

# Robust Implementation

Dirk Bergemann and Stephen Morris  
Stony Brook Conference

July 2005

# Robust Full Implementation

Two concerns about classic Bayesian incentive compatibility results:

1. "Robustness": want to relax common knowledge assumption about structure of the type space and prior (Wilson (1987), Bergemann and Morris (2004))
2. "Full Implementation": want all equilibria, not just some equilibria, to deliver desirable outcomes (Postlewaite and Schmeidler (1986), Jackson (1991)).

In this talk and paper, we require both. Gives new and distinctive economic insights.

---

# This Talk

Consider a rich environment with independent preferences.

1. Robust full implementation (RFI) is equivalent to implementation in iterated deletion of strictly dominated strategies
2. RFI is possible if and only if it is possible in the direct mechanism
3. RFI is possible if and only if not too much interdependence of preferences.

---

# The Paper

”Robust Implementation: The Role of Large Type Spaces”

Cowles Foundation Discussion Paper 1519

General environments

1. Robust full implementation (RFI) is equivalent to implementation in iterated deletion of never best responses
2. ”Robust monotonicity” condition is necessary and almost sufficient condition for RFI
3. Robust monotonicity is equivalent to Bayesian monotonicity on all type spaces

---

## 4. Sufficient conditions under which direct RFI is possible whenever RFI is possible

## Public Good Example

- $I$  agents
- Agent  $i$  has payoff type  $\theta_i \in [0, 1]$
- Utility of agent  $i$  is

$$u_i(x, y, \theta) = \left( \theta_i + \gamma \sum_{j \neq i} \theta_j \right) x + y_i$$

where  $x \in \mathbb{R}_+$  is the level of a public good,  $y \in \mathbb{R}^I$  is a vector of cash transfers and  $\gamma \in \mathbb{R}_+$  measures the degree of interdependence in preferences.

- Social cost of providing public good is  $\frac{1}{2}x^2$ , so social welfare is

$$(1 + \gamma(I - 1)) \left( \sum_i \theta_i \right) x - \frac{1}{2}x^2$$

- Efficient level of the public good is then

$$f_0(\theta) = (1 + \gamma(I - 1)) \left( \sum_i \theta_i \right)$$

- Planner does not care about transfers
- This is an example of "payoff environment"

## Robust Mechanism Design

- Does there exist a mechanism with the property that whatever agents beliefs and higher order beliefs about other agents' types, there is an equilibrium where the efficient level of the public good is chosen?
- Our earlier paper on "Robust Mechanism Design" shows that (since this environment is "separable") this is true if and only if there are ex post incentive compatible (EPIC) transfers, so that for each  $i$ , there exists  $f_i : \Theta \rightarrow \mathbb{R}^I$  with

$$\left( \theta_i + \gamma \sum_{j \neq i} \theta_j \right) f_0(\theta_i, \theta_{-i}) + f_i(\theta_i, \theta_{-i}) \geq \left( \theta_i + \gamma \sum_{j \neq i} \theta_j \right) f_0(\theta'_i, \theta_{-i}) + f_i(\theta'_i, \theta_{-i})$$

for all  $\theta_i$ ,  $\theta'_i$  and  $\theta_{-i}$ .

- EPIC transfers always exist in this exist. Up to constants, they are:

$$f_i(\theta) = - (1 + \gamma (I - 1)) \left( \gamma \left( \sum_{j \neq i} \theta_j \right) \theta_i + \frac{1}{2} \theta_i^2 \right)$$

## Iterated Deletion in Direct Mechanism

- Consider social choice function  $f$  (with EPIC transfers)
- Consider direct mechanism where each agent reports his payoff type and  $f$  is applied to reports as if truthful
- Let  $\gamma < \frac{1}{I-1}$
- Let  $S^0(\theta_i) = [0, 1]$
- Let  $S^{k+1}(\theta_i)$  be the set of rationalizable reports for agent with type  $\theta_i$  when it is common knowledge that each type  $\theta_j$  chooses a report in  $S^k(\theta_j)$ .

- if agent  $i$  has payoff type  $\theta_i$ , has a point conjecture that other agents have type profile  $\theta_{-i}$  and report their types to be  $\theta'_{-i}$ , and he reports himself to be type  $\theta'_i$ , his expected payoff is a constant  $(1 + \gamma(I - 1))$  times

$$\left( \theta_i + \gamma \sum_{j \neq i} \theta_j \right) \left( \theta'_i + \sum_{j \neq i} \theta'_j \right) - \left( \gamma \left( \sum_{j \neq i} \theta'_j \right) \theta'_i + \frac{1}{2} (\theta'_i)^2 \right).$$

so he would wish to set

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

- So we will have

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

where

$$\begin{aligned} \bar{\beta}^k(\theta_i) &= \min \left\{ 1, \theta_i + \gamma \max_{\{(\theta'_{-i}, \theta_{-i}) : \theta'_j \in S^k(\theta_j) \text{ for all } j \neq i\}} \sum_{j \neq i} (\theta_j - \theta'_j) \right\} \\ &= \min \left\{ 1, \theta_i + \gamma \max_{\theta_{-i}} \sum_{j \neq i} (\theta_j - \underline{\beta}^{k-1}(\theta_j)) \right\}. \end{aligned}$$

Thus

$$\bar{\beta}^k(\theta_i) = \min \left\{ 1, \theta_i + (\gamma(I-1))^k \right\},$$

and  $\bar{\beta}^k(\theta_i) \downarrow \theta_i$

- Symmetric argument shows  $\underline{\beta}^k(\theta_i) \uparrow \theta_i$ .

- Thus  $\theta'_i \neq \theta_i \Rightarrow \theta'_i \notin S^k(\theta_i)$  for sufficiently large  $k$ , provided that  $\gamma < \frac{1}{I-1}$ .

## Impossibility of Implementation with too much Interdependence

- Now suppose that  $\gamma > \frac{1}{I-1}$ .
- There exist type spaces the social choice function  $f$  is not virtually implementable in any refinement of Nash equilibrium:
- whenever agent  $i$  has type  $\theta_i$ , he is convinced that the types of other players  $\theta_{-i}$  are such that

$$\sum_{j \neq i} \theta_j = \frac{1}{\gamma} \left( \frac{1}{2} - \theta_i \right).$$

Observe that  $\gamma > \frac{1}{I-1}$  implies that we can choose the  $\theta_j$  to be in the interval  $[0, 1]$ . Agent  $i$ 's preferences are independent of his type on this type space. Now fix any mechanism and restrict each player to a pooling strategy, i.e., sending the same message independent of his type. Since all types now have identical preferences over outcomes, this pooling strategy is an equilibrium.

## Payoff Relevant Environment

- agent  $i \in \{1, 2, \dots, I\}$
- $i$ 's payoff relevant type  $\theta_i \in \Theta_i$
- payoff relevant type profile  $\theta \in \Theta = \Theta_1 \times \dots \times \Theta_I$
- social outcome  $a \in A$
- utility function  $u_i : A \times \Theta \rightarrow \mathbb{R}$
- social choice correspondence  $F : \Theta \rightarrow 2^A$

## Mechanism and Ex Post Implementation

- A mechanism  $\mathcal{M} = \left( (M_i)_{i=1}^I, g \right)$ ,  $g : M_1 \times \dots \times M_I \rightarrow A$
- A pure strategy in payoff types game:  $s_i : \Theta_i \rightarrow M_i$
- Strategy profile  $s$  is an *ex post equilibrium* if

$$u_i(g(s(\theta)), \theta) \geq u_i(g(m_i, s_{-i}(\theta_{-i})), \theta)$$

for all  $i$ ,  $\theta$  and  $m_i$ .

## Type Space

$$\mathcal{T} = \left\{ T_i, \hat{\theta}_i, \hat{\pi}_i \right\}_{i=1}^I$$

- $i$ 's type is  $t_i \in T_i$
- $\hat{\theta}_i(t_i)$  is  $i$ 's payoff relevant type of  $t_i$

$$\hat{\theta}_i : T_i \rightarrow \Theta_i$$

- $\hat{\pi}_i(t_i)$  is  $i$ 's belief type of  $t_i$

$$\hat{\pi}_i : T_i \rightarrow \Delta(T_{-i})$$

## Mechanism

- A mechanism  $\mathcal{M}$  and a type space  $\mathcal{T}$  is an incomplete information game
- A pure strategy  $s_i : T_i \rightarrow M_i$
- Strategy profile  $s$  is a (Bayesian) equilibrium if

$$\begin{aligned} & \sum_{t_{-i}} \hat{\pi}_i(t_i) [t_{-i}] u_i \left( g(s(t)), \hat{\theta}(t) \right) \\ & \geq \sum_{t_{-i}} \hat{\pi}_i(t_i) [t_{-i}] u_i \left( g(m_i, s_{-i}(t_{-i})), \hat{\theta}(t) \right) \end{aligned}$$

for all  $i, t_i$ .

## Ex Post Incentive Compatibility

DEFINITION: A social choice function  $f : \Theta \rightarrow A$  is ex post incentive compatible if, for all  $i$  and  $\theta \in \Theta$ ,

$$u_i (f (\theta) , \theta) \geq u_i (f (\theta'_i, \theta_{-i}) , \theta)$$

for all  $\theta'_i \in \Theta_i$ .

## Robust Mechanism Design

The environment is *separable* if

$$\begin{aligned}
 A &= X \times Y_1 \times \dots \times Y_I \\
 u_i((x, y), \theta) &= v_i(x, y_i, \theta) \\
 \text{and } F(\theta) &= f_0(\theta) \times F_1(\theta) \times \dots \times F_I(\theta)
 \end{aligned}$$

where  $f_0$  is a function.

**THEOREM.** If the environment is separable, there exists a mechanism  $\mathcal{M}$  such that for every type space  $\mathcal{T}$ , there is an equilibrium  $s$  of  $(\mathcal{M}, \mathcal{T})$  consistent with  $F$  (i.e.,  $g(s(t)) \in F(\hat{\theta}(t))$  for all  $t$ ) if and only if there exists  $f \in F$  that is ex post incentive compatible.

## Example

- $\Theta_1 = \{\theta_1, \theta'_1\}$
- $\Theta_2 = \{\theta_2, \theta'_2\}$
- $A = \{a, b, c\}$
- Payoffs

a	$\theta_2$	$\theta'_2$	b	$\theta_2$	$\theta'_2$	c	$\theta_2$	$\theta'_2$
$\theta_1$	1, 0	-1, 2	$\theta_1$	-1, 2	1, 0	$\theta_1$	0, 0	0, 0
$\theta'_1$	0, 0	0, 0	$\theta'_1$	0, 0	0, 0	$\theta'_1$	1, 1	1, 1

## Example

### Social Choice Correspondence

$F$	$\theta_2$	$\theta'_2$
$\theta_1$	$\{a, b\}$	$\{a, b\}$
$\theta'_1$	$\{c\}$	$\{c\}$

## Example

Agent 1's ex post incentive constraints require

$F$	$\theta_2$	$\theta'_2$
$\theta_1$	$\{a\}$	$\{b\}$
$\theta'_1$	$\{c\}$	$\{c\}$

This violates agent 2's ex post incentive constraints....

## Two Questions

- Robust Mechanism Design: Does there exist a mechanism  $\mathcal{M}$  such that for every type space  $\mathcal{T}$ , there is an equilibrium  $s$  of  $(\mathcal{M}, \mathcal{T})$  consistent with  $F$ , i.e.,

$$g(s(t)) \in F(\hat{\theta}(t))$$

for all  $t$ .

- Robust Full Implementation: Does there exist a mechanism  $\mathcal{M}$  such that for every type space  $\mathcal{T}$ , there exists an equilibrium of  $(\mathcal{M}, \mathcal{T})$  and every equilibrium is consistent with  $F$ .

## Nice Environment

- ASSUMPTION 1: Aggregator Single Crossing Preferences

$$u_i(y, \theta) = v_i(y, \mu_i(\theta))$$

where  $\mu_i$  is increasing (and strictly increasing in  $\theta_i$ ) and  $v$  satisfies the single crossing property: If  $\mu_i > \mu'_i$ ,  $v_i(y', \mu_i) > v_i(y, \mu_i)$  and  $v_i(y', \mu'_i) = v_i(y, \mu'_i)$ , then  $v_i(y', \mu''_i) > v_i(y, \mu''_i)$  for all  $\mu''_i \in (\mu'_i, \mu_i)$ .

Let  $\beta_i : \Theta_i \rightarrow 2^{\Theta_i}$  with  $\theta_i \in \beta_i(\theta_i)$ . A deception is a profile  $\beta = (\beta_1, \dots, \beta_I)$ . The trivial deception has  $\beta_i(\theta_i) = \{\theta_i\}$  for all  $i$  and  $\theta_i$ .

DEFINITION: Aggregator functions  $\mu$  satisfy the contraction property if and only for all non-trivial deceptions  $\beta$ , there exists  $i$ ,  $\theta'_i \in \beta_i(\theta_i)$  with

$\theta'_i \neq \theta_i$  such that

$$sn(\theta_i - \theta'_i) = sn(\mu_i(\theta_i, \theta_{-i}) - \mu_i(\theta'_i, \theta'_{-i}))$$

for all  $\theta'_{-i} \in \beta_{-i}(\theta_{-i})$ .

Note that if  $\mu_i(\theta) = \theta_i + \gamma \sum_{j \neq i} \theta_j$ , then the contraction property is satisfied only if  $\gamma < \frac{1}{I-1}$ .

- ASSUMPTION 2: Strict EPIC

$$u_i(f(\theta), \theta) > u_i(f(\theta'_i, \theta_{-i}), \theta)$$

for all  $i$ ,  $\theta_i \neq \theta'_i$  and  $\theta_{-i}$ .

ASSUMPTION 3: Strictly concave deviations:  $u_i (f (\theta'_i, \theta_{-i}), \theta)$  is strictly concave in  $\theta'_i$ .

PROPOSITION: Under assumptions 1, 2 and 3, RFI is possible if and only if RFI is possible in the direct mechanism. RFI is possible in the direct mechanism if and only if the contraction property is satisfied.