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), \( F: \Theta \to 2^A \setminus \emptyset \), or social choice function (SCF), \( f: \Theta \to A \)
  - A payoff environment consists of a set of agents, \( I = \{1,...,N\} \), a set of outcomes or alternatives, \( A \), a set of payoff types, \( \Theta = \Theta_1 \times \ldots \times \Theta_N \), and a set of utility functions, \( \{u_i: A \times \Theta \to \mathbb{R}\}_{i=1}^N \)
  - The objective of the mechanism designer (player 0) is to implement the social choice correspondence

- An environment can have many type spaces, \( T = (T_i, \hat{\theta}_i, \tilde{\pi}_i)_{i=1}^N \)
  - For each \( i \in I \), \( T \) specifies a payoff type, \( \hat{\theta}_i; T_i \to \Theta_i \), and beliefs about other players’ types, \( \tilde{\pi}_i: T_i \to \Delta(T_{-i}) \)
  - Larger type spaces require more ICs to hold—implementation is harder, but more “robust”

- The naive type space is the smallest possible type space, where \( T_i = \Theta_i \) and \( \theta_i = \text{id}_{\Theta_i} \)
  - The universal type space (allow any higher-order beliefs about other players’ payoff relevant type) is the largest
  - \( T \) is finite if each \( T_i \) is finite
  - \( T \) has full support if \( \tilde{\pi}_i(t_i) \mid t_{-i} \mid > 0 \), \( \forall i \) and \( \forall t \)
  - \( T \) has a common prior if \( \exists \pi \in \Delta(T) \) such that

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

and,

\[
\tilde{\pi}_i(t_i) \mid t_{-i} = \frac{\pi(t_i, t_{-i})}{\sum_{t'_{-i}} \pi(t_i, t'_{-i})} > 0
\]

- A mechanism is \( M = (M_1, ..., M_N, g) \); it specifies a message set for each player, \( M_i \), and an outcome function \( g: M \to A \)

- Given an environment, \( (T, M) \) induces an incomplete info game

\( \sigma \in \mathcal{E}(T, M) \) is an equilibrium if \( \sigma(m_i | t_i) > 0 \) implies \( m_i \) solves

\[
\max_{m_i} \sum_{t_{-i}, t_{-i}} \tilde{\pi}_i(t_i) \mid t_{-i} \left( \prod_{j \neq i} \sigma_j(m_j | t_j) \right) u_i(g(m_i, m_{-i}), \hat{\theta}(t))
\]

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

### Partial Implementation

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

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

- If other players are honest, then dominant to be honest

### 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 \( \beta_i: \Theta_i \to \Theta_i \); a misreport
    - An SCF \( f \) is ex post monotonic if for every deception \( \beta, \exists i, \theta, a \) s.t. \( u_i(a, \theta) > u_i(f(\beta(\theta)), \theta) \) and

\[
u_i(f(\theta_i', \beta_{-i}(\theta_{-i})), (\theta_i', \beta_{-i}(\theta_{-i}))) \geq u_i(a, (\theta_i', \beta_{-i}(\theta_{-i}))) \quad \forall \theta_i'
\]

- 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 Mechanism Design

- Full implementation in interim Bayes-Nash equil. on any \( T \)?
  - Converse of [1] fails in general, so doesn’t follow trivially; need \( \forall \sigma \in \mathcal{E}(T, M), g(m) \in F(\hat{\theta}(t)) \) 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 implemented, but nearly efficient allocations can be:
    - w.p. \( 1 - \varepsilon \) do a standard second-price auction
    - w.p. \( \frac{\varepsilon}{N} \) 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 \( \max_{j \neq i} b_j + \gamma \sum_{j \neq i} b_j \)
General results rely on incomplete information rationalizability

- Let $S_{i}^{M,0}(\theta_{i}) = M_{i}$ and define inductively $S_{i}^{M,k+1}(\theta_{i})$ as the set of $m_{i}$ for which $\exists \mu_{i} \subseteq \Delta(\Theta_{-i} \times M_{-i})$ such that
  
  \( \mu_{i}(\theta_{-i}, m_{-i}) > 0 \implies m_{i} \in S_{i}^{M,k+1}(\theta_{i}) \)
  
  (a) $m_{i} \in \arg\max_{m_{i}^{'} \in S_{i}^{M,k+1}(\theta_{i})} u(g(m_{i}, m_{-i}), \theta) \, d\mu_{i}$

- $S_{i}^{M}(\theta_{i}) = \bigcap_{k=0}^{\infty} S_{i}^{M,k}(\theta_{i})$ is set of rationalizable messages

$m_{i}$ is played in some $\sigma \in \mathcal{E}(T, M)$ for some $T \iff m_{i}$ is incomplete information rationalizable, i.e., $m_{i} \in S_{i}^{M}(\theta_{i})$.

**Proof.** ($\Leftarrow$) Let $m_{i} \in S_{i}^{M}(\theta_{i})$, then $\exists \mu_{i}^{m_{i}, \theta_{i}} \subseteq \Delta(\Theta_{-i} \times M_{-i})$ such that (a) and (b) hold. Consider the type space where $T_{i} = \{(\theta_{i}, m_{i}) \in \Theta_{i} \times M_{i} | m_{i} \in S_{i}^{M}(\theta_{i})\}$, $\tilde{\theta}_{i}(\theta_{i}, m_{i}) = \theta_{i}$ and $\tilde{\pi}_{i}(\theta_{i}, m_{i})[\theta_{-i}, m_{-i}] = \mu_{i}^{m_{i}, \theta_{i}}(\theta_{-i}, m_{-i})$. Note that there is a pure strategy equilibrium where type $(\theta_{i}, m_{i})$ plays $m_{i}$.

- $h_{i}: \Theta \rightarrow \mathbb{R}$ is an aggregator function if $u_{i}(a, \theta) = v_{i}(a, h_{i}(\theta))$ and $h_{i}$ is cts and strictly increasing in $\theta_{i}$, and $v_{i}$ is cts wrt $h_{i}$.

- Strict single crossing (SSC) holds if $\forall \theta_{i} \in \Theta_{i} \subseteq \mathcal{E} \,, \, v_{i}(a, \theta_{i}) > v_{i}(a', \theta_{i}) \land v_{i}(a, \theta_{i}) = v_{i}(a', \theta_{i}) \implies v_{i}(a, \theta_{i}) < v_{i}(a', \theta_{i})$.

- Aggregators $(h_{i})_{i \in I}$ satisfy the contraction property if $\forall \beta \neq \beta^{'}, \exists i, \theta_{i}^{'} \in \beta^{'}(\theta_{i})$ s.t. $\forall \theta_{i} \neq \theta_{i}^{'} \in \beta(\theta_{i})$, $\tilde{\theta}_{i}^{'} = \beta^{'}(\theta_{i})$.

- Assumptions there is not too much interdependence of preferences, as well as a whistleblower, i.e., in deception $\beta_{i}: \Theta_{i} \rightarrow 2^{\Theta_{i}}, \emptyset$, and $\beta^{*}_{i}(\theta_{i}) = \{\theta_{i}\}$ is truthful.

**Robust Virtual Implementation**

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

- Indistinguishability: $\theta_{i} \sim \theta_{i}^{'}$ if $S_{i}^{M}(\theta_{i}) \cap S_{i}^{M}(\theta_{i}^{'}) \neq \emptyset$, $\forall M$

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

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

**Robust Implementation in General Mechanisms**

- 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)

- If SCF $f$ satisfies robust monotonicity if $\forall \beta$ which is unacceptable, $\exists i, \theta_{i}, \theta_{i}^{'} \in \beta(\theta_{i})$ s.t. $\forall \theta_{i} \neq \theta_{i}^{'} \in \beta(\theta_{i})$.

- Rely on [2] for proof, but need to ensure a set of rationalizable messages exists; trivial if $M$ is finite, but given that we need infinite mechanisms it’s a strong assumption

**Robustness of Robust Implementation**

- Fix an SSC environment with contraction property and EPIC

- Payoff types $\Theta$, utility functions $u_{i}(a, \theta) = v_{i}(a, h_{i}(\theta))$

- Fix an interim type space, $T = (T_{i}, \pi_{i}, u_{i})_{i \in I}$, $\pi_{i}: T_{i} \rightarrow \Delta(T_{i})$ and $\tilde{u}_{i}: A \times T \rightarrow \mathbb{R}$

- SCC $F: T \rightarrow A$ is $\lambda$-optimal at type profile $t \in T$

- $\tilde{u}_{0}(a', t) \geq \sup_{y} \tilde{u}_{0}(a, t) - \lambda, \forall a \in F(t)$

- Payoff environment $(\Theta_{i}, (u_{i})_{i \in I})$ is $\gamma$-approximate common knowledge at type profile $t \in T$ if $\exists \varepsilon < T_{i} \rightarrow T_{i}$ s.t.

  \[ (a) \quad \left| u_{i}(a, \tilde{\theta}(\theta)) - \tilde{u}_{i}(a', \tilde{\theta}) \right| \leq \gamma, \forall i, a \in A, t \in E. \]

  \[ (b) \quad \pi_{i}(E \setminus t) \geq 1 - \gamma \]

- $f$ is uniformly EPIC if $\exists \Theta: \mathbb{R}^{+} \rightarrow \mathbb{R}^{+}$, $\alpha$ strictly increasing s.t.

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