15-780: Graduate AI
Lecture 21. Learning in Games

Geoff Gordon (this lecture)
Ziv Bar-Joseph
TAs Michael Benisch, Yang Gu

Recap

- *Matrix games*
  - *2 or more players choose action simultaneously*
  - *Each from a discrete set of choices*
  - *Payoff to each agent is a function of all agents' choices (write as a collection of matrices)*

Recap

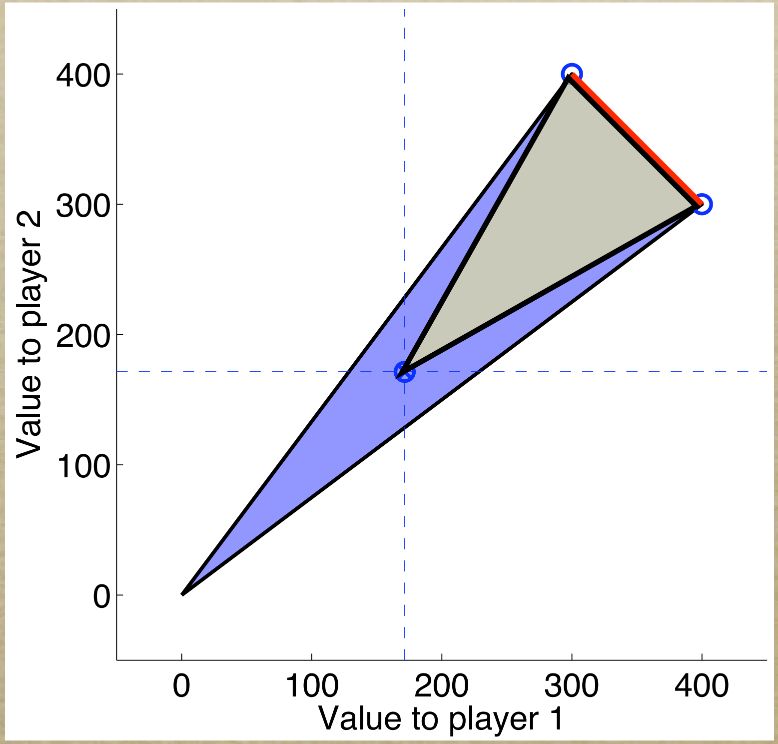
- ***Safety value** is the best I can guarantee myself with worst-case assumptions about opponent*
- *Also called maximin*
- *If we assume more about opponent (e.g., rationality) we might be able to get more reward*

Recap

- *Equilibrium = profile of strategies so that no one agent wants to deviate unilaterally*
  - *Nash: the one everyone talks about*
  - *Minimax: only makes sense in zero-sum two-player games, easier to compute*
  - *more later…*

Recap

- *Pareto dominance: not all equilibria are created equal*
- *For any in brown triangle, there is one on red line that's at least as good for **both** players*
- *Red line = **Pareto dominant***



Finding Nash

Shapley's game

	A	B	C
1	0,0	1,0	0,1
2	0,1	0,0	1,0
3	1,0	0,1	0,0

Support enumeration algorithm

	A	*B*	*C*
1	*0,0*	*1,0*	*0,1*
2	*0,1*	*0,0*	*1,0*
3	*1,0*	*0,1*	*0,0*

- *Enumerate all support sets for each player*
- *Row: 1, 2, 3, 12, 13, 23, 123*
- *Col: A, B, C, AB, AC, BC, ABC*
- $7 \times 7 = 49$ *possibilities*

Support enumeration

- *For each pair of supports, solve an LP*
- *Vars are P(action) for each action in support (one set for each player), and also expected value to each player*
- *Constraints:*
  - *All actions in support have value v*
  - *All not in support have value ≤ v*
  - *Probabilities in support ≥ 0, sum to 1*

Support enumeration

	A	B	C
1	0,0	1,0	0,1
2	0,1	0,0	1,0
3	1,0	0,1	0,0

- *Checking singleton supports is easy: sum-to-1 constraint means p=1 for action in support*
- *So just check whether actions out of support are worse*

Try 2-strategy supports: 12, AB

	A	B	C
1	0,0	1,0	0,1
2	0,1	0,0	1,0
3	1,0	0,1	0,0

- *Payoff of Row 1:* $0\, p(A) + 1\, p(B) = v$
- *Payoff of Row 2:* $0\, p(A) + 0\, p(B) = v$
- *Payoff of Col A:* $0\, p(1) + 1\, p(2) = w$
- *Payoff of Col B:* $0\, p(1) + 0\, p(2) = w$

Try 2-strategy supports: 12, AB

	A	*B*	*C*
1	*0,0*	*1,0*	*0,1*
2	*0,1*	*0,0*	*1,0*
3	*1,0*	*0,1*	*0,0*

- $0\ p(A) + 1\ p(B) = v = 0\ p(A) + 0\ p(B)$
- $0\ p(1) + 1\ p(2) = w = 0\ p(1) + 0\ p(2)$
- *Row payoff* $\geq$ *row 3:* $v \geq 1\ p(A) + 0\ p(B)$
- *Col payoff* $\geq$ *col C:* $w \geq 1\ p(1) + 0\ p(2)$

More supports

- *Other 2-vs-2 are similar*
- *We also need to try 1-vs-2, 1-vs-3, and 2-vs-3, but in interest of brevity: they don’t work either*
- *So, on the 49th iteration, we reach 123 vs ABC…*

123 vs ABC

	A	*B*	*C*
1	*0,0*	*1,0*	*0,1*
2	*0,1*	*0,0*	*1,0*
3	*1,0*	*0,1*	*0,0*

- *Row 1: 0 p(A) + 1 p(B) + 0 p(C) = v*
- *Row 2: 0 p(A) + 0 p(B) + 1 p(C) = v*
- *Row 3: 1 p(A) + 0 p(B) + 0 p(C) = v*
- *So, p(A) = p(B) = p(C) = v = 1/3*

123 vs ABC

	A	*B*	*C*
1	*0,0*	*1,0*	*0,1*
2	*0,1*	*0,0*	*1,0*
3	*1,0*	*0,1*	*0,0*

- *Col A:* $0\,p(1) + 0\,p(2) + 1\,p(3) = w$
- *Col B:* $1\,p(1) + 0\,p(2) + 0\,p(3) = w$
- *Col C:* $0\,p(1) + 1\,p(2) + 0\,p(3) = w$
- *So,* $p(1) = p(2) = p(3) = w = 1/3$

Nash of Shapley

- *There are nonnegative probs p(1), p(2), & p(3) for Row that equalize Col's payoffs for ABC*
- *There are nonnegative probs p(A), p(B), & p(C) for Col that equalize Row's payoffs for 123*
- *No strategies outside of supports to check*
- *So, we've found the (unique) NE*

Correlated equilibrium

Correlated equilibrium

If there is intelligent life on other planets, in a majority of them, they would have discovered correlated equilibrium before Nash equilibrium.

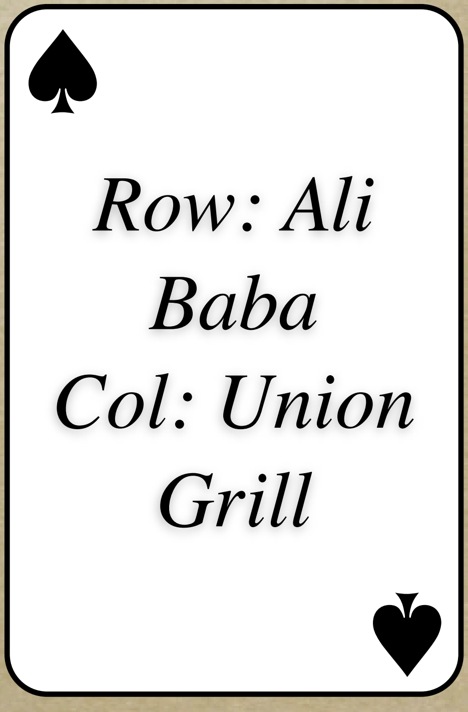
—Roger Myerson

The game of “Lunch”

	A	*U*
A	*4, 3*	*0, 0*
U	*0, 0*	*3, 4*

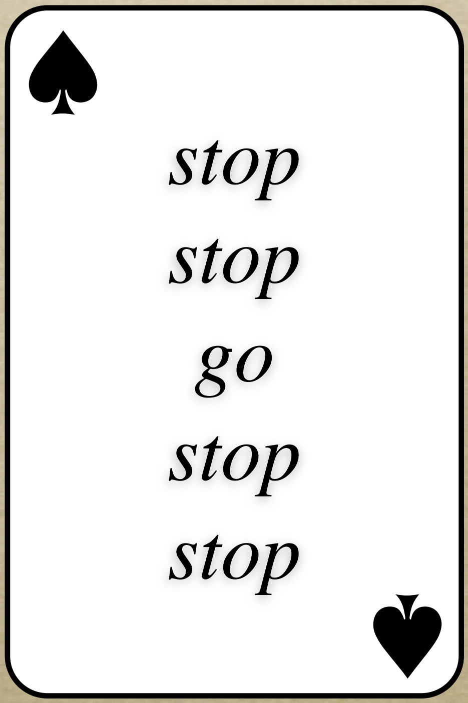
Moderator

- *A moderator has a big deck of cards*
- *Each card has written on it a recommended action for each player*
- *Moderator draws a card, whispers actions to corresponding players*
    - *actions may be correlated*
    - *only find out your own*



Correlated equilibrium

- *Since players can have correlated actions, an equilibrium with a moderator is called a **correlated equilibrium***
- *Example: 5-way stoplight*
- *All NE are CE*
- *At least as many CE as NE in every game (often strictly more)*



Realism?

- *Moderators are often available*
- *Sometimes have to be kind of clever*
- *E.g., can simulate a moderator using cheap talk and some crypto*
- *Or, can use private function of public randomness (e.g., headline of NY Times, # of sunspots, or even past history of play)*

Finding correlated equilibrium

	A	U
A	a	b
U	c	d

- *P(Row is recommended to play A) = a + b*
- *P(Col recommended A | Row recommended A) = a / (a + b)*
- *Rationality: when I'm recommended to play A, I don't want to play U instead*

Rationality constraint

$$4\frac{a}{a+b} + 0\frac{b}{a+b} \geq 0\frac{a}{a+b} + 3\frac{b}{a+b} \qquad \text{if } a+b>0$$

	A	U
A	a	b
U	c	d

	A	U
A	4,3	0,0
U	0,0	3,4

Rationality constraint

Rpayoff(A, A) P(col A | row A)

$$4\frac{a}{a+b} + 0\frac{b}{a+b} \geq 0\frac{a}{a+b} + 3\frac{b}{a+b} \qquad \text{if } a+b>0$$

	A	U
A	a	b
U	c	d

	A	U
A	4,3	0,0
U	0,0	3,4

Rationality constraint

Rpayoff(A, A) P(col A | row A)

$$4\frac{a}{a+b} + 0\frac{b}{a+b} \geq 0\frac{a}{a+b} + 3\frac{b}{a+b} \qquad \text{if } a+b>0$$

Rpay(A, U) P(U | A)

	A	U
A	a	b
U	c	d

	A	U
A	4,3	0,0
U	0,0	3,4

Rationality constraint

Rpayoff(A, A) P(col A | row A)

Rpay(U, A) P(A | A)

$$4\frac{a}{a+b} + 0\frac{b}{a+b} \geq 0\frac{a}{a+b} + 3\frac{b}{a+b} \qquad \text{if } a+b>0$$

Rpay(A, U) P(U | A)

	A	U
A	a	b
U	c	d

	A	U
A	4,3	0,0
U	0,0	3,4

Rationality constraint

Rpayoff(A, A) P(col A | row A)

Rpay(U, A) P(A | A)

$$4\frac{a}{a+b} + 0\frac{b}{a+b} \geq 0\frac{a}{a+b} + 3\frac{b}{a+b} \qquad \text{if } a+b>0$$

Rpay(A, U) P(U | A)

Rpay(U, U) P(U | A)

	A	*U*
A	*a*	*b*
U	*c*	*d*

	A	*U*
A	*4,3*	*0,0*
U	*0,0*	*3,4*

Rationality constraint is linear

$$4\frac{a}{a+b} + 0\frac{b}{a+b} \geq 0\frac{a}{a+b} + 3\frac{b}{a+b} \qquad \text{if } a+b>0$$

$$4a + 0b \geq 0a + 3b$$

All rationality constraints

	A	*U*
A	*a*	*b*
U	*c*	*d*

	A	*U*
A	*4,3*	*0*
U	*0*	*3,4*

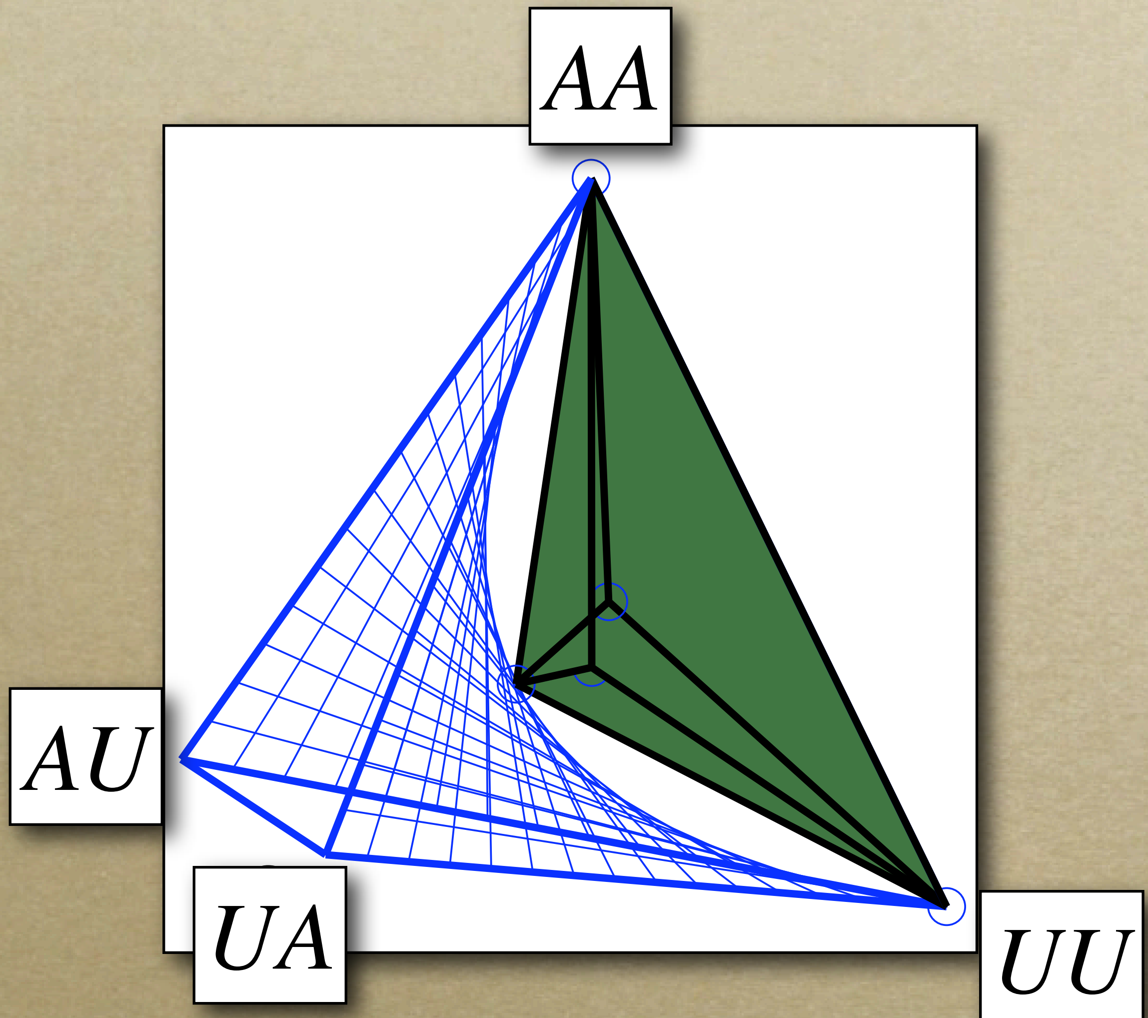
Row recommendation A $4a + 0b \geq 0a + 3b$

Row recommendation U $0c + 3d \geq 4c + 0d$

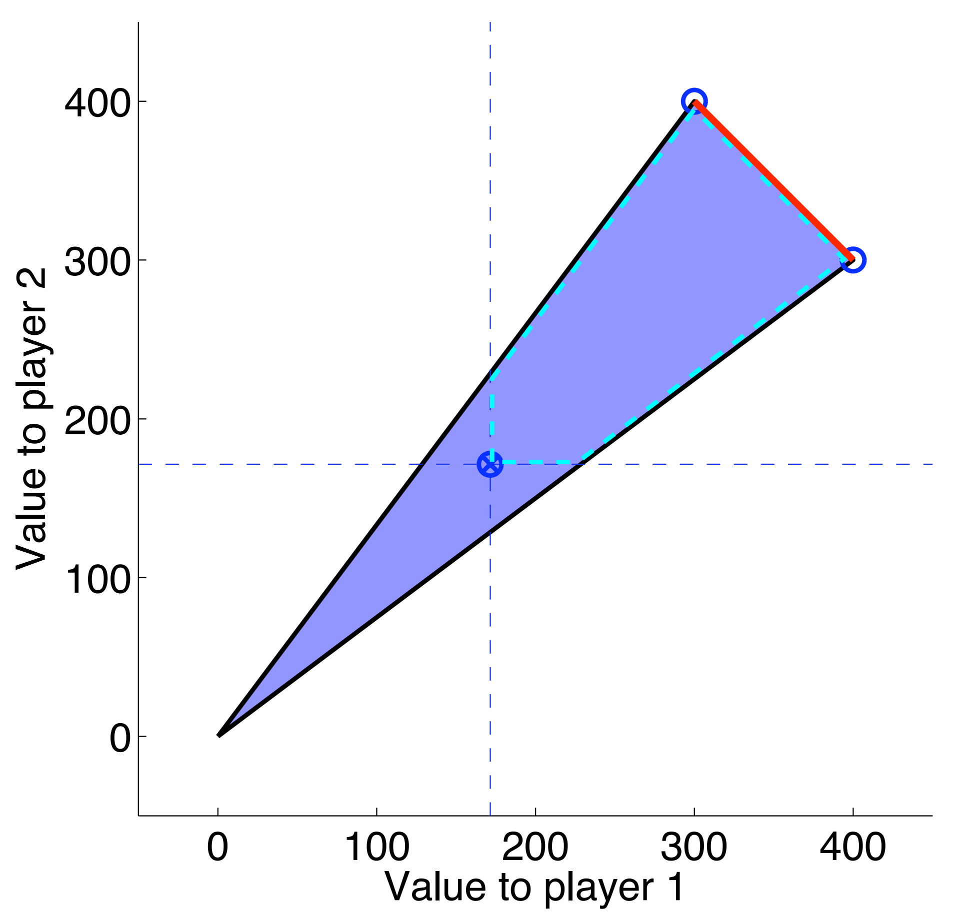
Col recommendation A $3a + 0c \geq 0a + 4c$

Col recommendation U $0b + 4d \geq 3b + 0d$

Correlated equilibrium



Correlated equilibrium payoffs



Bargaining

Predicting outcomes

- *We've talked about different things we might assume about "rational" agents*
- *Each assumption leads to different predictions about set of possible outcomes*
- *E.g., independent utility maximizers should reach a Nash equilibrium*
- *E.g., adding a moderator increases possible outcomes to set of CE*

Predicting outcomes

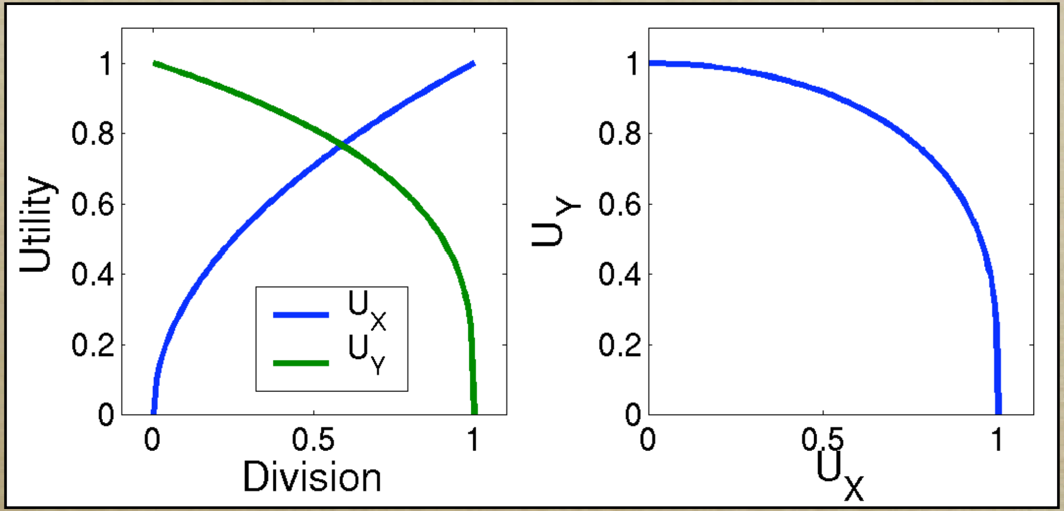
- *But so far we can't predict what will actually happen when "rational" agents play a game together*
- *Most specific prediction so far is Pareto frontier (of either set of Nash or set of CE)*
- *Next: try adding "cheap talk" to see whether we finally get a unique prediction*

Return of "Lunch"

	A	*U*
A	*4, 3*	*0, 0*
U	*0, 0*	*3, 4*

A = Ali Baba, U = Union Grill

Rubinstein's game



- *Two players split a pie*
- *Each has concave, increasing utility for a share in [0,1]*

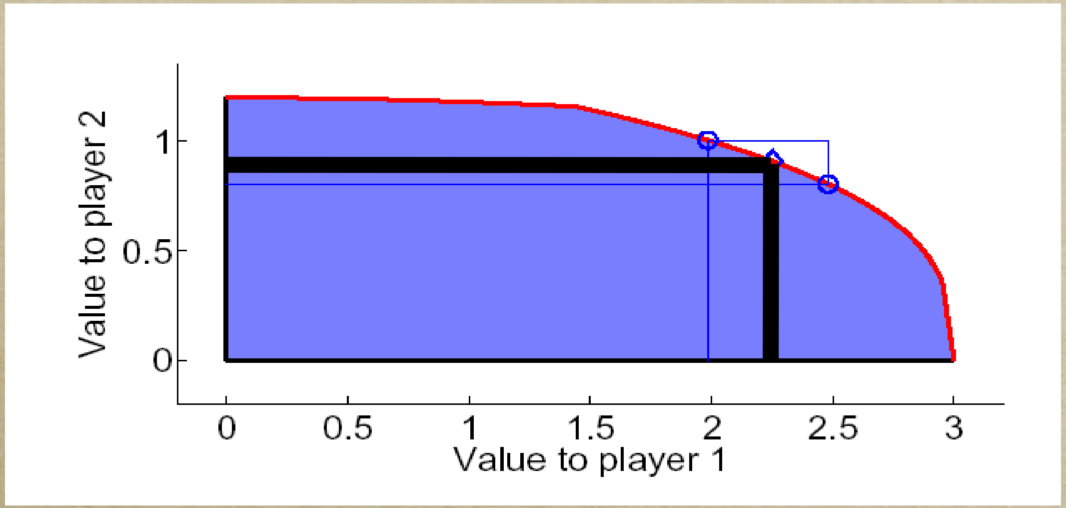
Rubinstein's game

- *Bargain by alternating offers:*
  - *Alice offers 60-40*
  - *Bob says no, how about 30-70*
  - *Alice says no, wants 55-45*
  - *Bob says OK*
- *Alice gets $\gamma^2 U_A(0.55)$, Bob: $\gamma^2 U_B(0.45)$*
- *In case of disagreement, no pie for anyone*

Rubinstein's game

- *Bargain by alternating offers:*
  - *Alice offers 60-40*
  - *Bob says no, how about 30-70*
  - *Alice says no, wants 55-45*
  - *Bob says OK*
- *Alice gets* $\gamma^2 U_A(0.55)$, *Bob:* $\gamma^2 U_B(0.45)$
- *In case of disagreement, no pie for anyone*

Theorem

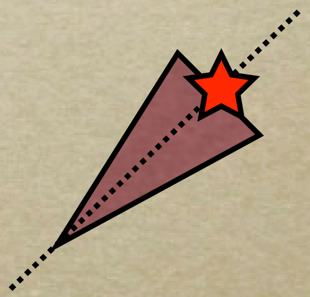


- *In this model, we can finally predict what "rational" players will do*
- *Will arrive (near)* ***Nash bargaining point****, which maximizes product of extra utilities*

$$(U_1 - min_1)\,(U_2 - min_2)$$

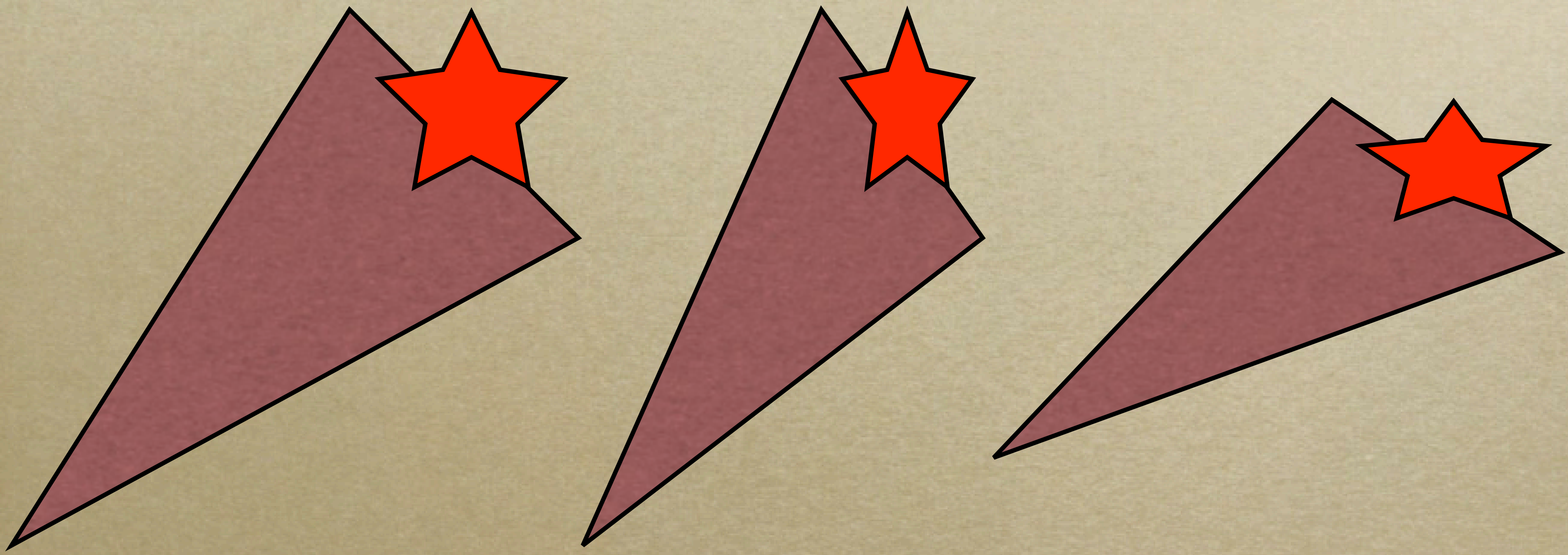
Theorem

- *NBP is* ***unique*** *outcome that is*
  - ***optimal*** *(on Pareto frontier)*
  - ***symmetric*** *(utilities are equal if possible outcomes are symmetric)*

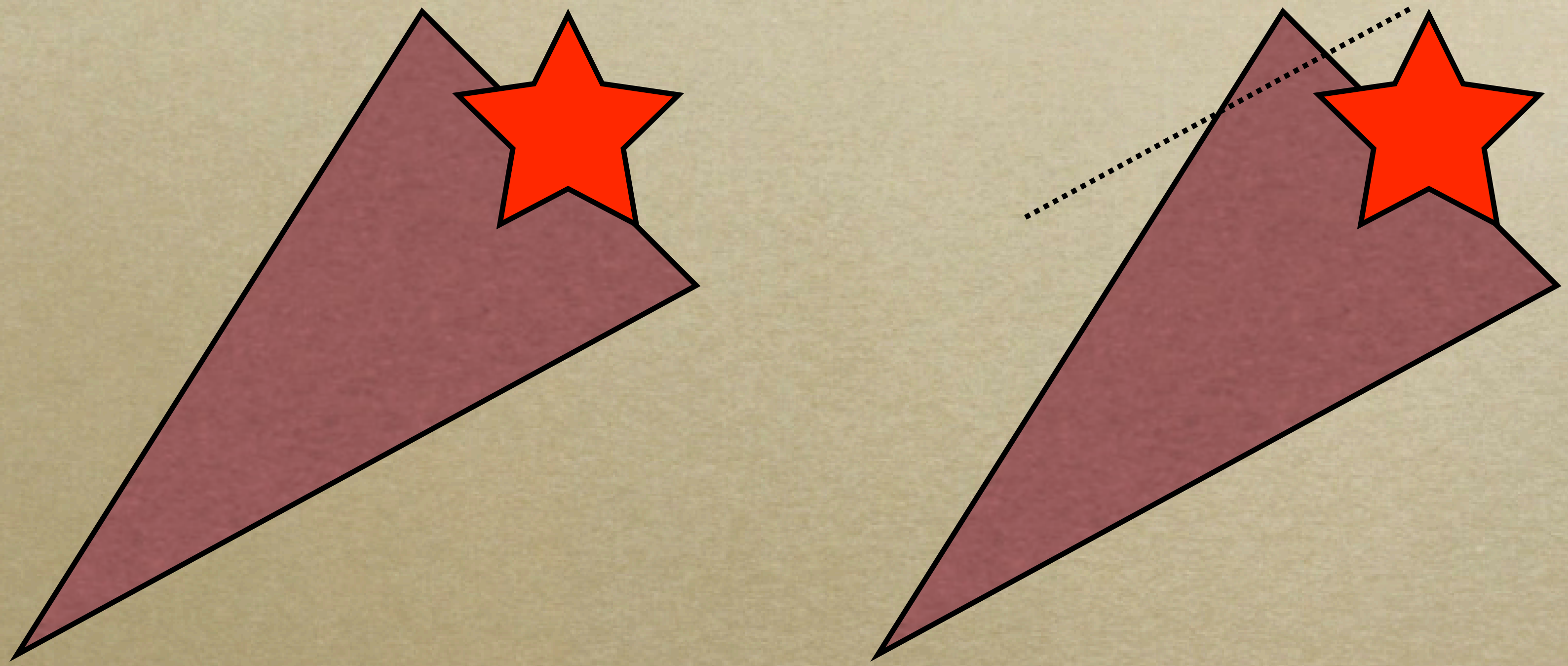


  - ***scale-invariant***
  - ***independent of irrelevant alternatives***

Scale invariance



Independence of irrelevant alternatives

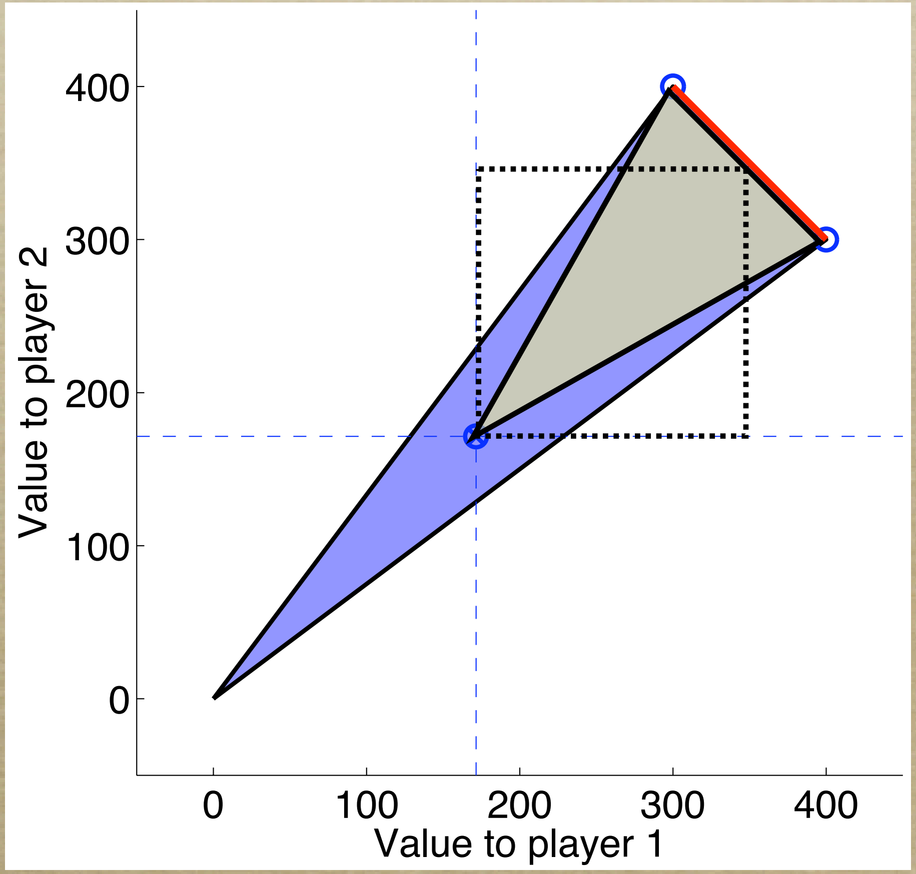


Lunch with Rubinstein

- *Can we use Rubinstein's game to predict outcome of Lunch?*
- *Now an offer = "let's play this equilibrium"*
- *Must at least assume communication*
- *What else?*

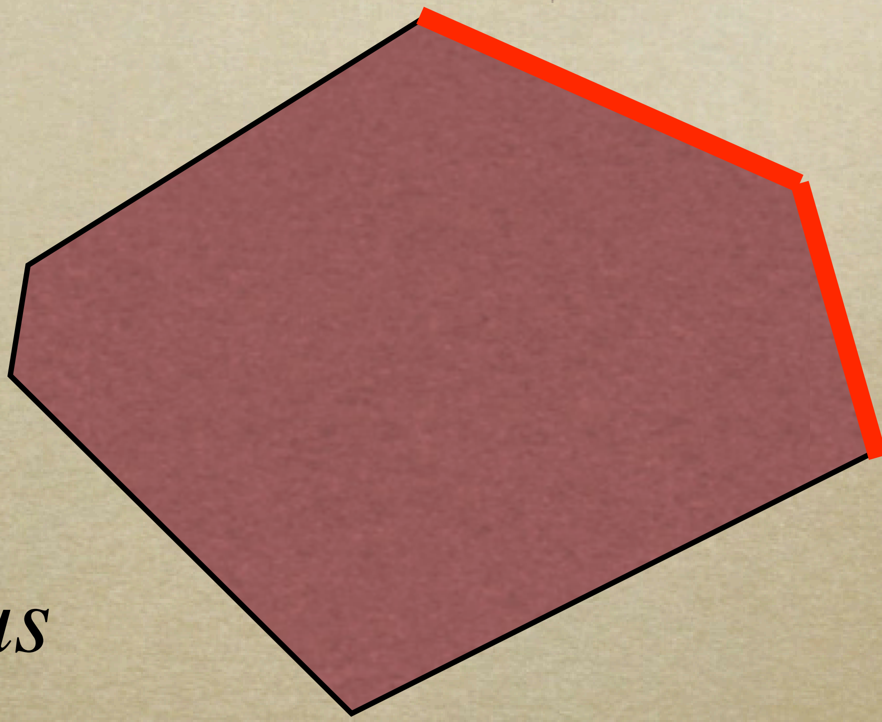
Lunch with Rubinstein

- *Can we use Rubinstein's game to predict outcome of Lunch?*
- *Now an offer = "let's play this equilibrium"*
- *Must at least assume communication*
- *What else?*



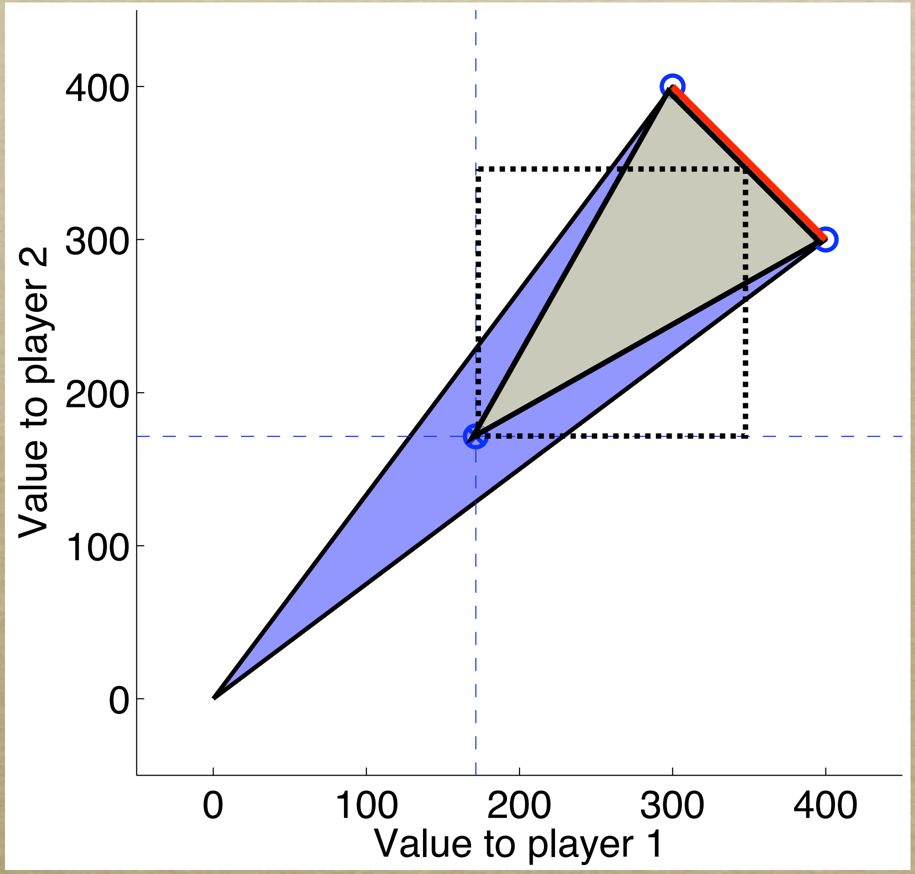
What else?

- *Rubinstein assumes that players know what will happen if they disagree*
- *In pie-splitting it's obvious*
- *In general, just as hard as agreeing in the first place*



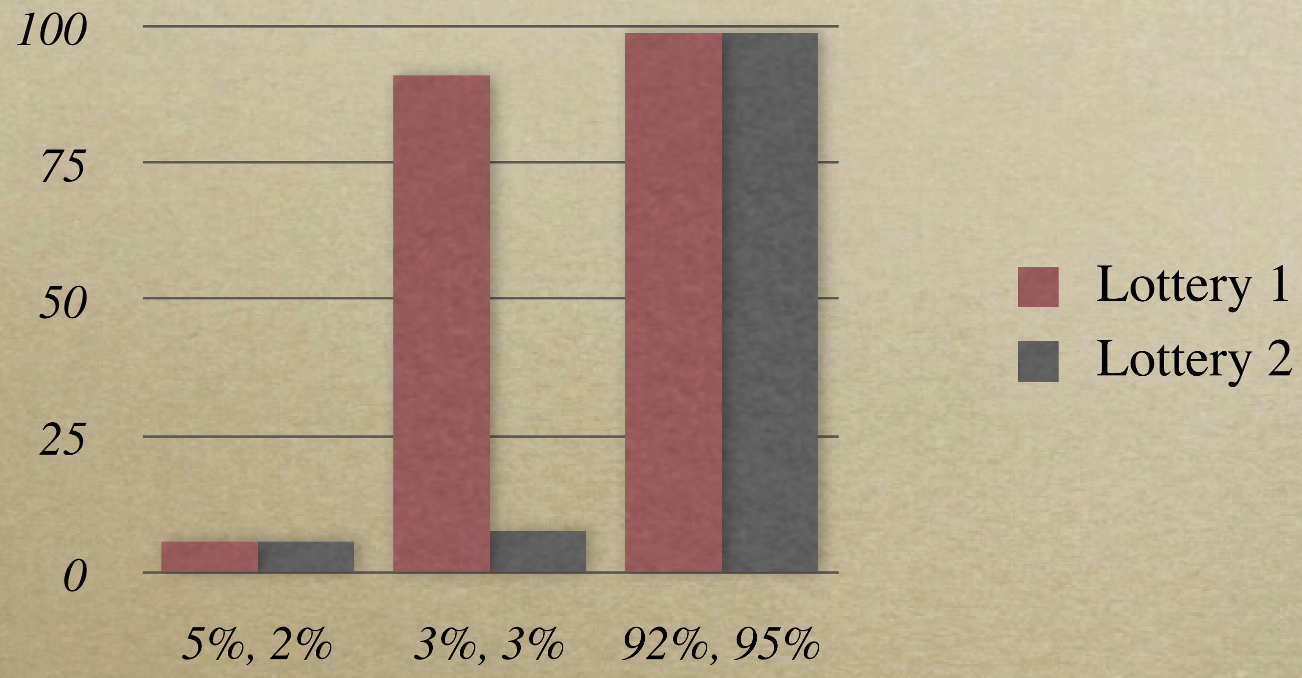
Disagreement over Lunch

- *In Lunch, one NE is an obvious disagreement point*
- *But even this isn't completely obvious: strategy isn't same as safety strategy w/ same payoff*



Another example

Let’s play the lottery



- *($6, .05; $91, .03; $99, .92)*
- *($6, .02; $8, .03; $99, .95)*
- *Which would you pick?*

Rationality

- *People often pick*
  - *($6, .05; $91, .03; $99, .92)*
- *over*
  - *($6, .02; $8, .03; $99, .95)*
- *But, note **stochastic dominance***

Stochastic dominance

Birnbaum & Navarrete. Testing Descriptive Utility Theories: Violations of Stochastic Dominance and Cumulative Independence

Stochastic dominance

Birnbaum & Navarrete. Testing Descriptive Utility Theories: Violations of Stochastic Dominance and Cumulative Independence

Learning in Games

Why study learning in games?

- *To predict what humans will do*
- *To predict what "rational" agents will do*
- *To compute an equilibrium*
- *To build an agent that plays "well" with minimal assumptions about others*
  - *this seems like the most AI-ish goal*

Learning

- *Start with beliefs / inductive bias (about other players, Nature, rules of game…)*
- *During repeated plays of the game*
  - *or during one long play of a game where we can revisit the same or similar states*
- *Adjust our own play to improve payoff*

First try

- *Run any standard supervised learning algorithm to predict*
    - *payoff of each of my actions, or*
    - *play of all other players*
- *Now act to maximize my predicted utility on next turn*

For example

- *In Rock-Paper-Scissors, suppose I tally opponent's past plays, and find:*
  - *173 Rock, 173 Paper, 174 Scissors*
  - *(or perhaps, tally opp's plays in situations "like" the current one)*

For example

- *173 Rock, 173 Paper, 174 Scissors*
- *Learning algorithm tells me Rock has slightly higher predicted payoff*
- *So I play Rock*

For example

- *Sadly, opponent played Paper.*

For example

- *Tally is now 173, 174, 174*
- *So learning algo tells us to play Scissors or Rock*
- *Say we break tie and pick Scissors*

For example

- *Sadly, opponent played Rock.*

For example

- *Tally is now 174, 174, 174*
- *So learning algo tells us everything's the same*
- *Say we break tie and pick Paper*

For example

- *Sadly, opponent played Scissors.*

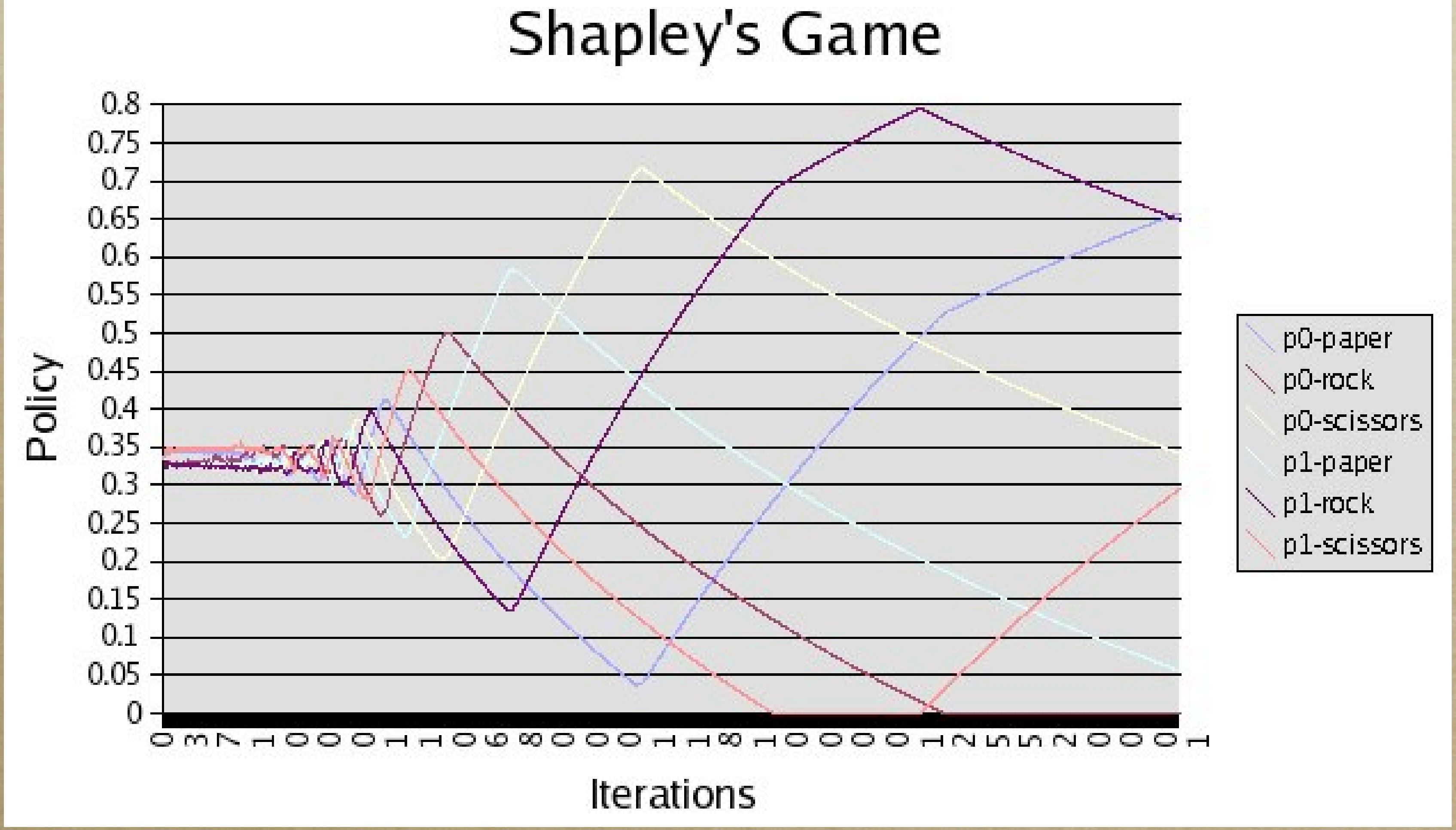
For example

- *Tally is now 174, 174, 175*
- *And cycle repeats*

Fictitious play

- *Algorithm we just ran was called* ***fictitious play***
- *Could it really do this badly?*
- *Yes, if opponent knows we're using FP*
- *Knowing tie-break rule helps but isn't essential*

Fictitious play



- *Even in self-play, FP can do badly*

Second try

- *We were kind of short-sighted when we chose to optimize our immediate utility*
- *What if we formulate a prior, not over single plays, but over (infinite) sequences of play (conditioned on our own strategy)?*
- *E.g., P(7th opp play is R, 12th is S | my first 11 plays are RRRPRPRSSSR) = 0.013*

Rational learner

- *Now we can look ahead: find best play considering all future effects*
- *R might garner more predicted reward now, but perhaps S will confuse opponent and let me get more reward later…*
- *This is called* ***rational learning***
- *A complete rational learner must also specify tie-break rule*

Rational learner: discussion

- *First problem: maximization over an uncountable set of strategies*
- *Second problem: our play is still deterministic, so if opponent gets a copy of our code we're still sunk*
- *What if we have a really big computer and can hide our prior?*

Theorem

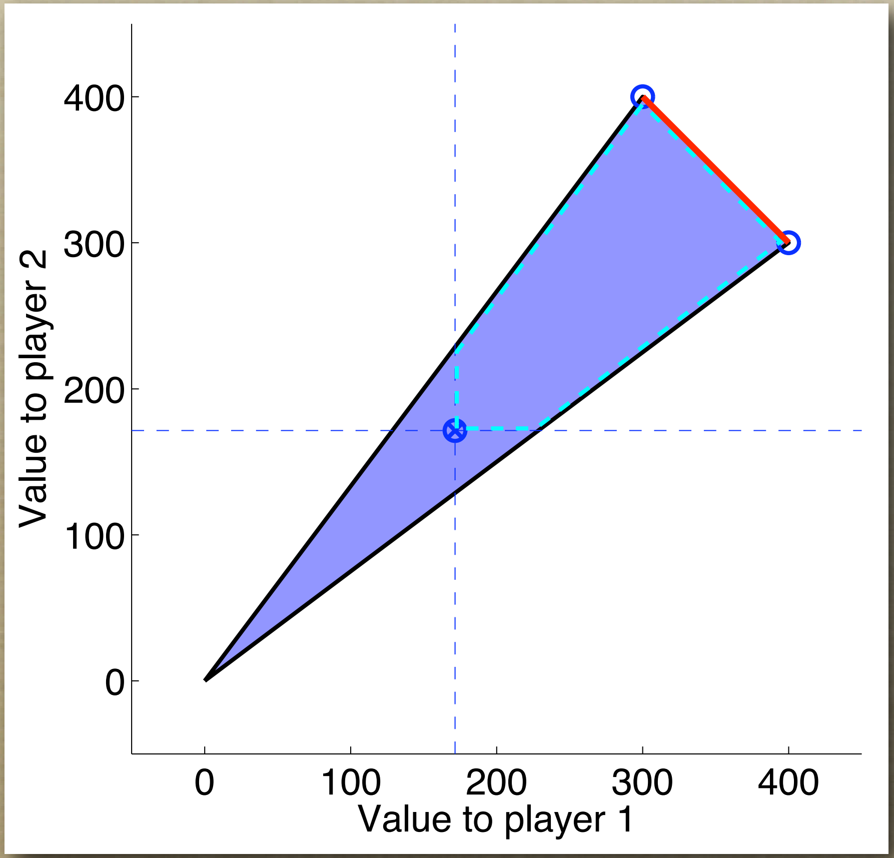
- *Any vector of rational learners which (mumble mumble) will, when playing each other in a repeated game, approach the play frequencies and payoffs of some Nash equilibrium arbitrarily closely in the limit*

  Ehud Kalai and Ehud Lehrer. Rational Learning Leads to Nash Equilibrium. Econometrica, Vol. 61, No. 5, 1993.

What does this theorem tell us?

- *Problem: "mumble mumble" actually conceals a condition that's difficult to satisfy in practice*
  - *for example, it was violated when we peeked at prior and optimized response*
  - *nobody knows whether there's a weaker condition that guarantees anything nice*

What does this theorem tell us?



- *Problem: there are often a lot of Nash equilibria*