# Robust Mechanism Design and Robust Implementation

joint work with Stephen Morris

August 2009
Barcelona

# Introduction

- mechanism design and implementation literatures are theoretical successes
- mechanisms often seem too complicated to use in practise
- successful applications of auctions and trading mechanisms commonly include ad hoc restrictions:
    - simplicity
    - non-parametric
    - belief-free
    - detail free

# Weaken Informational Assumptions

- if the optimal solution to the planner's problem is too complicated or sensitive to be used in practice, presumably the original description of the planner's problem was itself flawed
- weaken informational requirements
- specifically weaken common knowledge assumption in the description of the planner's problem
  - "Wilson doctrine"
- can improved modelling of the planner's problem endogenously generate the "robust" features of mechanisms that researchers have been tempted to assume?

# Weakening Common Knowledge

- in game theory, Harsanyi (1967), Mertens & Zamir (1985) establish that environments with incomplete information can be modeled as a Bayesian game
- in particular, in the universal type space there is without loss of generality common knowledge among players of
  - each player's type spaces
  - each type's beliefs over types of other players
- yet in economic analysis generally assumes smaller type spaces than universal type space *yet maintains common knowledge*

# Weakening Common Knowledge in Mechanism Design

- are the implicit common knowledge assumptions that come from working with small type spaces problematic?
- especially in mechanism design
  - Neeman (1999) on surplus extraction
  - "beliefs determine preferences"
- especially in auctions:
  - no strategic uncertainty among bidders
  - designer and bidder $i$ have identical information about all other bidders

# Agenda

- introduce rich (higher order belief) types and strategic uncertainty into mechanism design literature
- relax (implicit) common knowledge assumptions by going from "naive" type space to "universal" type space
- characterize social choice function/mechanism with robust incentive compatibility
    - ex post incentive compatibility as necessary and sufficient condition
    - ex post equilibrium as belief free solution concept
- characterize social choice function/mechanism with robust implementation
    - rationalizability as necessary and sufficient condition
    - for direct and augmented mechanism

# A Selection

- joint work Stephen Morris:

1. "Robust Mechanism Design", *ECTA 2005*
2. "An Ascending Auction for Interdependent Values" *AER 2007*
3. "Ex Post Implementation" *GEB 2008*
4. "The Role of the Common Prior Assumption in Robust Implementation" *JEEA 2008*
5. "Robust Virtual Implementation" *TE 2009*
6. "Robust Implementation in General Mechanisms" *2009*
7. "Robust Implementation in Direct Mechanisms" *REStud forthcoming*

# Payoff Environment

- agent $i \in \mathcal{I} = \{1, 2, ..., I\}$
- $i$'s "payoff type" $\theta_i \in \Theta_i$
- payoff type profile $\theta \in \Theta = \Theta_1 \times \cdots \times \Theta_I$
- social outcome $a \in A$
- utility function $u_i : A \times \Theta \to \mathbb{R}$
- social choice function $f : \Theta \to A$
- fix payoff types and social objective
- for fixed payoff environment, we can construct many type spaces in terms of beliefs and higher-order beliefs

# Type Spaces

- richer type space $T_i$ than payoff type space $\Theta_i$
- $i$'s *type* is $t_i \in T_i$, $t_i$ includes description of:
- payoff type $\widehat{\theta}_i(t_i)$ of $t_i$ :

$$\widehat{\theta}_i : T_i \to \Theta_i$$

- belief type $\widehat{\pi}_i(t_i)$ of $t_i$ :

$$\widehat{\pi}_i : T_i \to \Delta(T_{-i})$$

- *type space* is a collection $\mathcal{T} = \{T_i, \widehat{\theta}_i, \widehat{\pi}_i\}_{i=1}^{I}$
- type $t_i$ contains information about preferences and information of others agents, i.e. beliefs and higher-order beliefs

# Many Type Spaces

- smallest type space: "naive type space":
  - possible types equal to payoff types ($T_i = \Theta_i$)
  - standard construction in mechanism design
- largest type space: "universal type space"
  - allow any (higher order) beliefs about other players' payoff relevant type
  - without common prior
- many type spaces in between smallest and largest type space:
  - common prior payoff type space
  - common prior type space
- study role of common knowledge by comparative statics on type spaces, going from "naive" type space to "universal" type space

## Allocating a Single Object Efficiently

- agent $i = 1, ..., I$ has a payoff type $\theta_i \in \Theta_i = [0, 1]$
- agent $i$'s valuation of the object is

$$v_i(\theta_1, ..., \theta_I) = \theta_i + \gamma \sum_{j \neq i} \theta_j$$

- interdependent value model (Dasgupta and Maskin (1999))
- interdependence is represented by $\gamma$
- private value: $\gamma = 0$
- interdependent value: $\gamma \neq 0$ (negative or positive externality)
- principal/designer does not know anything about agent $i$'s beliefs and higher order beliefs about $\theta_{-i}$

- value of $i$ only depends on payoff type of agent $i$:

$$v_i(\theta) = \theta_i$$

- second price sealed bid auction, agent $i$ bids/reports $b_i \in [0,1]$
- highest bid wins, pays second highest bid
- truthful reporting leads to efficient allocation of object $q^*(\theta)$ :

$$q_i^*(\theta) = \begin{cases} \frac{1}{\#\{j:\theta_j \geq \theta_k \text{ for all } k\}}, & \text{if } \theta_i \geq \theta_k \text{ for all } k \\ 0, & \text{if otherwise} \end{cases}$$

- dominant strategy to truthfully report/bid

- with interdependence $\gamma \neq 0$:

$$v_i(\theta) = \theta_i + \gamma \sum_{j \neq i} \theta_j$$

- "generalized" VCG mechanism: agent $i$ bids/reports $b_i \in [0, 1]$,
- highest bid wins, pays the second highest bid plus $\gamma$ times the bid of others:

$$\max_{j \neq i} \{b_j\} \; + \; \gamma \sum_{j \neq i} b_j$$

- truthful reporting is an ex post equilibrium in direct mechanism if and only if $\gamma \leq 1$ (single crossing condition)

- robust incentive compatibility: for any beliefs and higher order beliefs
- when does there exist a mechanism with the property that for any beliefs and higher order beliefs that the agents may have, thruthtelling is an interim equilibrium in the direct mechanism?
- in single good example, consider efficient allocation $q^*$ of object and any suitable transfers

- type space $\mathcal{T} = \{T_i, \widehat{\theta}_i, \widehat{\pi}_i\}_{i=1}^{I}$

**Definition**

A scf $f : \mathcal{T} \to A$ is interim incentive compatible on type space $\mathcal{T}$ if

$$\int_{t_{-i}} u_i\left(f(t), \widehat{\theta}(t)\right) d\widehat{\pi}_i(t_{-i}|\, t_i) \geq \int_{t_{-i}} u_i\left(f(t_i', t_{-i}), \widehat{\theta}(t)\right) d\widehat{\pi}_i(t_{-i}|\, t_i)$$

for all $i$, $t \in T$ and $t_i' \in T_i$.

- "interim" to emphasize that $\widehat{\pi}_i(t_{-i}|\, t_i)$ are interim beliefs (without the necessity of a common prior)
- the larger the type space, the more incentive constraints there are, the harder it becomes to implement scc
- from smallest type space: "naive type space" to largest type space: "universal type space"

- a belief free solution concept requires strategies of players to remain an equilibrium for all possible beliefs and higher order beliefs

### Definition

A scf $f$ is ex post incentive compatible if, for all $i$, $\theta \in \Theta$, $\theta_i' \in \Theta_i$:

$$u_i \left( f \left( \theta \right), \theta \right) \geq u_i \left( f \left( \theta_i', \theta_{-i} \right), \theta \right).$$

- "ex post equilibrium": each type of each agent has an incentive to tell truth *if* he expects all other agents to tell the truth (whatever his beliefs about others' payoff types)
- compare: a scf $f$ is dominant strategy incentive compatible if for all $i$ and all $\theta, \theta'$ :

$$u_i \left( f \left( \theta_i, \theta_{-i}' \right), \theta \right) \geq u_i \left( f \left( \theta_i', \theta_{-i}' \right), \theta \right)$$

**Theorem (2005)**

*$f$ is interim incentive compatible on every type space $\mathcal{T}$ if and only if $f$ is ex post incentive compatible.*

- ex post equilibrium notion incorporates concern for robustness to higher-order beliefs
- robustness imposes simplicity: constraints are satisfied at every profile rather than for all possible expectations
- in private values case, ex post implementation is equivalent to dominant strategies implementation:
    - c.f. Ledyard (1978) in private value environments and dominant strategies

# Proof and Limits of Equivalence Result

- with rich type spaces and beliefs ex post incentive constraints are included
- equivalence result does not require universal type space
- truthtelling in direct mechanism: analyze incentives to reveal private, agent by agent, while presuming thruthtelling by other agents
- constructing a specific equilibrium in a specific mechanism...
- ...but for every specific type space and every specific mechanism there might be other equilibria which do not lead to the desired outcome

# Robust Implementation

- strengthening the question to cover all equilibria for all type spaces...
- when does there exist a mechanism with the property that for any beliefs and higher order beliefs that the agents may have, *every* interim equilibrium has the property that an acceptable outcome is chosen?
- we call this "robust implementation"

# An Aside: Ex Post versus Robust Implementation

- ex post implementation: to rule out bad equilibria, it is enough to make sure you could not construct a "bad" ex post equilibrium;
- when does there exist a mechanism such that, not only is there an ex post equilibrium delivering the right outcome, but *every* ex post equilibrium delivers the right outcome?
- for robust implementation, we must rule out bad Bayesian, or interim equilibria on all type spaces
- in addition to ex post incentive compatibility - an ex post monotonicity condition is necessary and almost sufficient

# Back to the Single Object Example....

- is robust implementation possible in single object auction?
- actually no: robust implementation fails *even in the private value model*
- truthtelling is only a weak best response and there are many equilibria leading to inefficient outcomes in second price sealed bid auctions
- but robust implementation is achievable for almost efficient allocations (and strict incentive compatibility)

# Private Values: A Modified Second Price Auction

- with probability
$$1 - \varepsilon$$
allocate object to highest bidder and pay second highest bid
- with probability
$$\varepsilon$$
assign object to agent $i$ with (conditional) probability
$$\frac{b_i}{I}$$
and agent $i$ pays $\frac{1}{2}b_i$
- truth-telling is now a strictly dominant strategy and $\varepsilon$-efficient allocation is robustly implemented

# Interdependent Values: A Modified VCG Mechanism

- with probability

$$1 - \varepsilon$$

allocate object to highest bidder $i$ and winner pays

$$\max_{j \neq i} \{b_j\} \; + \; \gamma \sum_{j \neq i} b_j$$

- with probability

$$\varepsilon$$

assign object to agent $i$ with (conditional) probability

$$\frac{b_i}{I}$$

and agent $i$ pays:

$$\frac{1}{2} b_i + \gamma \sum_{j \neq i} b_j$$

- truth telling is a strict ex post equilibrium

# The Modified Generalized VCG Mechanism

- but existence of strict ex post equilibrium does *not* imply robust implementation
- in fact, we show this mechanism robustly implements the efficient outcome if and only if

$$|\gamma| < \frac{1}{I-1}$$

- and no mechanism robustly implements efficient outcome if

$$|\gamma| \geq \frac{1}{I-1}$$

- contrast with single crossing condition

$$\gamma < 1$$

# Robustness and Rationalizability

- before: truthtelling in direct mechanism: analyze incentives to reveal private, agent by agent, while presuming thruthtelling by other agents
- now: we cannot suppose behavior of other agents but rather have to guarantee it
- identify restriction on rational behavior of each agent, and then use these restriction to inductively obtain further restrictions
- rationalizability with incomplete information

# Rationalizability with Incomplete Information

- an action is incomplete information rationalizable for a payoff type of an agent if it survives the process of iteratively elimination of dominated strategies
- as rationalizability with complete information it defines an inductive process:
    1. first suppose every payoff type $\theta_i$ could send any message $m_i$
    2. delete those messages $m_i$ that are not a best response to some conjecture over pairs of payoff type and message $(\theta_{-i}, m_{-i})$ of the opponents that have not yet been deleted
    3. repeat step 2 until converge is achieved
- the notion of incomplete information rationalizability is belief free as the candidate action needs only to be a best response to some beliefs about the other agents actions and payoff types

# Rationalizability: A Key Epistemic Result

**Theorem**

*A message $m_i$ can be sent by an agent with payoff type $\theta_i$ in an interim equilibrium on some type space if and only if $m_i$ is "incomplete information rationalizable"*

- incomplete information counterpart to Brandenburger and Dekel (1987)
- identify disjoint rationalizable strategic choices for all possible beliefs and higher order beliefs about others' types
- types are distinguishable

# Rationalizability in Direct Mechanism

- direct mechanism: message $m_i$ is report $\theta_i'$
- $i$ conjectures other agents have type $\theta_{-i}$ and report $\theta'_{-i}$ :

$$\lambda_i \left(\theta_{-i}, \theta'_{-i}\right) \in \Delta \left(\Theta_{-i} \times \Theta_{-i}\right)$$

- set of reports $i$ might send for some conjecture $\lambda_i \left(\theta_{-i}, \theta'_{-i}\right)$ over his opponents' types $\theta_{-i}$ and reports $\theta'_{-i}$:

$$\beta_i^k \left(\theta_i\right)$$

  with restriction on conjecture $\lambda_i \left(\theta_{-i}, \theta'_{-i}\right)$ that type $\theta_j$ sends message $\theta_j' \in \beta_i^{k-1} \left(\theta_j\right)$
- initialize at step $k = 0$ by allowing all reports $\beta_i^0 \left(\theta_i\right) = [0, 1]$

# Rationalizability in Generalized VCG mechanism

- with linear interdependence: $\gamma > 0,\ \theta_i \in [0,1]$

$$v_i(\theta) = \theta_i + \gamma \sum_{j \neq i} \theta_j$$

ex post compatible transfer $y_i^*(\theta)$ is quadratic in reports $\theta'$

- agent $i$ with type $\theta_i$ has linear best response $\theta_i'$:

$$\theta_i' = \theta_i + \gamma \sum_{j \neq i} \left(\theta_j - \theta_j'\right)$$

- linear best response leads to set of best responses $\beta_i^k(\theta_i)$:

$$\beta_i^k(\theta_i) = \left[\underline{\beta}_i^k(\theta_i), \overline{\beta}_i^k(\theta_i)\right]$$

- the bounds $\left\{\underline{\beta}_i^k(\theta_i), \overline{\beta}_i^k(\theta_i)\right\}$ in step $k$ are determined by restrictions of round $k-1$ :

$$\left\{\left(\theta'_{-i}, \theta_{-i}\right) : \theta'_j \in \beta_j^{k-1}(\theta_j),\ \forall j \neq i\right\}$$

- the upper bound $\overline{\beta}^k(\theta_i)$ is:

$$\overline{\beta}^k(\theta_i) = \theta_i + \ \gamma \max_{\left\{\left(\theta'_{-i}, \theta_{-i}\right) : \theta'_j \in \beta_j^{k-1}(\theta_j),\ \forall j \neq i\right\}} \sum_{j \neq i} (\theta_j - \theta'_j)\}$$

- using lower bound $\underline{\beta}_j^{k-1}(\theta_j)$ from round $k-1$ explicitly:

$$\overline{\beta}^k(\theta_i) = \theta_i + \ \gamma \max_{\theta_{-i}} \sum_{j \neq i} (\theta_j - \underline{\beta}_j^{k-1}(\theta_j))\}$$

rewriting:

$$\overline{\beta}^{k}(\theta_i) = \theta_i + \gamma \max_{\theta_{-i}} \sum_{j \neq i} (\theta_j - \underline{\beta}_j^{k-1}(\theta_j))\}$$

we obtain

$$\overline{\beta}^{k}(\theta_i) = \theta_i + (\gamma(I-1))^k,$$

and likewise the recursion for the lower bound:

$$\underline{\beta}^{k}(\theta_i) = \theta_i - (\gamma(I-1))^k$$

and thus

$$\theta_i' \neq \theta_i \Rightarrow \theta_i' \notin \beta^k(\theta_i)$$

for sufficiently large $k$, provided that

$$|\gamma|(I-1) < 1 \quad \Leftrightarrow \quad |\gamma| < \frac{1}{I-1}$$

- but now suppose that $\gamma \geq \frac{1}{I-1}$
- use rich type space to identify specific beliefs
- *each* type $\theta_i$ convinced that type $\theta_j$ is

$$\theta_j \triangleq \frac{1}{2} + \frac{1}{\gamma(I-1)}\left(\frac{1}{2} - \theta_i\right), \ \forall j$$

- now the expected value of the object for $i$ is independent of $\theta_i$

$$\theta_i + \gamma(I-1)\left[\frac{1}{2} + \frac{1}{\gamma(I-1)}\left(\frac{1}{2} - \theta_i\right)\right] = \frac{1}{2}\left[1 + \gamma(I-1)\right]$$

- types cannot be distinguished (and hence separated) in direct or any other mechanism, they are indistinguishable

- robust implementation possible (using the modified generalized VCG mechanism) if

$$|\gamma| < \frac{1}{I - 1}$$

- robust implementation impossible (in *any* mechanism) if

$$|\gamma| \geq \frac{1}{I - 1}$$

- in contrast (robust) incentive compatibility required (only)

$$\gamma < 1$$

- "contraction property" leads to robust implementation

- each $\Theta_i$ is a compact subset of the real line
- agent $i$'s preferences depend on $\theta$ through $h_i : \Theta \to \mathbb{R}$
- preferences are single crossing in $h_i(\theta)$
- as an example linear aggregator for each $i$:

$$h_i(\theta) = \theta_i + \sum_{j \neq i} \gamma_{ij} \theta_j$$

- $\gamma_{ij}$ measures the importance of payoff type $j$ for preference of agent $i$

# Contraction Property

- with linear aggregator for each $i$:

$$h_i(\theta) = \theta_i + \sum_{j \neq i} \gamma_{ij} \theta_j$$

- the interaction matrix:

$$\Gamma \triangleq \begin{bmatrix} 0 & |\gamma_{12}| & \cdots & |\gamma_{1I}| \\ |\gamma_{21}| & 0 & & \vdots \\ \vdots & & \ddots & |\gamma_{I-1I}| \\ |\gamma_{I1}| & \cdots & |\gamma_{II-1}| & 0 \end{bmatrix}$$

- the contraction property is satisfied if and only if largest eigenvalue of the interaction matrix is less than 1.

- possible reports: $\beta = (\beta_1, ..., \beta_I)$; $\beta_i : \Theta_i \to 2^{\Theta_i} / \varnothing$
- the aggregator functions $h$ satisfy the strict contraction property if, $\forall \beta$, $\exists i, \theta_i' \in \beta_i(\theta_i)$ with $\theta_i' \neq \theta_i$, such that

$$\operatorname{sign}(\theta_i - \theta_i') = \operatorname{sign}\left(h_i(\theta_i, \theta_{-i}) - h_i(\theta_i', \theta_{-i}')\right),$$

for all $\theta_{-i}$ and $\theta_{-i}' \in \beta_{-i}(\theta_{-i})$

**Theorem (2009)**

1. *Robust implementation is possible in the direct mechanism if strict EPIC and the contraction property hold.*
2. *Robust implementation is impossible in any mechanism if either strict EPIC or the contraction property fail.*

- robustness leads to simple mechanism, augmented mechanism loose their force

# The Role of the Common Prior

- in the analysis so far, no restrictions were placed on agents' beliefs and higher order beliefs
- consider the role of beliefs and hence intermediate notions of robustness
- what if we know that the common prior assumption holds?
- now the size but also sign of the interdependence matters

# Strategic Complements

- recall the linear best response in the auction model

$$\theta_i' = \theta_i + \gamma \sum_{j \neq i} (\theta_j - \theta_j')$$

- negative interdependence in agents' types,

$$\gamma < 0$$

  gives rise to strategic complementarity in direct mechanism
- restricting attention to common prior type spaces makes no difference, and the contraction property continues to play the same role as described earlier
- Milgrom and Roberts (1991): with strategic complementarities, there are multiple equilibria if and only if there are multiple rationalizable actions

- recall the linear best response in the auction model

$$\theta_i' = \theta_i + \gamma \sum_{j \neq i} (\theta_j - \theta_j')$$

- positive interdependence in agents' types,

$$\gamma > 0$$

gives rise to strategic substitutability in direct mechanism

- it is possible even if contraction property fails

$$\frac{1}{I-1} < \gamma < 1,$$

robust implementation is possible if we restrict attention to type spaces satisfying the common prior assumption

# The Role of the Common Prior

**Theorem (2008)**

1. *If the reports are strategic complements, then robust implementation with common prior implies robust implementation without common prior.*
2. *If the reports are strategic substitutes, then robust implementation with common prior fails to imply robust implementation without common prior.*

- given restriction to common prior, incomplete information rationalizable behavior is equivalent to incomplete information correlated equilibrium behavior

- local, intermediate notions of robustness (common prior, common payoff prior, etc.)
- robust predictions for revenue maximization problems
- beyond mechanism design: robust predictions in games with private information
- perhaps we cannot make unique predictions, can we provide robust bounds on the distribution of outcomes
- strategic revealed preference