Conditionals in Game Theory

Ilaria Canavotto, University of Maryland
Eric Pacuit, University of Maryland

Lecture 3

ESSLLI 2022

Yesterday: Bayes and counterfactual rationality

1. Bayesian rationality and counterfactual rationality
2. Stalnaker-Lewis semantics for counterfactuals
3. Bayesian rationality $\neq$ counterfactual rationality
4. Bayesian rationality $=$ counterfactual rationality given independence
5. Counterfactual rationality and ratifiability (started)

Shin H.S.. *A reconstruction of Jeffrey's notion of ratifiability in terms of counterfactual belief*. Theory and Decision 31, pp. 21-47, 1991.

Shin's theorems

Theorem 1. p_i is modestly ratifiable iff p_i is a correlated equilibrium.

Shin's theorems

Theorem 1. p_i is modestly ratifiable iff p_i is a correlated equilibrium.

- Assuming that the players have common prior probabilities (i.e. $p_i = p_{-i}$), correlated equilibria can be viewed as the result of Bayesian rationality.

R. Aumann. *Correlated equilibrium as an expression of Bayesian rationality*. Econometrica 55, pp. 1-18, 1987.

Shin's theorems

Theorem 1. p_i is modestly ratifiable iff p_i is a correlated equilibrium.

- Assuming that the players have common prior probabilities (i.e. $p_i = p_{-i}$), correlated equilibria can be viewed as the result of Bayesian rationality.

R. Aumann. *Correlated equilibrium as an expression of Bayesian rationality*. Econometrica 55, pp. 1-18, 1987.

Theorem 2. Player i is counterfactually rational at w iff i is Bayes rational at w.

Shin's theorems

Theorem 1. p_i is modestly ratifiable iff p_i is a correlated equilibrium.

- Assuming that the players have common prior probabilities (i.e. $p_i = p_{-i}$), correlated equilibria can be viewed as the result of Bayesian rationality.

R. Aumann. *Correlated equilibrium as an expression of Bayesian rationality*. Econometrica 55, pp. 1-18, 1987.

Theorem 2. Player i is counterfactually rational at w iff i is Bayes rational at w.

No need of independence??

Plan for today

1. Shin's notion of counterfactual rationality
2. Dropping independence, way 1: conditional choice rules & communication
3. Dropping independence, way 2: translucent agents

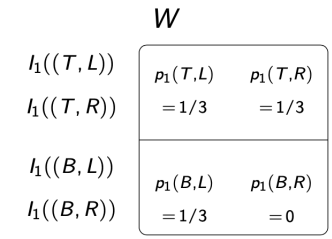
Shin's notion of counterfactual rationality

Intuitive idea

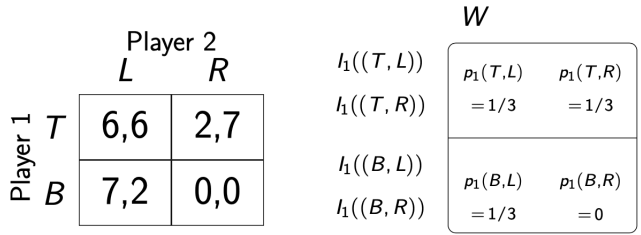
A player should never find herself at a possible world at which ... her payoff would be higher if she were to deviate from the strategy she has chosen. This is the principle which motivates our rationality criterion. (p. 29)

Another way to define a model of a game

		Player 2	
		L	R
Player 1	T	6,6	2,7
	B	7,2	0,0



Another way to define a model of a game



At world (T, L), player 1 believes that she is at a world where she plays T with probability 1 and player 2 plays L (R) with probability 0.5

Another way to define a model of a game

Player 1 \ Player 2	L	R
T	6,6	2,7
B	7,2	0,0

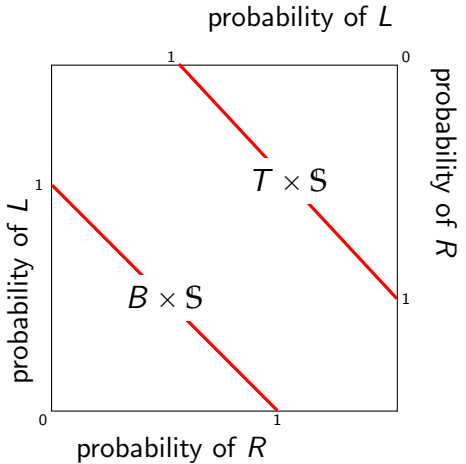
	W
$I_1((T,L))$	$p_1(T,L) = 1/3$
$I_1((T,R))$	$p_1(T,R) = 1/3$
$I_1((B,L))$	$p_1(B,L) = 1/3$
$I_1((B,R))$	$p_1(B,R) = 0$

At world (T, L), player 1 believes that she is at a world where she plays T with probability 1 and player 2 plays L (R) with probability 0.5

- Define $\beta^1 : W \to S_1 \times \mathbb{S}$, where $\mathbb{S}$ is the one dimensional unit simplex representing the set of all probability distributions over $\{L, R\}$

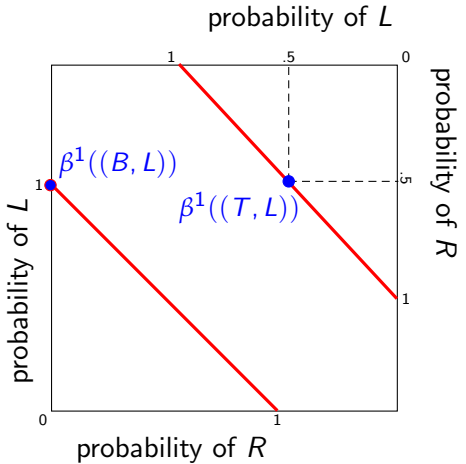
Belief space

The belief space of player 1 is $\{T, B\} \times \mathbb{S}$:



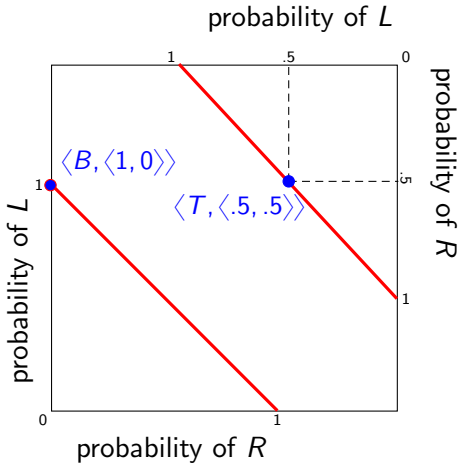
Belief space

The belief space of player 1 is $\{T, B\} \times \mathbb{S}$:



Belief space

The belief space of player 1 is $\{T, B\} \times \mathbb{S}$:



"Library stack metric"

We now define a distance measure, λ, to measure the distance (or closeness) between states in player 1's belief space.

"Library stack metric"

We now define a distance measure, λ, to measure the distance (or closeness) between states in player 1's belief space.

Let $\langle a, y \rangle$ and $\langle a', y' \rangle$ be two worlds in i's belief space:

- $a, a' \in \{T, B\}$
- $y = \langle y_1, y_2 \rangle \in \mathbb{R}^2$ with $y_1 + y_2 = 1$
- $y' = \langle y'_1, y'_2 \rangle \in \mathbb{R}^2$ with $y_1 + y_2 = 1$

"Library stack metric"

We now define a distance measure, λ, to measure the distance (or closeness) between states in player 1's belief space.

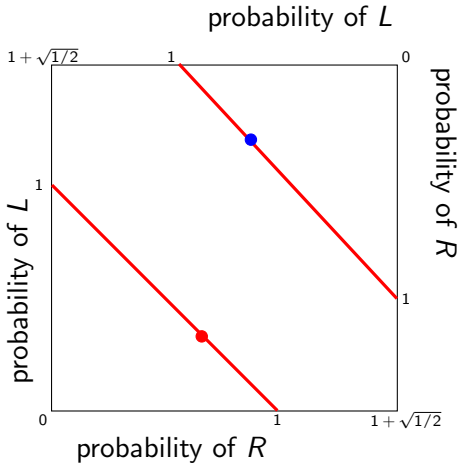
Let $\langle a, y \rangle$ and $\langle a', y' \rangle$ be two worlds in i's belief space:

- $a, a' \in \{T, B\}$
- $y = \langle y_1, y_2 \rangle \in \mathbb{R}^2$ with $y_1 + y_2 = 1$
- $y' = \langle y'_1, y'_2 \rangle \in \mathbb{R}^2$ with $y_1 + y_2 = 1$

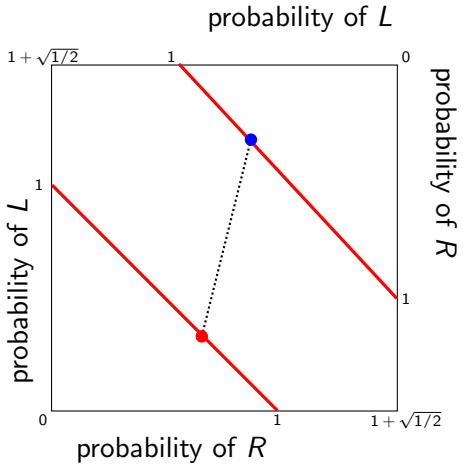
Then:

$$\lambda(\langle a, y \rangle, \langle a', y' \rangle) = \begin{cases} \sqrt{|y_1 - y'_1|^2 + |y_2 - y'_2|^2} & \text{if } a = a' \\ \sqrt{|y_1 - y'_1|^2 + |y_2 - y'_2|^2} + 1 & \text{if } a \neq a' \end{cases}$$

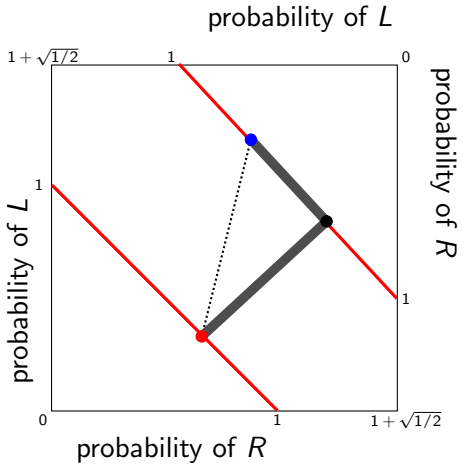
“Library stack metric”



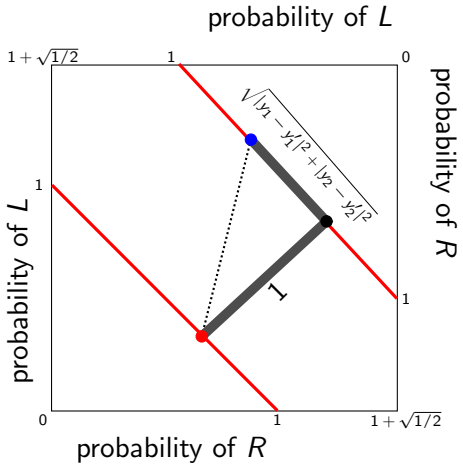
“Library stack metric”



“Library stack metric”

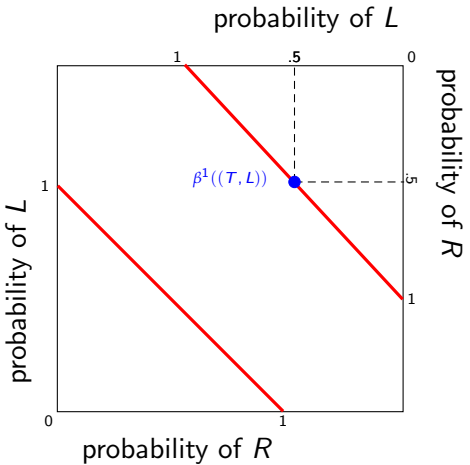


“Library stack metric”



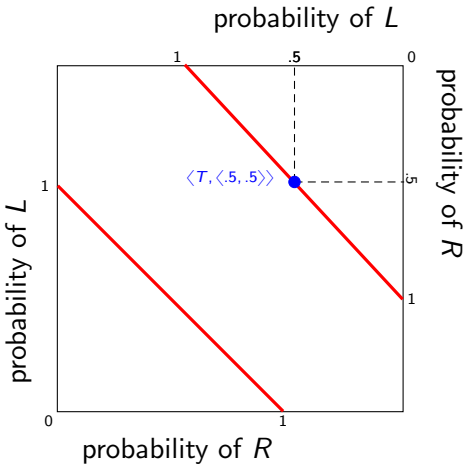
Counterfactual rationality

Player 1 is λ-rational at (T, L) if she believes that she is at a world at which, according to the metric λ, her payoff would not be higher if she were to play B.



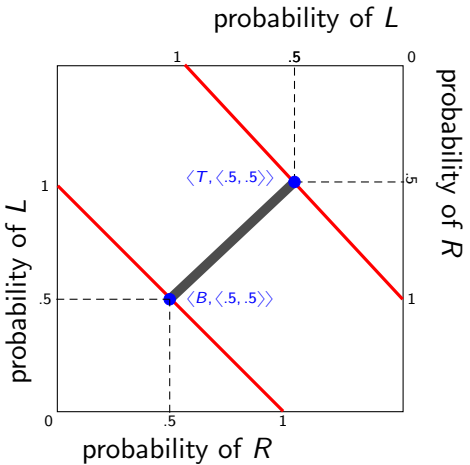
Counterfactual rationality

Player 1 is λ-rational at (T, L) if she believes that she is at a world at which, according to the metric λ, her payoff would not be higher if she were to play B.



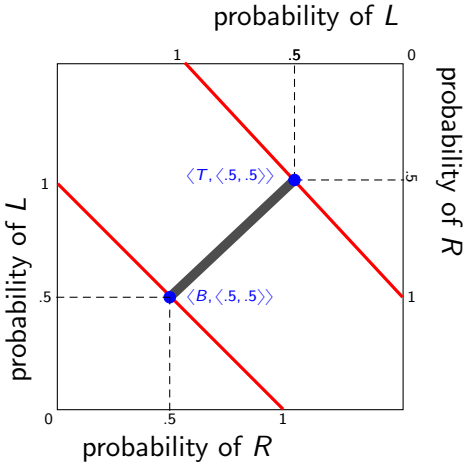
Counterfactual rationality

Player 1 is λ-rational at (T, L) if she believes that she is at a world at which, according to the metric λ, her payoff would not be higher if she were to play B.



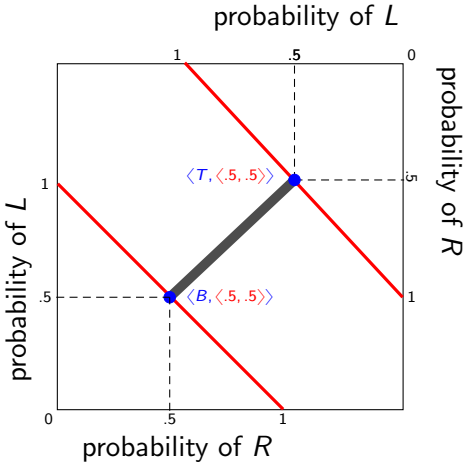
Counterfactual rationality

Player 1 is λ-rational at (T, L) if she believes that she is at a world at which, according to the metric λ, her payoff would not be higher if she were to play B. (Expected payoff at $\langle T, \langle .5, .5\rangle\rangle = 4$; expected payoff at $\langle B, \langle .5, .5\rangle\rangle = 3.5$.)



Counterfactual rationality

Player 1 is λ-rational at (T, L) if she believes that she is at a world at which, according to the metric λ, her payoff would not be higher if she were to play B. (Expected payoff at $\langle T, \langle .5, .5\rangle\rangle = 4$; expected payoff at $\langle B, \langle .5, .5\rangle\rangle = 3.5$.)



Take home messages

- The notion of Bayesian rationality is based on the (hidden) assumption that the players' choices are independent of one another—and that there is common belief that this is the case.

Take home messages

- The notion of Bayesian rationality is based on the (hidden) assumption that the players' choices are independent of one another—and that there is common belief that this is the case.
- If we allow (beliefs in) dependencies between the players' choices, then we can distinguish two notions of rationality: Bayesian rationality and counterfactual rationality.

  Keep in mind: some relations of relative closeness (like those defined by Shin) build in the assumption of independence of choices.

Take home messages

- The notion of Bayesian rationality is based on the (hidden) assumption that the players' choices are independent of one another—and that there is common belief that this is the case.
- If we allow (beliefs in) dependencies between the players' choices, then we can distinguish two notions of rationality: Bayesian rationality and counterfactual rationality.

  Keep in mind: some relations of relative closeness (like those defined by Shin) build in the assumption of independence of choices.
- Besides (common belief in) independence of choices, the notion of Bayesian rationality encodes the idea that the players' choices are rational when they are ratifiable (i.e., stable or non self-defeating).

A key question

Is it plausible to assume that the principle of independence of choices characterizes rational beliefs?

A key question

Is it plausible to assume that the principle of independence of choices characterizes rational beliefs?

Answer 1: YES

especially when we consider games in normal form, where the players are typically assumed to move simultaneously and be ignorant of each other's strategies.

A key question

> *[A] causal independence assumption is part of the idealization built into the normal form.*

W.L. Harper. *Causal decision theory and game theory: A classic argument for equilibrium solutions, a defense of weak equilibria, and a new problem for the normal form representation*. Causation in Decision, Belief Change and Statistics II, 1988.

> *[I]n a strategic form game, the assumption is that the strategies are chosen independently, which means that the choices made by one player cannot influence the beliefs or the actions of the other players.*

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

A key question

Is it plausible to assume that the principle of independence of choices characterizes rational beliefs?

Answer 1: YES
especially when we consider games in normal form, where the players are typically assumed to move simultaneously and be ignorant of each other's strategies.

Answer 2: NO
if we do not exclude that the players can communicate or be "translucent" to one another or when we consider games where the players move sequentially.

Conditional choice rules and communication

S.J. Brams. *Newcomb's problem and the Prisoner's Dilemma*. Journal of Conflict Resolution 19(4), pp. 596-612, 1975.

Newcomb's paradox

Ilaria		*pred_B*	*pred_AB*
	B	$1M$	0
	AB	$1M+1T$	$1T$

R. Nozick. *Newcomb's Problem and Two Principles of Choice*. 1969.

Newcomb's paradox

Ilaria		*pred_B*	*pred_AB*
	B	$1M$	0
	AB	$1M+1T$	$1T$

Principle of dominance: take both boxes.

R. Nozick. *Newcomb's Problem and Two Principles of Choice*. 1969.

Newcomb's paradox

Ilaria		*pred_B*	*pred_AB*
	B	$1M$	0
	AB	$1M+1T$	$1T$

Expected utility maximization: take box B.

$p_i(pred\_B \mid B)1M + p_i(pred\_AB \mid B)0 > p_i(pred\_B \mid AB)(1M + 1T) + p_i(pred\_AB \mid AB)1T$

R. Nozick. *Newcomb's Problem and Two Principles of Choice*. 1969.

One solution: make it a decision problem

Ilaria		*pred* ✓	*pred* ✗
	B	$1M$	0
	AB	$1T$	$1M+1T$

- If $p(\textit{pred}\ \checkmark) > 0.5005$, I should choose B
- If $p(\textit{pred}\ \checkmark) < 0.5005$, I should choose AB
- If $p(\textit{pred}\ \checkmark) = 0.5005$, I can choose either B or AB

Game-theoretic or decision-theoretic representation?

If you believe that the Being has no control over which state of nature obtains... then the being is not properly a player in a two-person game... hence, the appropriate representation of Newcomb's problem is decision-theoretic...

Game-theoretic or decision-theoretic representation?

If you believe that the Being has no control over which state of nature obtains... then the being is not properly a player in a two-person game... hence, the appropriate representation of Newcomb's problem is decision-theoretic...

On the other hand, if you believe that the Being has some control over which state of nature obtains... then he is not an entirely passive state of nature, at least with respect to being correct; hence the game-theoretic representation... is the appropriate one. (pp. 600-1)

Game-theoretic or decision-theoretic representation?

> *If you believe that the Being has no control over which state of nature obtains... then the being is not properly a player in a two-person game... hence, the appropriate representation of Newcomb's problem is decision-theoretic...*
>
> *On the other hand, if you believe that the Being has some control over which state of nature obtains... then he is not an entirely passive state of nature, at least with respect to being correct; hence the game-theoretic representation... is the appropriate one. (pp. 600-1)*

(According to Brams, the decision-theoretic representation is correct)

Game-theoretic or decision-theoretic representation?

> *If you believe that the Being has no control over which state of nature obtains... then the being is not properly a player in a two-person game... hence, the appropriate representation of Newcomb's problem is decision-theoretic...*
>
> *On the other hand, if you believe that the Being has some control over which state of nature obtains... then he is not an entirely passive state of nature, at least with respect to being correct; hence the game-theoretic representation... is the appropriate one. (pp. 600-1)*

(According to Brams, the decision-theoretic representation is correct)

> *... it is still intriguing to ask what consequences the predictive ability assumed on the part of the Being would have if both actors in the Newcomb's problem could make genuine choices as players in a game.*

Newcomb's paradox: game-theoretic representation

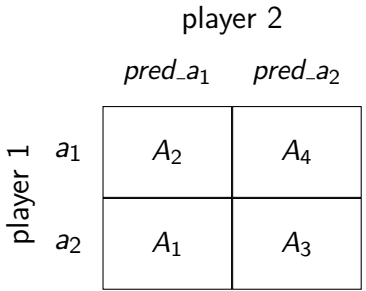
Ilaria		*pred_B*	*pred_AB*
	B	$1M$	0
	AB	$1M+1T$	$1T$

Newcomb's paradox: game-theoretic representation

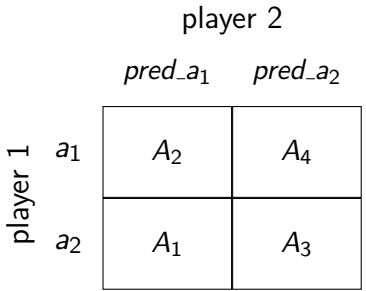
Ilaria	*pred_B*	*pred_AB*
B	$1M$	0
AB	$1M+1T$	$1T$

First, let us generalize the game.

Newcomb's paradox: game-theoretic representation



Newcomb's paradox: game-theoretic representation



Now, let us make the game symmetric so that player 1 can also make predictions.

Newcomb's paradox: game-theoretic representation

player 1 \ player 2	$pred\_a_1$	$pred\_a_2$
a_1	A_2	A_4
a_2	A_1	A_3

player 1 \ player 2	b_1	b_2
$pred\_b_1$	B_2	B_1
$pred\_b_2$	B_4	B_3

Newcomb's paradox: game-theoretic representation

player 1 \ player 2	$pred\_a_1$	$pred\_a_2$
a_1	A_2	A_4
a_2	A_1	A_3

player 1 \ player 2	b_1	b_2
$pred\_b_1$	B_2	B_1
$pred\_b_2$	B_4	B_3

Let us merge the two games...

Newcomb's paradox: game-theoretic representation

player 1 \ player 2	b_1	b_2
a_1	(A_2, B_2)	(A_4, B_1)
a_2	(A_1, B_4)	(A_3, B_3)

We obtained the classic Prisoner's Dilemma!

Newcomb's paradox: game-theoretic representation

player 1	player 2: b_1	player 2: b_2
a_1	(A_2, B_2)	(A_4, B_1)
a_2	(A_1, B_4)	(A_3, B_3)

But why is this interesting?

Newcomb's paradox: game-theoretic representation

[T]he condition in the symmetric version of Newcomb's problem that each player knows that the other player can predict—with a high degree of accuracy—which strategy he will choose does have a surprising consequence for the play of Prisoners' Dilemma: it provides an incentive for each player not to choose his second dominant strategy [i.e. defect]. (pp. 603-4)

Choice rules in the Prisoner's Dilemma

player 1 \ player 2	b_1	b_2
a_1	(A_2, B_2)	(A_4, B_1)
a_2	(A_1, B_4)	(A_3, B_3)

Let: p_1 be player 1's degree of belief that player 2 correctly predicts her choice
p_2 be player 2's degree of belief that player 1 correctly predicts her choice

Result: if p_1 and p_2 are sufficiently high, then there is a *choice rule* (i.e. a conditional strategy based on one's prediction) that either player can adopt that will induce the other player to choose his cooperative strategy.

Choice rules in the Prisoner's Dilemma

		player 2	
		b_1	b_2
player 1	a_1	(A_2, B_2)	(A_4, B_1)
	a_2	(A_1, B_4)	(A_3, B_3)

Let: p_1 be player 1's degree of belief that player 2 correctly predicts her choice
p_2 be player 2's degree of belief that player 1 correctly predicts her choice

Choice rule of conditional cooperation:

Player i cooperates if she predicts that player $-i$ cooperates;
Player i defects otherwise.

Choice rules in the Prisoner's Dilemma

player 1 \ player 2	b_1	b_2
a_1	(A_2, B_2)	(A_4, B_1)
a_2	(A_1, B_4)	(A_3, B_3)

Suppose that 2 assumes the choice rule of conditional cooperation

IF 1 chooses a_1 (cooperate)

2 predicts a_1 with probability p_2

So, given his choice rule, 2 plays

b_1 with probability p_2

b_2 with probability $1 - p_2$

So: $EU(a_1) = p_2 \cdot A_2 + (1 - p_2) \cdot A_4$

IF 1 chooses a_2 (defects)

2 predicts a_1 with probability $1 - p_2$

So, given his choice rule, 2 plays

b_1 with probability $1 - p_2$

b_2 with probability p_2

So: $EU(a_2) = (1 - p_2) \cdot A_1 + p_2 \cdot A_3$

Choice rules in the Prisoner's Dilemma

		player 2 b_1	player 2 b_2
player 1	a_1	(A_2, B_2)	(A_4, B_1)
player 1	a_2	(A_1, B_4)	(A_3, B_3)

Suppose that 2 assumes the choice rule of conditional cooperation

IF 1 chooses a_1 (cooperate)

2 predicts a_1 with probability p_2

So, given his choice rule, 2 plays

b_1 with probability p_2

b_2 with probability $1 - p_2$

IF 1 chooses a_2 (defects)

2 predicts a_1 with probability $1 - p_2$

So, given his choice rule, 2 plays

b_1 with probability $1 - p_2$

b_2 with probability p_2

$$p_2 \cdot A_2 + (1 - p_2) \cdot A_4 \; > \; (1 - p_2) \cdot A_1 + p_2 \cdot A_3 \qquad \boxed{?}$$

Choice rules in the Prisoner's Dilemma

player 1 \ player 2	b_1	b_2
a_1	(A_2, B_2)	(A_4, B_1)
a_2	(A_1, B_4)	(A_3, B_3)

Suppose that 2 assumes the choice rule of conditional cooperation

IF 1 chooses a_1 (cooperate)

2 predicts a_1 with probability p_2

So, given his choice rule, 2 plays

b_1 with probability p_2

b_2 with probability $1 - p_2$

IF 1 chooses a_2 (defects)

2 predicts a_1 with probability $1 - p_2$

So, given his choice rule, 2 plays

b_1 with probability $1 - p_2$

b_2 with probability p_2

Suppose that $A_1 = 4$, $A_2 = 3$, $A_3 = 2$, and $A_4 = 1$

Choice rules in the Prisoner's Dilemma

		player 2	
		b_1	b_2
player 1	a_1	(A_2, B_2)	(A_4, B_1)
	a_2	(A_1, B_4)	(A_3, B_3)

Suppose that 2 assumes the choice rule of conditional cooperation

IF 1 chooses a_1 (cooperate)

2 predicts a_1 with probability p_2

So, given his choice rule, 2 plays

b_1 with probability p_2

b_2 with probability $1 - p_2$

IF 1 chooses a_2 (defects)

2 predicts a_1 with probability $1 - p_2$

So, given his choice rule, 2 plays

b_1 with probability $1 - p_2$

b_2 with probability p_2

Then $EU(a_1) > EU(a_2)$ iff $p_2 > 3/4$

Choice rules in the Prisoner's Dilemma

player 1 \ player 2	b_1	b_2
a_1	(A_2, B_2)	(A_4, B_1)
a_2	(A_1, B_4)	(A_3, B_3)

Suppose that 2 assumes the choice rule of conditional cooperation

IF 1 chooses a_1 (cooperate)

2 predicts a_1 with probability p_2

So, given his choice rule, 2 plays

b_1 with probability p_2

b_2 with probability $1 - p_2$

IF 1 chooses a_2 (defects)

2 predicts a_1 with probability $1 - p_2$

So, given his choice rule, 2 plays

b_1 with probability $1 - p_2$

b_2 with probability p_2

If $p_2 > 3/4$, it is irrational for 1 not to cooperate

Choice rules in the Prisoner's Dilemma

player 1 \ player 2	b_1	b_2
a_1	(A_2, B_2)	(A_4, B_1)
a_2	(A_1, B_4)	(A_3, B_3)

Suppose that 1 assumes the choice rule of conditional cooperation

IF 2 chooses b_1 (cooperate)

1 predicts b_1 with probability p_1

So, given his choice rule, 1 plays

a_1 with probability p_1

a_2 with probability $1 - p_1$

IF 2 chooses b_2 (defects)

2 predicts b_1 with probability $1 - p_1$

So, given his choice rule, 1 plays

a_1 with probability $1 - p_1$

a_2 with probability p_1

If $p_1 > n$, it is irrational for 1 not to cooperate

Conclusion

- If the players
  1. can predict their opponent's choice with a sufficiently high probability **AND**
  2. they adopt the choice rule of conditional cooperation **AND**
  3. there is common knowledge of 1 and 2,

  then they will be better off by playing their pareto-dominant strategies

So, *if the players can communicate, they are better off by influencing each other*

Translucent agents

V. Capraro and J. Halpern. *Translucent Players: Explaining Cooperative Behavior in Social Dilemmas*. Proceedings of the 15th conference on Theoretical Aspects of Rationality and Knowledge, 2015.

Prisoner's Dilemma

Ann \ Bob	c	d
c	3,3	0,4
d	4,0	1,1

Social Dilemmas

1. There is a unique Nash equilibrium s^N, which is a pure strategy profile;
2. There is a unique welfare-maximizing profile s^W, again a pure strategy profile, such that each player's utility if s^W is played is higher than his utility if s^N is played.

Traveler's Dilemma

1. You and your friend write down an integer between 2 and 100 (without discussing).
2. If both of you write down the same number, then both will receive that amount in dollars from the airline in compensation.
3. If the numbers are different, then the airline assumes that the smaller number is the actual price of the luggage.
4. The person that wrote the smaller number will receive that amount plus \$2 (as a reward), and the person that wrote the larger number will receive the smaller number minus \$2 (as a punishment).

Suppose that you are randomly paired with another person from class. What number would you write down?

Expected Utility, Best Response

Suppose that $G = \langle N, (S_i)_{i \in N}, (u_i)_{i \in N} \rangle$ is a game in strategic form.
For $a \in S_i$ and $p \in \Delta(S_{-i})$, a is a best response to p when: for all $a' \in S_i$,

$$\sum_{s_{-i} \in S_{-i}} p_i(s_{-i}) u_i(a, s_{-i}) \geqslant \sum_{s_{-i} \in S_{-i}} p_i(s_{-i}) u_i(a', s_{-i})$$

Expected Utility, Best Response

Suppose that $G = \langle N, (S_i)_{i \in N}, (u_i)_{i \in N} \rangle$ is a game in strategic form.
For $a \in S_i$ and $p \in \Delta(S_{-i})$, a is a best response to p when: for all $a' \in S_i$,

$$\sum_{s_{-i} \in S_{-i}} p_i(s_{-i}) u_i(a, s_{-i}) \geqslant \sum_{s_{-i} \in S_{-i}} p_i(s_{-i}) u_i(a', s_{-i})$$

Implicitly assumes that i's beliefs about what other agents are doing do not change if i switches from s_i, the strategy he was *intending* to play, to a different strategy.

$p_i^{s_i, s_i'}$: i's beliefs if he intends to play s_i but instead deviates to s_i'

$p_i^{s_i, s_i'}$: i's beliefs if he intends to play s_i but instead deviates to s_i'

Strategy $a \in S_i$ is a best response for i with respect to the beliefs $\{p_i^{a,a'} : a' \in S_i\}$ if for all strategies $a' \in S_i$

$$\sum_{s_{-i} \in S_{-i}} p_i^{a,a}(s_{-i}) u_i(a, s_{-i}) \geqslant \sum_{s_{-i} \in S_{-i}} p_i^{a,a'}(s_{-i}) u_i(a', s_{-i})$$

A player is **translucently rational—** if he best responds to his beliefs.

Translucency will be used to determine $p_i^{a,a'}$:

Suppose that G is a two-player game, player 1 believes that, if he were to switch from a to a', this would be detected by player 2 with probability α, and if player 2 did detect the switch, then player 2 would switch to b.

Translucency will be used to determine $p_i^{a,a'}$:

Suppose that G is a two-player game, player 1 believes that, if he were to switch from a to a', this would be detected by player 2 with probability α, and if player 2 did detect the switch, then player 2 would switch to b.

Then $p_i^{a,a'}$ is $(1-\alpha)p_i^{a,a} + \alpha p'$, where p' assigns probability 1 to b: that is, player 1 believes that with probability $1-a$, player 2 continues to do what he would have done all along (as described by $p_i^{a,a}$) and with probability α, player 2 switches to b.

Explaining Cooperation

Say that an player i has type (α, β, C) if i intends to cooperate and believes that

1. if he deviates from that, then each other agent will independently realize this with probability α;
2. if a player j realizes that i is not going to cooperate, then j will defect; and
3. all other players will either cooperate or defect, and they will cooperate with probability β.

	C	D
C	$b-c,\ b-c$	$-c,\ b$
D	$b,\ -c$	$0,\ 0$

Proposition In the Prisoner's Dilemma, it is translucently rational for a player of type (α, β, C) to cooperate if and only if $\alpha\beta b \geqslant c$.

J. Halpern and R. Pass. *Game theory with translucent players*. International Journal of Game Theory, 47:3, pp. 949 - 976, 2018.

Given a strategic-form game $G = \langle N, (S_i)_{i \in N}, (u_i)_{i \in N} \rangle$, a model of G is a triple

$$\langle W, f, (P_i)_{i \in N}, \sigma \rangle$$

where W is a non-empty set of states, $\sigma : W \to \Pi_{i \in N} S_i$, and:

$$\text{For each } i \in N,\ P_i : W \to \Delta(W).$$

- For all $w \in W$, $P_i(w)([\sigma_i(w)]) = 1$.
- For all $w \in W$, $P_i(w)(\{v \mid P_i(v) = P_i(w)\}) = 1$.

Given a strategic-form game $G = \langle N, (S_i)_{i \in N}, (u_i)_{i \in N} \rangle$, a model of G is a triple

$$\langle W, f, (P_i)_{i \in N}, \sigma \rangle$$

where W is a non-empty set of states, $\sigma : W \to \Pi_{i \in N} S_i$, and:

$$\text{For each } i \in N,\ P_i : W \to \Delta(W).$$

- For all $w \in W$, $P_i(w)([\sigma_i(w)]) = 1$.
- For all $w \in W$, $P_i(w)(\{v \mid P_i(v) = P_i(w)\}) = 1$.
- f associates with each state w, player i and strategy a a state $f(w, i, a)$ where player i plays a. If $f(w, i, a) = w'$, then
  - $\sigma_i(w') = a$.
  - If $\sigma_i(w) = a$, then $w' = w$.

$$P^c_{i,a}(w)(w') = \sum_{\{w'' \in W \mid f(w'',i,a)=w'\}} P_i(w)(w'')$$

$$P^c_{i,a}(w)(w') = \sum_{\{w'' \in W \mid f(w'',i,a)=w'\}} P_i(w)(w'')$$

- $P^c_{i,a}$ is i's counterfactual beliefs at state w: what i believes would happen if she switched to s at w
- $P^c_{i,a}(w)([a]) = 1$
- It may *not* be the case that $P^c_{i,a}(w)([P^c_{i,a}(w), i]) = 1$: players do not in general know their counterfactual beliefs in state w
- A model is a *strongly appropriate counterfactual structure* if at every state w, every player i knows his counterfactual beliefs.

$$P^c_{i,a}(w)(w') = \sum_{\{w'' \in W \mid f(w'',i,a)=w'\}} P_i(w)(w'')$$

Claim. For all $w \in W$, $P^c_{i,\sigma_i(w)}(w)(w) = P_i(w)(w)$.

Proof. By the definition of $P^c_{i,a}$, we have that:

$$P^c_{i,\sigma_i(w)}(w) = \sum_{\{w'' \in W \mid f(w'',i,\sigma_i(w))=w\}} P_i(w)(w'')$$

Recall two properties of f and P_i:

1. for all states x, $f(x, i, \sigma_i(x)) = x$.
2. for all states x, if $\sigma_i(w') \neq \sigma_i(w)$, then $P_i(w)(w') = 0$.

$$P^c_{i,a}(w)(w') = \sum_{\{w'' \in W \mid f(w'',i,a)=w'\}} P_i(w)(w'')$$

Claim. For all $w \in W$, $P^c_{i,\sigma_i(w)}(w)(w) = P_i(w)(w)$.

Proof, continued. Recall that the definition of $P^c_{i,a}$, we have that:

$$P^c_{i,\sigma_i(w)}(w) = \sum_{\{w'' \in W \mid f(w'',i,\sigma_i(w))=w\}} P_i(w)(w'')$$

Suppose that $w'' \in W$ such that $f(w'', i, \sigma_i(w)) = w$. If $\sigma_i(w'') = \sigma_i(w)$, then $f(w'', i, \sigma_i(w)) = f(w'', i, \sigma_i(w'')) = w''$ (the last equality is by part 1). Hence, $w = w''$. Other the other hand, if $\sigma_i(w'') \neq \sigma_i(w)$, then, by part 2, we have that $P_i(w)(w'') = 0$. Putting everything together, we have that:

$$P^c_{i,\sigma_i(w)}(w) = P_i(w)(w)$$

$$B_i(E) = \{w \mid P_i(w)(E) = 1\}$$

$$B_i^*(E) = \{w \mid \text{for all } s' \in S_i, P_{i,s'}^c(w)(E) = 1\}$$

Characterize solution concepts in terms of the players beliefs, common beliefs, counterfactual beliefs and common counterfactual beliefs.