Interdependent Preferences and Strategic Distinguishability

Dirk Bergemann, Stephen Morris and Satoru Takahashi

Bocconi, Milan
15 March 2011

Questions

1. Economists assume interdependence of preferences for informational and/or psychological reasons.
2. Economists model differences in beliefs as deriving from differences in priors and differences in information.

Questions

1. Economists assume interdependence of preferences for informational and/or psychological reasons.
2. Economists model differences in beliefs as deriving from differences in priors and differences in information.

- What are the operational (observable) content of these modelling choices?

1. Interdependence

- Agent i has type $\theta_i \in [0, 1]$
- Agent i's valuation of an object is $\theta_i + \gamma \sum_{j \neq i} \theta_j$
- Interesting mechanism design problem allocating the object "efficiently"
- Informational story: I have a signal which is more relevant about its private value to me than about its private value to you
- Psychological story: I want to own a painting that everyone (but more especially I) think is pretty
- Can we tell the difference? Do we care?

2. Priors and Information

- Sometime in the fifteen years between 1967 (Harsanyi) and 1982 (Milgrom-Stokey), economists internalized a key distinction between differences in beliefs due to differences in priors and differences due to asymmetric information.
- If you are an expert, and knowing your belief leads me to change my belief, this must be modelled as asymmetric information (even if there is no physical counterpart of your informaiton)
- Candidate operational definition of "information": something you do would lead me to change my action

This Paper: Main Result

- Reports a canonical description of interdependent preference types (universal EU preference hierarchy space)
- Gives an operational meaning to this space:
    - Two types are "strategically indistinguishable" if they have an equilibrium action in common in every "mechanism"
    - We show that two types are strategically indistinguishable if and only if they correspond to the same preference hierarchy

Answers

1. Operational definition of "private values": common certainty that agents' choices do not depend on others' choices (in *some* equilibrium)
2. Without private values, no useful distinction between "beliefs" and "utility" and so
    1. no operational distinction between informational and psychological interdependence
    2. no meaning to priors versus information
3. With private values, state independence separates "beliefs" and "utility"
    1. interdependence is naturally interpreted as informational
    2. all information structures embedded in universal preference hierarchy space

Related Literature 1 (preview)

- Mertens-Zamir (1985) constructed universal of higher order beliefs
  - our universal space formally isomorphic to MZ space, but removes new "redundancy" once payoffs are added
- Abreu-Matsushima (1992) identified a measurability condition necessary for virtual Bayesian implementation
  - strategic distinguishability is a re-writing of the idea of the AM measurability condition

Related Literature 1 (preview)

- Dekel-Fudenberg-Morris (2007) showed that two types have the same rationalizable actions in all games if and only if they have the same Mertens-Zamir type
  - DFM corresponds formally to a special case of this paper with common certainty of vNM indices
- Gul-Pesendorfer (2007) constructed canonical space of interdependent preferences
  - We have incomplete information and static games/solution concepts, so cannot extract counterfactual information contained in GP types

Paper Outline 1: Main Result

1. Take any "Harsanyi" type space (any and all interdependence)
2. Define strategic distinguishability of Harsanyi types
3. Step 1: Remove two kinds of "decision theoretic redundancy" by mapping to "preference type space"
4. Step 2: Remove "strategic redundancy" by mapping to universal preference hierarchy space
5. Main result

Paper Outline 2: Extensions

1. Two types are "strategically equivalent" if they have the same equilibrium actions in every "mechanism"
2. Strategic equivalence strictly more demanding than strategic indistinguishability
3. Many versions of rationalizability depending on what you may believe others actions are correlated with....
4. Universal preference hierarchy characterizes strategic distinguishability for all solution concepts
5. Universal preference hierarchy characterizes strategic equivalence for most permissive version of rationalizability

Outline of Talk

1. Introduction
2. Formal Description of Set up and Main Question
3. Example
4. Universal Preference Hierarchy Construction
5. Main Result
6. Back to Example and Extensions

Environment

An outside observer will see an *environment* consisting of:

- Agents $1, .., I$
- Set of Outcomes Z (finite)
- For each player i, a worst outcome $w_i \in Z$ (relaxation will be discussed later)
- Set of Observable States Θ (general metric space)

Mechanism

A strategic situation or *mechanism* is $\mathcal{M} = \left((A_i)_{i=1}^{I}, g \right)$ where

- each A_i is a finite set of actions available to i
  - $A = A_1 \times \ldots \times A_I$
- an outcome function $g : A \times \Theta \rightarrow \Delta(Z)$

Harsanyi Type Spaces

A Harsanyi type space $\mathcal{T} = \left((T_i, u_i, \nu_i)_{i=1}^{I}, \Omega \right)$ where

- Ω is a set of unobservable states
- each agent i is characterized by
  - a set of types T_i
  - an (interdependent) vNM utility index, $u_i : Z \times T \times \Theta \times \Omega \to \mathbb{R}_+$
  - beliefs $\nu_i : T_i \to \Delta (T_{-i} \times \Theta \times \Omega)$
- respecting worst outcome:

$$u_i (z, t, \theta, \omega) \geq u_i (w_i, t, \theta, \omega)$$

for all i, z, t, θ and ω.

Equilibrium

- A pair $(\mathcal{T}, \mathcal{M})$ is a game of incomplete information
- A behavioral strategy for agent i is $\sigma_i : T_i \to \Delta(A_i)$
- A strategy profile $\sigma = (\sigma_i)_{i=1}^{I}$ is an equilibrium if, for each i and t_i, $\sigma_i(a_i|t_i) > 0$ only if a_i maximizes

$$\int_{T_{-i}\times\Theta\times\Omega} u_i(g((a_i, \sigma_{-i}(t_{-i})), \theta), (t_i, t_{-i}), \theta, \omega) d\nu_i(t_i)$$

- Equilibrium may fail to exist on whole type space but exist on a belief-closed subset.
- $E_i(t_i, \mathcal{T}, \mathcal{M})$: set of actions type t_i may play in some equilibrium on a belief-closed subset of the type space

Defining Strategic Distinguishability

DEFINITION. Types t_i (in $\mathcal{T}$) and t_i' in $(\mathcal{T}')$, are *strategically indistinguishable* if, for every mechanism $\mathcal{M}$, there exists some action that can be chosen by both types, i.e. for every $\mathcal{M}$,

$$E_i(t_i, \mathcal{T}, \mathcal{M}) \cap E_i(t_i', \mathcal{T}', \mathcal{M}) \neq \emptyset$$

DEFINITION. Type t_i and t_i' are *strategically distinguishable* if there exists a mechanism in which no action can be chosen by both types, i.e., for some $\mathcal{M}^*$

$$E_i(t_i, \mathcal{T}, \mathcal{M}^*) \cap E_i(t_i', \mathcal{T}', \mathcal{M}^*) = \emptyset$$

- Our main result will be a characterization of strategic (in)distinguishability.

Example 1: Unobservable States and Outcomes

- Two detectives, 1 and 2
- Three equally likely unobservable states $\Omega = \{I, M, A\}$
  - state I, suspect innocent
  - state M, suspect committed crime in morning
  - state A, suspect committed crime in afternoon
- Three outcomes, $Z = \{N, C, A\}$
  - "no verdict" (N)
  - "conviction" (C)
  - "acquital" (A)

Example 1: Signals and Beliefs

- Each detective i observes alibi $t_i \in \{m, a\}$
- $T_1 = T_2 = \{m, a\}$
- if innocent, each signal equally likely
- if guilty, signal not "equal to" state
- no observable states ($\Theta = \{\theta_0\}$)
- (adding a little asymmetry) if suspect committed crime in the morning, $\varepsilon > 0$ chance detective 2 misremembers the alibi as morning

Example 1: Signals and Beliefs

- beliefs on $T_1 \times T_2 \times \Omega$ consistent with common prior:

$\omega = I$:

$t_1 \backslash t_2$	m	a
m	$\frac{1}{12}$	$\frac{1}{12}$
a	$\frac{1}{12}$	$\frac{1}{12}$

$\omega = M$:

$t_1 \backslash t_2$	m	a
m	0	0
a	$\frac{\varepsilon}{3}$	$\frac{1-\varepsilon}{3}$

$\omega = A$:

$t_1 \backslash t_2$	m	a
m	$\frac{1}{3}$	0
a	0	0

Example 1: Payoffs

- Correct verdict gives payoff of 1,
- Incorrect or no verdict gives payoff of 0

$$u_i\left(z,\left(t_1,t_2\right),\omega\right)=\begin{cases}1, \text{ if } \left(z,\omega\right)=\left(C,M\right),\left(C,A\right),\left(A,I\right)\\ 0, \text{ otherwise}\end{cases};$$

Removing Decision Theoretic Redundancy 1: Integrating Out Unobserved States

- $T_1 = T_2 = \{l, h\}$;
- beliefs on $T_1 \times T_2$ consistent with common prior:

$t_1 \backslash t_2$	m	a
m	$\frac{5}{12}$	$\frac{1}{12}$
a	$\frac{1+4\varepsilon}{12}$	$\frac{5-4\varepsilon}{12}$

- (expected) payoffs

$u\left(C \middle\| \left(t_1, t_2\right)\right)$	m	a
m	$\frac{4}{5}$	0
a	$\frac{4\varepsilon}{1+4\varepsilon}$	$\frac{4-4\varepsilon}{5-4\varepsilon}$

Behavioral Interpretation

- types m or a correspond to different backgrounds
- conviction is more fun if like minded

Removing Decision Theoretic Redundancy 2: State Dependent Expected Utility

- $T_1 = T_2 = \{l, h\}$;
- beliefs on $T_1 \times T_2$ consistent with common prior:

$t_1 \backslash t_2$	m	a
m	$\frac{1}{4}$	$\frac{1}{4}$
a	$\frac{1}{4}$	$\frac{1}{4}$

- (expected) payoffs

$u((C, A) \mid (t_1, t_2))$	m	a
m	$4, 1$	$0, 1$
a	$4\varepsilon, 1$	$4 - 4\varepsilon, 1$

Dealing with DT Redundancies: Preference Type Space

- Identify each type of detective 1 with an EU preference on AA lotteries contingent on detective 2's type, $f : T_2 \rightarrow \Delta(Z)$
- Type m of agent 1 has preference $f \succeq f'$ if and only if

$$\begin{aligned} & 4f(m)(C) + f(m)(A) + f(a)(A) \\ \geq \quad & 4f'(m)(C) + f'(m)(A) + f'(a)(A) \end{aligned}$$

Canonical Representation of Types

- "First level" observation for detective 2: unconditional preferences ($=$ marginal rate of substitution of acquital for conviction) is $2(1+\varepsilon)$ for type m and $2(1-\varepsilon)$ for type a.
- "First level" observation for detective 1: unconditional preferences ($=$ marginal rate of substitution of acquital for conviction) is 2 for both types, so cannot be distinguished.

Canonical Representation of Types

- "Second level" observation for detective 1: willingness to "pay" (in units of unconditional prob of acquital) for conviction/acquital is
  - conditional on detective 1 being type m: $2/\frac{1}{2}$
  - conditional on detective 1 being type a: $0/\frac{1}{2}$
- Our universal preference hierarchy space is the natural formalization of this

Strategic Redundancy

- Suppose $\varepsilon = 0$, so we have type space
- beliefs on $T_1 \times T_2$ consistent with common prior:

$t_1 \backslash t_2$	m	a
m	$\frac{1}{4}$	$\frac{1}{4}$
a	$\frac{1}{4}$	$\frac{1}{4}$

- (expected) payoffs

$u((C, A) \mid (t_1, t_2))$	m	a
m	$4, 1$	$0, 1$
a	$0, 1$	$4, 1$

Strategic Redundancy

- each of the two types m and a of each player is equivalent to complete information type with common certainty of mrs of 2.

$t_1 \backslash t_2$	$*$
$*$	1

- (expected) payoffs

$u\left((C, A) \mid (t_1, t_2)\right)$	$*$
$*$	$2, 1$

Strategic Redundancy

- consistent with our main result as there will always be a "pooling equilibrium" where types m and a behave as the complete information type
- strategic redundancy analogous to (but different from) the redundancy of Mertens and Zamir (1985) and Dekel, Fudenberg and Morris (2007).
- but suggests more demanding notion of strategic equivalence, we will return to this later

Anscombe-Aumann Acts

- Z: finite set of outcomes
- $f : X \to \Delta(Z)$: measurable function (Anscombe-Aumann act)
- $F(X)$: set of all acts over X

State-Dependent EU Preferences

- X finite
- State dependent EU representation: $\nu \in \Delta(X)$, $u_x \in \mathbb{R}^Z$

$$f \succsim f' \Leftrightarrow \sum_{x \in X, z \in Z} f(x)(z)\, u_x(z)\, \nu(x) \geq \sum_{x \in X, z \in Z} f'(x)(z)\, u_x(z)\, \nu($$

- letting $w \in Z$ be worst outcome, normalize for each x, $u_x(w) = 0$ and each $u_x \in \Delta(Z / \{w\})$
- define $\mu \in \Delta(X \times Z / \{w\})$ by $\mu(x, z) = u_x(z)\nu(x)$:

$$f \succsim f' \Leftrightarrow \sum_{x \in X, z \in Z} f(x)(z)\, \mu(x, z) \geq \sum_{x \in X, z \in Z} f'(x)(z)\, \mu(x, z)$$

Worst Outcome State-Dependent EU Preferences

- $P_w(X)$: set of all binary relations $\succsim$ over $F(X)$ that are represented by $\mu \in \Delta(X \times Z / \{w\})$:

$$f \succsim f' \Leftrightarrow \int f(x)(z) d\mu(x,z) \geq \int f'(x)(z) d\mu(x,z).$$

- Anscombe-Aumann's axiomatization for state-dependent EU, replacing monotonicity with worst outcome property, i.e., $z \succeq w$ for all z.
- We will be imposing common knowledge that (for any X) i has preferences in $P_i(X) \equiv P_{w_i}(X)$

Step 1: Harsanyi Type Spaces to Preference Type Spaces

- Preference Type Space $\mathcal{T} = (T_i, \pi_i)_{i \in \mathcal{I}}$
  - T_i: measurable space of player i's types
  - $\pi_i : T_i \to P_i(\Theta \times T_{-i})$: measurable function that maps each type to his interdependent preference

Step 1: Harsanyi Type Spaces to Preference Type Spaces

- Preference Type Space $\mathcal{T} = (T_i, \pi_i)_{i \in \mathcal{I}}$
- Natural mapping from Harsanyi Type Space into a Preference Type Space (removes "decision theoretic redundancy"), replacing $(\nu_i(t_i), u_i(t_i))$ with $\pi_i(t_i)$, where for acts

$$f, f' : \Theta \times \Omega \times T_{-i} \to \Delta(Z)$$

$$f \;\; \pi_i(t_i) \;\; f' \Leftrightarrow$$

$$\begin{aligned} & \int_{T_{-i} \times \Theta \times \Omega} u_i(g((a_i, f(\theta, \omega, t_{-i})), \theta), (t_i, t_{-i}), \theta, \omega) d\nu_i(t_i) \\ \geq & \int_{T_{-i} \times \Theta \times \Omega} u_i(g((a_i, f'(\theta, \omega, t_{-i})), \theta), (t_i, t_{-i}), \theta, \omega) d\nu_i(t_i) \end{aligned}$$

Induced Preferences, Marginal Preferences

- $\varphi : X \to Y$ *induces* $\varphi^P : P(X) \to P(Y)$ by

$$\succsim \in P_i(X),\ f\ \phi^P(\succsim)\ f' \Leftrightarrow f \circ \varphi \succsim f' \circ \varphi.$$

- $proj_X : X \times Y \to X$ induces

$$marg_X = (proj_X)^P : P_i(X \times Y) \to P_i(X).$$

- $\text{marg}_X \succsim$ is the restriction of $\succsim$ to acts that are independent of the Y coordinate.

Step 2: Preference Types Spaces to Hierarchies of Higher Order Preferences

- For simplicity, state for $I = 2$
- each $\mathcal{T} = (T_i, \pi_i)_{i=1,2}$ and $t_i \in T_i$,

$$
\begin{aligned}
\hat{\pi}_{i,1}(t_i) &= \operatorname{marg}_{\Theta}\pi_i(t_i) \in P_i(\Theta), \\
\hat{\pi}_{i,2}(t_i) &= (\mathrm{id}_{\Theta} \times \hat{\pi}_{-i,1})^P(\pi_i(t_i)) \in P_i(\Theta \times P_{-i}(\Theta)), \\
\hat{\pi}_{i,3}(t_i) &= (\mathrm{id}_{\Theta} \times (\hat{\pi}_{-i,1}, \hat{\pi}_{-i,2}))^P(\pi_i(t_i)) \\
&\in P_i(\Theta \times P_{-i}(\Theta) \times P_{-i}(\Theta \times P_i(\Theta))), \\
&\vdots \\
\hat{\pi}_{i,n}(t_i) &= (\mathrm{id}_{\Theta} \times (\hat{\pi}_{-i,1}, \ldots, \hat{\pi}_{-i,n-1}))^P(\pi_i(t_i)), \\
&\vdots
\end{aligned}
$$

Construction of Hierarchies

- For each $\mathcal{T} = (T_i, \pi_i)_{i \in \mathcal{I}}$, $i \in \mathcal{I}$, and $t_i \in T_i$,

$$\begin{aligned}
\hat{\pi}_{i,1}(t_i) &= \text{marg}_{\Theta} \pi_i(t_i) \in P_i(\Theta), \\
\hat{\pi}_{i,2}(t_i) &= (\mathrm{id}_{\Theta} \times \hat{\pi}_{-i,1})^P (\pi_i(t_i)) \in P_i(\Theta \times P_{-i}(\Theta)), \\
&\vdots \\
\hat{\pi}_{i,n}(t_i) &= (\mathrm{id}_{\Theta} \times (\hat{\pi}_{-i,1}, \dots, \hat{\pi}_{-i,n-1}))^P (\pi_i(t_i)), \\
&\vdots
\end{aligned}$$

- $\hat{\pi}_{i,n}(t_i)$: the *n-th order preference of* t_i.
- $\hat{\pi}_i(t_i) = (\hat{\pi}_{i,1}(t_i), \hat{\pi}_{i,2}(t_i), \dots)$: the *preference hierarchy of* t_i.

The Universal Type Space

- write T_i^* for the set of all preference hierarchies for agent i that can arise from type spaces (satisfies a "coherence" condition)
- back to $I \geq 2$

PROPOSITION: For each agent, there is a natural preference preserving isomorphism $\pi_i^* : T_i^* \to P_i(T_{-i}^* \times \Theta)$

- $\mathcal{T}^* = (T_i^*, \pi_i^*)_{i=1}^I$: the *universal type space*.

Compare

- Epstein-Wang 96:
  - universal preference hierarchy without independence (expected utility) but with monotonicity
- Di Tillio 08:
  - universal preference hierarchy without independence or monotonicity but restricted to finite preferences

Main Result

DEFINITION. Types t_i and t_i' are *strategically indistinguishable* if, for every mechanism $\mathcal{M}$, there exists some action that can be chosen by both types, i.e. for every $\mathcal{M}$,

$$E_i(t_i, \mathcal{T}, \mathcal{M}) \cap E_i(t_i', \mathcal{T}', \mathcal{M}) \neq \varnothing$$

THEOREM 1 (Equilibrium Strategic Distinguishability). Two countable types are strategically indistinguishable if and only if they are higher order preference equivalent

$$E_i(t_i, \mathcal{T}, \mathcal{M}) \cap E_i(t_i', \mathcal{T}', \mathcal{M}) \neq \varnothing \text{ for all } \mathcal{M} \Leftrightarrow \widehat{\pi}_i\left(t_i\right) = \widehat{\pi}_i\left(t_i'\right)$$

Main Result Proof

THEOREM 1 (Equilibrium Strategic Distinguishability). Two countable types are strategically indistinguishable if and only if they are higher order preference equivalent

$$E_i(t_i, \mathcal{T}, \mathcal{M}) \cap E_i(t_i', \mathcal{T}', \mathcal{M}) \neq \varnothing \text{ for all } \mathcal{M} \Leftrightarrow \widehat{\pi}_i\left(t_i\right) = \widehat{\pi}_i\left(t_i'\right)$$

PROOF. "If" Find "pooling" equilibria where types with same higher order preferences behave the same.
"Only If" Construct a game where agents report higher order preference types.

Proof

Construct a game where agents report higher order preference types.

PROPOSITION. For every $\varepsilon > 0$, there exists a mechanism $\mathcal{M}$ such that

$$d_i^*\left(\widehat{\pi}_i\left(t_i\right), \widehat{\pi}_i\left(t_i'\right)\right) > \varepsilon \Rightarrow E_i(t_i, \mathcal{T}, \mathcal{M}) \cap E_i(t_i', \mathcal{T}', \mathcal{M}) = \varnothing$$

Issues in the Proof of Sufficient Condition

- compare Abreu-Matsushima 92, DFM 07, BM 09 and this paper
- all will construct canonical mechanism with players reporting 1st level preferences/beliefs, 2nd level preferences/beliefs, etc...
- for each player i and each $k = 1, 2, ...,$ there will be (with some positive probability) a lottery y_{ik} chosen that depends on kth level report of player i and the $(k - 1)$th reports of players other than i
- this should give player i an incentive to report his kth level preferences/beliefs correctly if he thinks others are reporting their $(k - 1)$th level preferences/beliefs correctly.

Issues in the Proof of Sufficient Condition

- all will construct canonical mechanism with players reporting 1st level preferences/beliefs, 2nd level preferences/beliefs, etc...
- for each player i and each $k = 1, 2, \ldots$, there will be (with some positive probability) a lottery y_{ik} chosen that depends on kth level report of player i and the $(k-1)$th reports of players other than i
- this should give player i an incentive to report his kth level preferences/beliefs correctly if he thinks others are reporting their $(k-1)$th level preferences/beliefs correctly.
- key problem: ensuring that player i does not have incentive to mis-report his kth level preferences/beliefs in order to manipulate $y_{j,k+1}$ for $j \neq i$

Issues in the Proof

- key problem: ensuring that player i does not have incentive to mis-report his kth level preferences/beliefs in order to manipulate $y_{j,k+1}$ for $j \neq i$
- Abreu-Matsushima 92: exploit finiteness of types
- BM 09: exploit finiteness of "payoff types"
- DFM 07 exploit private goods (agent i is indifferent about $y_{j,k+1}$)
- this paper: worst outcome restriction gives compactness required for continuity argument, care in order of limits

Motivating Questions

- A type's preference is monotonic if conditional preferences on lotteries in universal preference hierarchy space are equal to unconditional preferences on lotteries.
- A type is monotonic if it belongs to a preference-closed subspace of the universal preference space where all types' preferences are monotonic
- Monotonicity implies state independent expected utility representation and meaningful "beliefs"
- Common prior can be characterized in terms of those beliefs

Strategic Redundancy Example

- beliefs on $T_1 \times T_2$ consistent with common prior:

$t_1 \backslash t_2$	m	a
m	$\frac{1}{4}$	$\frac{1}{4}$
a	$\frac{1}{4}$	$\frac{1}{4}$

- (expected) payoffs

$u((C,A) \mid (t_1,t_2))$	m	a
m	$4,1$	$0,1$
a	$0,1$	$4,1$

Strategic Redundancy Example

- each of the two types m and a of each detective is equivalent to complete information type with common certainty of mrs of 2.

$t_1 \backslash t_2$	$*$
$*$	1

- (expected) payoffs

$u((C, A) \mid (t_1, t_2))$	$*$
$*$	2, 1

- Strategic redundancy analogous to (but different from) the redundancy of Mertens and Zamir (1985) and Dekel, Fudenberg and Morris (2007).

Strategic Non-Equivalence

- While types m, a and $*$ are strategically indistinguishable, it is easy construct mechanisms where they have different sets of equilibrium actions:

	m	a	opt out
m	$\delta : C$	$\delta : A$	C
a	$\delta : A$	$\delta : C$	C
opt out	C	C	C

- if $\frac{4}{5} < \delta < 1$,
  - opting out is an equilibrium action for all types
  - there is strict "truth-telling" eq. for "redundant" types, but opt out is unique equilibrium action for CI type

Strategic Equivalence

DEFINITION. Types t_i and t_i' are *strategically indistinguishable* if, for every mechanism $\mathcal{M}$, there exists some action that can be chosen by both types, i.e. for every $\mathcal{M}$,

$$E_i(t_i, \mathcal{T}, \mathcal{M}) \cap E_i(t_i', \mathcal{T}', \mathcal{M}) \neq \emptyset$$

DEFINITION. Types t_i and t_i' are *strategically equivalent* if, for every mechanism $\mathcal{M}$, they have the same equilibrium actions, i.e. for every $\mathcal{M}$,

$$E_i(t_i, \mathcal{T}, \mathcal{M}) = E_i(t_i', \mathcal{T}', \mathcal{M})$$

- In above mechanism, types are strategically indistinguishable but not strategically equivalent.

Rationalizability

- "opt out" remains also the unique interim correlated rationalizable (ICR) action (Dekel, Fudenburg and Morris (2007)) for the complete information type
- but *m* and *a* are "interim preference correlated rationalizable" for complete information type..... if we allowed "complete information type" detective 1 with unconditional mrs 2 to believe that detective 2's action is appropriately correlated with suspect's guilt.

Rest of the Paper in Words

1. Introduce various notions of rationalizability for this setting: most permissive is "interim preference correlated rationalizability"
2. Can discuss strategic distinguishability and strategic equivalence for different solution concepts, i.e., equilibrium and versions of rationalizability
3. Preference hierarchy space characterizes strategic distinguishability for equilibrium and all versions of rationalizability
4. Preference hierarchy space characterizes strategic equivalence for interim preference correlated rationalizability only

Special Case: Common Certainty of "Payoffs" (vN-M Indices)

- Universal preference hierarchy reduces to Mertens-Zamir belief hierarchy
- As Corollary of our results in this paper: the following are equivalent
  - two types have the same MZ belief hierarchy
  - two types are strategically distinguishable (under "any" solution concept)
  - two types are strategically equivalent under "interim correlated rationalizability"
- These results were shown / easily implied by Dekel-Fudenberg-Morris 06+07
- Failure of universal preference hierarchy to characterize equilibrium strategic equivalence analogous to failure of

Conclusion

- There were strong maintained assumptions: common certainty of expected utility maximization with worst outcome
- Conceptual framework for thinking about strategic revealed preference
- Natural language for expressing operation characteristics of agents' types

Relaxing Worst Outcome Property

- Worst Outcome Property delivered two properties:
  1. Impossibility of complete indifference
  2. Compactness of Preferences
- Alternative ways of ensuring these properties exist....

Bounded Preferences

- $\succeq \in P(X)$ is ε-bounded if there exist z and z' with
  1. $z \succeq z'$
  2. for every $f, f' \in F(X)$,

$$(1-\varepsilon)z + \varepsilon f \succ (1-\varepsilon)z' + \varepsilon f'$$

- All preferences are ε-bounded for some $\varepsilon > 0$: writing $P_\varepsilon(X)$ for ε-bounded preferences,

$$P(X) = \bigcup_{\varepsilon>0} P_\varepsilon(X)$$

Bounded Preferences

- for each $\varepsilon > 0$, can construct universal space of ε-bounded preferences
- ε is uniform on that spaces
- can work with union of such universal spaces...
- can define rationalizability respected ε-bounded property

Rationalizability: Four Reasons to Think about...

1. Countability restriction not required for existence....
2. Natural solution concept in absence of common prior assumption
3. Will help understand strategic equivalence...
4. Will help understand relation to the literature...

Review: Environment

An outside observer will see an *environment* consisting of:

- Agents 1, .., I
- Set of Outcomes Z (finite)
- For each player i, a worst outcome $w_i \in Z$
- Set of Observable States Θ (general metric space)

Review: Mechanism

A strategic situation or *mechanism* is $\mathcal{M} = \left((A_i)_{i=1}^{I}, g \right)$ where

- each A_i is a finite set of actions available to i
  - $A = A_1 \times ... \times A_I$
- an outcome function $g : A \times \Theta \to \Delta(Z)$

Review: Harsanyi Type Spaces

A Harsanyi type space $\mathcal{T} = \left((T_i, u_i, \nu_i)_{i=1}^{I}, \Omega \right)$ where

- Ω is a set of unobservable states
- each agent i is characterized by
  - a set of types T_i
  - an (interdependent) vNM utility index, $u_i : Z \times T \times \Theta \times \Omega \to \mathbb{R}_+$
  - beliefs $\nu_i : T_i \to \Delta\left(T_{-i} \times \Theta \times \Omega\right)$
- respecting worst outcome:

$$u_i\left(z, t, \theta, \omega\right) \geq u_i\left(w_i, t, \theta, \omega\right)$$

for all i, z, t, θ and ω.

Interim Preference Correlated Rationalizability

- A pair $(\mathcal{T}, \mathcal{M})$ is a game of incomplete information
- $R_{i,0}(t_i, \mathcal{T}, \mathcal{M}) = A_i$
- $a_i \in R_{i,n+1}(t_i, \mathcal{T}, \mathcal{M})$ if.....

Rationalizability

- $a_i \in R_{i,n+1}(t_i, \mathcal{T}, \mathcal{M})$ if.....
- there exists $\succeq \in P_i(A_{-i} \times T_{-i} \times \Theta \times \Omega)$ such that
  1. $\{(a_{-i}, t_{-i}, \theta, \omega) \,|\, a_j \notin R_{j,n}(t_j, \mathcal{T}, \mathcal{M}) \text{ for some } j\}$ is null
  2. $\text{marg}_{T_{-i} \times \Theta \times \Omega} \succeq = \pi_i(t_i)$
  3. $g(\cdot | (a_i, a_{-i}), \theta) \succeq g(\cdot | (a_i', a_{-i}), \theta)$ for all a_i'
- $R_i(t_i, \mathcal{T}, \mathcal{M}) = \bigcap_{n \geq 1} R_{i,n}(t_i, \mathcal{T}, \mathcal{M})$

Strategic Indistinguishability

DEFINITION. Types t_i and t_i' in *rationalizable strategically indistinguishable* if, in every mechanism, there exists a rationalizable action that can be chosen by both types, i.e. for every $\mathcal{M}$,

$$R_i(t_i, \mathcal{T}, \mathcal{M}) \cap R_i(t_i', \mathcal{T}', \mathcal{M}) \neq \varnothing$$

THEOREM 2 (Rationalizable Strategic Indistinguishability). Two types are rationalizable strategically indistinguishable if and only if they are higher order preference equivalent

$$R_i(t_i, \mathcal{T}, \mathcal{M}) \cap R_i(t_i', \mathcal{T}', \mathcal{M}) \neq \varnothing \text{ for all } \mathcal{M} \Leftrightarrow \widehat{\pi}_i\left(t_i\right) = \widehat{\pi}_i\left(t_i'\right)$$

Strategic Indistinguishability

THEOREM 2 (Rationalizable Strategic Indistinguishability). Two types are rationalizable strategically indistinguishable if and only if they are higher order preference equivalent

$$R_i(t_i, \mathcal{T}, \mathcal{M}) \cap R_i(t_i', \mathcal{T}', \mathcal{M}) \neq \varnothing \text{ for all } \mathcal{M} \Leftrightarrow \widehat{\pi}_i(t_i) = \widehat{\pi}_i(t_i')$$

COROLLARY. Higher order preference equivalence characterizes strategically distinguishability for any solution concept that refines R_i and coarsens E_i.

Proof

Pooling argument:

$$\begin{aligned} \widehat{\pi}_i\left(t_i\right) &= \widehat{\pi}_i\left(t_i'\right) \\ &\Rightarrow E_i(t_i, \mathcal{T}, \mathcal{M}) \cap E_i(t_i', \mathcal{T}', \mathcal{M}) \neq \varnothing \\ &\Rightarrow R_i(t_i, \mathcal{T}, \mathcal{M}) \cap R_i(t_i', \mathcal{T}', \mathcal{M}) \neq \varnothing \end{aligned}$$

Proof

Converse:

$$\begin{aligned} \widehat{\pi}_i(t_i) &\neq \widehat{\pi}_i(t_i') \\ &\Rightarrow R_i(t_i, \mathcal{T}, \mathcal{M}) \cap R_i(t_i', \mathcal{T}', \mathcal{M}) = \varnothing \\ &\Rightarrow E_i(t_i, \mathcal{T}, \mathcal{M}) \cap E_i(t_i', \mathcal{T}', \mathcal{M}) = \varnothing \end{aligned}$$

Proof

PROPOSITION. For every $\varepsilon > 0$, there exists a mechanism $\mathcal{M}$ such that

$$d_i^*\left(\widehat{\pi}_i\left(t_i\right), \widehat{\pi}_i\left(t_i'\right)\right) > \varepsilon \Rightarrow R_i(t_i, \mathcal{T}, \mathcal{M}) \cap R_i(t_i', \mathcal{T}', \mathcal{M}) = \varnothing$$

Ex Post Restrictions on Preferences

- $U_i : \Theta \longrightarrow 2^{\Delta(Z/\{w_i\})}$; each $U_i(\theta)$ linear independent; $U = (U_i)_{i=1}^{I}$
- Interpretation: even contingent on others' actions and types, ex post preferences must be representable by $conv\,(U_i(\theta))$
- Harsanyi type space is U-consistent if all types' preferences, conditional on other θ and t_{-i}, are consistent with $U_i(\theta)$.

Two Special Cases of U

- $\underline{U}_i(\theta)$ is a singleton
  - this gives "interim correlated rationalizability" of DFM
- $\overline{U}_i(\theta)$ is $\Delta(Z / w_i)$
  - this gives our earlier definition of "interim preference correlated rationalizability" (IPCR)
  - very permissive

Rationalizability

Ex Post Preference Restriction U

- A pair $(\mathcal{T}, \mathcal{M})$ is a game of incomplete information
- $R^U_{i,0}(t_i, \mathcal{T}, \mathcal{M}) = A_i$
- $a_i \in R^U_{i,n+1}(t_i, \mathcal{T}, \mathcal{M})$ if.....

Rationalizability

- $a_i \in R^U_{i,n+1}(t_i, \mathcal{T}, \mathcal{M})$ if.....
- there exists $\succeq \in P_i(A_{-i} \times T_{-i} \times \Theta \times \Omega)$ such that
  1. $\left\{(a_{-i}, t_{-i}, \theta, \omega) \,|\, a_j \notin R^U_{j,n}(t_j, \mathcal{T}, \mathcal{M}) \text{ for some } j\right\}$ is null
  2. conditional preferences $\succeq_{a_{-i}, t_{-i}, \theta, \omega}$ has representation in $\text{conv} U_i(\theta)$
  3. $\text{marg}_{T_{-i} \times \Theta \times \Omega} \succeq = \pi_i(t_i)$
  4. $g(\cdot | (a_i, a_{-i}), \theta) \succeq g(\cdot | (a_i', a_{-i}), \theta)$ for all a_i'

Strategic Equivalence

DEFINITION. Types t_i and t_i' in are R^U- *strategically indistinguishable* if, in every mechanism, there exists a U-rationalizable action that can be chosen by both types, i.e. for every $\mathcal{M}$,

$$R_i^U(t_i, \mathcal{T}, \mathcal{M}) \cap R_i^U(t_i', \mathcal{T}', \mathcal{M}) \neq \emptyset$$

DEFINITION. Type t_i and t_i' are R^U- *strategically equivalent* if, for every mechanism $\mathcal{M}$, the same actions are U-rationalizable, i.e., for every $\mathcal{M}$

$$R_i^U(t_i, \mathcal{T}, \mathcal{M}) = R_i^U(t_i', \mathcal{T}', \mathcal{M})$$

Strategic Equivalence Result

Fix U.

THEOREM 3 (Strategic Equivalence). If t_i and t_i' are U-consistent, then

$$\widehat{\pi}_i\left(t_i\right) = \widehat{\pi}_i\left(t_i'\right) \Leftrightarrow R_i^U(t_i, \mathcal{T}, \mathcal{M}) = R_i^U(t_i', \mathcal{T}', \mathcal{M})$$

Idea of Proof: extra (U-consistent) detail in type space (beyond higher order preference types) can be replicated within solution concept.

Strategic Equivalence Result

THEOREM 3 (Strategic Equivalence). If t_i and t_i' are U-consistent, then

$$\widehat{\pi}_i\left(t_i\right)=\widehat{\pi}_i\left(t_i'\right) \Leftrightarrow R_i^U(t_i,\mathcal{T},\mathcal{M})=R_i^U(t_i',\mathcal{T}',\mathcal{M})$$

COROLLARY. Higher order preference equivalence characterizes IPCR strategic equivalence:

$$\widehat{\pi}_i\left(t_i\right)=\widehat{\pi}_i\left(t_i'\right) \Leftrightarrow R_i(t_i,\mathcal{T},\mathcal{M})=R_i(t_i',\mathcal{T}',\mathcal{M})$$

Results Summary in Words

The following statements are equivalent....

1. Types t_i and t_i' have the same higher order preference type
2. Types t_i and t_i' are IPCR strategically equivalent
3. Types t_i and t_i' are IPCR strategically indistinguishable
4. Types t_i and t_i' are (equilibrium) strategically indistinguishable

...but not equivalent to

- Types t_i and t_i' are equilibrium strategically equivalent

Common Certainty of vN-M Indices (singleton U)

- Dekel-Fudenberg-Morris 06+07 show that two types are "interim correlated rationalizability" (ICR) strategically equivalent if and only if they have same Mertens-Zamir type
- Ely-Peski 06 gives a characterization of when two types are "interim independent rationalizability" (IIR) strategically equivalent (in terms of a richer hierarchy)
- Sadzik 07 gives charactization of when two types are equilibrium strategically equivalent
- "Redundant types" are key to these distinctions

Common Certainty of vN-M Indices

OBSERVATION. The following are equivalent:

1. Two types are equilibrium strategically indistinguishable
2. Two types are IIR strategically indistinguishable
3. Two types are ICR strategically indistinguishable
4. Two types map to the same MZ type

"PROOF" (1) $\Rightarrow$ (2) because equilibrium is refinement of IIR; (2) $\Rightarrow$ (3) because IIR is refinement of ICR; (3) $\Rightarrow$ (4) follows an adaption of DFM argument; (4) $\Rightarrow$ (1) because there always exists an equilibrium where strategies are measurable w.r.t. MZ types.

Common Certainty of vN-M Indices

	Strategic Equivalence	Strategic Indistinguishability
ICR	MZ space	MZ space
IIR	EP space	MZ space
Equilibrium	Liu/Sadzik	MZ space

Without Common Certainty of vN-M Indices

	Strategic Equivalence	Strategic Indistinguishability
WPCR	BMT space	BMT space
ICR	?	BMT space
Equilibrium	?	BMT space

More Related Literature

- Abreu-Matsushima 93
  - essentially characterize interim correlated rationalizability strategic distinguishability for finite types
  - also show that characterization is unchanged with equilibrium
  - their characterization depends on the finite type space in which types live, i.e., not "universal"
  - we do not encompass their result because of worst outcome restriction

More Related Literature

- Bergemann-Morris 09
    - consider an environment without beliefs but commonly known set of possible "payoff types" for each agent
    - ask when two payoff types θ_i and θ_i' are "strategically distinguishable"
    - equivalent to asking if the union of rationalizable actions of all types consistent with θ_i has a non-empty intersection with union of rationalizable actions of all types consistent with θ_i'
    - natural interpretation: when there is not too much interdependence of payoffs