Robust Mechanism Design

Setephen Morris' ECO512 Lectures summary by N. Antić

Incomplete information environments can be modeled as Bayesian games where there is common knowledge of each player's type space and each type's beliefs over types of other players. Much work in mechanism design assumes smaller type spaces than the universal, yet maintains common knowledge—this makes strong implicit assumptions. Robustness requires mechanisms to work on rich type spaces with less common knowledge assumptions; this is sometimes equivalent to using stronger solution concepts.

- Fix an environment and a social choice correspondence (SCC), $F\colon \Theta \to 2^A \setminus \emptyset$, or social choice function (SCF), $f\colon \Theta \to A$
  - A payoff environment consists of a set of agents, $I \in \{1, ..., N\}$, a set of outcomes or alternatives, A, a set of payoff types, $\Theta = \Theta_1 \times ... \times \Theta_N$, and a set of utility functions, $(u_i\colon A \times \Theta \to \mathbb{R})_{i=1}^N$
  - The objective of the mechanism designer (player 0) is to implement the social choice correspondence
- An environment can have many type spaces, $\mathcal{T} = \left(T_i, \widehat{\theta}_i, \widehat{\pi}_i\right)_{i=1}^N$
  - For each $i \in I$, $\mathcal{T}$ specifies a payoff type, $\widehat{\theta}_i\colon T_i \to \Theta_i$, and beliefs about other players' types, $\widehat{\pi}_i\colon T_i \to \Delta(T_{-i})$
  - Larger type spaces require more ICs to hold—implemention is harder, but more "robust"
- The *naive type space* is the smallest possible type space, where $T_i = \Theta_i$ and $\widehat{\theta}_i = \mathrm{id}_{\Theta_i}$
  - The *universal type space* (allow any higher-order beliefs about other players' payoff relevant type) is the largest
  - $\mathcal{T}$ is finite if each T_i is finite
  - $\mathcal{T}$ has full support if $\widehat{\pi}_i(t_i)[t_{-i}] > 0$, $\forall i$ and $\forall t$
  - $\mathcal{T}$ has a common prior if $\exists \pi \in \Delta(T)$ such that

$$\sum_{t_{-i}} \pi(t_i, t_{-i}) > 0, \quad \forall i \in I, \forall t \in T$$

$$\text{and,} \qquad \widehat{\pi}_i(t_i)[t_{-i}] = \frac{\pi(t_i, t_{-i})}{\sum_{t'_{-i}} \pi(t_i, t'_{-i})} > 0$$

- A mechanism is $\mathcal{M} = (M_1, ..., M_N, g)$; it specifies a message set for each player, M_i, and an outcome function $g\colon M \to A$
- Given an environment, $(\mathcal{T}, \mathcal{M})$ induces an incomplete info game
- $\sigma \in \mathcal{E}(\mathcal{T}, \mathcal{M})$ is an equilibrium if $\sigma(m_i|t_i) > 0$ implies m_i solves

$$\max_{m'_i} \sum_{t_{-i}, m_{-i}} \widehat{\pi}_i(t_i)[t_{-i}] \left(\prod_{j \neq i} \sigma_j(m_j|t_j)\right) u_i\left(g(m'_i, m_{-i}), \widehat{\theta}(t)\right)$$

Note that this is an interim definition of Bayes-Nash equilibrium

Partial Implementation

- An SCF f is ex post incentive compatible (EPIC) if $\forall i \in I$

$$u_i(f(\theta), \theta) \geq u_i(f(\theta'_i, \theta_{-i}), \theta), \quad \forall \theta \in \Theta \text{ and } \forall \theta'_i \in \Theta_i$$

  - If other players are honest, then dominant to be honest
- An SCC F is ex post weakly implementable if there is an SCF f such that f is EPIC and $f(\theta) \in F(\theta)$ for all $\theta \in \Theta$

$\boxed{1}$ If F is ex post weakly implementable, then $\exists \mathcal{M}$ such that $\forall \mathcal{T}$, $\exists \sigma \in \mathcal{E}(\mathcal{T}, \mathcal{M})$ for which $\sigma(m|t) > 0$ implies $g(m) \in F\left(\widehat{\theta}(t)\right)$.

- The converse is not generally true, but it holds in *separable* environments, i.e., payoff environments and SCCs for which
  - $A = A_0 \times A_1 \times ... \times A_N$
  - $u_i(a, \theta) = \widetilde{u}_i(a_0, a_i, \theta)$ for some $\widetilde{u}_i$, $\forall i$ and $\forall a$
  - $F(\theta) = f_0(\theta) \times F_1(\theta) \times ... \times F_N(\theta)$, where $f_0\colon \Theta \to A_0$

Proof. ($\Leftarrow$) Let $\mathcal{T}$ be such that it is common knowledge that agents other than i are θ_{-i}. Then $\exists g^{i,\theta_{-i}}\colon \Theta_i \to F_i(\Theta)$ such that

$$\widetilde{u}_i\left(f_0(\theta), g^{i,\theta_{-i}}(\theta_i), \theta\right) \geq \widetilde{u}_i\left(f_0(\theta'_i, \theta_{-i}), g^{i,\theta_{-i}}(\theta'_i), \theta\right), \forall \theta_i, \theta'_i.$$

Let $f(\theta) = \left(f_0(\theta), g_1^{1,\theta_{-1}}(\theta_{-1}), ..., g_N^{N,\theta_{-N}}(\theta_{-N})\right)$ and note that f is EPIC. Further, by separability $f(\theta) \in F(\theta)$ for all $\theta \in \Theta$. ∎

- The converse also holds if the environment is quasi-linear with no budget balance (in this case only need that a mechanism exists for all $\mathcal{T}$ with full support and common prior)
- Quasi-linear environment with budget balance is not separable
  - But converse is true for $N = 2$ or for $N > 2$ if $|\Theta_i| \leq 2$

Ex post Implementation

- Full implementation in ex post equilibrium?
  - Necessary for every selection of F to be EPIC and for F to be *ex post monotonic*
    - In the direct mechanism a deception is a strategy profile $\beta = (\beta_i)_{i=1}^n$, with $\beta_i\colon \Theta_i \to \Theta_i$; a misreport
    - An SCF f is *ex post monotonic* if for every deception β, $\exists i, \theta, a$ s.t. $u_i(a, \theta) > u_i(f(\beta(\theta)), \theta)$ and

$$u_i\left(f\left(\theta'_i, \beta_{-i}(\theta_{-i})\right), \left(\theta'_i, \beta_{-i}(\theta_{-i})\right)\right) \geq u_i\left(a, \left(\theta'_i, \beta_{-i}(\theta_{-i})\right)\right) \ \forall \theta'_i$$

  - Sufficiency for $N \geq 3$ further requires:
    - An economic environment—$\forall \theta \in \Theta$, $a \in A$ $\exists i, j, a_i, a_j$ s.t. $i \neq j$, $u_i(a_i, \theta) > u_i(a, \theta)$ and $u_j(a_j, \theta) > u_j(a, \theta)$
    - A "no veto power" assumption

Robust Implementation in Direct Mechanism

- Full implementation in interim Bayes-Nash equil. on any $\mathcal{T}$?
  - Converse of $\boxed{1}$ fails in general, so doesn't follow trivially; need $\forall \sigma \in \mathcal{E}(\mathcal{T}, \mathcal{M})$, $g(m) \in F\left(\widehat{\theta}(t)\right)$ if $\sigma(m|t) > 0$
  - Consider single-good auction examples with payoff type $\theta_i \in [0, 1]$ for all i and valuations $v_i(\theta_i) = \theta_i + \gamma \sum_{j \neq i} \theta_j$
  - For $\gamma = 0$, no efficient allocation can be robustly implmented, but nearly effcient allocations can be:
    - w.p. $1 - \varepsilon$ do a standard second-price auction
    - w.p. $\frac{\varepsilon b_i}{N}$ allocate good to i and ask for transfer $\frac{1}{2} b_i$
  - For interdependent values use modified generalized VCG:
    - w.p. $1 - \varepsilon$ do generalized VCG—allocate object to highest bidder, i, who pays $\max_{j \neq i} b_j + \gamma \sum_{j \neq i} b_j$

- w.p. $\frac{\varepsilon b_i}{N}$, i gets object and pays $\frac{1}{2}b_i + \gamma \sum_{j \neq i} b_j$
- Truth-telling is a strict ex post equilibrium, but this does not imply robust implmentation
- This mechanism works if $\gamma < \frac{1}{N-1}$; no mechanism robustly implements the efficient outcome if $\gamma \geq \frac{1}{N-1}$

■ General results rely on *incomplete information rationalizability*

- Let $S_i^{\mathcal{M},0}(\theta_i) = M_i$ and define inductively $S_i^{\mathcal{M},k+1}(\theta_i)$ as the set of m_i for which $\exists \mu_i \in \Delta(\Theta_{-i} \times M_{-i})$ such that

  (a) $\mu_i(\theta_{-i}, m_{-i}) > 0 \;\Rightarrow\; m_{-i} \in S_{-i}^{\mathcal{M},k+1}(\theta_i)$

  (b) $m_i \in \arg\max_{m_i'} \int_{\theta_{-i}, m_{-i}} u_i(g(m_i', m_{-i}), \theta)\; d\mu_i$

- $S_i^{\mathcal{M}}(\theta_i) = \bigcap_{k=0}^{\infty} S_i^{\mathcal{M},k}(\theta_i)$ is set of rationalizable messages

$\boxed{2}$ m_i is played in some $\sigma \in \mathcal{E}(\mathcal{T}, \mathcal{M})$ for some $\mathcal{T} \Leftrightarrow m_i$ is *incomplete information rationalizable*, i.e., $m_i \in S_i^{\mathcal{M}}(\theta_i)$

Proof. ($\Leftarrow$) Let $m_i \in S_i^{\mathcal{M}}(\theta_i)$, then $\exists \mu_i^{m_i,\theta_i} \in \Delta(\Theta_{-i} \times M_{-i})$ such that (a) and (b) hold. Consider the type space where $T_i = \{(\theta_i, m_i) \in \Theta_i \times M_i | m_i \in S_i^{\mathcal{M}}(\theta_i)\}$, $\widehat{\theta}_i(\theta_i, m_i) = \theta_i$ and $\widehat{\pi}_i(\theta_i, m_i)[\theta_{-i}, m_{-i}] = \mu_i^{m_i,\theta_i}(\theta_{-i}, m_{-i})$. Note that there is a pure strategy equilibrium where type (θ_i, m_i) plays m_i. ■

■ $h_i \colon \Theta \to \mathbb{R}$ is an *aggregator function* if $u_i(a,\theta) = v_i(a, h_i(\theta))$ and h_i is cts and strictly increasing in θ_i, and v_i is cts wrt h_i.

■ *Strict single crossing* (SSC) holds if $\forall \underline{\phi} < \phi < \overline{\phi}$

$$v_i(a, \underline{\phi}) > v_i(a', \underline{\phi}),\; v_i(a,\phi) = v_i(a',\phi) \;\Rightarrow\; v_i(a, \overline{\phi}) < v_i(a', \overline{\phi})$$

■ Aggregators $(h_i)_{i \in I}$ satisfy the *contraction property* if $\forall \beta \neq \beta^*$, $\exists i, \theta_i' \in \beta_i(\theta_i)$ with $\theta_i' \neq \theta_i$ s.t. $\forall \theta_{-i} \in \Theta_{-i}$, $\theta_{-i}' \in \beta_{-i}(\theta_{-i})$

$$\operatorname{sgn}(\theta_i - \theta_i') = \operatorname{sgn}(h_i(\theta_i, \theta_{-i}) - h_i(\theta_i', \theta_{-i}'))$$

- Assumes there is not too much interdependence of preferences, as well as a whistleblower, i, in deception β, where $\beta_i \colon \Theta_i \to 2^{\Theta_i} \setminus \emptyset$, and $\beta_i^*(\theta_i) = \{\theta_i\}$ is truthful

$\boxed{3}$ In an SSC environment there is robust implementation in the direct mechanism, if strict EPIC and contraction are satisfied

- Use $\boxed{2}$ to simplify proof, since f can be robustly implemented by $\mathcal{M}$ if $m \in S^{\mathcal{M}}(\theta) \Rightarrow g(m) = f(\theta)$

■ SCF f is *responsive* if for $\theta_i \neq \theta_i'$, $\exists \theta_{-i}$ s.t. $f(\theta) \neq f(\theta_i', \theta_{-i})$

$\boxed{4}$ In an SSC environment f satisfies strict EPIC and contraction, if f is responsive and robustly implementable in *any* mechanism

■ Can simplify the contraction property if each $h_i(\theta)$ is linear, i.e., $h_i(\theta) = \theta_i + \sum_{j \neq i} \gamma_{ij}\theta_j$

- Linear $(h_i)_{i \in I}$ satisfy the contraction property iff $\forall c \in \mathbb{R}_+^N$ with $c \neq \overrightarrow{0}$, $\exists i$ s.t. $c_i > + \sum_{j \neq i} |\gamma_{ij}| c_j$
- Define the interdependence matrix Γ as:

$$\Gamma = \begin{bmatrix} 0 & |\gamma_{12}| & \cdots & |\gamma_{1N}| \\ |\gamma_{21}| & 0 & \cdots & |\gamma_{2N}| \\ \vdots & \vdots & \ddots & \vdots \\ |\gamma_{N1}| & |\gamma_{N2}| & \cdots & 0 \end{bmatrix}$$

- Contraction is satisfied if Γ's largest eigenvalue $\lambda < 1$

Robust Virtual Implementation

■ Relax implementation requirement, i.e., we have robust virtual implementation if $m \in S^{\mathcal{M}}(\theta) \Rightarrow \|g(m) = f(\theta)\| < \varepsilon$

■ Indistinguishability: $\theta_i \sim \theta_i'$ if $S_i^{\mathcal{M}}(\theta_i) \cap S_i^{\mathcal{M}}(\theta_i') \neq \emptyset$, $\forall \mathcal{M}$

■ f is *robustly measurable* (RM) if $\theta_i \sim \theta_i' \Rightarrow f(\theta) = f(\theta_i', \theta_{-i})$

- In SSC environment, RM $\Leftrightarrow$ Contraction property $\Leftrightarrow$ Abreu-Matsushima measurability on every type space $\mathcal{T}$

$\boxed{5}$ f is EPIC and RM, if f is robustly virtually implementable

- EPIC and RM are almost sufficient—need to assume that one can create sufficiently bad lotteries for each i (mild)

Robust Implementation in General Mecahnisms

■ If restrictions on the environment (Robust Implementation in Direct Mechanisms) or a weaker implementation condition (Robust Virtual Implementation) are undesirable, need to rely on badly behaved mechanisms (as in the non-robust literature)

■ SCF f satisfies *robust monotonicity* if $\forall \beta$ which is unacceptable, $\exists i, \theta_i, \theta_i' \in \beta_i(\theta_i)$ s.t. $\forall \theta_{-i}' \in \Theta_{-i}$, $\exists a$ s.t. $\forall \theta_{-i} \in \beta_{-i}^{-1}(\theta_{-i}')$

$$u_i(a, \theta) > u_i(f(\theta'), \theta) \qquad \text{and}$$
$$u_i(f(\theta_i'', \theta_{-i}'), (\theta_i'', \theta_{-i}')) \geq u_i(a, (\theta_i'', \theta_{-i}')),\;\; \forall \theta_i''$$

- Deception β is *unacceptable* if $\theta' \in \beta(\theta)$ and $f(\theta') \neq f(\theta)$

■ In SSC environment with responsive SCF, robustly monotonicity $\Leftrightarrow$ contraction property $\Leftrightarrow$ Bayesian monotonicity on all $\mathcal{T}$

$\boxed{6}$ f satisfies robust monotonicity, if f is robustly implementable

- Rely on $\boxed{2}$ for proof, but need to ensure a set of rationalizable messages exists; trivial if $\mathcal{M}$ is finite, but given that we need infinite mechanisms it's a strong assumption
- Robust monotonicity and no total indifference sufficient

Robustness of Robust Implementation

■ Fix an SSC environment with contraction property and EPIC

- Payoff types Θ, utility functions $u_i(a,\theta) = v_i(a, h_i(\theta))$

■ Fix an *interim type space*, $\mathcal{T} = (T_i, \pi_i, u_i)_{i \in I}$, $\pi_i \colon T_i \to \Delta(T_{-i})$ and $\widetilde{u}_i \colon A \times T \to \mathbb{R}$

■ SCC $F : T \to A$ is λ-optimal at type profile $t \in T$

$$\widetilde{u}_0(a', t) \geq \sup_y \widetilde{u}_0(a, t) - \lambda, \quad \forall a' \in F(t)$$

where $\widetilde{u}_0$ denotes desginer's utility, e.g. $\widetilde{u}_0(\theta) = \sum_i \widetilde{u}_i(\theta)$

■ Payoff environment $(\Theta, (u_i)_{i \in I})$ is γ-approximate common knowledge at type profile $t \in T$ if $\exists E \subset T$ and $\widetilde{\theta}_i \colon E_i \to \Theta_i$ s.t.

(a) $\left| u_i\left(a, \widetilde{\theta}(t)\right) - \widetilde{u}_i(a', t) \right| \leq \gamma,\;\; \forall i, a \in A, t \in E.$

(b) $\pi_i(E_{-i} | t_i) \geq 1 - \gamma$

■ f is uniformly EPIC if $\exists \alpha \colon \mathbb{R}^+ \to \mathbb{R}^+$, α strictly increasing s.t.

$$u_i(f(\theta), \theta) - u_i(f(\theta_i', \theta_{-i}), \theta) \geq \alpha(|\theta_i - \theta_i'|),\;\; \forall i, \theta_i \neq \theta_i', \theta_{-i}$$

$\boxed{7}$ Let SCF f be λ-optimal $\forall \theta \in \Theta$ and uni-EPIC. $\forall \varepsilon > 0$, $\exists \gamma > 0$ s.t. direct mech implements an SCC which is $(\lambda + \varepsilon)$-optimal $\forall t \in T$ with γ-approx common knowledge of $(\Theta, (u_i)_{i \in I})$