Epistemic Game Theory

Lecture 12

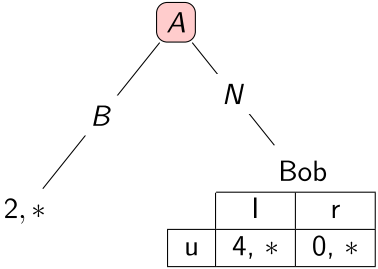
Eric Pacuit

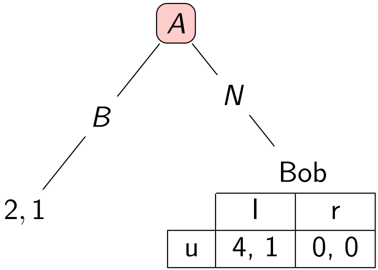
University of Maryland, College Park
pacuit.org
epacuit@umd.edu

April 28, 2014

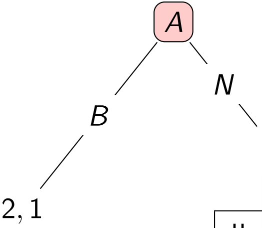
Rationalizability Assumption

Instead of Bob changing his opinion about Ann's rationality, maintaining his belief about her passive beliefs, he might have maintained his belief in her rationality, changing his beliefs about her beliefs about him.

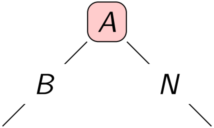

A
B
N
2, ∗
Bob
l
r
u
4, ∗
0, ∗



A
B
N
2, 1
Bob
l
r
u
4, 1
0, 0



		Bob	
		l	r
Ann	u	4, 1	0, 0
	d	0, 0	1, 4



Bob

Ann	l	r
u	2, 1	-2, 0
d	-2, 0	-1, 4

Bob

Ann	l	r
u	4, 1	0, 0
d	0, 0	1, 4

Rationalization Principle

A player should believe that all players are perfectly rational, and this belief should be robust relative to any compatible information about the behavior of any player....

Rationalization Principle

A player should believe that all players are perfectly rational, and this belief should be robust relative to any compatible information about the behavior of any player....If you are surprised by the actions of some player, you should change your beliefs about that player's passive beliefs, rather than about her rationality. If possible, find an alternative hypothesis about her beliefs about other players that will make what she does perfectly rational.

Eliminate weakly dominated strategies for *just two* rounds, and then eliminate *strictly* dominated strategies iteratively.

"Theorem": It can be proved that al and only strategies that survive this process are realizable in sufficiently rich models in which it is common belief that all players are rational, and that all revise their beliefs in conformity with the rationalization principle.

Eliminate weakly dominated strategies for *just two* rounds, and then eliminate *strictly* dominated strategies iteratively.

"Theorem": It can be proved that al and only strategies that survive this process are realizable in sufficiently rich models in which it is common belief that all players are rational, and that all revise their beliefs in conformity with the rationalization principle.

Sufficiently rich: For any players i and j, for any possible world w and any admissible strategy s for i, there is a possible world v such that $w \approx_j v$, $\mathbf{s}_i(v) = s$ and i is perfectly rational in v.

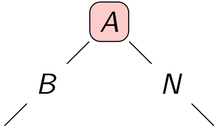
If it is logically possible for i to play s rationally, then it is *conceivable* for j that i should have the beliefs that make it rational for i to play s.

Eliminate weakly dominated strategies for *just two* rounds, and then eliminate *strictly* dominated strategies iteratively.

"Theorem": It can be proved that all and only strategies that survive this process are realizable in sufficiently rich models in which it is common belief that all players are rational, and that all revise their beliefs in conformity with the rationalization principle.

Eliminate weakly dominated strategies for *just two* rounds, and then eliminate *strictly* dominated strategies iteratively.

"Theorem": It can be proved that all and only strategies that survive this process are realizable in sufficiently rich models in which it is common belief that all players are rational, and that all revise their beliefs in conformity with the rationalization principle.



		Bob	
		l	r
Ann	u	2, 1	-2, 0
	d	-2, 0	-1, 4

		Bob	
		l	r
Ann	u	4, 1	0, 0
	d	0, 0	1, 4

		Bob			
		ll	lr	rl	rr
Ann	Bu	2, 1	2, 1	-2, 0	-2, 0
	Bd	-2, 0	-2, 0	-1, 4	-1, 4
	Nu	4, 1	0, 0	4, 1	0, 0
	Nd	0, 0	1, 4	0, 0	1, 4

		Bob			
		ll	lr	rl	rr
Ann	Bu	2, 1	2, 1	-2, 0	-2, 0
	Bd	-2, 0	-2, 0	-1, 4	-1, 4
	Nu	4, 1	0, 0	4, 1	0, 0
	Nd	0, 0	1, 4	0, 0	1, 4

		Bob			
		ll	lr	rl	rr
Ann	Bu	2, 1	2, 1	-2, 0	-2, 0
	Bd	-2, 0	-2, 0	-1, 4	-1, 4
	Nu	4, 1	0, 0	4, 1	0, 0
	Nd	0, 0	1, 4	0, 0	1, 4

		Bob			
		ll	lr	rl	rr
Ann	Bu	2, 1	2, 1	-2, 0	-2, 0
	Bd	-2, 0	-2, 0	-1, 4	-1, 4
	Nu	4, 1	0, 0	4, 1	0, 0
	Nd	0, 0	1, 4	0, 0	1, 4

		Bob			
		ll	lr	rl	rr
Ann	Bu	2, 1	2, 1	-2, 0	-2, 0
	Bd	-2, 0	-2, 0	-1, 4	-1, 4
	Nu	4, 1	0, 0	4, 1	0, 0
	Nd	0, 0	1, 4	0, 0	1, 4

		Bob			
		ll	lr	rl	rr
Ann	Bu	2, 1	2, 1	-2, 0	-2, 0
	Bd	-2, 0	-2, 0	-1, 4	-1, 4
	Nu	4, 1	0, 0	4, 1	0, 0
	Nd	0, 0	1, 4	0, 0	1, 4

		Bob			
		ll	lr	rl	rr
Ann	Bu	2, 1	2, 1	-2, 0	-2, 0
	Bd	-2, 0	-2, 0	-1, 4	-1, 4
	Nu	4, 1	0, 0	4, 1	0, 0
	Nd	0, 0	1, 4	0, 0	1, 4

Bob

Ann	ll	lr	rl	rr
Bu	2, 1	2, 1	-2, 0	-2, 0
Bd	-2, 0	-2, 0	-1, 4	-1, 4
Nu	4, 1	0, 0	4, 1	0, 0
Nd	0, 0	1, 4	0, 0	1, 4

"...in general, a payer's beliefs about what another player will do are based on an inference from two other kinds of beliefs: beliefs about the passive beliefs of that player, and beliefs about her rationality.

"...in general, a payer's beliefs about what another player will do are based on an inference from two other kinds of beliefs: beliefs about the passive beliefs of that player, and beliefs about her rationality. If one's prediction based on these beliefs is defeated, one must choose whether to revise one's belief about the other players's beliefs or one's belief that she is rational...

"...in general, a payer's beliefs about what another player will do are based on an inference from two other kinds of beliefs: beliefs about the passive beliefs of that player, and beliefs about her rationality. If one's prediction based on these beliefs is defeated, one must choose whether to revise one's belief about the other players's beliefs or one's belief that she is rational...But the assumption that the rationalization principle is common belief is itself an assumption about the passive beliefs of other players, and so it is itself something that (according to the principle) might have to be given up in the face of surprising behavioral information. So the rationalization principle undermines its own stability." (pg. 51, Stalnaker)

		Bob			
		ll	lr	rl	rr
Ann	Bu	2, 1	2, 1	-2, 0	-2, 0
	Bd	-2, 0	-2, 0	-1, 4	-1, 4
	Nu	4, 1	0, 0	4, 1	0, 0
	Nd	0, 0	1, 4	0, 0	1, 4

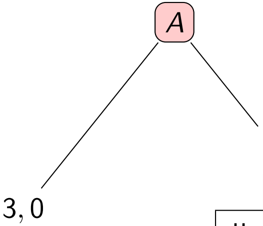
		Bob			
		ll	lr	rl	rr
Ann	Bu	2, 1	2, 1	-2, 0	-2, 0
	Bd	-2, 0	-2, 0	-1, 4	-1, 4
	Nu	4, 1	0, 0	4, 1	0, 0
	Nd	0, 0	1, 4	0, 0	1, 4

- Bob believes that Ann will choose *Bu*, because she believes that Bob will play *lr*
- If Bob were surprised by Ann choosing *N*, he is disposed to infer that she believed instead that he was choosing strategy *rr*, and so would make the rational response to this belief, *Nd*.
- Ann is still rational in the world in which she chooses *Nd*, and so Bob's belief revision conforms to the raitonalizability principle
- Ann is perfectly rational since, regardless of her belief revision policy, *Bu* is the *only* rational response to her belief about Bob.

		Bob			
		ll	lr	rl	rr
Ann	Bu	2, 1	2, 1	-2, 0	-2, 0
	Bd	-2, 0	-2, 0	-1, 4	-1, 4
	Nu	4, 1	0, 0	4, 1	0, 0
	Nd	0, 0	1, 4	0, 0	1, 4

- The final steps in the forward induction argument are blocked, since we cannot assume that belief in the rationalization principle itself will be robust.

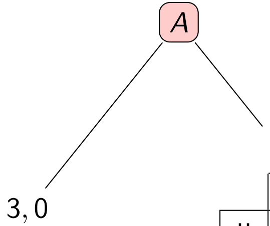
Backwards vs. Forwards Induction



Bob

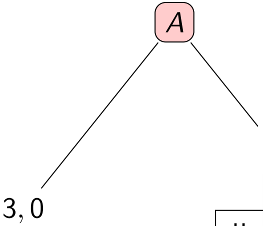
		l	c	r
	u	2, 2	2, 1	0, 0
Ann	d	1, 1	1, 2	4, 0

Backwards vs. Forwards Induction



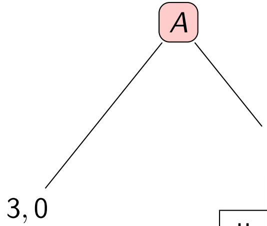
		Bob		
		l	c	r
Ann	u	2, 2	2, 1	0, 0
	d	1, 1	1, 2	4, 0

Backwards vs. Forwards Induction



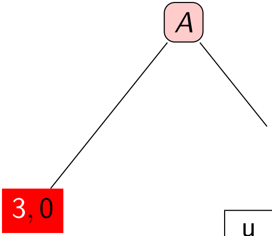
		Bob		
		l	c	r
Ann	u	2, 2	2, 1	0, 0
	d	1, 1	1, 2	4, 0

Backwards vs. Forwards Induction



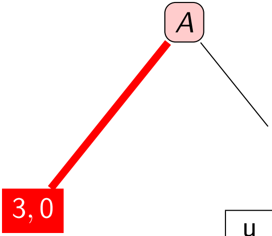
		Bob		
		l	c	r
Ann	u	2, 2	2, 1	0, 0
	d	1, 1	1, 2	4, 0

Backwards vs. Forwards Induction



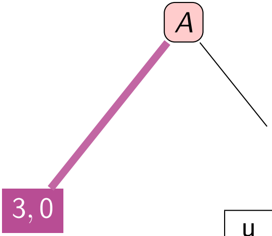
		Bob		
		l	c	r
Ann	u	2, 2	2, 1	0, 0
	d	1, 1	1, 2	4, 0

Backwards vs. Forwards Induction



		Bob		
		l	c	r
Ann	u	2, 2	2, 1	0, 0
	d	1, 1	1, 2	4, 0

Backwards vs. Forwards Induction



		Bob		
		l	c	r
	u	2, 2	2, 1	0, 0
Ann	d	1, 1	1, 2	4, 0

P. Battigalli and M. Siniscalchi. *Strong Belief and Forward Induction Reasoning*. Journal of Economic Theory, 106, pgs. 356 - 391, 2002.

Resiliency

B. Skyrms. *Resiliency, propensities, and causal necessity*. Journal of Philosophy, 74:11, pgs. 704 - 713, 1977.

A. Baltag and S. Smets. *Probabilistic Belief Revision*. Synthese, 2008.

H. Leitgeb. *Reducing belief simpliciter to degrees of belief*. Annals of Pure and Applied Logic, 16:4, pgs. 1338 - 1380, 2013.

Probability

Let W be a set of states and $\mathfrak{A}$ a σ-algebra: $\mathfrak{A} \subseteq \wp(W)$ such that

- $W, \emptyset \in \mathfrak{A}$
- if $X \in \mathfrak{A}$ then $W - X \in \mathfrak{A}$
- if $X, Y \in \mathfrak{A}$ then $X \cup Y \in \mathfrak{A}$
- if $X_0, X_1, \ldots \in \mathfrak{A}$ then $\bigcup_{i \in \mathbb{N}} X_i \in \mathfrak{A}$.

Probability

$P : \mathfrak{A} \rightarrow [0, 1]$ satisfying the usual constraints

- $P(W) = 1$
- (finite additivity) If $X_1, X_2 \in \mathfrak{A}$ are pairwise disjoint, then $P(X_1 \cup X_2) = P(X_1) + P(X_2)$

$P(Y|X) = \frac{P(Y \cap X)}{P(X)}$ whenever $P(X) > 0$. So, $P(Y|W)$ is $P(Y)$.

- P is countably additive (σ-additive): if $X_1, X_2, \ldots, X_n, \ldots$ are pairwise disjoint members of $\mathfrak{A}$, then $P(\bigcup_{n \in \mathbb{N}} X_n) = \sum_{n \in \mathbb{N}} P(X_n)$

CPS (Popper Space)

A **conditional probability space** (CPS) over $(W, \mathfrak{A})$ is a tuple $(W, \mathfrak{A}, \mathfrak{A}', \mu)$ such that $\mathfrak{A}$ is an algebra over W, $\mathfrak{A}'$ is a set of subsets of W (not necessarily an algebra) that does not contain $\emptyset$ and $\mu : \mathfrak{A} \times \mathfrak{A}' \to [0, 1]$ satisfying the following conditions:

1. $\mu(U \mid U) = 1$ if $U \in \mathfrak{A}'$
2. $\mu(E_1 \cup E_1 \mid U) = \mu(E_1 \mid U) + \mu(E_2 \mid U)$ if $E_1 \cap E_2 = \emptyset$, $U \in \mathfrak{A}'$ and $E_1, E_2 \in \mathfrak{A}$
3. $\mu(E \mid U) = \mu(E \mid X) \times \mu(X \mid U)$ if $E \subseteq X \subseteq U$, $U, X \in \mathfrak{A}'$ and $E \in \mathfrak{A}$.

Certainty: $P(H) = 1$

Absolute Certainty: $P(H \mid E) = 1$ for all E

Strong Belief: $w \in SB(H)$ iff for all $E \in \mathfrak{A}$ with $H \cap E \neq \emptyset$ and $P(E) \neq 0$: $P(H \mid E) \geq t$

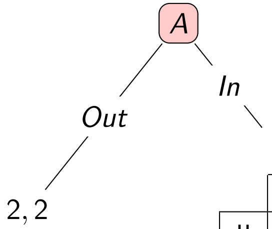
Strong Defeasible Belief: $w \in U(H)$ iff for all $E \in \mathfrak{A}$ with $w \in E$ and $P(E) \neq 0$: $P(H \mid E) \geq t$

Certainty: $P(H) = 1$

Absolute Certainty: $P(H \mid E) = 1$ for all E

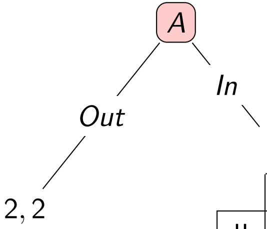
Strong Belief: $w \in SB(H)$ iff for all $E \in \mathfrak{A}$ with $H \cap E \neq \emptyset$ and $P(E) \neq 0$: $P(H \mid E) \geq t$

Strong Defeasible Belief: $w \in U(H)$ iff for all $E \in \mathfrak{A}$ with $w \in E$ and $P(E) \neq 0$: $P(H \mid E) \geq t$



		Bob	
		l	r
Ann	u	3, 1	0, 0
	d	0, 0	1, 3

ω_1	$g_{1,\emptyset}(t_1)$	$g_{1,(In)}(t_1)$	ω_2	$g_{2,\emptyset}(t_1)$	$g_{2,(In)}(t_1)$
$(In\ d, t_1^1)$	$0, 1, 0$	$0, 1, 0$	(l, t_2^1)	$0, 1, 0, 0, 0$	$0, 1, 0, 0, 0$
$(In\ u, t_1^1)$	$0, 1, 0$	$0, 1, 0$	(r, t_2^2)	$0, 0, 1, 0, 0$	$1, 0, 0, 0, 0$
$(Out\ d, t_1^1)$	$0, 1, 0$	$0, 1, 0$	(l, t_2^3)	$0, 0, 0, 0, 1$	$0, 0, 0, 0, 1$
$(Out\ u, t_1^1)$	$0, 1, 0$	$0, 1, 0$			
$(In\ u, t_1^2)$	$0, 0, 1$	$0, 0, 1$			

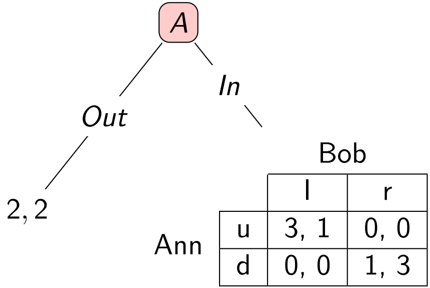


Bob

Ann	l	r
u	3, 1	0, 0
d	0, 0	1, 3

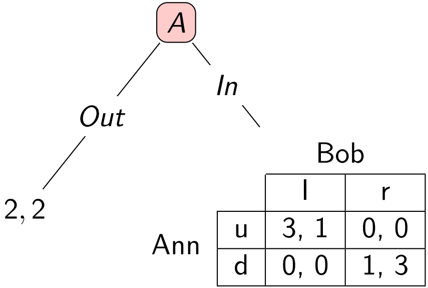
ω_1	$g_{1,\emptyset}(t_1)$	$g_{1,(In)}(t_1)$	ω_2	$g_{2,\emptyset}(t_1)$	$g_{2,(In)}(t_1)$
$(In\ d, t_1^1)$	$0,1,0$	$0,1,0$	(l, t_2^1)	$0,1,0,0,0$	$0,1,0,0,0$
$(In\ u, t_1^1)$	$0,1,0$	$0,1,0$	(r, t_2^2)	$0,0,1,0,0$	$1,0,0,0,0$
$(Out\ d, t_1^1)$	$0,1,0$	$0,1,0$	(l, t_2^3)	$0,0,0,0,1$	$0,0,0,0,1$
$(Out\ u, t_1^1)$	$0,1,0$	$0,1,0$			
$(In\ u, t_1^2)$	$0,0,1$	$0,0,1$			

Player 1 is rational (R_1)



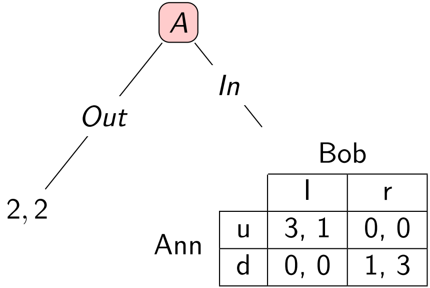
ω_1	$g_{1,\emptyset}(t_1)$	$g_{1,(In)}(t_1)$	ω_2	$g_{2,\emptyset}(t_1)$	$g_{2,(In)}(t_1)$
$(In\ d, t_1^1)$	$0,1,0$	$0,1,0$	(l, t_2^1)	$0,1,0,0,0$	$0,1,0,0,0$
$(In\ u, t_1^1)$	$0,1,0$	$0,1,0$	(r, t_2^2)	$0,0,1,0,0$	$1,0,0,0,0$
$(Out\ d, t_1^1)$	$0,1,0$	$0,1,0$	(l, t_2^3)	$0,0,0,0,1$	$0,0,0,0,1$
$(Out\ u, t_1^1)$	$0,1,0$	$0,1,0$			
$(In\ u, t_1^2)$	$0,0,1$	$0,0,1$			

Ann initially believes Bob would play R after observing In $B_{A,\emptyset}([s_2 = r])$



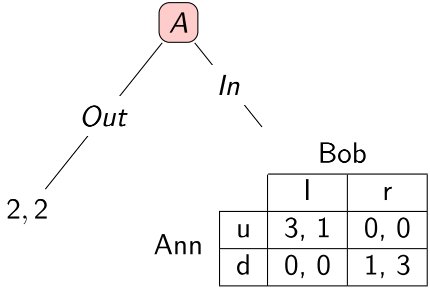
ω_1	$g_{1,\emptyset}(t_1)$	$g_{1,(In)}(t_1)$	ω_2	$g_{2,\emptyset}(t_1)$	$g_{2,(In)}(t_1)$
$(In\ d, t_1^1)$	$0, 1, 0$	$0, 1, 0$	(l, t_2^1)	$0, 1, 0, 0, 0$	$0, 1, 0, 0, 0$
$(In\ u, t_1^1)$	$0, 1, 0$	$0, 1, 0$	(r, t_2^2)	$0, 0, 1, 0, 0$	$1, 0, 0, 0, 0$
$(Out\ d, t_1^1)$	$0, 1, 0$	$0, 1, 0$	(l, t_2^3)	$0, 0, 0, 0, 1$	$0, 0, 0, 0, 1$
$(Out\ u, t_1^1)$	$0, 1, 0$	$0, 1, 0$			
$(In\ u, t_1^2)$	$0, 0, 1$	$0, 0, 1$			

$$R_1 \cap B_{1,\emptyset}([s_2 = R]) = [Out]$$



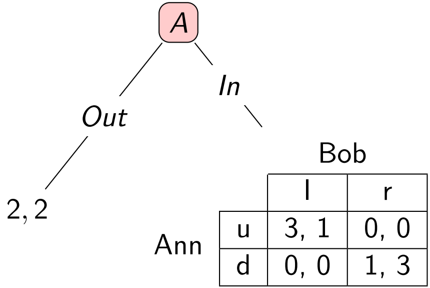
ω_1	$g_{1,\emptyset}(t_1)$	$g_{1,(In)}(t_1)$	ω_2	$g_{2,\emptyset}(t_1)$	$g_{2,(In)}(t_1)$
$(In\ d, t_1^1)$	$0, 1, 0$	$0, 1, 0$	(l, t_2^1)	$0, 1, 0, 0, 0$	$0, 1, 0, 0, 0$
$(In\ u, t_1^1)$	$0, 1, 0$	$0, 1, 0$	(r, t_2^2)	$0, 0, 1, 0, 0$	$1, 0, 0, 0, 0$
$(Out\ d, t_1^1)$	$0, 1, 0$	$0, 1, 0$	(l, t_2^3)	$0, 0, 0, 0, 1$	$0, 0, 0, 0, 1$
$(Out\ u, t_1^1)$	$0, 1, 0$	$0, 1, 0$			
$(In\ u, t_1^2)$	$0, 0, 1$	$0, 0, 1$			

R_1 and $B_{1,\emptyset}([s_2 = R])$ are consistent with $[In]$



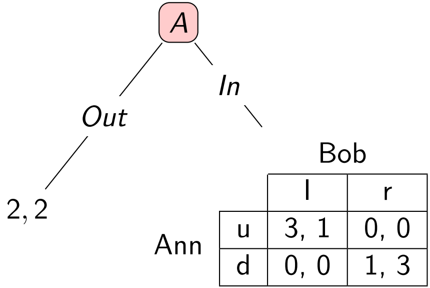
ω_1	$g_{1,\emptyset}(t_1)$	$g_{1,(In)}(t_1)$	ω_2	$g_{2,\emptyset}(t_1)$	$g_{2,(In)}(t_1)$
$(In\ d, t_1^1)$	$0,1,0$	$0,1,0$	(l, t_2^1)	$0,1,0,0,0$	$0,1,0,0,0$
$(In\ u, t_1^1)$	$0,1,0$	$0,1,0$	(r, t_2^2)	$0,0,1,0,0$	$1,0,0,0,0$
$(Out\ d, t_1^1)$	$0,1,0$	$0,1,0$	(l, t_2^3)	$0,0,0,0,1$	$0,0,0,0,1$
$(Out\ u, t_1^1)$	$0,1,0$	$0,1,0$			
$(In\ u, t_1^2)$	$0,0,1$	$0,0,1$			

$SB_2(R_1) \subseteq B_{2,(In)}(R_1)$ and $SB_2(B_{1,\emptyset}([s_2 = R])) \subseteq B_{2,(In)}(B_{1,\emptyset}([s_2 = R]))$



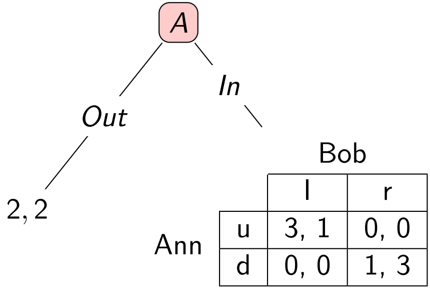
ω_1	$g_{1,\emptyset}(t_1)$	$g_{1,(In)}(t_1)$	ω_2	$g_{2,\emptyset}(t_1)$	$g_{2,(In)}(t_1)$
$(In\ d, t_1^1)$	$0, 1, 0$	$0, 1, 0$	(l, t_2^1)	$0, 1, 0, 0, 0$	$0, 1, 0, 0, 0$
$(In\ u, t_1^1)$	$0, 1, 0$	$0, 1, 0$	(r, t_2^2)	$0, 0, 1, 0, 0$	$1, 0, 0, 0, 0$
$(Out\ d, t_1^1)$	$0, 1, 0$	$0, 1, 0$	(l, t_2^3)	$0, 0, 0, 0, 1$	$0, 0, 0, 0, 1$
$(Out\ u, t_1^1)$	$0, 1, 0$	$0, 1, 0$			
$(In\ u, t_1^2)$	$0, 0, 1$	$0, 0, 1$			

$$SB_2(R_1) \cap SB_2(B_{1,\emptyset}([s_2 = R])) \subseteq B_{2,(In)}([Out]) = \emptyset$$



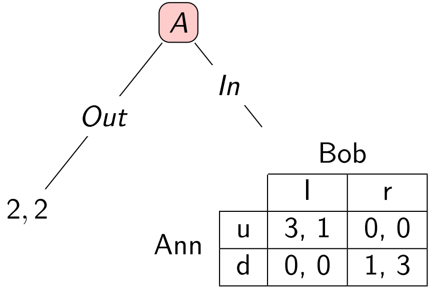
ω_1	$g_{1,\emptyset}(t_1)$	$g_{1,(In)}(t_1)$	ω_2	$g_{2,\emptyset}(t_1)$	$g_{2,(In)}(t_1)$
$(In\ d, t_1^1)$	$0,1,0$	$0,1,0$	(l, t_2^1)	$0,1,0,0,0$	$0,1,0,0,0$
$(In\ u, t_1^1)$	$0,1,0$	$0,1,0$	(r, t_2^2)	$0,0,1,0,0$	$1,0,0,0,0$
$(Out\ d, t_1^1)$	$0,1,0$	$0,1,0$	(l, t_2^3)	$0,0,0,0,1$	$0,0,0,0,1$
$(Out\ u, t_1^1)$	$0,1,0$	$0,1,0$			
$(In\ u, t_1^2)$	$0,0,1$	$0,0,1$			

$SB_2(R_1 \cap B_{1,\emptyset}([s_2 = R]))$



ω_1	$g_{1,\emptyset}(t_1)$	$g_{1,(In)}(t_1)$	ω_2	$g_{2,\emptyset}(t_1)$	$g_{2,(In)}(t_1)$
$(In\ d, t_1^1)$	$0,1,0$	$0,1,0$	(l, t_2^1)	$0,1,0,0,0$	$0,1,0,0,0$
$(In\ u, t_1^1)$	$0,1,0$	$0,1,0$	(r, t_2^2)	$0,0,1,0,0$	$1,0,0,0,0$
$(Out\ d, t_1^1)$	$0,1,0$	$0,1,0$	(l, t_2^3)	$0,0,0,0,1$	$0,0,0,0,1$
$(Out\ u, t_1^1)$	$0,1,0$	$0,1,0$			
$(In\ u, t_1^2)$	$0,0,1$	$0,0,1$			

$R_1 \cap R_2 \cap SB_2(R_1)$



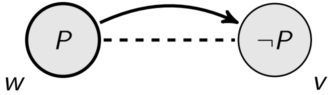
ω_1	$g_{1,\emptyset}(t_1)$	$g_{1,(In)}(t_1)$	ω_2	$g_{2,\emptyset}(t_1)$	$g_{2,(In)}(t_1)$
$(In\ d, t_1^1)$	$0, 1, 0$	$0, 1, 0$	(l, t_2^1)	$0, 1, 0, 0, 0$	$0, 1, 0, 0, 0$
$(In\ u, t_1^1)$	$0, 1, 0$	$0, 1, 0$	(r, t_2^2)	$0, 0, 1, 0, 0$	$1, 0, 0, 0, 0$
$(Out\ d, t_1^1)$	$0, 1, 0$	$0, 1, 0$	(l, t_2^3)	$0, 0, 0, 0, 1$	$0, 0, 0, 0, 1$
$(Out\ u, t_1^1)$	$0, 1, 0$	$0, 1, 0$			
$(In\ u, t_1^2)$	$0, 0, 1$	$0, 0, 1$			

$R_1 \cap R_2 \cap SB_2(R_1) \cap B_{1,\emptyset}(R_2 \cap SB_2(R_1))$

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.

Recall...



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.

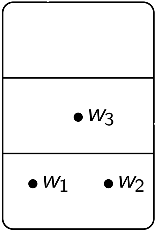
Language: $\varphi := p \mid \neg\varphi \mid \varphi \wedge \psi \mid K_i\varphi \mid B^\varphi\psi \mid [\preceq_i]\varphi$

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

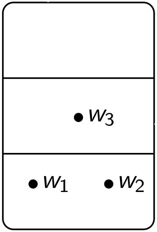
Recall...

- $w_1 \sim w_2 \sim w_3$



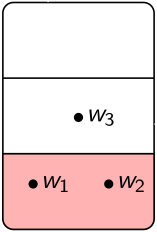
Recall...

- $w_1 \sim w_2 \sim w_3$
- $w_1 \preceq w_2$ 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 $w_3 \not\preceq w_2$)

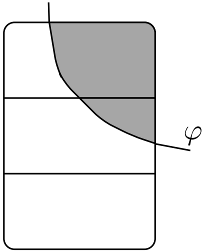


Recall...

- $w_1 \sim w_2 \sim w_3$
- $w_1 \preceq w_2$ 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 $w_3 \not\preceq w_2$)
- $\{w_1, w_2\} \subseteq Min_{\preceq}([w_i])$

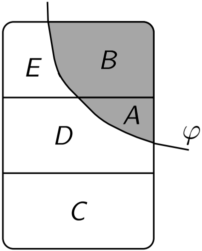


Recall...



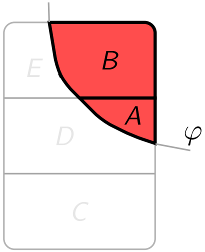
Incorporate the new information φ

Recall...



Incorporate the new information φ

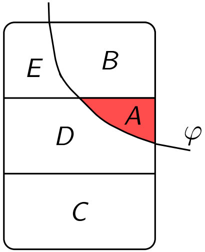
Recall...



Public Announcement: Information from an infallible source

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

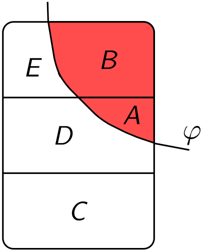
Recall...



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

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

Recall...



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

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

Radical Upgrade: Information from a strongly trusted source
$(\Uparrow\varphi)$: $A \prec_i B \prec_i C \prec_i D \prec_i E$

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

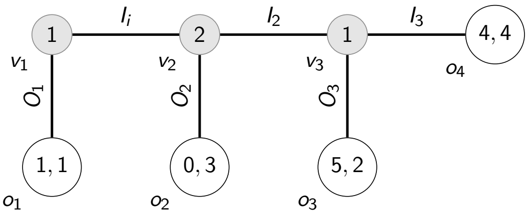
The structure of the game and past moves are ‘hard information’: *irrevocably known*

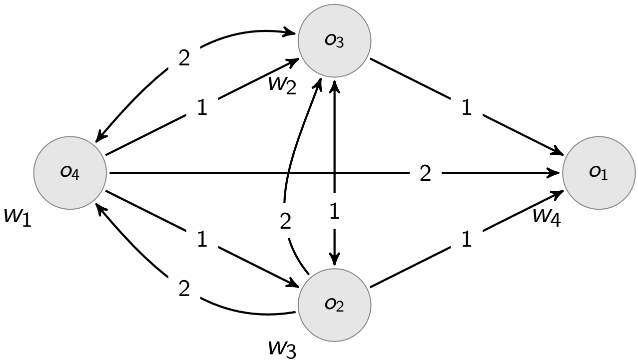
Players’ ‘knowledge’ of other players’ rationality and ‘knowledge’ of her own future moves at nodes not yet reached are not of the same degree of certainty.

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.


1
2
1
4, 4
I_i
I_2
I_3
v_1
v_2
v_3
O_1
O_2
O_3
1, 1
0, 3
5, 2
o_1
o_2
o_3
o_4


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

There are atomic propositions for each possible outcome o_i is true only at state o_i).

There are atomic propositions for each possible outcome $\mathsf{o_i}$ is true only at state o_i).

The non-terminal nodes $v \in V$ are then identified with the set of outcomes reachable from that node:

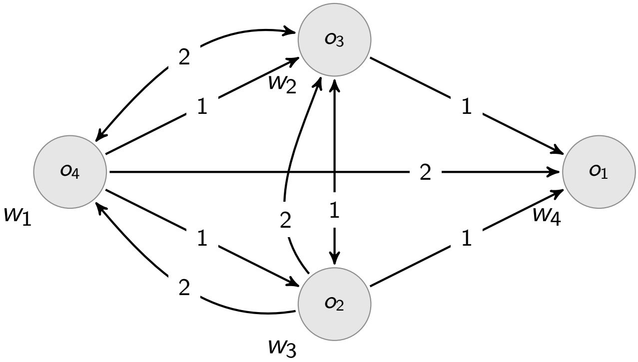
$$\mathsf{v} := \bigvee_{v \rightsquigarrow o} \mathsf{o}$$

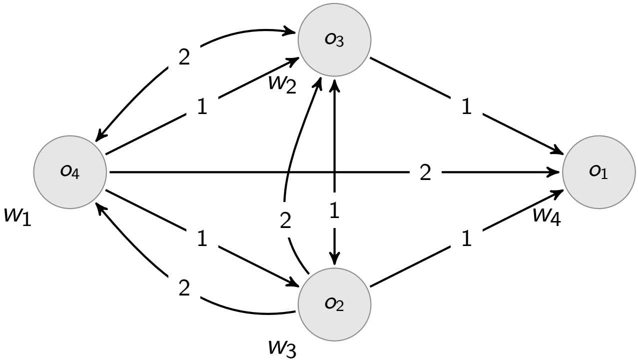
There are atomic propositions for each possible outcome $\mathsf{o_i}$ is true only at state o_i).

The non-terminal nodes $v \in V$ are then identified with the set of outcomes reachable from that node:

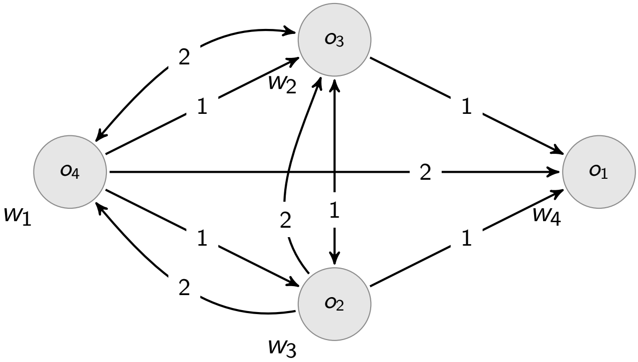
$$\mathsf{v} := \bigvee_{v \rightsquigarrow o} \mathsf{o}$$

Open future: none of the players have "hard information" that an outcome is ruled out


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



Player 1 is committed to the BI strategy is encoded in the conditional beliefs of the player: both $B_1^{v_1} o_1$ and $B_1^{v_3} o_3$ are true in the previous model.



Player 1 is committed to the BI strategy is encoded in the conditional beliefs of the player: both $B_1^{v_1} o_1$ and $B_1^{v_3} o_3$ are true in the previous model.

For player 2, $B_2^{v_2}(o_3 \vee o_4)$ is true in the above model, which implies player 2 plans on choosing action l_2 at node v_2.

The players' belief change as they learn (irrevocably) which of the nodes in the game are reached:

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

The players' belief change as they learn (irrevocably) which of the nodes in the game are reached:

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

The assumption that the players are "incurably optimistic" is represented as follows: no matter what true formula is publicly announced (i.e., no matter how the game proceeds), there is common belief that the players will make a rational choice (when it is their turn to move).

The players' belief change as they learn (irrevocably) which of the nodes in the game are reached:

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

The assumption that the players are "incurably optimistic" is represented as follows: no matter what true formula is publicly announced (i.e., no matter how the game proceeds), there is common belief that the players will make a rational choice (when it is their turn to move).

$\mathcal{M}, w \models [\ !\ ]\varphi$ provided for all formulas ψ if $\mathcal{M}, w \models \psi$ then $\mathcal{M}, w \models [!\psi]\varphi$.

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.

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

J. Halpern and R. Pass. *Iterated Regret Minimization: A New Solution Concept*. Games and Economic Behavior, 2012.

Traveler's Dilemma

Suppose that two travelers have identical luggage, for which they both paid the same price. Their luggage is damaged in an identical way by an airline.

Traveler's Dilemma

Suppose that two travelers have identical luggage, for which they both paid the same price. Their luggage is damaged in an identical way by an airline.

The airline offers to pay them for their damaged luggage. They may ask for any dollars amount between \$2 and \$100.

Traveler's Dilemma

Suppose that two travelers have identical luggage, for which they both paid the same price. Their luggage is damaged in an identical way by an airline.

The airline offers to pay them for their damaged luggage. They may ask for any dollars amount between \$2 and \$100.

There is one catch: If they ask for the same amount, then that is what they will both receive. However, if they ask for different amounts—say one asks for $\$m$ and the other for $\$m'$ with $m < m'$ then two ever asks for $\$m$ will get $\$(m + p)$ while the other traveler will get $\$(m - p)$, where p is a reward (assume $p > 1$).

Traveler's Dilemma, continued

- $(\$2, \$2)$ is the *only* Nash equilibrium

Traveler's Dilemma, continued

- $(\$2, \$2)$ is the *only* Nash equilibrium
- Becker, Carter, and Naeve [2005] asked members of the Game Theory Society (presumably, all experts in game theory) to submit a strategy for the game. Fifty-one of them did so. Of the 45 that submitted pure strategies, 31 submitted a strategy of 96 or higher, and 38 submitted a strategy of 90 or higher; only 3 submitted the "recommended" strategy of 2.

Traveler's Dilemma, continued

- ($\$2, \$2$) is the *only* Nash equilibrium
- Becker, Carter, and Naeve [2005] asked members of the Game Theory Society (presumably, all experts in game theory) to submit a strategy for the game. Fifty-one of them did so. Of the 45 that submitted pure strategies, 31 submitted a strategy of 96 or higher, and 38 submitted a strategy of 90 or higher; only 3 submitted the "recommended" strategy of 2.
- Another sequence of experiments by Capra et al. [1999] showed, among other things, that this result was quite sensitive to the choice of p. For low values of p, people tended 1to play high values, and keep playing them when the game was repeated. By way of contrast, for high values of p, people started much lower, and converged to playing 2 after a few rounds of repeated play

Traveler's Dilemma, continued

- ($\$2, \$2$) is the *only* Nash equilibrium
- Becker, Carter, and Naeve [2005] asked members of the Game Theory Society (presumably, all experts in game theory) to submit a strategy for the game. Fifty-one of them did so. Of the 45 that submitted pure strategies, 31 submitted a strategy of 96 or higher, and 38 submitted a strategy of 90 or higher; only 3 submitted the "recommended" strategy of 2.
- Another sequence of experiments by Capra et al. [1999] showed, among other things, that this result was quite sensitive to the choice of p. For low values of p, people tended 1to play high values, and keep playing them when the game was repeated. By way of contrast, for high values of p, people started much lower, and converged to playing 2 after a few rounds of repeated play

Let S be a (finite) set of states, A a (finite) set of actions and $u(s, a)$ the utility associated with the outcome of doing a in state s.

$u^*(s) = \max_{a \in A} u(s, a)$

$regret_u(a, s) = u^*(s) - u(a, s)$

$regret_u(a) = \max_{s \in S} regret_u(a, s)$

The **minimax-regret decision rule** orders acts by their regret; the "best" act is the one that minimizes regret.

Minmax Regret

	w_1	w_2	w_3	w_4
a_1	12	8	20	20
a_2	10	15	16	8
a_3	30	6	25	14
a_4	20	4	30	10

Minmax Regret

	w_1	w_2	w_3	w_4
a_1	12	8	20	20
a_2	10	15	16	8
a_3	30	6	25	14
a_4	20	4	30	10

Minmax Regret

	w_1	w_2	w_3	w_4
a_1	12	8	20	20
a_2	10	15	16	8
a_3	30	6	25	14
a_4	20	4	30	10

	w_1	w_2	w_3	w_4
a_1	18	7	10	0
a_2	20	0	14	12
a_3	0	9	5	6
a_4	10	11	0	10

Minmax Regret

	w_1	w_2	w_3	w_4
a_1	12	8	20	20
a_2	10	15	16	8
a_3	30	6	25	14
a_4	20	4	30	10

	w_1	w_2	w_3	w_4
a_1	18	7	10	0
a_2	20	0	14	12
a_3	0	9	5	6
a_4	10	11	0	10

Minmax Regret

	w_1	w_2	w_3	w_4
a_1	12	8	20	20
a_2	10	15	16	8
a_3	30	6	25	14
a_4	20	4	30	10

	w_1	w_2	w_3	w_4
a_1	18	7	10	0
a_2	20	0	14	12
a_3	0	9	5	6
a_4	10	11	0	10

Minmax Regret

	w_1	w_2	w_3	w_4
a_1	12	8	20	20
a_2	10	15	16	8
a_3	30	6	25	14
a_4	20	4	30	10
a_5	-10	10	10	39

	w_1	w_2	w_3	w_4
a_1	18	7	10	19
a_2	20	0	14	31
a_3	0	9	5	25
a_4	10	11	0	29
a_5	40	5	20	0

Minmax Regret

	w_1	w_2	w_3	w_4
a_1	12	8	20	20
a_2	10	15	16	8
a_3	30	6	25	14
a_4	20	4	30	10
a_5	-10	10	10	39

	w_1	w_2	w_3	w_4
a_1	18	7	10	19
a_2	20	0	14	31
a_3	0	9	5	25
a_4	10	11	0	29
a_5	40	5	20	0

Traveler's Dilemma

If the penalty/rewards p is such that $2 \leq p \leq 48$, then the acts that minimize regret are the ones in the interval $[100 - 2p, 100]$.

Traveler's Dilemma

If the penalty/rewards p is such that $2 \leq p \leq 48$, then the acts that minimize regret are the ones in the interval $[100-2p, 100]$.

Suppose player 1 asks for an amount $m \in [100-2p, 100]$

1. If player 2 asks for $m' < m$, then the payoff to 1 is $m' - p$. The best response is $(m'-1)+p$, so her regret is $(m'-1)+p-(m'-p) = 2p-1$.

Traveler's Dilemma

If the penalty/rewards p is such that $2 \leq p \leq 48$, then the acts that minimize regret are the ones in the interval $[100 - 2p, 100]$.

Suppose player 1 asks for an amount $m \in [100 - 2p, 100]$

1. If player 2 asks for $m' < m$, then the payoff to 1 is $m' - p$. The best response is $(m' - 1) + p$, so her regret is $(m' - 1) + p - (m' - p) = 2p - 1$.

2. If player 2 asks for $m' \geq m$, then player 1 gets $m' + p$, and the best possible payoff in the game is $99 + p$, so his regret is $99 + p - (m' + p) = 99 - m$. Note that $99 - m \leq 2p - 1$ for $m \in [100 - 2p, 100]$.

Traveler's Dilemma

- If player 1 chooses $m < 100 - 2p$, then his regret will be $99 - m > 2p - 1$ if player 2 plays 100.

Traveler's Dilemma

- If player 1 chooses $m < 100 - 2p$, then his regret will be $99 - m > 2p - 1$ if player 2 plays 100.
- If $49 \leq p \leq 96$, then the unique act that minimizes regret is asking for \$3.

Traveler's Dilemma

- If player 1 chooses $m < 100 - 2p$, then his regret will be $99 - m > 2p - 1$ if player 2 plays 100.

- If $49 \leq p \leq 96$, then the unique act that minimizes regret is asking for \$3.

- When $p = 97$, both \$3 and \$2 minimizes regret.

Traveler's Dilemma

- If player 1 chooses $m < 100 - 2p$, then his regret will be $99 - m > 2p - 1$ if player 2 plays 100.
- If $49 \leq p \leq 96$, then the unique act that minimizes regret is asking for \$3.
- When $p = 97$, both \$3 and \$2 minimizes regret.
- When $p \geq 98$, then only \$2 minimizes regret.

Suppose that $2 \leq p \leq 48$, then applying regret minimization once suggests using a strategy in the interval $[100 - 2p, 100]$.

Suppose that $2 \leq p \leq 48$, then applying regret minimization once suggests using a strategy in the interval $[100 - 2p, 100]$. If we assume that both players use a strategy in this interval, then the strategy that minimizes regret is that of asking form \$$(100 - 2p + 1)$.

Suppose that $2 \leq p \leq 48$, then applying regret minimization once suggests using a strategy in the interval $[100 - 2p, 100]$. If we assume that both players use a strategy in this interval, then the strategy that minimizes regret is that of asking form $\$(100 - 2p + 1)$.

For $p = 2$, this suggests the strategy \$97.

If rationality means minimize regret, then if the players are confident the other is "rational", then she should play \$96...

If rationality means minimize regret, then if the players are confident the other is "rational", then she should play \$96...but (contrary to the Brandenburger, Friedenberg and Keisler approach), assigning infinitesimal probability to deleted strategies where \$97 is given very high probability will not make \$97 a best response.

Traveler's Dilemma, continued

Rather than assuming common knowledge of rationality, assign successively lower probability to higher orders of rationality: with overwhelming probability no assumptions are made about the choice of the players, with probability ϵ the player are assumed to be rational, with ϵ^2 the players are rational and believe they are playing rational players, etc.

(related to *cognitive hierarchy theory*)

Pure Coordination

		Bob	
		L	R
Ann	U	1,1	0,0
	D	0,0	1,1

Hi-Low

		Bob	
		L	R
Ann	U	3,3	0,0
	D	0,0	1,1

Focal Points

"There are these two broad empirical facts about Hi-Lo games, people almost always choose A [Hi] and people with common knowledge of each other's rationality think it is obviously rational to choose A [Hi]."
[Bacharach, *Beyond Individual Choice*, 2006, pg. 42]

See also chapter 2 of:

C.F. Camerer. *Behavioral Game Theory*. Princeton UP, 2003.

N. Bardsley, J. Mehta, C. Starmer and R. Sugden. *The Nature of Salience Revisited: Cognitive Hierarchy Theory* versus *Team Reasoning*. Economic Journal.

‘primary salience’: players’ psychological propensities to play particular strategies by default, when there are no other reasons for choice.

‘primary salience’: players’ psychological propensities to play particular strategies by default, when there are no other reasons for choice.

level-n theory/ *cognitive hierarchy* theory

‘primary salience’: players’ psychological propensities to play particular strategies by default, when there are no other reasons for choice.

level-n theory/ *cognitive hierarchy* theory

‘team reasoning’: assumes that each player chooses the decision rule which, if used by all players, would be optimal for each of them.

Do the two approaches make different predictions?

What do the experiments support?

pickers: choose between labels without any incentive to choose one rather than the other

pickers: choose between labels without any incentive to choose one rather than the other

guessers: guess how pickers have behaved

pickers: choose between labels without any incentive to choose one rather than the other

guessers: guess how pickers have behaved

coordinators: try to coordinate their choices

pickers: choose between labels without any incentive to choose one rather than the other

guessers: guess how pickers have behaved

coordinators: try to coordinate their choices

labels vs. *options*

$$\{water, beer, sherry, whisky, wine\}$$

$$\{water, beer, sherry, whisky, wine\}$$

Task 1: pick an option

$$\{\textbf{water}, \textit{beer}, \textit{sherry}, \textit{whisky}, \textit{wine}\}$$

Task 1: pick an option

$$\{\textbf{water}, \textit{beer}, \textit{sherry}, \textit{whisky}, \textit{wine}\}$$

Task 1: pick an option

Task 2: guess what your opponent picked

$$\{\textbf{water}, \textit{beer}, \textit{sherry}, \textit{whisky}, \textit{wine}\}$$

Task 1: pick an option
Task 2: guess what your opponent picked
Task 3: try to coordinate with your (unknown) partner

$\{$**water**, *beer*, *sherry*, *whisky*, *wine*$\}$

Task 1: pick an option

Task 2: guess what your opponent picked

Task 3: try to coordinate with your (unknown) partner

	pick	guess	coordinate
water	20	15	38
beer	13	26	11
sherry	4	1	0
whisky	6	6	5
wine	10	4	2

"The main aim of the two experiments was to test cognitive hierarchy theory and the theory of team reasoning as rival explanations of behaviour in pure coordination and Hi-Lo games.

"The main aim of the two experiments was to test cognitive hierarchy theory and the theory of team reasoning as rival explanations of behaviour in pure coordination and Hi-Lo games. Formally, our conclusion must be that each theory failed at least one test."

"The main aim of the two experiments was to test cognitive hierarchy theory and the theory of team reasoning as rival explanations of behaviour in pure coordination and Hi-Lo games. Formally, our conclusion must be that each theory failed at least one test."

" The implication is that our subjects were able to use subtle features of the experimental environment to solve the problem of coordinating on a common mode of reasoning. This behaviour reveals an ability to solve coordination problems at a conceptual level above that of the theories of cognitive hierarchy and team reasoning that we have been examining. Each of those theories captures certain aspects of focal-point reasoning, but some essential feature of the human ability to solve coordination problems seems to have escaped formalisation."

"The basic intellectual premise, or working hypothesis, for rational players in this game seems to be the premise that some rule must be used if success is to exceed coincidence, and that the best rule to be found, whatever its rationalization, is consequently a rational rule." (Thomas Schelling)

A. Brojndahl, J. Halpern and R. Pass. *Language-Based Games*. manuscript, 2013.

Surprise Proposal

Ann and Bob have been dating for a while now, and Bob has decide tha that time is right to pop the big question. Though he is not one for fancy proposals, he does want it to be a surprise. In fact, if Ann expects the proposal, Bob would prefer to postpone it entirely until such time as it might be a surprise. Otherwise if Ann is not expecting it, Bob's preference is to take the opportunity.

Surprise Proposal

Ann and Bob have been dating for a while now, and Bob has decide tha that time is right to pop the big question. Though he is not one for fancy proposals, he does want it to be a surprise. In fact, if Ann expects the proposal, Bob would prefer to postpone it entirely until such time as it might be a surprise. Otherwise if Ann is not expecting it, Bob's preference is to take the opportunity.

	p	$\neg p$
$B_A p$	0	1
$\neg B_A p$	1	0

Language-Based Games

Let $G = \langle N, \{S_i\}_{i \in N} \rangle$ be a game form.

Let Φ be a set of primitive propositions, let $\mathcal{L}(\Phi)$ denote the propositional language generated by Φ.

$\Phi_G = \{play_i(s_i) \mid i \in N, s_i \in S_i\}$.

A $\mathcal{L}(\Phi)$-situation is a *maximal* satisfiable set of formulas. Let $\mathcal{S}(\mathcal{L}(\Phi))$ denote the set of $\mathcal{L}(\Phi)$-**situations** (maximally consistent sets of sentences)

Let $\mathcal{L}_B(\Phi_G)$be the language generated by the following grammar:

$$p \mid \neg\varphi \mid \varphi \wedge \psi \mid B_i\varphi$$

where $p \in \Phi_G$ and $i \in N$.

Let $\mathcal{L}_B(\Phi_G)$be the language generated by the following grammar:

$$p \mid \neg\varphi \mid \varphi \wedge \psi \mid B_i\varphi$$

where $p \in \Phi_G$ and $i \in N$.

A G-structure is a tuple $\mathcal{M} = \langle W, \vec{s}, Pr_1, \ldots, Pr_n \rangle$ such that

1. W is a nonempty topological space
2. each Pr_i assigns to each $w \in W$ a probability measure $Pr_i(w)$ on W
3. If $w' \in Pr_i[w]$, then $Pr_i(w) = Pr_i(w')$, where $Pr[w] = supp(Pr_i(w))$
4. $\vec{s} : W \to \Pi_{i \in N} S_i$ satisfies $Pr_i[w] \subseteq \{w' \mid s_i(w') = s_i(w)\}$ where $s_i(w)$ is player i's strategy in the profile $\vec{s}(w)$.

- $[\![play_i(\sigma_i)]\!]_{\mathcal{M}} = \{w \in W \mid s_i(w) = \sigma_i\}$
- $[\![\neg\varphi]\!]_{\mathcal{M}} = W - [\![\varphi]\!]_{\mathcal{M}}$
- $[\![\varphi \wedge \psi]\!]_{\mathcal{M}} = [\![\varphi]\!]_{\mathcal{M}} \cap [\![\psi]\!]_{\mathcal{M}}$
- $[\![B_i\varphi]\!]_{\mathcal{M}} = \{w \in W \mid Pr_i[w] \subseteq [\![\varphi]\!]_{\mathcal{M}}\}$

$u_i : \mathcal{S}(\mathcal{L}_B(\Phi_G)) \to \mathbb{R}$

Fix a language $\mathcal{L}$, a function $u : \mathcal{S}(\mathcal{L}) \to \mathbb{R}$ is **finitely specified** if there is a finite set of formula $F \subseteq \mathcal{L}$ and a function $f : F \to \mathbb{R}$ such that every situation $S \in \mathcal{S}(\mathcal{L})$ contains exactly one formula from F, and whenever $\varphi \in S \cap P$, $U(S) = F(\varphi)$.

Indignant altruism

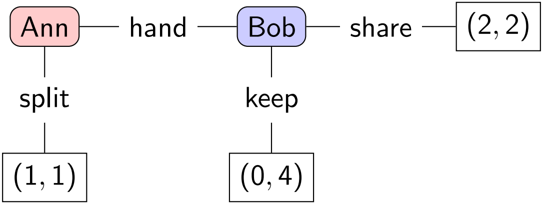
Ann and Bib sit down to play a classic game of Prisoner's dilemma, with one twist: neither wishes to live up to low expectations. Specifically, if Bob expects the worst of Ann (i.e., expects her to defect), then Ann, indignant at Bob's opinion of her, prefers to cooperate. Likewise for Bob. On the other hand, in the absence of such low expectations from their opponent, each will revert to the classical behavior:

	c	d
c	$(3,3)$	$(0,5)$
d	$(5,0)$	$(1,1)$

Let u_A and u_B be Ann and Bob's classic utility functions, define $u'_A : \mathcal{S}(\mathcal{L}_B(\Phi_G)) \to \mathbb{R}$ as follows

$$u'_A(S) = \begin{cases} -1 & \text{if } play_A(d) \in S \text{ and } B_B play_A(d) \in S \\ u_A(\rho_A(S), \rho_B(S)) & \text{otherwise} \end{cases}$$

Ann is handed \$2 and given a choice: either split the money with Bob, or hand him all of it. If she splits the money, the game is over and they each walk away with \$1. If she hands the money to Bob, it is doubled to \$4, and Bob is offered a choice: either share the money equally with Ann, or keep it all for himself. However, if Bob chooses to keep the money for himself, then he suffers from guilt to the extent that he feels he let Ann down.



Let m_A and m_B be the monetary payoffs.

$$u_B(S) = \begin{cases} -1 & \text{if } \mathit{play}(\mathit{hand}, \mathit{keep}) \in S \\ & \text{and } B_A \mathit{play}_B(\mathit{share}) \in S \\ m_B(\rho_A(S), \rho_B(S)) & \text{otherwise} \end{cases}$$

$$u_A(S) = m_A(\rho_A(S), \rho_B(S))$$

Consider the language $\mathcal{L}_B^5(\Phi_G)$ with a semantics given by $p_k = k/5$. A graded notion of Bob's guilt:

$$u'_B(S) = \begin{cases} 4 - k' & \text{if } play(hand, keep) \in S \\ & \text{and } B_A^1 play_B(share) \in S \\ m_B(\rho_A(S), \rho_B(S)) & \text{otherwise} \end{cases}$$

where $k' = \max\{k \mid B_A^k play_B(share) \in S\}$

$\mathcal{L}_B^k(\Phi_G)$ is generated by the following grammar

$$p \mid \neg\varphi \mid \varphi \wedge \psi \quad B_i^k \varphi$$

where $1 \leq k \leq l$ and $p \in \Phi_G$.

The semantics is a model $\mathcal{M}$ where we fix a sequence of real numbers $0 \leq p_1 < p_2 < \cdots < p_l \leq 1$

$$[\![B_i^k \varphi]\!]_{\mathcal{M}} = \{w \in W \mid Pr_i(w)([\![\varphi]\!]_{\mathcal{M}} \geq p_k\}$$

Rationality

Let $\langle G, \{u_i\}_{i \in N}\} \rangle$ be a $\mathcal{L}_B(\Phi_G)$-game and fix a G-structure $\mathcal{M} = \langle W, \vec{s}, \vec{Pr} \rangle$.

For each $w \in W$, there is a unique situation S such that $\mathcal{M}, w \models S$. Denote this situation by $\mathcal{S}(\mathcal{M}, w)$ (or $\mathcal{S}(w)$ when $\mathcal{M}$ is clear from context).

A formula $\varphi \in \mathcal{L}_B(\Phi_G)$ is *i*-**independent** if for each $\sigma_i \in S_i$, every occurrence of $play_i(\sigma_i)$ in φ falls within the scope of some B_j for $j \neq i$.

$\rho_{-i}(S) = \{\varphi \in S \mid \varphi \text{ is } i\text{-independent}\}$

Let $\mathcal{S}_{-i}$ denote the image of $\mathcal{S}$ under ρ_{-i}

$\mathcal{S}_{-i}$ are complete descriptions of states of affairs that are out of player i's control.

Proposition. For each $i \in N$, the map $\vec{\rho}_i : S \to S_i \times \mathcal{S}_{-i}$ defined by $\vec{\rho}_i = (\rho_i(S), \rho_{-i}(S))$ is a bijection.

Write $u_i(\sigma_i, S_{-i})$ to denote $u_i(S)$ where S is the unique situation corresponding to the pair (σ_i, S_{-i}), that is $\vec{\rho}_i(S) = (\sigma_i, S_{-i})$.

For each $w \in W$, there is a unique set S_{-i} such that $w \models S_{-i}$, which is denoted $S_{-i}(w)$.

Define $\hat{u}_i : S_i \times W \to \mathbb{R}$ as follows

$$\hat{u}_i(\sigma_i, w) = u_i(\sigma_i, S_{-i}(w))$$

For each $i \in N$, let $EU_i : S_i \times W \to \mathbb{R}$ be the expected utility of playing σ_i according to i's beliefs at w:

$$EU_i(\sigma_i, w) = \int_W \hat{u}_i(\sigma_i, w')dPr_i(w)$$

When W is finite

$$EU_i(\sigma_i, w) = \sum_{w' \in W} \hat{u}_i(\sigma_i, w') \times Pr_i(w)(w')$$

$BR_i : W \to \wp(S_i)$

$$BR_i(w) = \{\sigma_i \in S_i \mid \forall \sigma_i' \in S_i(EU_i(\sigma_i, w) \geq EU_i(\sigma', w))\}$$

Extend the language with atomic propositions RAT_i for each $i \in N$:
$\Phi_G^{rat} = \Phi_G \cup \{RAT_i \mid i \in N\}$

$[\![RAT_i]\!]_{\mathcal{M}} = \{w \mid s_i(w) \in BR_i(w)\}$

Let $RAT \equiv RAT_1 \wedge \cdots \wedge RAT_n$.

Note: $\mathcal{L}_B(\Phi_G^{rat})$ is not meant to *replace* $\mathcal{L}_B(\Phi_G)$.

Let $\mu = (\mu_1, \ldots, \mu_n) \in \Delta(S_1) \times \cdots \times \Delta(S_n)$ be a mixed-strategy profile.

Let $\mathcal{M}_\mu = \langle W_\mu, id_{W_\mu}, \vec{Pr}_\mu \rangle$ where

- $W_\mu = supp(\mu_1) \times \cdots \times supp(\mu_n) \subseteq S_1 \times \cdots \times S_n$
- Define a probability π on W_μ by $\pi(\sigma_1, \ldots, \sigma_n) = \Pi_{i \in N} \mu_i(\sigma_i)$
- For each $\sigma, \sigma' \in W_\mu$, let

$$Pr_{\mu,i}(\sigma)(\sigma') = \begin{cases} \pi(\sigma')/\mu_i(\sigma_i) & \text{if } \sigma_i = \sigma'_i \\ 0 & \text{otherwise} \end{cases}$$

μ is a **Nash equilibrium** iff $\mathcal{M}_\mu \models RAT$.

Proposition. There is no Nash equilibrium in the indignant altruism game.

Proposition. There is no Nash equilibrium in the indignant altruism game.

- Let $\mu = (\mu_A, \mu_B) \in \Delta(\{c, d\}) \times \Delta(\{c, d\})$ be a mixed strategy

Proposition. There is no Nash equilibrium in the indignant altruism game.

- Let $\mu = (\mu_A, \mu_B) \in \Delta(\{c, d\}) \times \Delta(\{c, d\})$ be a mixed strategy
- Claim: $\mathcal{M}_\mu \not\models RAT$

Proposition. There is no Nash equilibrium in the indignant altruism game.

- Let $\mu = (\mu_A, \mu_B) \in \Delta(\{c, d\}) \times \Delta(\{c, d\})$ be a mixed strategy
- Claim: $\mathcal{M}_\mu \not\models RAT$
- $\mu_A(c) > 0$.

Proposition. There is no Nash equilibrium in the indignant altruism game.

- Let $\mu = (\mu_A, \mu_B) \in \Delta(\{c, d\}) \times \Delta(\{c, d\})$ be a mixed strategy
- Claim: $\mathcal{M}_\mu \not\models RAT$
- $\mu_A(c) > 0$. Then, $\mathcal{M}, \mu \models \neg B_B play_A(d)$.

Proposition. There is no Nash equilibrium in the indignant altruism game.

- Let $\mu = (\mu_A, \mu_B) \in \Delta(\{c, d\}) \times \Delta(\{c, d\})$ be a mixed strategy
- Claim: $\mathcal{M}_\mu \not\models RAT$
- $\mu_A(c) > 0$. Then, $\mathcal{M}, \mu \models \neg B_B play_A(d)$. Hence, $\mathcal{M}_\mu \not\models RAT_A$, so μ is not a Nash equilibrium

Proposition. There is no Nash equilibrium in the indignant altruism game.

- Let $\mu = (\mu_A, \mu_B) \in \Delta(\{c, d\}) \times \Delta(\{c, d\})$ be a mixed strategy
- Claim: $\mathcal{M}_\mu \not\models RAT$
- $\mu_A(c) > 0$. Then, $\mathcal{M}, \mu \models \neg B_B play_A(d)$. Hence, $\mathcal{M}_\mu \not\models RAT_A$, so μ is not a Nash equilibrium
- $\mu_A(c) = 0$.

Proposition. There is no Nash equilibrium in the indignant altruism game.

- Let $\mu = (\mu_A, \mu_B) \in \Delta(\{c, d\}) \times \Delta(\{c, d\})$ be a mixed strategy
- Claim: $\mathcal{M}_\mu \not\models RAT$
- $\mu_A(c) > 0$. Then, $\mathcal{M}, \mu \models \neg B_B play_A(d)$. Hence, $\mathcal{M}_\mu \not\models RAT_A$, so μ is not a Nash equilibrium
- $\mu_A(c) = 0$. Then, $\mathcal{M}_\mu \models B_B play_A(d)$,

Proposition. There is no Nash equilibrium in the indignant altruism game.

- Let $\mu = (\mu_A, \mu_B) \in \Delta(\{c, d\}) \times \Delta(\{c, d\})$ be a mixed strategy
- Claim: $\mathcal{M}_\mu \not\models RAT$
- $\mu_A(c) > 0$. Then, $\mathcal{M}, \mu \models \neg B_B play_A(d)$. Hence, $\mathcal{M}_\mu \not\models RAT_A$, so μ is not a Nash equilibrium
- $\mu_A(c) = 0$. Then, $\mathcal{M}_\mu \models B_B play_A(d)$, and so Alice being sure of this is not rational at any state where she defects (chooses d) since she receives -1 in that case.

Proposition. There is no Nash equilibrium in the indignant altruism game.

- Let $\mu = (\mu_A, \mu_B) \in \Delta(\{c, d\}) \times \Delta(\{c, d\})$ be a mixed strategy
- Claim: $\mathcal{M}_\mu \not\models RAT$
- $\mu_A(c) > 0$. Then, $\mathcal{M}, \mu \models \neg B_B play_A(d)$. Hence, $\mathcal{M}_\mu \not\models RAT_A$, so μ is not a Nash equilibrium
- $\mu_A(c) = 0$. Then, $\mathcal{M}_\mu \models B_B play_A(d)$, and so Alice being sure of this is not rational at any state where she defects (chooses d) since she receives -1 in that case. Since $\mathcal{M}_\mu$ contains states where Ann defects, we have $\mathcal{M}_\mu \not\models RAT_A$.

Proposition. There is no Nash equilibrium in the indignant altruism game.

- Let $\mu = (\mu_A, \mu_B) \in \Delta(\{c, d\}) \times \Delta(\{c, d\})$ be a mixed strategy
- Claim: $\mathcal{M}_\mu \not\models RAT$
- $\mu_A(c) > 0$. Then, $\mathcal{M}, \mu \models \neg B_B play_A(d)$. Hence, $\mathcal{M}_\mu \not\models RAT_A$, so μ is not a Nash equilibrium
- $\mu_A(c) = 0$. Then, $\mathcal{M}_\mu \models B_B play_A(d)$, and so Alice being sure of this is not rational at any state where she defects (chooses d) since she receives -1 in that case. Since $\mathcal{M}_\mu$ contains states where Ann defects, we have $\mathcal{M}_\mu \not\models RAT_A$.

The problem: the utility function exhibits a "discontinuity".

Proposition In the trust game, the only Nash equilibrium in which Ann places positive weight on *hand* is the pure equilibrium $(\textit{hand}, \textit{share})$.

Rationalizability

Let $\mathcal{L}_{CB}(\Phi_G^{rat})$ be the language generated as follows:

$$p \mid \neg\varphi \mid \varphi \wedge \psi \mid B_i\varphi \mid CB\varphi$$

where $p \in \Phi_G^{rat}$ and $i \in N$.

Rationalizability

Let $\mathcal{L}_{CB}(\Phi_G^{rat})$ be the language generated as follows:

$$p \mid \neg\varphi \mid \varphi \wedge \psi \mid B_i\varphi \mid CB\varphi$$

where $p \in \Phi_G^{rat}$ and $i \in N$.

$$[\![CB\varphi]\!]_{\mathcal{M}} = \bigcap_{k=1}^{\infty} [\![EB^k\varphi]\!]_{\mathcal{M}}$$

where $EB\varphi = B_1\varphi \wedge \cdots B_n\varphi$, and $EB^k = EB(EB^{k-1}\varphi)$ and $EB^0\varphi = \varphi$.

Rationalizability

Let $\mathcal{L}_{CB}(\Phi_G^{rat})$ be the language generated as follows:

$$p \mid \neg\varphi \mid \varphi \wedge \psi \mid B_i\varphi \mid CB\varphi$$

where $p \in \Phi_G^{rat}$ and $i \in N$.

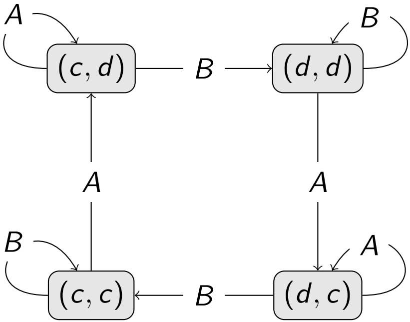
$$[\![CB\varphi]\!]_{\mathcal{M}} = \bigcap_{k=1}^{\infty} [\![EB^k\varphi]\!]_{\mathcal{M}}$$

where $EB\varphi = B_1\varphi \wedge \cdots B_n\varphi$, and $EB^k = EB(EB^{k-1}\varphi)$ and $EB^0\varphi = \varphi$.

A strategy $\sigma_i \in S_i$ is **rationalizable** in an $\mathcal{L}_B(\Phi_G^{rat})$-game if the formula $play_i(\sigma_i) \wedge CB(RAT)$ is satisfiable in some G-structure.

Proposition. Every strategy in the indignant altruism game is rationalizable.

Proposition. Every strategy in the indignant altruism game is rationalizable.



A deeply surprising proposal

Bob hopes to propose to Ann, but she wants it to be a surprise. HE knows that she would be upset if it were not a surprise, so he would prefer not to propose to if Ann so much as suspects it. Worse (for Bob), even if Ann does not suspect a proposal, if she suspects that Bob thinks she does, then she will also be upset, since in this case a proposal would indicate Bob's willingness to disappoint her. Of course, like the giant tortoise on whose back the world rests, this reasoning continues "all the way down"...

Let $P_i\varphi$ denote $\neg B_i \neg\varphi$

$$u_B(S) = \begin{cases} 1 & \text{if } play_B(p) \in S \text{ and} \\ & \forall k \in \mathbb{N}(P_A(P_B P_A)^k play_B(p) \notin S \\ 1 & \text{if } play_B(q) \in S \text{ and} \\ & \exists k \in \mathbb{N}(P_A(P_B P_A)^k play_B(p) \in S \\ 0 & \text{otherwise} \end{cases}$$

Proposition. The deeply surprising proposal game has no rationalizable strategies.

Proposition. The deeply surprising proposal game has no rationalizable strategies.

(CR) For all $S \in \mathcal{S}$, if $S \models \neg RAT$ then there is a finite subset $F \subseteq S$ such that $F \models \neg RAT$.

Theorem (CR) implies that raitonalizable strategies exist