

CAP 5993/CAP 4993

Game Theory

Instructor: Sam Ganzfried
sganzfri@cis.fiu.edu

Trembling-hand perfect equilibrium

	L	R
U	1, 1	2, 0
D	0, 2	2, 2

- Two pure strategy equilibria (U,L) and (D,R).
- Assume row player is playing $(1 - \varepsilon, \varepsilon)$ for $0 < \varepsilon < 1 \dots$

- In game theory, **trembling hand perfect equilibrium** is a refinement of Nash equilibrium due to Reinhard Selten. A trembling hand perfect equilibrium is an equilibrium that takes the possibility of off-the-equilibrium play into account by assuming that the players, through a “slip of the hand” or **tremble**, may choose unintended strategies, albeit with negligible probability.

- First we define a **perturbed game**. A perturbed game is a copy of a base game, with the restriction that only totally mixed strategies are allowed to be played. A totally mixed strategy is a mixed strategy where *every* pure strategy is played with non-zero probability. This is the "trembling hands" of the players; they sometimes play a different strategy than the one they intended to play. Then we define a strategy set S (in a base game) as being trembling hand perfect if there is a sequence of perturbed games that converge to the base game in which there is a series of Nash equilibria that converge to S .

Evolutionarily stable strategies

- A mixed strategy x^* in a two-player symmetric game is an evolutionarily stable strategy (ESS) if for every mixed strategy x that differs from x^* there exists $\varepsilon_0 = \varepsilon_0(x) > 0$ such that, for all ε in $(0, \varepsilon_0)$,

$$(1 - \varepsilon)u_1(x, x^*) + \varepsilon u_1(x, x) < (1 - \varepsilon)u_1(x^*, x^*) + \varepsilon u_1(x^*, x)$$

- Interpret x^* as distribution of types among “normal” individuals. Consider a mutation making use of strategy x , and assume that the proportion of this mutation in the population is ε .
- In ESS, the expected payoff of the mutation is smaller than the expected payoff of a normal individual, and hence the proportion of mutations will decrease and eventually disappear over time, with the composition of the population returning to being mostly x^* . An ESS is therefore a mixed strategy of the column player that is immune to being overtaken by mutations.

Sequential equilibrium

- **Sequential equilibrium** is a refinement of Nash Equilibrium for extensive form games due to David M. Kreps and Robert Wilson. A sequential equilibrium specifies not only a strategy for each of the players but also a **belief** for each of the players. A belief gives, for each information set of the game belonging to the player, a probability distribution on the nodes in the information set. A profile of strategies and beliefs is called an **assessment** for the game. Informally speaking, an assessment is a perfect Bayesian equilibrium if its strategies are sensible given its beliefs **and** its beliefs are confirmed on the outcome path given by its strategies. The definition of sequential equilibrium further requires that there be arbitrarily small perturbations of beliefs and associated strategies with the same property.

Proper equilibrium

- **Proper equilibrium** is a refinement of Nash Equilibrium due to Roger B. Myerson. Proper equilibrium further refines Reinhard Selten's notion of a trembling hand perfect equilibrium by assuming that more costly trembles are made with significantly smaller probability than less costly ones.
- Given a normal form game and a parameter $\epsilon > 0$, a totally mixed strategy profile σ is defined to be ϵ -proper if, whenever a player has two pure strategies s and s' such that the expected payoff of playing s is smaller than the expected payoff of playing s' (that is $u(s, \sigma_{-i}) < u(s', \sigma_{-i})$), then the probability assigned to s is at most ϵ times the probability assigned to s' . A strategy profile of the game is then said to be a proper equilibrium if it is a limit point, as ϵ approaches 0, of a sequence of ϵ -proper strategy profiles.

Matching pennies with a twist

	Guess heads up	Guess Tails up	Grab penny
Hide Heads Up	-1,1	0,0	-1,1
Hide Tails Up	0,0	-1,1	-1,1

- The Nash equilibria of the game are the strategy profiles where Player 2 grabs the penny with probability 1. Any mixed strategy of Player 1 is in (Nash) equilibrium with this pure strategy of Player 2. Any such pair is even trembling hand perfect. Intuitively, since Player 1 expects Player 2 to grab the penny, he is not concerned about leaving Player 2 uncertain about whether it is heads up or tails up. However, it can be seen that the unique proper equilibrium of this game is the one where Player 1 hides the penny heads up with probability $1/2$ and tails up with probability $1/2$ (and Player 2 grabs the penny). This unique proper equilibrium can be motivated intuitively as follows: Player 1 fully expects Player 2 to grab the penny. However, Player 1 still prepares for the unlikely event that Player 2 does not grab the penny and instead for some reason decides to make a guess. Player 1 prepares for this event by making sure that Player 2 has no information about whether the penny is heads up or tails up, exactly as in the original Matching Pennies game.

Critiques of Nash equilibrium

- Is it too strict?
 - Does not exist in all games
 - Might rule out some more “reasonable” strategies (e.g., a “safer” maxmin strategy)
- Not strict enough?
 - Potentially many equilibria to select through
 - Refinements: subgame perfect, trembling-hand perfect, Sequential equilibrium, proper equilibrium, evolutionarily stable strategy, ...
- Just right?
- It depends??

Repeated games

- $\Gamma = (N, (S_i)_{i \in N}, (u_i)_{i \in N})$
- Players play Γ over and over.
- Three cases:
 - Finite number of stages T , and every player wants to maximize his average payoff.
 - The game lasts an infinite number of stages, and every player wants to maximize the upper limit of his average payoffs
 - The game lasts an infinite number of stages, and each player wants to maximize the time-discounted sum of his payoffs.
- Let $M = \max_{i \in N} \max_{s \in S} |u_i(s)|$

	D	C
D	1, 1	4, 0
C	0, 4	3, 3

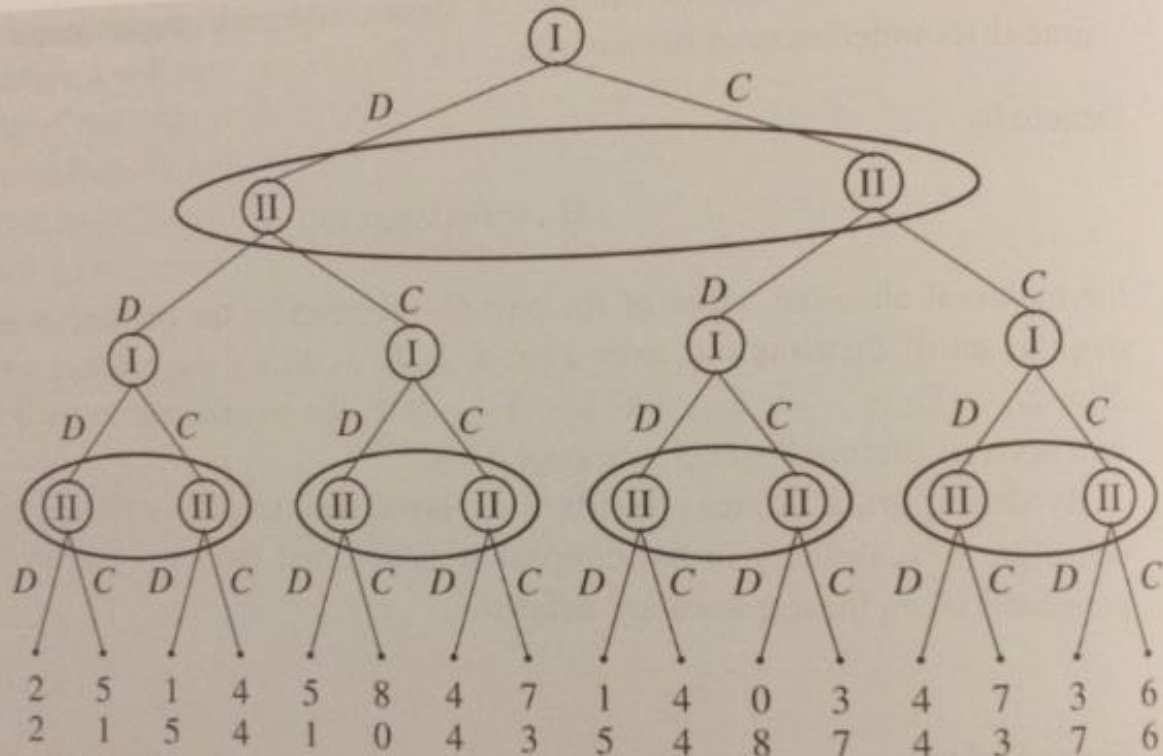


Figure 13.2 The two-stage Prisoner's Dilemma, represented as an extensive-form game

- At every equilibrium of the two-stage repeated game, the players play (D,D) in both stages.
- Proof:
 - Suppose instead there exists an equilibrium in which the players do not play (D,D) with positive probability in some stage. Let t in $\{1,2\}$ be the last stage in which there is positive probability they do not play (D,D) and suppose that in this event, Player I does not play D at stage t . This means that if the game continues after stage t the players will play (D,D). We will show that this strategy cannot be an equilibrium strategy.

- Case 1: $t = 1$.
 - Consider the strategy of Player I at which he plays D in both stages. We will show that this strategy grants him a higher payoff. Since D strictly dominates C, Player I's payoff rises if he switches from C to D in the first stage. And since, by assumption, after stage t the players play (D,D) (since stage t is the last stage in which they may not play (D,D)), Player I's payoff in the second stage was supposed to be 1. By playing D in the second stage, Player I's payoff is either 1 or 4 (depending on whether Player II plays D or C); in either case, Player I cannot lose in the second stage. The sum total of Player I's payoffs therefore rises.

- Case 2: $t = 2$.
 - Consider the strategy of Player I at which he plays in the first stage what the original strategy tells him to play, and in the second stage he plays D. Player I's payoff in the first stage does not change, but because D strictly dominates C, his payoff in the second stage does increase. The sum total of Player I's payoffs therefore increases.

- Note that despite the fact that at every equilibrium of the two-stage repeated game the players play (D,D) in every stage, it is possible that at equilibrium, the strategy C is used off the equilibrium path; that is, if a player does deviate from the equilibrium strategy, the other player may play C with positive probability. For example, consider the following strategy σ_1 :
 - Play D in the first stage.
 - In the second stage, play as follows: if in the first stage the other player played D, play D in the second stage; otherwise play $[1/8(C), 7/8(D)]$ in the second stage.
- Direct inspection shows that the strategy vector (σ_1, σ_2) is an equilibrium of the two-stage repeated game.¹⁸

- As we saw, in the finitely repeated Prisoner's Dilemma, at every equilibrium the players play (D,D) in every stage. Does this extend to every repeated game? That is, does every equilibrium strategy of a repeated game call on the players to play a one-stage equilibrium in every stage?

	D	C	P
D	1, 1	4, 0	-1, 0
C	0, 4	3, 3	-1, 0
P	0, -1	0, -1	-2, -2

- This game is similar to the Prisoner's Dilemma, with the addition of a third action P to each player, yielding low payoffs for both players. Note that action P (Punishment) is strictly dominated by action D. After eliminating P for both players, we are left with the one-stage Prisoner's Dilemma, whose only equilibrium is (D,D). It follows that the equilibrium is (D,D). It follows that playing (D,D) in both stages of repeated game is an equilibrium. In contrast with standard repeated Prisoner's Dilemma, there are additional equilibria in the repeated game:
 - Play C in the first stage
 - If your opponent played C in the first stage, play D in the second stage. Otherwise, play P in the second stage.

- If both players play this strategy, they will both play C in the first stage and D in the second stage, and each player's total payoff will be 4 (in contrast to the total payoff 2 that they receive under the equilibrium of playing (D,D) in both stages).

- Folk Theorem: Under some technical conditions the set of equilibrium payoffs is (or approximates) the set of feasible and individually rational payoffs of the base game. Can be extended for discounted infinitely repeated games and to uniform ε -equilibria for finitely repeated games.

- Any Nash equilibrium payoff in a repeated game must satisfy two properties:
- 1. **Individual rationality (IR)**: the payoff must weakly dominate the minmax payoff profile of the constituent stage game. I.e, the equilibrium payoff of each player must be at least as large as the minmax payoff of that player. This is because a player achieving less than his minmax payoff always has incentive to deviate by simply playing his minmax strategy at every history.
- 2. **Feasibility**: the payoff must be a convex combination of possible payoff profiles of the stage game. This is because the payoff in a repeated game is just a weighted average of payoffs in the basic games.

- Folk theorems are partially converse claims: they say that, under certain conditions (are different in each folk theorem), *every* payoff that is both IR and feasible can be realized as a Nash equilibrium payoff profile in the repeated game.
- There are various folk theorems, some relate to finitely-repeated games while others relate to infinitely-repeated games.

- For example, in the one-shot Prisoner's Dilemma, if both players cooperate that is not a Nash equilibrium. The only Nash equilibrium that both players defect, which is also a mutual minmax profile. One folk theorem says that, in the infinitely repeated version of the game, provided players are sufficiently patient, there is a Nash equilibrium such that both players cooperate on the equilibrium path. But in finitely repeated game by using backward induction it can be determined that players play Nash equilibrium in last period of the game which is defecting.

Fictitious play

- Simple “learning” update rule
- Initially proposed as an iterative method for computing Nash equilibria in zero-sum games, not as a learning model!
- Brown, G.W. (1951) “Iterative Solutions of Games by Fictitious Play”
- Algorithm:
 - Initialize beliefs about the opponent’s strategy
 - Repeat:**
 - 1) Play a best response to the assessed strategy of the opponent
 - 2) Observe the opponent’s actual play and update beliefs accordingly

- In fictitious play, the agent believes that his opponent is playing the mixed strategy given by the empirical distribution of the opponent's previous actions. That is, if A is the set of the opponent's actions, and for every a in A we let $w(a)$ be the number of times that the opponent has played action a , then the agent assess the probability of a in the opponent's mixed strategy as

$$- P(a) = w(a) / \sum_{a' \text{ in } A} w(a')$$

- For example, in a repeated Prisoner's Dilemma game, if the opponent has played C, C, D, C, D in the first five games, before the sixth game he is assumed to be playing the mixed strategy (0.6, 0.4).
- In general the tie-breaking rule chosen has little effect on the results of fictitious play.
- On the other hand, fictitious play is very sensitive to the players' initial beliefs. This choice, which can be interpreted as action counts that were observed before the start of the game, can have a radical impact on the learning process. Note that one must pick some nonempty prior belief for each agent; the prior beliefs cannot be $(0, \dots, 0)$, since this does not define a meaningful mixed strategy.

	Heads	Tails
Heads	1, -1	-1, 1
Tails	-1, 1	1, -1

Round	1's action	2's action	1's beliefs	2's beliefs
0			(1.5,2)	(2,1.5)
1	T	T	(1.5,3)	(2,2.5)
2	T	H	(2.5,3)	(2,3.5)
3	T	H	(3.5,3)	(2,4.5)
4	H	H	(4.5,3)	(3,4.5)
5	H	H	(5.5,3)	(4,4.5)
6	H	H	(6.5,3)	(5,4.5)
7	H	T	(6.5,4)	(6,4.5)
⋮	⋮	⋮	⋮	⋮

Table 7.1: Fictitious play of a repeated game of Matching Pennies.

- As the number of rounds tends to infinity, the empirical distribution of the play of each player will converge to $(0.5,0.5)$. If we take this distribution to be the mixed strategy of each player, the play converges to the unique Nash equilibrium of the normal form stage game, that in which each player plays the mixed strategy $(0.5,0.5)$.

- Definition: An action profile a is a **steady state** (or **absorbing state**) of fictitious play if it is the case that whenever a is played at round t it is also played at round $t+1$ (and hence in all future rounds as well).
- Theorem: If a pure-strategy profile is a strict Nash equilibrium of a stage game, then it is a steady state of fictitious play in the repeated game.
- Theorem: If a pure-strategy profile is a steady state of fictitious play in the repeated game, then it is a (possibly weak) Nash equilibrium in the stage game.

- Note that one cannot guarantee that fictitious play always converges to a Nash equilibrium, if only because agents can only play pure strategies and a pure-strategy Nash equilibrium may not exist in a given game. However, while the stage game strategies may not converge, the empirical distribution of the stage game strategies over multiple iterations may. And indeed this was the case in the Matching Pennies example given earlier, where the empirical distribution of each player's strategy converged to their mixed strategy (in the unique Nash equilibrium of the game).

- Theorem: If the empirical distribution of each player's strategies converges in fictitious play, then it converges to a Nash equilibrium.
- Note that this result does not make any claims about the distribution of the particular outcomes played, only about the final strategy profile.

Anti-Coordination game

- Two pure Nash equilibria, (A,B) and (B,A) with payoffs of 1, and one mixed where both do 0.5
A/B with payoffs of 0.5. Suppose the agents use fictitious play with weights (1, 0.5).

	A	B
A	0, 0	1, 1
B	1, 1	0, 0

	<i>A</i>	<i>B</i>
<i>A</i>	0, 0	1, 1
<i>B</i>	1, 1	0, 0

Figure 7.5: The Anti-Coordination game.

Now let us see what happens when we have agents play the repeated Anti-Coordination game using fictitious play. Let us assume that the weight function for each player is initialized to $(1, 0.5)$. The play of the first few rounds is shown in Table 7.2.

Round	1's action	2's action	1's beliefs	2's beliefs
0			(1,0.5)	(1,0.5)
1	B	B	(1,1.5)	(1,1.5)
2	A	A	(2,1.5)	(2,1.5)
3	B	B	(2,2.5)	(2,2.5)
4	A	A	(3,2.5)	(3,2.5)
⋮	⋮	⋮	⋮	⋮

Table 7.2: Fictitious play of a repeated Anti-Coordination game.

- The play of each player converges to the mixed strategy $(0.5, 0.5)$, which is the mixed strategy Nash equilibrium. However, the payoff received by each player is 0, since the players never hit the outcomes with positive probability. Thus, although the empirical distribution of the strategies converges to the mixed strategy Nash equilibrium, the players may not receive the expected payoff of the Nash equilibrium, because their actions are miscorrelated.

	Rock	Paper	Scissors
Rock	0, 0	0, 1	1, 0
Paper	1, 0	0, 0	0, 1
Scissors	0, 1	1, 0	0, 0

Figure 7.6: Shapley's Almost-Rock-Paper-Scissors game.

Round	1's action	2's action	1's beliefs	2's beliefs
0			(0,0,0.5)	(0,0.5,0)
1	Rock	Scissors	(0,0,1.5)	(1,0.5,0)
2	Rock	Paper	(0,1,1.5)	(2,0.5,0)
3	Rock	Paper	(0,2,1.5)	(3,0.5,0)
4	Scissors	Paper	(0,3,1.5)	(3,0.5,1)
5	Scissors	Paper	(0,1.5,0)	(1,0,0.5)
⋮	⋮	⋮	⋮	⋮

Table 7.3: Fictitious play of a repeated game of the Almost-Rock-Paper-Scissors game.

Theorem 7.2.5 *Each of the following is a sufficient condition for the empirical frequencies of play to converge in fictitious play:*

- *The game is zero sum;*
- *The game is solvable by iterated elimination of strictly dominated strategies;*
- *The game is a potential game;*⁵
- *The game is $2 \times n$ and has generic payoffs.*⁶

No-regret learning

- Let α^t be the average per-period reward the agent received up until time t , and let $\alpha^t(s)$ be the average per-period reward the agent *would have* received up until time t had he played pure strategy s instead, assuming all other agents continue to play as they did.

- Definition: The **regret** an agent experiences at time t for not having played s is

$$R^t(s) = \alpha^t(s) - \alpha^t.$$

- A learning rule is said to exhibit *no regret* if it guarantees that with high probability the agent will experience no positive regret.
- Definition: A learning rule exhibits *no regret* if for any pure strategy of the agent s it holds that $\Pr([\liminf R^t(s)] \leq 0) = 1$.
- This “in hindsight” requirement ignores the possibility that the opponents’ play might change as a result of the agent’s own play. For example, in finite-repeated Prisoner’s Dilemma, the only no-regret strategy is to always defect.

Examples of no-regret learning rules

- Regret matching: At each time step each action is chosen with probability proportional to its regret.
- Smooth Fictitious Play: Instead of playing the best response to the empirical frequency of the opponent's play, as fictitious play prescribes, one introduces a perturbation that gradually diminishes over time.

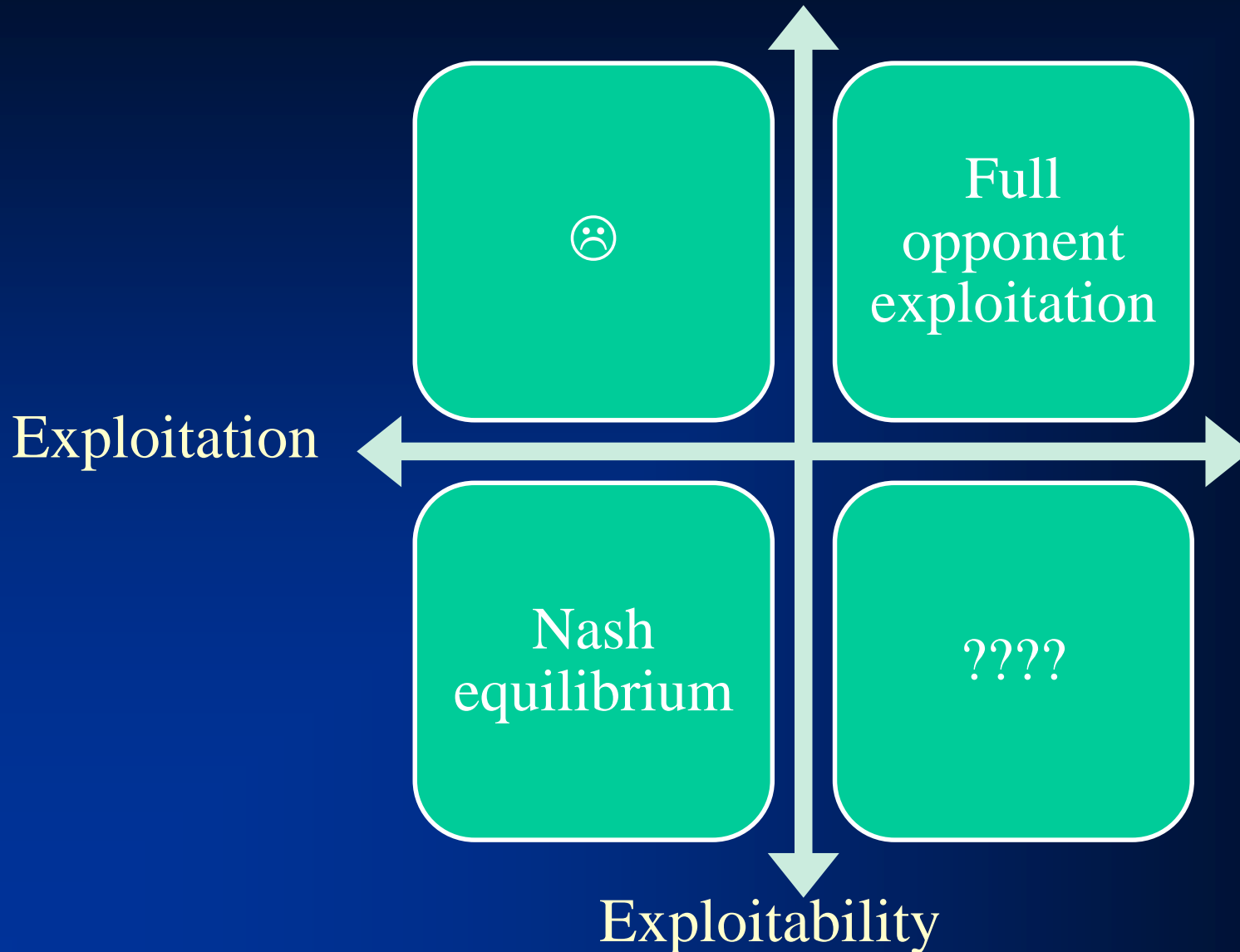
Counterfactual regret minimization

- Both players simultaneously play no-regret strategies.
- Shown in limit to converge to Nash equilibrium in many classes of games (even some classes of games with imperfect recall – *skew well-formed games*).
- Can be run in multiplayer games (produced a strong agent for three-player limit Texas hold ‘em), though no significant theoretical guarantees
 - Guarantees that strictly dominated actions and strategies won’t be played
- Used by strongest computer poker agents (all variants)
- Recently applied to medicine (robust diabetes management) and national security.

Opponent modeling and exploitation

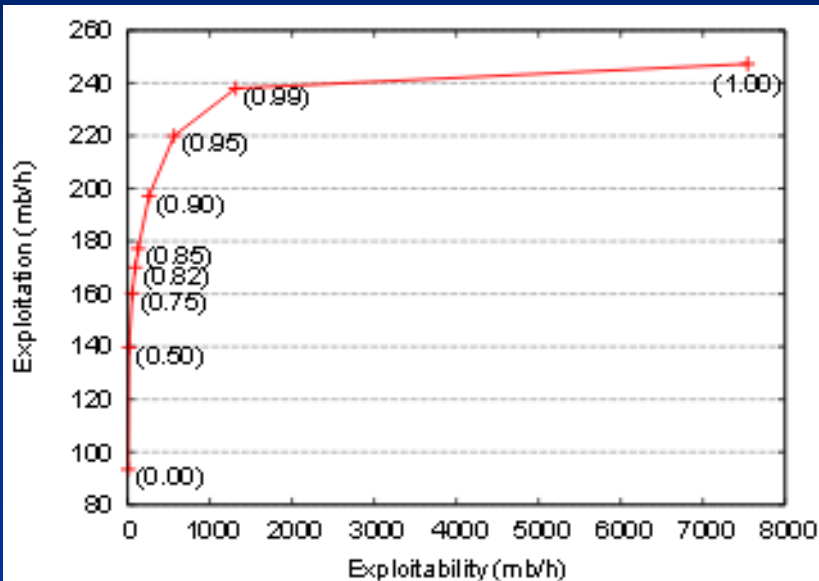
- Opponent modeling: construct prediction of opponent's strategy
 - Prior, how to integrate observations with prior to construct posterior
- Opponent exploitation:
 - Rule for responding to the opponent model
- Assuming a Dirichlet prior with multinomial sampling, Fictitious play is optimal.
- Much more challenging in imperfect-information games
 - Can't apply fictitious play because we don't know which information set action was taken at, so don't know which node to increase counter for.

Exploitation-exploitability tradeoff

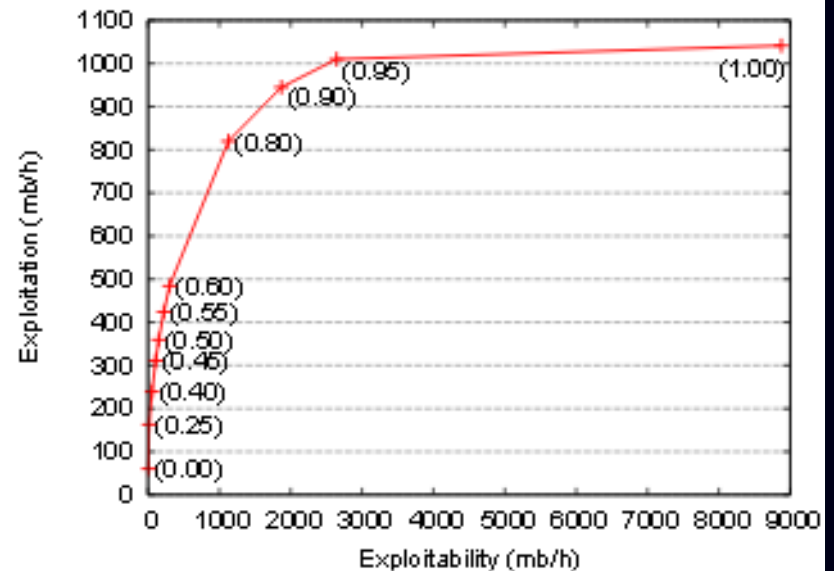


Robust opponent exploitation

- Restricted Nash Response: assume that opponent follows model σ^* with probability p , and plays best response to our strategy with probability $1-p$. Compute Nash equilibrium in this game, where we pick strategy x that is a best response to $p\sigma^* + (1-p)y$, and he picks y that is best response to x .



(a) Versus PsOpti4



(b) Versus A80

- A strategy which can be exploited for no more than ε is ε -safe.
 - The exploitability of a strategy is the difference between the game value and the performance against a nemesis. E.g., in Rock-Paper-Scissors, always playing Rock has exploitability 1, and the Nash equilibrium has exploitability 0.
- Theorem: For all σ_2 in Σ_2 , for all p in $(0,1]$, if σ_1 is a p -RNR to σ_2 , then there exists an epsilon such that σ_1 is an ε -safe best response to σ_2 .
 - $\varepsilon = \text{expl}(\sigma_1)$

Safe opponent exploitation

- Definition. *Safe* strategy achieves at least the value of the (repeated) game in expectation
- Is safe exploitation possible (beyond selecting among equilibrium strategies in the one-shot game)?

Rock-Paper-Scissors

- Suppose the opponent has played Rock in each of the first 10 iterations, while we have played the equilibrium σ^*
- Can we exploit him by playing pure strategy Paper in the 11th iteration?
 - Yes, but this would not be safe!
- By similar reasoning, any deviation from σ^* will be unsafe
- So safe exploitation is not possible in Rock-Paper-Scissors

Rock-Paper-Scissors-Toaster

	rock	paper	scissors	toaster
Rock	0,0	-1, 1	1, -1	4, -4
Paper	1,-1	0, 0	-1,1	3, -3
Scissors	-1,1	1,-1	0,0	3, -3

- t is *strictly dominated*
 - s does strictly better than t regardless of P1's strategy
- Suppose we play NE in the first round, and he plays t
 - Expected payoff of $10/3$
- Then we can play R in the second round and guarantee at least $7/3$ between the two rounds
- Safe exploitation is possible in RPST!
 - Because of presence of 'gift' strategy t

When can opponent be exploited safely?

- ~~Opponent played an (iterated weakly) dominated strategy?~~

R is a gift
but not iteratively weakly dominated

	L	M	R
U	3	2	10
D	2	3	0



- ~~Opponent played a strategy that isn't in the support of any eq?~~

R isn't in the support of any equilibrium
but is also not a gift

	L	R
U	0	0
D	-2	1

- Definition.** We received a *gift* if opponent played a strategy such that we have an equilibrium strategy for which the opponent's strategy isn't a best response
- Theorem.** Safe exploitation is possible iff the game has gifts

Exploitation algorithms

1.  Risk what you've won so far
 2.  Risk what you've won so far in expectation (over nature's & own randomization), i.e., risk the gifts received
 - Assuming the opponent plays a nemesis in states we don't observe
- **Theorem.** A strategy for a two-player zero-sum game is safe iff it never risks more than the gifts received according to #2
 - Can be used to make any opponent model / exploitation algorithm safe
 - No prior (non-eq) opponent exploitation algorithms are safe
 - We developed several new algorithms that are safe
 - Present analogous results and algorithms for extensive-form games of perfect and imperfect-information

Risk What You've Won in Expectation (RWYWE)

- Set $k^1 = 0$
- for $t = 1$ to T do
 - Set π_i^t to be k^t -safe best response to M
 - Play action a_i^t according to π_i^t
 - Update M with opponent's action a_{-i}^t
 - Set $k^{t+1} = k^t + u_i(\pi_i^t, a_{-i}^t) - v^*$

Assignment

- HW3 due 3/2
- Midterm on 3/7 (midterm review on 3/2).
 - Will cover material from lectures and homeworks (will not cover material from the textbooks that was not covered in lectures or homeworks).
 - 3 parts: multiple choice, true/false with explanation, analytical exercises
- No class 2/28