Higher-Order Beliefs in Game Theory

Oxford Mini-Course Lecture 2

Stephen Morris

November 2025

higher order beliefs game theory literature and PME

1. higher-order beliefs literature: lightning survey
2. three perturbations selecting the PME (and its generalizations)

higher order beliefs game theory literature and PME

1. higher-order beliefs literature: lightning survey
    1.1 e-mail game (Rubinstein 1989)
    1.2 approximate common knowledge (common **p**-belief) Monderer and Samet (1989)
    1.3 robustness to incomplete information Kajii and Morris (1997)
    1.4 global games
        1.4.1 one dimensional types (Frankel, Morris and Pauzner (2003))
        1.4.2 multidimensional types (Carlsson and van Damme (1993), Oury (2013) and Veiel (2025))
2. three perturbations selecting the PME (and its generalizations)
    2.1 exogenous noise about state (global games)
    2.2 endogenous information about state (information design)
    2.3 endogenous noise about payments (noisy contracts)

complete information game

- set of players $I = \{1, 2, ..., |I|\}$
- for each player i,
  - action set A_i (finite)
  - payoff function $g_i : A \longrightarrow \mathbb{R}$
    - where $A = A_1 \times ... \times A_{|I|}$
- this is a complete information game $\mathbf{g} = (g_1, ..., g_{|I|})$.

p-dominance

- let $\mathbf{p} = \left(p_{1,}p_2, ..., p_{|I|}\right) \in [0,1]^I$

Definition

A (pure strategy) Nash equilibrium a^* of $\mathbf{g}$ is $\mathbf{p}$-dominant if

$$p_i g_i\left(a_i^*, a_{-i}^*\right) + (1-p_i)\, g_i\left(a_i^*, a_{-i}\right) \geq p_i g_i\left(a_i, a_{-i}^*\right) + (1-p_i)\, g_i\left(a_i, a_{-i}\right)$$

for all i, a_i and a_{-i}

- if a^* is $\mathbf{0}$-dominant Nash equilibrium, then a^* is a (weakly) dominant strategy equilibrium
- if a^* is a $\mathbf{1}$-dominant Nash equilibrium, then a^* is a Nash equilibrium
- if a^* is a $\mathbf{p}$-dominant Nash equilibrium for any $\mathbf{p} < \mathbf{1}$, then a^* is a strict Nash equilibrium
- if a^* is a (p_1, p_2)-dominant Nash equilibrium in a 2×2 game with $p_1 + p_2 \leq 1$, then a^* is a risk dominant and potential maximiziing equilibrium

incomplete information game

- fix players and actions as before
- add for each player i,
  - a countable set of types: T_i
  - a payoff function $u_i : A \times T \rightarrow \mathbb{R}$
- common prior on $T = T_1 \times T_2 \times .... \times T_{|I|}$
- incomplete information game: $(T; P, \mathbf{u})$

incomplete information game equilibrium

- i's strategy $\sigma_i : T_i \to \Delta(A_i)$; set of all strategies Σ_i
- interim utility

$$U_i(a_i, \sigma_{-i}|t_i) = \sum_{t_{-i}} P(t_{-i}|t_i)\, u_i((a_i, \sigma_{-i}(t_{-i})), (t_i, t_{-i}))$$

- best responses

$$BR_i(\sigma_{-i}|t_i) = \left\{ a_i \in A_i | U_i(a_i, \sigma_{-i}|t_i) \geq U_i\left(a_i', \sigma_{-i}|t_i\right) \text{ for all } a_i' \right\}$$

- $\sigma \in \Sigma$ is a Bayes Nash equilibrium of $(T; P, \mathbf{u})$ if, for all $i \in I$, all $a_i \in A_i$ and $t_i \in T_i$, $\sigma_i(a_i|t_i) > 0 \Rightarrow a_i \in BR_i(\sigma_{-i}|t_i)$.

equilibrium action distribution

- Any $(T; P, \mathbf{u})$ has at least one BNE
- Action distribution $\xi \in \Delta(A)$ is induced by strategy profile σ of $(T; P, \mathbf{u})$ if
$$\xi(a) = \sum_t P(t)\,\sigma(a|t)$$
- $\xi \in \Delta(A)$ is an *equilibrium action distribution* of $(T; P, \mathbf{u})$ if it is induced by a Bayes Nash equilibrium σ of $(T; P, \mathbf{u})$

elaborations

- given $\mathbf{g}$ and $(T; P, \mathbf{u})$, let

$$T_i^{\mathbf{g}_i} = \left\{ t_i \in T_i \;\; \begin{array}{l} u_i\left(a, t_i, t_{-i}\right) = g_i\left(a\right) \text{ for all } a \in A \text{ and} \\ \text{for all } t_{-i} \in T_{-i} \text{ with } P\left(t_{-i} | t_i\right) \end{array} \right\}$$

  and $T^{\mathbf{g}} = \prod_{i \in I} T_i^{\mathbf{g}_i}$

- $(T; P, \mathbf{u})$ is an ε-elaboration of $\mathbf{g}$ if $P\left[T^{\mathbf{g}}\right] \geq 1 - \varepsilon$

electronic mail game

(Rubinstein 1989) variation

- if the "state" is bad, two players have a dominant strategy to not invest

bad	invest	not invest
invest	$-c$	$-c$
not invest	0	0

where $c > \frac{1}{2}$.

- if the "state" is good, each player has a best response if the other firm invests:

good	invest	not invest
invest	$1-c$	$-c$
not invest	0	0

information structure 1

- with probability ε, the state is bad; with probability $1 - \varepsilon$ the state is good
- if the state is bad, player 1 knows and sends a message telling player 2
- if player 2 receives a message, who sends a confirmation of receipt to player 1
- if player 1 receives a message, who sends a confirmation of receiving the confirmation to player 1
- and so on....
- each message gets lost with probability ε

information structure 2

- let t_1^k be the type of player 1 who has sent k messages
- let t_2^k be the type of player 2 who has received k messages

	t_2^0	t_2^1		t_2^k	t_2^{k+1}	
t_1^0	ε	0	$\cdot$	0	0	$\cdot$
t_1	$\varepsilon\left(1-\varepsilon\right)$	$\varepsilon\left(1-\varepsilon\right)^2$	$\cdot$	0	0	$\cdot$
	$\cdot$	$\cdot$	$\cdot$	$\cdot$	$\cdot$	$\cdot$
t_1^k	0	0	$\cdot$	$\varepsilon\left(1-\varepsilon\right)^{2k}$	0	$\cdot$
t_1^{k+1}	0	0	$\cdot$	$\varepsilon\left(1-\varepsilon\right)^{2k+1}$	$\varepsilon\left(1-\varepsilon\right)^{2k+2}$	$\cdot$
	$\cdot$	$\cdot$	$\cdot$	$\cdot$	$\cdot$	$\cdot$

unique equilibrium

- game has unique equilibrium where both players never invest
- in fact, this is the unique strategy profile surviving iterated deletion of strictly dominated strategies
- argument by induction:

unique equilibrium

- initial observation: if you assign probability $\frac{1}{2-\varepsilon} < \frac{1}{2}$ to the state being good and your co-player investing, you will not invest since the payoff is at most $\frac{1}{2} - c < 0$
- now type t_1^0 of player 1 will not invest because she knows that the state is bad
- type t_2^0 of player 2 will not invest because he knows that the state is bad with probability $\frac{1-\varepsilon}{2-\varepsilon}$
- now argue by induction if that types $t_1^0, \ldots., t_1^k$ of player 1 do not invest and types $t_2^0, \ldots., t_2^k$ of player 2 do not invest.....
  - type t_1^{k+1} of player 1 will never invest because he assigns at least probability $\frac{1-\varepsilon}{2-\varepsilon}$ to player 2 not investing
  - type t_2^{k+1} of player 2 will never invest because she assigns at least probability $\frac{1-\varepsilon}{2-\varepsilon}$ to player 1 not investing
- claim follows by induction

paradox?

- types t_1^k and t_2^k do not invest in spite of have $(k-1)$th level knowledge that the state is good
- for large k, Rubinstein (1989) says
  - there is "approximate common knowledge" that the state is good
  - it is a paradoxical prediction that players do not invest; they surely will in practise
- but maybe the confirmation problem is real?
- experiments mixed....

approximate common knowledge: a "correct" definition?

- high levels of knowledge is not close to common knowledge in terms of game theory. what is the "right" definition of approximate common knowledge?
- Monderer and Samet's (1989) alternative definition of approximate common knowledge (slight generalization):
  - let $\mathbf{p} = \left(p_{1,}p_{2}, ..., p_{|I|}\right) \in [0, 1]^{I}$
  - an event is **p**-believed if everyone assigns it probability at least **p**.....
  - an event is common **p**-belief if it is **p**-believed, it is **p**-believed that it is **p**-believed, and so on.
  - an event is "approximate common knowledge" if it is common **p**-belief with each p_i close to 1

approximate common knowledge: formal definition

- write $B^{\mathbf{p}}(E)$ for the event where the event $E \subseteq T$ is **p**-believed
- iterative defination:

$$C^{\mathbf{p}}(E) = \cap$$

- fixed point characterization (or alternative definition):
  - say that E is **p**-evident if $E \subseteq B^{\mathbf{p}}(E)$
  - $t \in C^{\mathbf{p}}(E)$ if and only if there exists **p**-evident event F such that $t \in F$ and $F \subseteq B^{\mathbf{p}}(E)$

approximate common knowledge

Lemma

Suppose that a^* is a $\mathbf{p}$-dominant Nash equilibrium of $\mathbf{g}$. Any game of incomplete information has a Bayes-Nash equilibrium where a^* is played whenever there is common $\mathbf{p}$-belief of $T^{\mathbf{g}}$.

so this is the correct way of defining being close to common knowledge?.

topologies on information structures

- we would like to know the coarsest topology on information structures that generates continuity of economic outcomes; i.e., in two information structures are close, then for any game and any fixed equilibrium of that game and information structure, there is an approximate equilibrium of that game and the other information structure generating behavior close to the fixed equilibrium.
- **meta-claim**: if two information structures (type spaces) are close if there is approximate common knowledge that their interim beliefs are close
- Monderer and Samet (1996) and Kajii and Morris (1998) prove results of this form for ad hoc classes of information structures
- Bergemann, Morris and Veiel (2025) prove "correct" version of meta-claim

another email game paradox?

- although the probability that payoffs were given by $\mathbf{g}$ was arbitrarily close to 1 (as $\varepsilon \to 0$), there was no equilibrium where (I, I) was played with positive probability.
- but (I, I) was a risk dominated equilibrium
- the argument wouldn't have worked if (I, I) was risk dominant
- more generally, are there some equilibria of the complete information game that are not sensitive to small probability (if large) perturbations?

robustness to incomplete information: definition

- for ξ, ξ', write $\left\|\xi - \xi'\right\| = \max_{a \in A} \left|\xi\left(a\right) - \xi'\left(a\right)\right|$

Definition

Nash equilibrium $\xi \in \Delta\left(A\right)$ is robust to incomplete information in $\mathbf{g}$ if, for any $\delta > 0$, there exists $\overline{\varepsilon} > 0$ such that for all $\varepsilon \leq \overline{\varepsilon}$, and $\varepsilon-$elaboration of $\mathbf{g}$ has an equilibrium action distribution $\xi' \in \Delta\left(A\right)$ such that $\left\|\xi - \xi'\right\| \leq \delta$.

- we have already seen a risk dominated Nash equilibrium that is not robust

robustness to incomplete information: existence

- robust equilibria do not exist in some complete information games (consider two player three action cyclic matching pennies)
- if $\mathbf{g}$ has a unique correlated equilibrium, it is (a Nash equilibrium) that is robust to incomplete information.

critical path result

For any information structure (T, P) and any event $E \subseteq T$

$$P\left[C^{\mathbf{p}}(E)\right] \geq 1 - \frac{1 - \min_i p_i}{1 - \sum_{i \in I} p_i}\left[1 - P(E)\right]$$

Oyama and Takahashi (2020) provide a more understandable proof that Kajii and Morris (1997)

- so if $\sum_{i \in I} p_i < 1$, if $P(E) \approx 1$, then $P\left[C^{\mathbf{p}}(E)\right] \approx 1$

intuition for critical path result

	t_2^0	t_2^1		t_2^k	t_2^{k+1}
t_1^0	ε	0	·	0	0
t_1	$\frac{p_1}{1-p_1}\varepsilon$	$\frac{p_1}{1-p_1}\frac{p_2}{1-p_2}\varepsilon$	·	0	0
	·	·	·	·	·
t_1^k	0	0	·	$\left(\frac{p_1}{1-p_1}\right)^k \left(\frac{p_2}{1-p_2}\right)^k \varepsilon$	0
t_1^{k+1}	0	0	·	$\left(\frac{p_1}{1-p_1}\right)^{k+1} \left(\frac{p_2}{1-p_2}\right)^k \varepsilon$	$\left(\frac{p_1}{1-p_1}\right)^{k+1} \left(\frac{p_2}{1-\right.$
	·	·	·	·	·

adding up probabilities

if $p_1 + p_2 < 1$, then $\frac{p_1}{1-p_1}\frac{p_2}{1-p_2} < 1$ and infinite sum of probabilities is bounded.
if $p_1 + p_2 > 1$, then $\frac{p_1}{1-p_1}\frac{p_2}{1-p_2} > 1$ and infinite sum of probabilities is unbounded.

more sufficient conditions for robustness

- [Ui (2000)] If $\mathbf{g}$ is a potential game, and a is a potential maximizing equilibrium of $\mathbf{g}$, then a is robust to incomplete information [Ui (2000)]
- [Morris and Ui (2005)] If A^* is a generalized potential maximizing action set profile of $\mathbf{g}$, then A^* is robust to incomplete information (encompasses Kajii-Morris and Ui 2000 results)
- and more generalizations....
- these results provide sufficient conditions for robustness, but are not necessary conditions

monotone potential games

- All play 1 is a *monotone potential maximizer* of BAS game $\mathbf{g}$ if there exist a *monotone potential* $P : \{0,1\}^I \to \mathbb{R}$ and weights $w \in \mathbb{R}^I_{++}$ such that for all $i \in I$,

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

  for all $S \subseteq I \backslash \{i\}$; and $P (I) > P(S)$ for all $S \neq I$
- All play 1 is robust to incomplete information in BAS game $\mathbf{g}$ if and only if it is a monotone potential maximizer (Oyama and Takahashi (2020))
- "Does one Soros make a difference..." is not a monotone potential game

one dimensional global games

1. action sets A_i are ordered
2. state space $\Theta = \mathbb{R}$
3. state monotonic payoffs: $u_i(a_i, a_{-i}, \theta) - u_i(a'_i, a_{-i}, \theta)$ is increasing in θ for all i, $a_i > a'_i$ and a_{-i}
4. supermodular payoffs: $u_i(a_i, a_{-i}, \theta) - u_i(a'_i, a_{-i}, \theta)$ is increasing in a_{-i} for all i, $a_i > a'_i$ and θ
5. dominance regions: action $\overline{a}_i$ ($\underline{a}_i$) is dominant for sufficiently large (small) θ,
6. let θ be distributed according to smooth density $g(.)$
7. let players observe signals $x_i = \theta + \sigma\varepsilon_i$, where $\sigma > 0$ and each ε_i is distributed according to density $f_i(.)$
8. add some strong continuity + strictness properties

one-dimensional global games (Frankel, Morris and Pauzner (2003))

1. action sets A_i are ordered
2. state space $\Theta = \mathbb{R}$
3. state monotonic payoffs: $u_i(a_i, a_{-i}, \theta) - u_i(a_i', a_{-i}, \theta)$ is increasing in θ for all i, $a_i > a_i'$ and a_{-i}
4. supermodular payoffs: $u_i(a_i, a_{-i}, \theta) - u_i(a_i', a_{-i}, \theta)$ is increasing in a_{-i} for all i, $a_i > a_i'$ and θ
5. dominance regions: action $\overline{a}_i$ $(\underline{a}_i)$ is dominant for sufficiently large (small) θ,
6. let θ be distributed according to smooth density $g(.)$
7. let players observe signals $x_i = \theta + \sigma\varepsilon_i$, where $\sigma > 0$ and each ε_i is distributed according to density $f_i(.)$
8. add some strong continuity + strictness properties

one-dimensional global games

game is parameterized by level of noise $\sigma > 0$ and $\mathbf{f} = \left(f_1, f_2, ..., f_{|I|}\right)$

- Two important properties:
    1. LIMIT UNIQUENESS: in the limit as $\sigma \rightarrow 0$, there is an essentially unique strategy profile $\left(s_i^{\mathbf{f}}\right)_{i \in I}$ surviving iterated deletion of dominated strategies
    2. NOISE INDEPENDENT SELECTION: the limit equilibrium $\left(s_i^{\mathbf{f}}\right)_{i \in I}$ does not depend on the structure of noise

one-dimensional global games

- limit uniqueness always holds
- noise independent selection does not hold in general
- noise independent selection holds if $u(\cdot, \theta)$ has an equilibrium that is robust to incomplete information, i.e., if $u(\cdot, \theta)$ has a (perhaps generalized) potential maximizing equilibrium
- if there is noise independent selection, we will say that the induced selection is the FMP selection

two key steps in proof

- as noise goes to zero, as if there is a uniform prior
- if not unique, there is a largest and smallest strategy profile
  - show contradiction by translation

applications

- "most" applications are symmetric across players
- in this case, noise independent selection holds

global games and email game

- Rani Spiegler book observes that email game is treated as a curiosity while global games are treated as substantive applied theory applications
- what do we conclude from this?

multi-dimensional global games

- Carllson and van Damme (1993) considered 8 dimensional payoff space (for 2×2 games) and multidimensional noise
- Oury (2013) and Veiel (2025) return to multidimensional global game models
  1. Oury (2013): fixed FMP setting but allowed multidimensional noise
     - limit uniqueness holds only if there is noise independent selection in the one-dimensional game
     - example shows failure of limit uniqueness otherwise
  2. Veiel (2025) limit uniqueness only if there exists generalized ordinal potential in two player games

two perturbations of a complete information game

- Fix a supermodular complete information game $\mathbf{g}$
- Add two perturbations:
  1. Small uncertainty about payoffs: Fix global game payoff function $u(., \theta)$, where $u(., \theta^*) = \mathbf{g}$. Let θ be drawn according to a smooth distribution h_ε which assigns probability $1 - \varepsilon$ to the interval $[\theta^* - \varepsilon, \theta^* + \varepsilon]$
  2. Noisy information: players observe noisy signal $x_i = \theta + \sigma \eta_i$ where $\eta_i \sim f_i(\cdot)$ for small σ
- If a^* is a PME of $\mathbf{g}$, a^* will be played with probability 1 in limit as $\varepsilon \to 0$ and $\sigma \to 0$.
  - also true for generalizations of PME

three interpretations of perturbations

1. both perturbations are exogenous (Carlsson and van Damme (1993))
2. exogenous payoff uncertainty but information structure is chosen by designer (Morris, Oyama and Takahashi (2024)
   - information design with adversarial equilibrium selection
3. both perturbations are endogenous (Halac, Lipnowski and Rappaport 2021)
   - design chooses perturbations to payoffs and noise structure