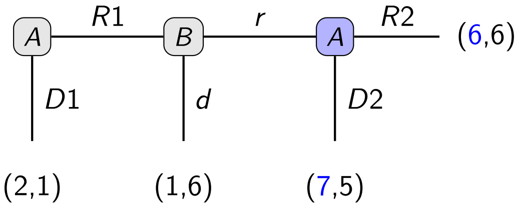
Planning stage: While planning his trip home at the bar, the decision maker is faced with a choice between "Continue; Continue" and "Exit". Since he cannot distinguish between the two intersections, he cannot plan to "Exit" at the second intersection (he must plan the same behavior at both X and Y). Since "Exit" will lead to the worst outcome (with a payoff of 0), the optimal strategy is "Continue; Continue" with a guaranteed payoff of 1.

Action stage: When arriving at an intersection, the decision maker is faced with a local choice of either “Exit” or “Continue” (possibly followed by another decision). Now the decision maker knows that since he committed to the plan of choosing “Continue” at each intersection, it is possible that he is at the second intersection. Indeed, the decision maker concludes that he is at the first intersection with probability $1/2$. But then, his expected payoff for “Exit” is 2, which is greater than the payoff guaranteed by following the strategy he previously committed to. Thus, he chooses to “Exit”.

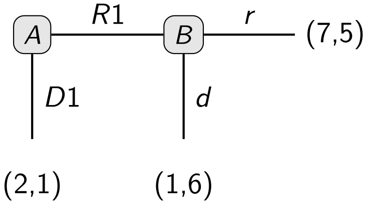
BI Puzzle



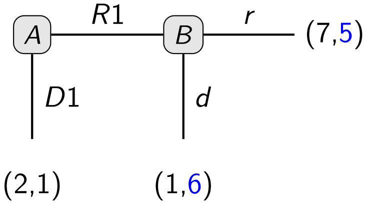
BI Puzzle



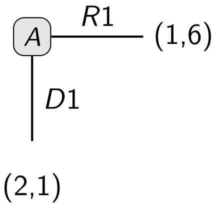
BI Puzzle



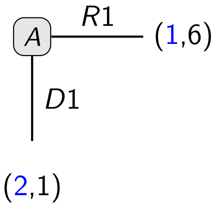
BI Puzzle



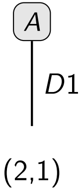
BI Puzzle



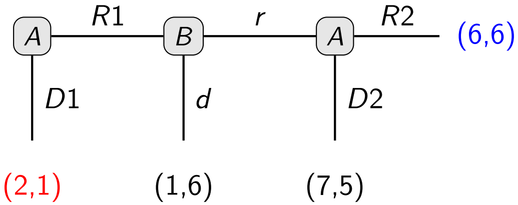
BI Puzzle



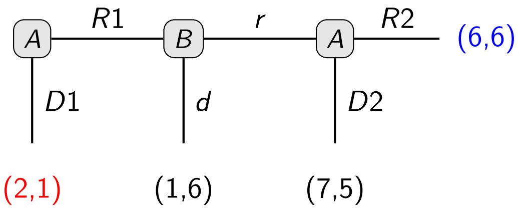
BI Puzzle



But what if...



But what if...



"On the one hand, Under common knowledge of rationality, *A* *must* go out on the first move. On the other hand, the backward induction argument for this is based on what the players *would* do if *A* stayed in. But, if she did stay in, then common knowledge of rationality is violated, so the argument that she will go out no longer has a basis."

R. Aumann. *Backwards induction and common knowledge of rationality*. Games and Economic Behavior, 8, pgs. 6 - 19, 1995.

R. Stalnaker. *Knowledge, belief and counterfactual reasoning in games*. Economics and Philosophy, 12, pgs. 133 - 163, 1996.

J. Halpern. *Substantive Rationality and Backward Induction*. Games and Economic Behavior, 37, pp. 425-435, 1998.

Models of Extensive Games

Let Γ be a *non-degenerate* extensive game with perfect information. Let Γ_i be the set of nodes controlled by player i.

Models of Extensive Games

Let Γ be a *non-degenerate* extensive game with perfect information. Let Γ_i be the set of nodes controlled by player i.

A strategy profile σ describes the choice for each player i at all vertices where i can choose.

Models of Extensive Games

Let Γ be a *non-degenerate* extensive game with perfect information. Let Γ_i be the set of nodes controlled by player i.

A strategy profile σ describes the choice for each player i at all vertices where i can choose.

Given a vertex v in Γ and strategy profile σ, σ specifies a unique path from v to an end-node.

Models of Extensive Games

Let Γ be a *non-degenerate* extensive game with perfect information. Let Γ_i be the set of nodes controlled by player i.

A strategy profile σ describes the choice for each player i at all vertices where i can choose.

Given a vertex v in Γ and strategy profile σ, σ specifies a unique path from v to an end-node.

$\mathcal{M}(\Gamma) = \langle W, \sim_i, \sigma \rangle$ where $\sigma : W \rightarrow Strat(\Gamma)$ and $\sim_i \subseteq W \times W$ is an equivalence relation.

Models of Extensive Games

Let Γ be a *non-degenerate* extensive game with perfect information. Let Γ_i be the set of nodes controlled by player i.

A strategy profile σ describes the choice for each player i at all vertices where i can choose.

Given a vertex v in Γ and strategy profile σ, σ specifies a unique path from v to an end-node.

$\mathcal{M}(\Gamma) = \langle W, \sim_i, \sigma \rangle$ where $\sigma : W \to Strat(\Gamma)$ and $\sim_i \subseteq W \times W$ is an equivalence relation.

If $\sigma(w) = \sigma$, then $\sigma_i(w) = \sigma_i$ and $\sigma_{-i}(w) = \sigma_{-i}$

Models of Extensive Games

Let Γ be a *non-degenerate* extensive game with perfect information. Let Γ_i be the set of nodes controlled by player i.

A strategy profile σ describes the choice for each player i at all vertices where i can choose.

Given a vertex v in Γ and strategy profile σ, σ specifies a unique path from v to an end-node.

$\mathcal{M}(\Gamma) = \langle W, \sim_i, \sigma \rangle$ where $\sigma : W \to Strat(\Gamma)$ and $\sim_i \subseteq W \times W$ is an equivalence relation.

If $\sigma(w) = \sigma$, then $\sigma_i(w) = \sigma_i$ and $\sigma_{-i}(w) = \sigma_{-i}$

(A1) If $w \sim_i w'$ then $\sigma_i(w) = \sigma_i(w')$.

Rationality

$h_i^v(\sigma)$ denote "i's payoff if σ is followed from node v"

Rationality

$h_i^v(\sigma)$ denote "i's payoff if σ is followed from node v"

i **is rational at** v **in** w provided for all strategies $s_i \neq \sigma_i(w)$, $h_i^v(\sigma(w')) \geq h_i^v((\sigma_{-i}(w'), s_i))$ for some $w' \in [w]_i$.

Substantive Rationality

i is **substantively rational** in state w if i is rational at a vertex v in w of every vertex in $v \in \Gamma_i$

Stalnaker Rationality

For every vertex $v \in \Gamma_i$, *if* i *were to actually reach* v*, then what he would do in that case would be rational*.

Stalnaker Rationality

For every vertex $v \in \Gamma_i$, *if* i *were to actually reach* v, *then what he would do in that case would be rational*.

$f : W \times \Gamma_i \to W$, $f(w, v) = w'$, then w' is the "closest state to w where the vertex v is reached.

Stalnaker Rationality

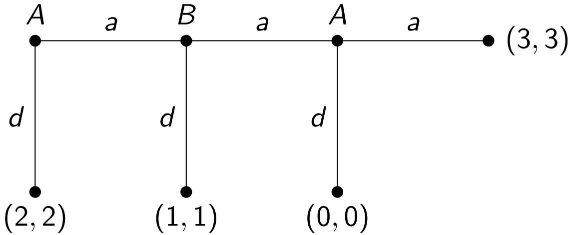
For every vertex $v \in \Gamma_i$, *if* i *were to actually reach* v, *then what he would do in that case would be rational*.

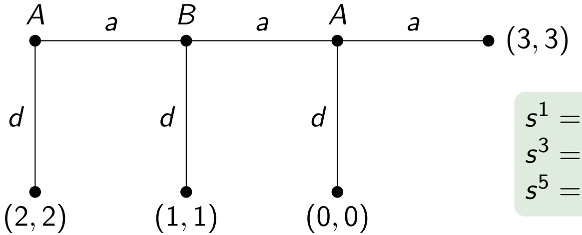
$f : W \times \Gamma_i \to W$, $f(w, v) = w'$, then w' is the "closest state to w where the vertex v is reached.

(F1) v is reached in $f(w, v)$ (i.e., v is on the path determined by $\sigma(f(w, v)))$

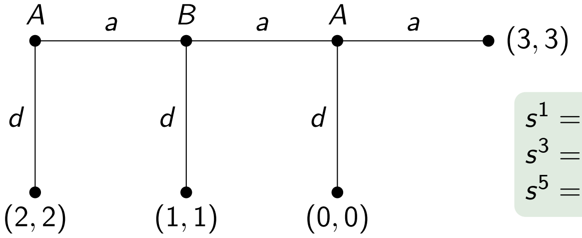
(F2) If v is reached in w, then $f(w, v) = w$

(F3) $\sigma(f(w, v))$ and $\sigma(w)$ agree on the subtree of Γ below v


A
a
B
a
A
a
(3, 3)
d
d
d
(2, 2)
(1, 1)
(0, 0)

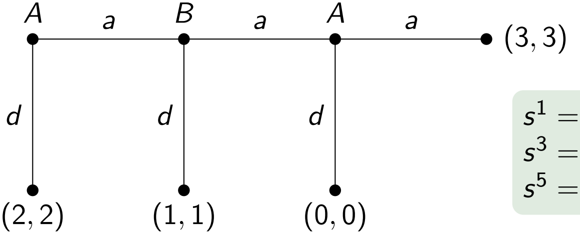


$s^1 = (da, d)$, $s^2 = (aa, d)$,
$s^3 = (ad, d)$, $s^4 = (aa, a)$,
$s^5 = (ad, a)$

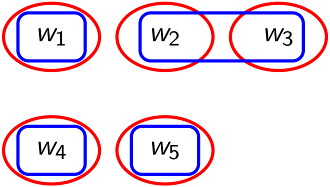


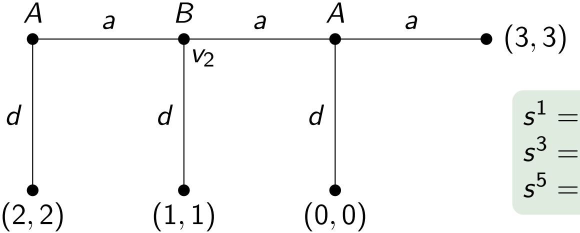
$s^1 = (da, d)$, $s^2 = (aa, d)$,
$s^3 = (ad, d)$, $s^4 = (aa, a)$,
$s^5 = (ad, a)$

- $W = \{w_1, w_2, w_3, w_4, w_5\}$ with $\sigma(w_i) = s^i$
- $[w_i]_A = \{w_i\}$ for $i = 1, 2, 3, 4, 5$
- $[w_i]_B = \{w_i\}$ for $i = 1, 4, 5$ and $[w_2]_B = [w_3]_B = \{w_2, w_3\}$

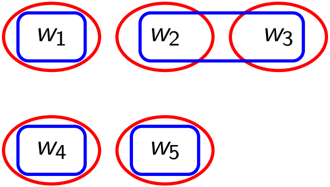


$s^1 = (da, d)$, $s^2 = (aa, d)$,
$s^3 = (ad, d)$, $s^4 = (aa, a)$,
$s^5 = (ad, a)$

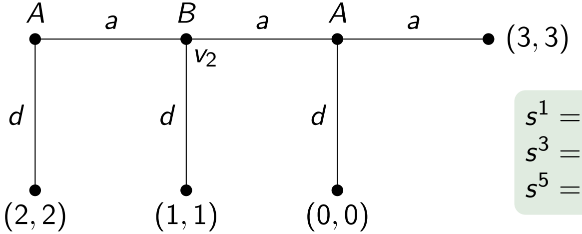




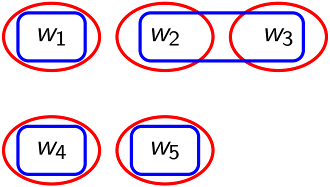
$s^1 = (da, d)$, $s^2 = (aa, d)$,
$s^3 = (ad, d)$, $s^4 = (aa, a)$,
$s^5 = (ad, a)$



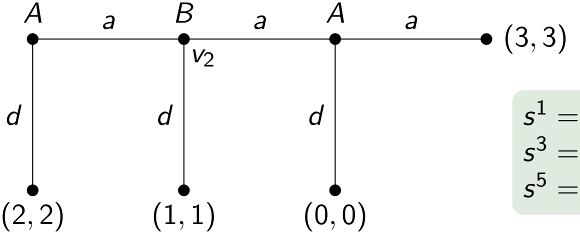
It is **common knowledge** at w_1 that if vertex v_2 were reached, Bob would play down.



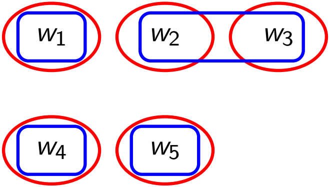
$s^1 = (da, d)$, $s^2 = (aa, d)$,
$s^3 = (ad, d)$, $s^4 = (aa, a)$,
$s^5 = (ad, a)$



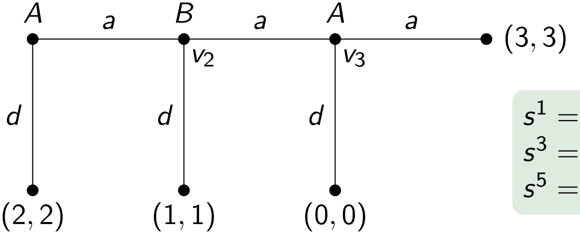
Bob is not rational at v_2 in w_1



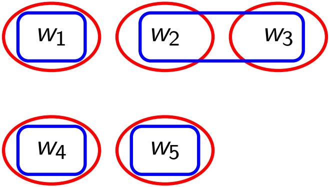
$s^1 = (da, d)$, $s^2 = (aa, d)$,
$s^3 = (ad, d)$, $s^4 = (aa, a)$,
$s^5 = (ad, a)$



Bob is rational at v_2 in w_2



$s^1 = (da, d)$, $s^2 = (aa, d)$,
$s^3 = (ad, d)$, $s^4 = (aa, a)$,
$s^5 = (ad, a)$



Note that $f(w_1, v_2) = w_2$ and $f(w_1, v_3) = w_4$, so there is common knowledge of S-rationality at w_1.

Aumann's Theorem: If Γ is a non-degenerate game of perfect information, then in all models of Γ, we have $C(A - Rat) \subseteq BI$

Stalnaker's Theorem: There exists a non-degenerate game Γ of perfect information and an extended model of Γ in which the selection function satisfies F1-F3 such that $C(S - Rat) \not\subseteq BI$.

Aumann's Theorem: If Γ is a non-degenerate game of perfect information, then in all models of Γ, we have $C(A - Rat) \subseteq BI$

Stalnaker's Theorem: There exists a non-degenerate game Γ of perfect information and an extended model of Γ in which the selection function satisfies F1-F3 such that $C(S - Rat) \not\subseteq BI$.

Revising beliefs during play:

"the rationality of choices in a game depends not only on what players believe, but also on their policies for revising their beliefs" (p. 31)

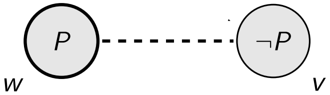
R. Stalnaker. *Belief revision in games: Forward and backward induction*. Mathematical Social Sciences, 36, pgs. 31 - 56, 1998.

F4. For all players i and vertices v, if $w' \in [f(w,v)]_i$ then there exists a state $w'' \in [w]_i$ such that $\sigma(w')$ and $\sigma(w'')$ agree on the subtree of Γ below v.

Theorem (Halpern). If Γ is a non-degenerate game of perfect information, then for every extended model of Γ in which the selection function satisfies F1-F4, we have $C(S-Rat) \subseteq BI$. Moreover, there is an extend model of Γ in which the selection function satisfies F1-F4.

J. Halpern. *Substantive Rationality and Backward Induction*. Games and Economic Behavior, 37, pp. 425-435, 1998.

Taking Stock



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

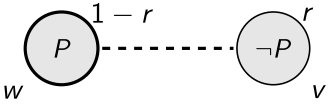
- wR_iv means v is compatible with everything i knows at w.

Language: $\varphi := p \mid \neg\varphi \mid \varphi \wedge \psi \mid K_i\varphi$

Truth:

- $\mathcal{M}, w \models p$ iff $w \in V(p)$ (p an atomic proposition)
- Boolean connectives as usual
- $\mathcal{M}, w \models K_i\varphi$ iff for all $v \in W$, if $w \sim_i v$ then $\mathcal{M}, v \models \varphi$

Taking Stock



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

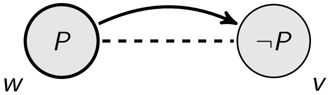
- $p_i : W \to [0,1]$ are probabilities, $\sim_i$ is an equivalence relation

Language: $\varphi := p \mid \neg\varphi \mid \varphi \wedge \psi \mid K_i\varphi \mid B^r\psi$

Truth:

- $[\![\varphi]\!]_{\mathcal{M}} = \{w \mid \mathcal{M}, w \models \varphi\}$
- $\mathcal{M}, w \models B^r\varphi$ iff $p_i([\![\varphi]\!]_{\mathcal{M}} \mid [w]_i) = \frac{p_i([\![\varphi]\!]_{\mathcal{M}} \cap [w]_i)}{\pi_i([w]_i)} \geq r$
- $\mathcal{M}, w \models K_i\varphi$ iff for all $v \in W$, if $w \sim_i v$ then $\mathcal{M}, v \models \varphi$

Taking Stock



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

- $w \preceq_i v$ means v is at least as plausibility as w for agent i.

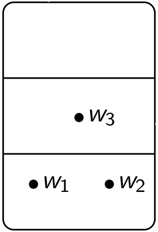
Language: $\varphi := p \mid \neg\varphi \mid \varphi \wedge \psi \mid K_i\varphi \mid B^\varphi\psi \mid [\preceq_i]\varphi$

Truth:

- $[\![\varphi]\!]_\mathcal{M} = \{w \mid \mathcal{M}, w \models \varphi\}$
- $\mathcal{M}, w \models B_i^\varphi\psi$ iff for all $v \in Min_{\preceq_i}([\![\varphi]\!]_\mathcal{M} \cap [w]_i)$, $\mathcal{M}, v \models \psi$
- $\mathcal{M}, w \models [\preceq_i]\varphi$ iff for all $v \in W$, if $v \preceq_i w$ then $\mathcal{M}, v \models \varphi$

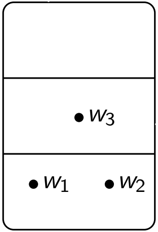
More on Plausibility Structures

- $w_1 \sim w_2 \sim w_3$



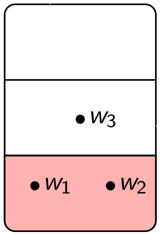
More on Plausibility Structures

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

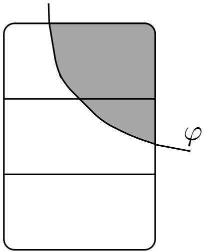


More on Plausibility Structures

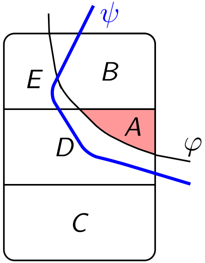
- $w_1 \sim w_2 \sim w_3$
- $w_1 \preceq w_2$ and $w_2 \preceq w_1$ (w_1 and w_2 are equi-plausbile)
- $w_1 \prec w_3$ ($w_1 \preceq w_3$ and $w_3 \not\preceq w_1$)
- $w_2 \prec w_3$ ($w_2 \preceq w_3$ and $w_3 \not\preceq w_2$)
- $\{w_1, w_2\} \subseteq Min_{\preceq}([w_i])$



More on Plausibility Structures

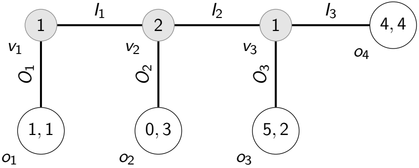


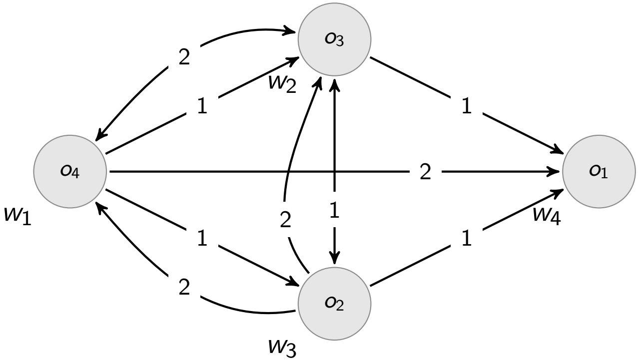
More on Plausibility Structures

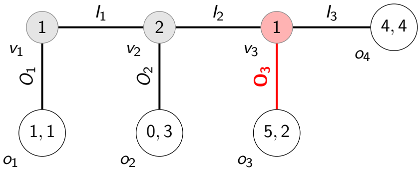


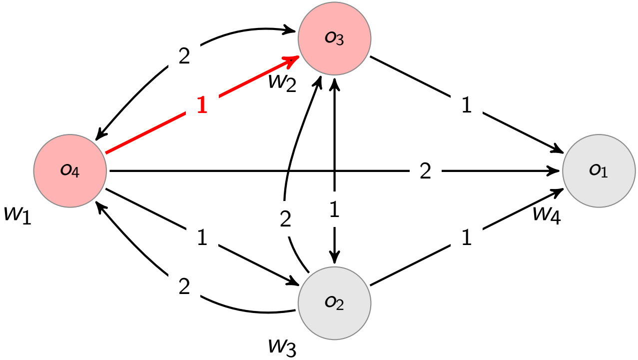
Conditional Belief: $B^{\varphi}\psi$

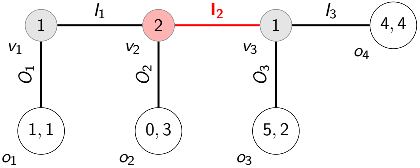
$$Min_{\preceq}(W \cap [\![\varphi]\!]_{\mathcal{M}}) \subseteq [\![\psi]\!]_{\mathcal{M}}$$

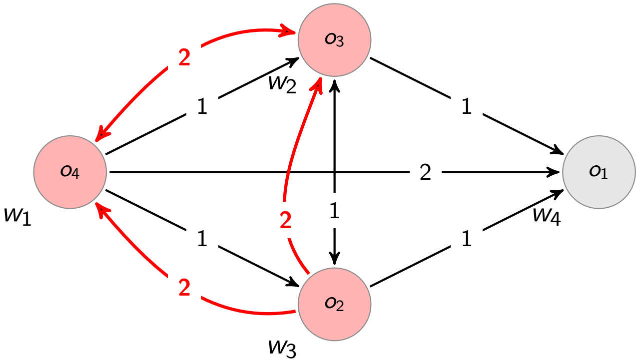

1
2
1
4, 4
l1
l2
l3
v1
v2
v3
o4
O1
O2
O3
1, 1
0, 3
5, 2
o1
o2
o3


o3
o4
o1
o2
w1
w2
w3
w4
2
1
1
2
2
1
1
2
1
1
2


1
2
1
4, 4
l_1
l_2
l_3
v_1
v_2
v_3
o_4
O_1
O_2
O_3
1, 1
0, 3
5, 2
o_1
o_2
o_3


o_3
o_4
o_1
o_2
w_1
w_2
w_3
w_4
2
1
1
2
1
2
1
1
2


1
l_1
2
l_2
1
l_3
4, 4
v_1
v_2
v_3
o_4
O_1
O_2
O_3
1, 1
0, 3
5, 2
o_1
o_2
o_3


o_3
o_4
o_1
o_2
w_1
w_2
w_3
w_4
2
1
1
2
1
2
1
1
2
2
1

Game play as public announcemnets

$$\mathsf{v} := \bigvee_{v \rightsquigarrow o} \mathsf{o}$$

$$\mathcal{M} = \mathcal{M}^{!\mathsf{v}_1}; \mathcal{M}^{!\mathsf{v}_2}; \mathcal{M}^{!\mathsf{v}_3}; \mathcal{M}^{!\mathsf{o}_4}$$

The Dynamics of Rational Play

A. Baltag, S. Smets and J. Zvesper. *Keep 'hoping' for rationality: a solution to the backward induction paradox*. Synthese, 169, pgs. 301 - 333, 2009.

Hard vs. Soft Information in a Game

The structure of the game and past moves are ‘hard information: *irrevocably known*

Hard vs. Soft Information in a Game

The structure of the game and past moves are 'hard information: *irrevocably known*

Players' 'knowledge' of other players' rationality and 'knowledge' of her own future moves at nodes not yet reached are not of the same degree of certainty.

Hard vs. Soft Information in a Game

The structure of the game and past moves are ‘hard information: *irrevocably known*

Players’ ‘knowledge’ of other players’ rationality and ‘knowledge’ of her own future moves at nodes not yet reached are not of the same degree of certainty.

What belief revision policy leads to BI?

Dynamic Rationality: The event R that all players are *rational* changes during the play of the game.

Players are assumed to be "incurably optimistic" about the rationality of their opponents.

What belief revision policy leads to BI?

Dynamic Rationality: The event R that all players are *rational* changes during the play of the game.

Players are assumed to be "incurably optimistic" about the rationality of their opponents.

Theorem (Baltag, Smets and Zvesper). Common knowledge of the game structure, of open future and *common stable belief* in dynamic rationality implies common belief in the backward induction outcome.

$$Ck(Struct_G \wedge F_G \wedge [\,!\,]CbRat) \rightarrow Cb(BI_G)$$