Potential Games and Equilibrium Selection

Oxford Mini-Course Lecture 1

Stephen Morris

October 2025

potential games

- "potential games" are games where best responses are the same as they would be if players had a common payoff function (the "potential" P).
- potential games will generally have multiple equilibria
- but there will "typically" be only one that maximizes the potential function (the potential maximizing equilibrium, or PME)

potential games

Monderer and Shapley (1996) introduced the "potential game" language and wrote that the existence of a potential function

> *raises the natural question about the economic content (or interpretation) of P: What do firms try to maximize? We do not have an answer to this question. However, it is clear that the mere existence of a potential function helps us (and the players) to better analyze the game.*

potential games

- my two lectures review Monderer and Shapley (1996) and experimental and theoretical (learning and higher-order belief) literatures that suggest selection of PME and generalizations thereof
  1. today: potential games overview, experiments, some selection
  2. tomorrow: review "higher-order belief" theoretical justification for PME: overview of literature on "robustness to incomplete information" and global games (as well as a connection to information design and noisy contract design)
- perhaps you would consider joining me in taking potential maximizing equilibria (PME) as a selection criterion?

outline

- lecture 1
    1. definition of "potential games"
    2. examples
    3. myopic learning
    4. experiments
    5. binary action supermodular (BAS) games
    6. generalized potential games
- lecture 2

outline

- lecture 1
- lecture 2
  1. higher order beliefs in game theory
  2. email game and global games (Rubinstein (1989) and Carlsson and van Damme (1993))
  3. robustness to incomplete information (Kajii and Morris (1997)) and global games with multidimensional types (Oury (2013) and Veiel (2025))

game

- set of players $I = \{1, 2, ..., |I|\}$
- for each player i,
  - action set A_i (sometimes but not always finite)
  - payoff function $g_i : A \longrightarrow \mathbb{R}$
    - where $A = A_1 \times ... \times A_{|I|}$
- this is (finite) game $\mathbf{g} = \left(g_1, ..., g_{|I|}\right)$.

notation

- we will write expressions like $A_{-i} = \underset{j \neq i}{\times} A_i$, $a_{-i} \in A_{-i}$
- a mixed strategy is a probability distribution over actions

$$\alpha_i \in \Delta\left(A_i\right)$$

- so a mixed strategy profile is:

$$\boldsymbol{\alpha} \in \underset{i \in I}{\times} \Delta\left(A_i\right)$$

- we will abuse notion by writing expressions like

$$g_i\left(a_i, \alpha_{-i}\right) = \sum_{a_{-i} \in A_{-i}} \left(\prod_{j \neq i} \alpha_j\left(a_j\right)\right) g_i\left(a_i, a_{-i}\right)$$

equilibrium

- a mixed strategy profile $\boldsymbol{\alpha} \in \underset{i \in I}{\times} \Delta(A_i)$ is a Nash equilibrium of $\mathbf{g}$ if $\alpha_i(a_i) > 0$ implies

$$g_i(a_i, \alpha_{-i}) \geq g_i\left(a_i', \alpha_{-i}\right)$$

for all $a_i' \in A_i$

- a pure strategy profile $a \in A$ is a pure strategy Nash equilibrium of $\mathbf{g}$ if the corresponding degenerate mixed strategy profile is a Nash equilibrium

ordinal potential games

- function $P : A \longrightarrow \mathbb{R}$ is an *ordinal potential* for the game $\mathbf{g}$ if, for all $i \in I$ and every $a_{-i} \in A_{-i}$,

$$g_i\left(a_i, a_{-i}\right) - g_i\left(a_i', a_{-i}\right) > 0 \Leftrightarrow P\left(a_i, a_{-i}\right) - P\left(a_i', a_{-i}\right) > 0$$

  for all a_i and a_i'.
- a game is an *ordinal potential game* if it admits an *ordinal potential*
- Monderer and Shapley (1996) *GEB*
- **lemma**: If game $\mathbf{g}$ admits an ordinal potential P, then the set of pure strategy Nash equilibria of $\mathbf{g}$ is the same as the set of pure strategy Nash equilibria of the common interest game with common payoff function P
  - corollary 1: if the potential P of an ordinal potential game $\mathbf{g}$ has a maximum, it is a pure strategy Nash equilibrium of $\mathbf{g}$
  - corollary 2: If a finite game $\mathbf{g}$ admits an ordinal potential P, then there exists a pure strategy Nash equilibrium of $\mathbf{g}$

best response potential games

- Function $P : A \to \mathbb{R}$ is a *best response potential* for the game if, for all $i \in I$ and $\lambda_i \in \Delta(A_{-i})$,

$$\arg\max_{a_i \in A_{-i}} \sum \lambda_i(a_{-i})\, g_i(a_i, a_{-i}) = \arg\max_{a_i \in A_{-i}} \sum \lambda_i(a_{-i})\, P(a_i, a_{-i})$$

  for all a_i and a_i'.
- A game is a best response potential game if it admits a best response potential.
- Morris and Ui (2004) *GEB*
- If a game $\mathbf{g}$ admits a best response potential P, the set of Nash equilibria of $\mathbf{g}$ is the same as the set of Nash equilibria of the common interest game with common payoff function P.

weighted potential games

- Function $P : A \to \mathbb{R}$ is an *weighted potential* for the game $\mathbf{g}$ if there exist weights $(w_i)_{i \in I} \in \mathbb{R}^I$ such that, for all $i \in I$ and every $a_{-i} \in A_{-i}$,

$$g_i(a_i, a_{-i}) - g_i(a_i', a_{-i}) = w_i \left(P(a_i, a_{-i}) - P(a_i', a_{-i})\right)$$

  for all a_i and a_i'.
- A game is a weighted potential game if it admits an weighted potential.
- Monderer and Shapley (1996) *GEB*
- If a game $\mathbf{g}$ admits a weighted potential P, the set of Nash equilibria of $\mathbf{g}$ is the same as the set of Nash equilibria of the common interest game with common payoff function P.

weighted potential games and best response potential games

- A weighted potential is a best response potential
- A best response potential may not be a weighted potential; (i) a best response potential puts no restrictions on dominated strategies (will discuss when this matters later); (ii) weights can also depend on actions for best response potential
  - will discuss when the distinction matters later

exact potential games

- function $P : A \to \mathbb{R}$ is an *(exact) potential* for the game $\mathbf{g}$ if, for all $i \in I$ and every $a_{-i} \in A_{-i}$,

$$g_i(a_i, a_{-i}) - g_i(a_i', a_{-i}) = P(a_i, a_{-i}) - P(a_i', a_{-i})$$

  for all a_i and a_i'.
- an (exact) potential is a weighted potential
- an (exact) potential may not be a weighted potential
- a game is an (exact) potential game if it admits an (exact) potential.
- Monderer and Shapley (1996) *GEB*
- if a game $\mathbf{g}$ admits an exact potential, then the set of Nash equilibria of $\mathbf{g}$ is the same as the set of Nash equilibria of the common interest game with common payoff function P.

Potential Maximizing Equilibria

- if game $\mathbf{g}$ is an ordinal potential game, there may be many potential functions establishing it
- game $\mathbf{g}$ has at most one (exact) potential, up to affine transformations
- this means that the following observation is well-defined (i.e., independent of the choice of potential)
- a mixed strategy profile $\alpha \in \underset{i \in I}{\times} \Delta(A_i)$ is a *potential maximizing (Nash) equilibrium* (PME) of (exact) potential game $\mathbf{g}$ if

$$\alpha \in \underset{\alpha' \in \underset{i \in I}{\times} \Delta(A_i)}{\arg\max}\, P(\alpha')$$

Cournot game

- actions are (strictly positive) output choices $A_i = \mathbb{R}_{++}$
- price depends on aggregate output, $Q = a_1 + .. + a_{|I|}$, with inverse demand curve $F : \mathbb{R}_{++} \to \mathbb{R}_{++}$
- note: no restrictions on F
- payoffs are

$$g_i(a) = \left(F\left(a_1 + .. + a_{|I|}\right) - c\right) a_i$$

- thus there is a constant and common marginal cost for each firm/player

ordinal potential

- an ordinal potential function is

$$P\left(a_1, a_2, ..., a_{|I|}\right) = a_1 a_2 ... a_{|I|}\left(F\left(a_1 + .. + a_{|I|}\right) - c\right)$$

- this is *not* exact potential (or best response potential)

linear demand Cournot game with general differentiable cost

- let's add requirement that inverse demand curve be linear, i.e.,

$$F(Q) = d - bQ$$

  for some $d, b > 0$

- but let's generalize assumption about costs: let $c_i : A_i \longrightarrow \mathbb{R}_+$ be arbitrary heterogeneous convex differentiable cost functions
- payoffs are

$$g_i(a) = \left(d - b\left(a_1 + .. + a_{|I|}\right)\right) a_i - c_i(a_i)$$

exact potential

- exact potential function

$$P(a) = d \sum_{i \in I} a_i - b \sum_{i \in I} a_i^2 - b \sum_{1 \leq i < j \leq |I|} a_i a_j - \sum_{i \in I} c_i(a_i)$$

- check best response: a_i solves FOC

$$d - b\mathbb{E}_i\left(\sum_{j \neq i} a_j\right) - 2ba_i = c_i'(a_i)$$

two by two games

- consider two player two action games
- i.e., consider game

	L	R
U	a, b	c, d
D	e, f	g, h

- suppose that there are two strict Nash equilibria
- without loss assume that (U, L) and (D, R) are strict Nash equilibria
  - so $a > e$, $g > c$, $b > d$ and $h > f$
- this game is "best response equivalent" to

	L	R
U	$a - e, b - d$	$0, 0$
D	$0, 0$	$g - c, h - f$

two by two games

- action U is a best response for player 1 if and only if he assigns probability at least $p_1 = \frac{g-c}{(g-c)+(a-e)} \in (0,1)$ to player 2 choosing L
- action L is a best response for player 2 if and only if he assigns probability at least $p_2 = \frac{h-f}{(h-f)+(b-d)} \in (0,1)$ to player 1 choosing U
- so the game is best response equivalent to

	L	R
U	$1-p_1, 1-p_2$	$0, 0$
D	$0, 0$	p_1, p_2

two by two games

- Harsanyi and Selten (1988) said that (U, L) was *risk dominant* if $p_1 + p_2 \leq 1$
- Harsanyi and Selten (1988) had an axiomatic justification for risk dominant selection in two by two games; and a (somewhat baroque and not widely used) "tracing procedure" generalizing it to all games
- risk dominant equilibrium in 2x2 game is potential maximizing equilibrium
- suppose (without loss) that (U, L) is risk dominant

two by two games

- our game is an (exact) potential game with potential

P	L	R
U	$1 - p_1 - p_2$	$-p_1$
D	$-p_2$	0

- so the risk dominant equilibrium (U, L) is the unique potential maximizing equilibrium if $p_1 + p_2 < 1$ If $p_1 + p_2 = 1$, both equilibria are risk dominant and potential maximizing

myopic learning foundations of Nash equilibrium

- does a myopic learning process converge to an equilibrium?
- if the game has multiple equilibria, which gets selected?

myopic learning 1: finite improvement property

- a *path* is a sequence of action profiles $\gamma = (a^0, a^1, a^2, ...)$ such that for all $k \geq 1$, there exists exactly one player i such that $a_i^k \neq a_i^{k-1}$
- path γ is finite if it is a finite sequence
- a^0 is the *initial point* of γ
- if γ is finite, the last action profile is the *terminal point*
- game $\mathbf{g}$ satisfies the *finite improvement property* (FIP) if every improvement path is finite
- **lemma**: every finite ordinal potential game has the FIP

finite improvement property

- ordinal potential not necessary for FIP
- a function P is a *generalized ordinal potential* for $\mathbf{g}$ if, for all $i \in I$ and every $a_{-i} \in A_{-i}$,

  $$g_i(a_i, a_{-i}) - g_i(a_i', a_{-i}) > 0 \Rightarrow P(a_i, a_{-i}) - P(a_i', a_{-i}) > 0$$

  for all a_i and a_i'.
- **lemma**: FIP holds if and only if a game is a generalized ordinal potential game.
- **lemma**: if $g_i(a_i, a_{-i}) \neq g_i(a_i', a_{-i})$ for all i, a_i, a_i' and a_{-i}; and $\mathbf{g}$ satisfies FIP, then $\mathbf{g}$ is an ordinal potential game

myopic learning 2: fictitious play

- a fictitious play process is a sequence of (perhaps mixed) strategy profiles such that each player's mixed strategy in each period is a best response to the average mixed strategy profile in previous periods
- early suggestion for myopic learning process converging to Nash equilibrium
- but there are famous counterexamples in general (giving rise to cycles)
- Monderer and Samet (1989b) **theorem**: fictitious play always converges in an ordinal potential game

myopic learning 3: stochastic stability

- Kandori, Mailath and Rob (1993) *Ecta* and Young (2003) *Ecta*
- players choose best response to past population play but play is subject to mutations or "noise"
- as noise becomes small, converge to an equilibrium
- potential maximizing equilibrium is selected

experiments: "stag hunt" games

- van Huyck, Battalio and Beil (1990)
- actions $A_i = \{1, 2, ..., 7\}$ are interpreted as effort levels
- payoffs are given by

$$g_i(a) = d \min \left\{ a_1, a_2, ..., a_{|I|} \right\} - ba_i + c$$

where $d > b \geq 0$ and

$$d - 7b + c > 0$$

- a potential function is

$$P(a) = d \min \left\{ a_1, a_2, ..., a_{|I|} \right\} - b \sum_{i \in I} a_i$$

- the PME is all play 1 if $d < |I| b$ and PME is all play 7 if $d > |I| b$
- PME is played

guessing game

- same action sets
- payoffs are given by

$$g_i(a) = dM(a) - b(M(a) - a_i)^2 + c$$

  where d, b, c are positive constants and $M(a)$ is the median of $\{a_1, a_2, ..., a_{|I|}\}$
- van Huyck, Battalio and Beil (1991) conducted experiments on this game
- if $M(a)$ were the mean, can show this is a potential game and all play 7 is the PME

experimental question

- how general is PME selection?
- known to occur in symmetric games, less is known about symmetric games

binary-action supermodular games

- many economic and social problems take the form of "coordination games"
- a leading model of "coordination games": binary-action supermodular (**BAS**) game
  - e.g., participate (invest, revolt, short the currency) or not participate (not invest, stay home, sit out the market turmoil)

binary-action supermodular games

- $A_i = \{0, 1\}$: the binary-action set for player i.
  - participate (1) or not participate (0)
- sufficient statistic for payoffs in binary-action game: payoff differences $d_i : 2^{I/\{i\}} \to \mathbb{R}$

$$\begin{aligned} d_i(S) &= g_i(\mathbf{1}_{S\cup\{i\}}, \mathbf{0}_{I/(S\cup\{i\})}) - g_i(\mathbf{1}_S, \mathbf{0}_{I/S}) \\ &= g_i\left(S \cup (\{i\})\right) - g_i\left(S\right) \end{aligned}$$

- will sometimes (as above) write $S \subseteq I$ to represent the strategy profile where S is the set of players choosing action 1
- this reduced form of game is parameterized by $\mathbf{d} = (d_i)_{i\in I}$
- payoffs supermodular in action profiles (strategic complementarities) is the requirement that payoff differences $d_i(S)$ increasing in $S \subseteq I$

many BAS games are (weighted) potential games (and thus BASP games)

1. all two player BAS games are potential games (and the risk dominant equilibrium is potential maximizing equilibria when there are multiple equilibria
2. all symmetric BAS games many player BAS games are potential games
   - in a large population, all participate is the PME if participate is the *Laplacian* action (best response to uniform belief about number of other players' participating)
3. many interesting asymmetric BAS games are potential games (lots of interesting economics here!) and, if not, suitably generalized potential games
4. not all BAS games are potential games ("Does one Soros make a difference?" REStud 2003 Corsetti, Dasupta, Morris and Shin is BAS regime change game that does not have a potentially maximizing equilibrium)

"investment game": a BASP game

- write $n(S)$ for the number of elements of S
- let
$$d_i(S) = h_{n(S)+1} - c_i$$
where h_k is increasing in k
- let c_i be the private cost of investing
  - assume without loss that $c_1 \leq c_2 \leq .... \leq c_{|I|}$

investment game is a BASP game

- all invest is a Nash equilbrium if $h_{|I|} > c_i$ for all i
- no one invest is a Nash equilibrium if $h_1 < c_i$ for all i
- intermediate equilibria may also exist although we can rule these out if we want
- both both extreme equilibria will often co-exist (when $h_I >> h_1$)
- potential function is

$$P(S) = \sum_{k=1}^{n(S)} h_k - \sum_{i \in S} c_i$$

- now "all invest" is potential maximizing among two extreme equilibrium if $P(I) > P(\varnothing)$, i.e.,

$$\sum_{k=1}^{|I|} h_k > \sum_{i \in I} c_i$$

three player experiments?

- let $h_1 = 5$, $h_2 = 10$ and $h_3 = 15$
- assume w.l.log. $c_1 \leq c_2 \leq c_3$
- assume $5 < c_1 \leq c_2 \leq c_3 < 15$
- all invest is PME if $30 > c_1 + c_2 + c_3$

payoff matrices

matrix invest	Invest	Not Invest
Invest	$15 - c_1, 15 - c_2, 15 - c_3$	$10 - c_1, 0, 10 - c_3$
Not Invest	$0, 10 - c_2, 10 - c_3$	$0, 0, 5 - c_3$

matrix not invest]	Invest	Not Invest
Invest	$10 - c_1, 10 - c_2, 0$	$5 - c_1, 0, 0$
Not Invest	$0, 5 - c_2, 0$	0

1 chooses row, 2 chooses column, 3 chooses matrix

examples for experiment: symmetric

- If $c_1 = c_2 = c_3 = 9$,

matrix invest	Invest	Not Invest
Invest	$6, 6, 6$	$1, 0, 1$
Not Invest	$0, 1, 1$	$0, 0, -4$

matrix not invest]	Invest
Invest	$1, 1, 1$
Not Invest	$0, -4, 0$

all invest is *PME*

- If $c_1 = c_2 = c_3 = 11$,

matrix invest	Invest	Not Invest
Invest	$4, 4, 4$	$-1, 0, -1$
Not Invest	$0, -1, -1$	$0, 0, -6$

matrix not invest]	Inves
Invest	$1, 1, 1$
Not Invest	$0, -6$

all invest is *PME*

examples for experiment: asymmetric

- If $c_1 = 6$, $c_2 = 9$ and $c_3 = 14$,

matrix invest	Invest	Not Invest
Invest	$9, 6, 1$	$4, 0, -4$
Not Invest	$0, 1, -4$	$0, 0, -9$

matrix not invest]	Invest
Invest	$4, 1, -4$
Not Invest	$0, -4, 0$

all invest is *PME*

- If $c_1 = 6$, $c_2 = 11$ and $c_3 = 14$,

matrix invest	Invest	Not Invest
Invest	$9, 4, 1$	$4, 0, -4$
Not Invest	$0, -1, -4$	$0, 0, -9$

matrix not invest]	Inves
Invest	$4, -1$
Not Invest	$0, -6$

no investment is *PME*

new experiment

Frank Heinemann (2024) "An experimental test of the global-game selection in coordination games with asymmetric players" *JEBO.*

- not potential games
- but bottom line is that players choose Laplacian actions in symmetric games (where they would be potential maximizing in potential games) but also in asymmetric ones.

Frank's predictions

- If $c_1 = c_2 = c_3 = 9$,

matrix invest	Invest	Not Invest
Invest	$6, 6, 6$	$1, 0, 1$
Not Invest	$0, 1, 1$	$0, 0, -4$

matrix not invest]	Invest
Invest	$1, 1, 1$
Not Invest	$0, -4, 0$

all invest is *PME*, Frank predicts all invest

- If $c_1 = c_2 = c_3 = 11$,

matrix invest	Invest	Not Invest
Invest	$4, 4, 4$	$-1, 0, -1$
Not Invest	$0, -1, -1$	$0, 0, -6$

matrix not invest]	Invest
Invest	$1, 1, 1$
Not Invest	$0, -6$

no investment is *PME*, Frank predicts no investment

examples for experiment: asymmetric

If $c_1 = 6$, $c_2 = 9$ and $c_3 = 14$,

matrix invest	Invest	Not Invest
Invest	$9, 6, 1$	$4, 0, -4$
Not Invest	$0, 1, -4$	$0, 0, -9$

matrix not invest]	Invest	N
Invest	$4, 1, -4$	$-$
Not Invest	$0, -4, 0$	0

all invest is *PME*, Frank predicts all invest

- If $c_1 = 6$, $c_2 = 11$ and $c_3 = 14$,

matrix invest	Invest	Not Invest
Invest	$9, 4, 1$	$4, 0, -4$
Not Invest	$0, -1, -4$	$0, 0, -9$

matrix not invest]	Inves
Invest	$4, -1$
Not Invest	$0, -6$

no investment is *PME*, **Frank predicts no investment**

one last example: team problem

- interpret participation as unobserved effort decision of a team member in a project
- let $\pi(S)$ be the probability of success if S is the set of players who exert effort
- let $b_i > 0$ be a bonus offered to team member i if the project is successful
- let c_i be the private cost of exerting effort for player i
- so payoffs are:

$$g_i(S) = \begin{cases} b_i \pi(S) - c_i, \text{ if } i \in S \\ b_i \pi(S), \text{ if } i \notin S \end{cases}$$

- payoff differences

$$d_i(S) = b_i (\pi(S \cup \{i\}) - \pi(S)) - c_i$$

team problem

- not an exact potential game
- but best response equivalent to game with

$$d_i(S) = \pi(S \cup \{i\}) - \pi(S) - \frac{c_i}{b_i}$$

- thus this is weighted potential game with weights $w_i = b_i$ and

$$P(S) = \pi(S) - \sum_{i \in S} \frac{c_i}{b_i}$$

best response potential games

- Function $P : A \to \mathbb{R}$ is a *best response potential* for the game if, for all $i \in I$ and $\lambda_i \in \Delta\left(A_{-i}\right)$,

$$\arg\max_{a_i \in A_{-i}} \sum \lambda_i\left(a_{-i}\right) g_i\left(a_i, a_{-i}\right) = \arg\max_{a_i \in A_{-i}} \sum \lambda_i\left(a_{-i}\right) P\left(a_i, a_{-i}\right)$$

  for all a_i and a_i'.
- A game is a best response potential game if it admits a best response potential.

better response potential games

- one more kind of potential game
- function $P : A \rightarrow \mathbb{R}$ is a *better response potential* for the game if, for all $i \in I$ and $\lambda_i \in \Delta(A_{-i})$,

  $$\sum \lambda_i (a_{-i}) \left[g_i (a_i, a_{-i}) - g_i \left(a_i', a_{-i} \right) \right] \geq 0 \Leftrightarrow \sum \lambda_i (a_{-i}) \left[P (a_i, a_{-i}\right.$$

  for all a_i and a_i'.
- a game is a best response potential game if it admits a better response potential.

better response potential games

- fix a game **g**; say that the undominated version of **g** is the game where we delete all strictly dominated actions for all players
- a game is a better-response potential game if and only if the undominated version of the game is a weighted potential game
- duality argument: by Farkas' lemma

best response potential games

- now restrict attention to games without strictly dominated strategies
- if two actions best response regions in $\Delta\left(A_{-i}\right)$ touch each other, this imposes a linear algebraic restriction on payoff differences
- but if few regions meet, there are fewer restrictions
- consider games with strategic complementarities and strictly concave-in-own-action payoffs

monotone potential games

- Fix a BAS game
- Let S maximize a potential V
- S is a monotone potential maximizer if

$$d_i(S) \geq P(S \cup \{i\}) - P(S) \geq 0$$

if $i \notin S$ and

$$d_i(S) \leq P(S) - P(S \setminus \{i\}) \leq 0$$

if $i \in S$