Robust Implementation: The Role of Large Type Spaces*

Dirk Bergemann† Stephen Morris‡

First Version: March 2003
This Version: April 2004

Abstract

We analyze the problem of fully implementing a social choice function when the planner does not know the agents' beliefs about other agents' types.

We identify an *ex post monotonicity* condition that is necessary and - in economic environments - sufficient for full implementation in ex post equilibrium; we also identify an ex post monotonicity no veto condition that is sufficient. These results are the ex post equilibrium analogues of Jackson's (1991) results about Bayesian implementation.

We show by example that ex post monotonicity implies neither Maskin monotonicity (necessary and almost sufficient for complete information implementation) nor - for some type spaces - interim monotonicity (i.e., the Bayesian monotonicity condition that is necessary and almost sufficient for Bayesian implementation). We identify a *robust monotonicity* condition that is equivalent to interim monotonicity on all type spaces; robust monotonicity implies both Maskin monotonicity and ex post monotonicity.

Robust monotonicity is necessary for interim implementation on all type spaces and is sufficient for interim implemention on all common support type spaces when there are at least three agents and an economic condition is satisfied. Without a common support restriction, we show that interim implementation on all type spaces is equivalent to implementation under an ex post version of dominance solvability.

KEYWORDS: Mechanism Design, Implementation, Common Knowledge, Universal Type Space, Interim Equilibrium, Ex-Post Equilibrium, Dominant Strategies.
JEL CLASSIFICATION: C79, D82

*This research is supported by NSF Grant #SES-0095321. The first author gratefully acknowledges support through a DFG Mercator Research Professorship at the Center of Economic Studies at the University of Munich. We benefited from discussion with Amanda Friedenberg, Matt Jackson and Mike Riordan. We would like to thank seminar audiences at Caltech, Columbia University, Cornell Univerity, New York University and the University of Michigan for helpful comments. Parts of this paper were reported in early drafts of our work on Robust Mechanism Design (Bergemann and Morris (2001)).

†Department of Economics, Yale University, 28 Hillhouse Avenue, New Haven, CT 06511, dirk.bergemann@yale.edu.

‡Department of Economics, Yale University, 30 Hillhouse Avenue, New Haven, CT 06511, stephen.morris@yale.edu.

Contents

1 Introduction

This paper looks at the problem of *fully* implementing a social choice function when agents have interdependent values. Thus each agent has a payoff type. The agents have preferences over outcomes that depend on the profile of payoff types. The planner does not know the agents' types but must choose a mechanism such that in *every* equilibrium of the mechanism, agents play of the game results in the outcome specified by the social choice function at every payoff type profile. This problem has been analyzed under the assumption of complete information, i.e., there is common knowledge among the agents of their payoff types (e.g., Maskin (1999)). It has also been analyzed under the assumption of incomplete information, on the assumption that there is a fixed type space and there is common knowledge among the agents of the prior (or the priors) according to which agents form their beliefs (e.g., Jackson (1991)). We want to analyze the problem of full implementation under the assumption that the planner knows nothing about what agents know or believe about other agents' payoff types, or their higher order beliefs. We believe that by fixing a small type space and assuming common knowledge among the agents of the type space and agents' beliefs on the type space, researchers have been making very strong implicit assumptions. We would like to relax those assumptions.

There has recently been much interest in the literature on using the concept of ex post equilibrium since it seems unrealistic to allow the mechanism to depend on the planner's knowledge of the type space (e.g., Dasgupta and Maskin (2000)). We provide a complete analysis of full implementation in ex post equilibrium. We introduce an ex post monotonicity condition that - along with ex post incentive compatibility - is necessary for ex post implementation. We show that a slight strengthening of ex post monotonicity - the ex post monotonicity no veto condition - is sufficient for implementation with at least three agents. The latter condition reduces to ex post monotonicity in economic environments. These results are the ex post analogues of the Bayesian implementation results of Jackson (1991), and we employ similar arguments to establish our results.

However, for full implementation using a strong solution concept does not necessarily imply stronger results: the fact that non truth-telling behavior may fail the stringent requirement of being an ex post equilibrium may make implementation easier. We show in an economic example that ex post monotonicity may hold even when both Maskin monotonicity (the necessary condition for complete information implementation) and interim monotonicity on a fixed type space (the necessary condition for interim implementation) fail. Thus ex post implementation is possible even when complete information implementation and interim incomplete information implementation are impossible.

We therefore find a condition - robust monotonicity - that is equivalent to requiring interim monotonicity on every type space. Suppose that we fix a "deception" specifiying, for each payoff type of each agent, a set of types that he might misreport himself to be. We require that for some agent i and a type misreport of agent i under the deception, for every misreport θ'_{-i} that that the other agents might make under the deception, there exists an outcome y which is strictly preferred by agent i to the outcome he would receive under the social choice function for *every* possible payoff type profile that might misreport θ'_{-i}; where this outcome y satisfies the extra restriction that no payoff type of agent i prefers outcome y to the social choice function if the other agents were really types θ'_{-i}. This condition - while a little convoluted - is a somewhat easier to interpret than the interim (Bayesian) monotonicity conditions. It is very strong and implies both Maskin monotonicity and ex post monotonicity conditions (but is strictly weaker than dominant strategies).

Robust monotonicity is necessary for interim implementation on all type spaces and is sufficient for interim implementation on all common support type spaces when there are at least three agents and an economic condition is satisfied. We show that interim implementation on all type spaces is possible if and only if it is possible to implement the social choice function using an ex post iterative deletion procedure: we fix a mechanism and iteratively delete messages for each payoff type that

are strictly dominated by another message for each payoff type profile and message profile that has survived the procedure. This requirement is stronger than robust monotonicity.

This last result about iterative deletion illustrates a general point well-known from the literature on epistemic foundations of game theory (e.g., Brandenburger and Dekel (1987), Battigalli and Siniscalchi (2003)): equilibrium solution concepts only have bite if we make strong assumptions about type spaces, i.e., we assume small type spaces where the common prior assumption holds. Our uniform implementation result says that equilibrium has no bite (relative to iterated deletion of strictly dominated strategies) if we allow for sufficiently rich type spaces.

The results in this paper concern full implementation. An earlier companion paper of ours (Bergemann and Morris (2003)) addresses the analogous questions of robustness to rich type spaces, but looking at the question of partial implementation, i.e., does there exist a mechanism such that *some* equilibrium implements the social choice function. We showed that ex post (partial) implementation of the social choice function is a necessary and sufficient condition for partial implementation on all type spaces. This paper establishes that an analogous result does not hold for full implementation. In that paper, we also looked at the partial implementation of social choice correspondences, but showed that partial implementation on all type spaces was sometimes easier than ex post partial implementation. We leave for future work the question of full implementation of social choice correspondences on large type spaces.

In the special case of private values, ex post incentive compatibility is equivalent to dominant strategies incentive compatibility and thus partial implementation on all type spaces implies dominant strategy implementation. But strictly dominant strategy implementation is a sufficient condition for full implementation. Thus in the private values case, moving to the stronger solution concept of ex post equilibrium / dominant strategies is always (up to the dominant / strictly dominant strategies distinction) a more stringent requirement. This paper shows that this well known observation does not translate to an interdependent values setting.

The paper is organized as follows. Section 2 describes a simple example that illustrates some of the key points in the paper. Section 3 describes the formal environment and solution concepts. Section 4 introduces our notion of ex post monotonicity and compares it to Maskin monotonicity. Section 5 reports our analysis of the ex post implementation problem. Section 6 introduces interim monotonicity and robust monotonicity, and characterizes how the monotonicity conditions relate to each other using propositions and examples. Section 7 presents our results on interim implementation on all type spaces and reports results on uniform implementability. Section 9 concludes.

2 Example A

Consider the following interdependent values social choice setting. There are two agents 1 and 2. Each agent has two possible payoff types, $\Theta_1 = \left\{\theta_1, \theta_1'\right\}$ and $\Theta_2 = \left\{\theta_2, \theta_2'\right\}$. There are four possible social outcomes, $A = \{a, b, c, d\}$. The payoffs of the two agents are given by:

a	θ_2	θ_2'
θ_1	$3,3$	$0,0$
θ_1'	$0,0$	$1,1$

b	θ_2	θ_2'
θ_1	$0,0$	$3,3$
θ_1'	$1,1$	$0,0$

c	θ_2	θ_2'
θ_1	$0,0$	$1,1$
θ_1'	$3,3$	$0,0$

d	θ_2	θ_2'
θ_1	$1,1$	$0,0$
θ_1'	$0,0$	$3,3$

Notice that the agents have identical interests and, for each payoff type profile, have a unique preferred outcome. The social choice function f will select that outcome:

f	θ_2	θ_2'
θ_1	a	b
θ_1'	c	d

We are interested in a setting where all this information is common knowledge among the agents and the planner, but the planner knows nothing about the agents' beliefs and higher order beliefs about each others' types. What can the planner do?

2.1 Ex Post Implementation

One approach to this problem is to focus attention on ex post implementation. That is, suppose the planner seeks a mechanism whose ex post equilibria implement f. Since ex post equilibria are independent of agents' beliefs about other agents' types, this is one way of dealing with the lack of common knowledge. We first analyze this approach.

Observe that the social choice function is ex post incentive compatible. Thus if the planner simply invites the agents to announce their payoff types, each agent will have an incentive to tell the truth as long as he expect others to do so, whatever his beliefs about the other agents' types. Thus truth telling is an ex post equilibrium of the payoff type direct mechanism.

However, this game also has another ex post equilibrium where each type of each agent always misreports his type. But it is easy to construct a simple augmented mechanism where all (pure strategy) ex post equilibria yield desirable outcomes.[1] Consider the mechanism where agent 2 simply announces his payoff type; and agent 1 announces his payoff type and also announces either "truth" or "lie" (with the interpretation that the latter announcement is agent 1's announcement about whether he believes agent 2 has told the truth). This mechanism can be represented by the following table:

	θ_2	θ_2'
(θ_1, truth)	a	b
(θ_1', truth)	c	d
(θ_1, lie)	b	a
(θ_1', lie)	d	c

(1)

What are the (pure strategy) ex post equilibria of this game? In any ex post equilibrium, type θ_2 of agent 2 must announce θ_2 or θ_2'. If type θ_2 of agent 2 announces θ_2, then type θ_1 of agent 1 must announce (θ_1, truth) and type θ_1' of agent 1 must announce (θ_1', truth); so type θ_2' of agent 2 must announce θ_2'.

On the other hand, if type θ_2 of agent 2 announces θ_2', then type θ_1 of agent 1 must announce (θ_1, lie) and type θ_1' of agent 1 must announce (θ_1', lie); so type θ_2' of agent 2 must announce θ_2. Thus there are two possible ex post equilibria and both implement the social choice function.

Thus for this example, we have shown the possibility of ex post implementation. Theorem 1 in Section 5 identifies an ex post monotonicity condition that is necessary for ex post implementation; we also show that this condition is sufficient if there are at least three agents in an economic environment and that a slightly stronger ex post monotonicity no veto condition is sufficient in non-economic environments.

[1] Mechanisms of this form - where the augmented mechanism contains a copy of the direct mechanism - are common in the implementation literature; Mookerjee and Reichelstein (1990) refer to them as "augmented direct mechanisms."

2.2 Interim Implementation

We can also analyze whether interim implementation is possible on different type spaces. Suppose that agents had the following type space:

	t_2	t_2'	t_2''	t_2'''	
t_1	$\frac{1}{8}(1-\varepsilon)$	$\frac{1}{8}(1-\varepsilon)$	$\frac{1}{8}\varepsilon$	$\frac{1}{8}\varepsilon$	θ_1
t_1'	$\frac{1}{8}(1-\varepsilon)$	$\frac{1}{8}(1-\varepsilon)$	$\frac{1}{8}\varepsilon$	$\frac{1}{8}\varepsilon$	θ_1'
t_1''	$\frac{1}{8}\varepsilon$	$\frac{1}{8}\varepsilon$	$\frac{1}{8}(1-\varepsilon)$	$\frac{1}{8}(1-\varepsilon)$	θ_1
t_1'''	$\frac{1}{8}\varepsilon$	$\frac{1}{8}\varepsilon$	$\frac{1}{8}(1-\varepsilon)$	$\frac{1}{8}(1-\varepsilon)$	θ_1'
	θ_2	θ_2'	θ_2	θ_2'	

where $\varepsilon < \frac{1}{2}$. The four types of agent 1 are represented as rows, the four types of agent 2 are represented as columns and the numbers represent the prior on type profiles. The payoff type of a given type is recorded at the end of his row/column. If this is the true type space and agents are invited to play the augmented mechanism (1), then there is clearly a strict pure strategy interim equilibrium where agents follow strategies:

$$s_1(\cdot) = \begin{cases} (\theta_1, \text{truth}) & \text{if} \quad t_1 \\ (\theta_1', \text{truth}) & \text{if} \quad t_1' \\ (\theta_1, \text{lie}) & \text{if} \quad t_1'' \\ (\theta_1', \text{lie}) & \text{if} \quad t_1''' \end{cases}$$

and

$$s_2(\cdot) = \begin{cases} \theta_2 & \text{if} \quad t_2 \\ \theta_2' & \text{if} \quad t_2' \\ \theta_2' & \text{if} \quad t_2'' \\ \theta_2 & \text{if} \quad t_2''' \end{cases}$$

To see why this is an equilibrium, note that if $\varepsilon = 0$, then we have disjoint type spaces consisting of types $(t_1, t_1'; t_2, t_2')$; and types $(t_1'', t_1'''; t_2'', t_2''')$, respectively and the above type space reduces to:

	t_2	t_2'	t_2''	t_2'''	
t_1	$\frac{1}{8}$	$\frac{1}{8}$	0	0	θ_1
t_1'	$\frac{1}{8}$	$\frac{1}{8}$	0	0	θ_1'
t_1''	0	0	$\frac{1}{8}$	$\frac{1}{8}$	θ_1
t_1'''	0	0	$\frac{1}{8}$	$\frac{1}{8}$	θ_1'
	θ_2	θ_2'	θ_2	θ_2'	

In this new type space, the types in the first disjoint type space $(t_1, t_1'; t_2, t_2')$ play according to one ex post equilibrium of the augmented mechanism (1), whereas the types in the second disjoint type space $(t_1'', t_1'''; t_2'', t_2''')$ play according to the other ex post equilibrium. Given the strict incentives, allowing ε to be positive but small does not stop these strategies being an equilibrium. But now, with probability ε, there is miscoordination.

This example illustrates one important message of this paper: there is a significant gap between ex post implementation and interim implementation. It is sometimes easier to ex post implement than to interim implement. In this example, there is no mechanism that interim implements f on every type space. Here is an informal argument by contradiction (the formal argument appears in Section 7).

We first claim that a mechanism interim implements f on every type space if and only if it iteratively implements f in the following sense. Iteratively delete for each payoff type all messages that were not best responses to some belief over payoff type - message pairs of the opponent that have not yet been deleted. There is iterative implementation of f if, for any payoff type profile,

every surviving message profile is consistent with f. To prove the harder "only if" part of the claim, construct a type space where each player has a type corresponding to every payoff type - message pair that survive the iterated elimination. For each payoff type-message pair surviving the iterated deletion, there is a belief over the surviving payoff type - message pairs of the opponent such that that message is a best response for that payoff type. Thus we can construct beliefs on the type space such that it is an equilibrium for each type to send the message with which it is labelled.

Now we argue that iterative implementation is not possible for our example. First, note that for each type θ_i, there is at least one message (call it $m_i^*(\theta_i)$) with the property that $g(m^*(\theta)) = f(\theta)$ for each θ. Also observe that message $m_i^*(\theta_i)$ is never deleted for type θ_i. There must be a first round - call it round n - when message $m_i^*(\theta_i)$ is deleted for type θ_i', for some i. Thus in the previous round, $m_j^*(\theta_j)$ had not been deleted for type θ_j'. Now if type θ_i' conjectures that his opponent is type θ_j' sending message $m_j^*(\theta_j)$, then his payoff to sending message $m_i^*(\theta_i)$ is 1. Since this is not a best response, there must exist another message $\widehat{m}_i$ such that $g_i\left(\widehat{m}_i, m_j^*(\theta_j)\right) = f\left(\theta_i', \theta_j'\right)$. But now this message $\widehat{m}_i$ can never be deleted for type θ_i', a contradiction.

3 The Implementation Problem

We fix a finite set of agents, $1, 2, ..., I$. Agent i's *payoff type* is $\theta_i \in \Theta_i$, where Θ_i is a finite set. We write $\theta \in \Theta = \Theta_1 \times ... \times \Theta_I$. There is a set of outcomes Y. Each agent has utility function $u_i : Y \times \Theta \to \mathbb{R}$. Thus we are in the world of interdependent types, where an agent's utility depends on other agents' payoff types. A social choice function is a mapping $f : \Theta \to Y$. If the true payoff type profile is θ, the planner would like the outcome to be $f(\theta)$. In this paper, we restrict our analysis to the implementation of a social choice function rather than a social choice correspondence or set.

We are interested in analyzing behavior in a variety of type spaces, including richer sets of types than payoff types. For this purpose, we shall refer to agent i's *type* as $t_i \in T_i$, where T_i is a finite set.[2] A type of agent i must include a description of his payoff type. Thus there is a function $\widehat{\theta}_i : T_i \to \Theta_i$ with $\widehat{\theta}_i(t_i)$ being agent i's payoff type when his type is t_i. A type of agent i must also include a description of his beliefs about the types of the other agents; thus there is a function $\widehat{\pi}_i : T_i \to \Delta(T_{-i})$ with $\widehat{\pi}_i(t_i)$ being agent i's *belief type* when his type is t_i. Thus $\widehat{\pi}_i(t_i)[t_{-i}]$ is the probability that type t_i of agent i assigns to other agents having types t_{-i}. A *type space* is a collection:

$$\mathcal{T} = \left(T_i, \widehat{\theta}_i, \widehat{\pi}_i\right)_{i=1}^{I}.$$

The type space is a *common support* type space if there exists $T^* \subseteq T$ such that

$$\widehat{\pi}_i(t_i)[t_{-i}] > 0 \Leftrightarrow (t_i, t_{-i}) \in T^*.$$

The type space if a *common prior* type space if there exists $p \in \Delta(T)$ such that

$$\widehat{\pi}_i(t_i)[t_{-i}] = \frac{p(t_i, t_{-i})}{\sum\limits_{t_{-i}'} p\left(t_i, t_{-i}'\right)}.$$

The type space is a *payoff* type space if, for each i, $T_i = \Theta_i$ and $\widehat{\theta}_i$ is the identity map.

A planner must choose a *game form* or *mechanism* for the agents to play in order the determine the social outcome. Let M_i be the countably infinite set of messages available to agent i.[3] Let

[2] The finite set restriction clarifies the relation to the existing literature. In Section 9, we discuss what happens if we allow for uncountable type spaces.

[3] This assumption clarifies the relation with the existing literature. We discuss in Section 9 what happens if we restrict attention to finite messages or allow larger sets of messages.

$g(m)$ be the outcome if action profile m is chosen. Thus mechanisms do not involve randomization contingent on the message profile. But randomization can be built into the outcome space Y. Thus a mechanism is a collection

$$\mathcal{M} = (M_1, ..., M_I, g(\cdot)),$$

where $g : M \to Y$.

Now holding fixed the payoff environment, we can combine a type space $\mathcal{T}$ with a mechanism $\mathcal{M}$ to get an incomplete information game $(\mathcal{T}, \mathcal{M})$.

We are interested in a setting where the planner does not know the payoff types of the agents and knows nothing about agents' beliefs and higher order beliefs about other agents' types. Two approaches to this problem are to look at ex post equilibria of the game with payoff types; or we can look at interim (Bayesian Nash) equilibria on a variety of richer type spaces. We consider each in turn.

3.1 Ex Post Equilibrium

Consider the "payoff types game" where each agent's possible types are Θ_i. Thus we have an incomplete information game where agent i's payoff if message profile m is sent and payoff type profile θ is realized is

$$u_i(g(m), \theta).$$

A pure strategy in this game is a function $s_i : \Theta_i \to M_i$.

Definition 1 (Ex post equilibrium)
A pure strategy profile $s = (s_1, ..., s_I)$ is an ex post equilibrium of the payoff types game if

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

for all i, θ and m_i.[4]

Definition 2 (Ex post implementation)
Social choice function f is ex post implementable if there exists a mechanism $\mathcal{M}$ such that every (pure strategy) ex post equilibrium s of the game $\mathcal{M}$ satisfies

$$g(s(\theta)) = f(\theta).$$

We restrict attention to pure strategy equilibria. This helps make comparisons with the existing literature (where the assumption is standard). However, when we conduct analysis allowing rich type spaces, the restriction will not bite. This issue is discussed in detail in Section 9.

3.2 Interim Equilibrium

Next we consider an incomplete information game with an arbitrary type space $\mathcal{T}$ and a mechanism $\mathcal{M}$. The payoff of agent i if message profile m is chosen and type profile t is realized is then given by

$$u_i\left(g(m), \widehat{\theta}(t)\right).$$

A pure strategy for agent i in the incomplete information game $(\mathcal{T}, \mathcal{M})$ is given by

$$s_i : T_i \to M_i.$$

Pure strategy (interim, or Bayesian Nash) equilibria are defined in the usual way.

[4] Ex post incentive compatibility was discussed as "uniform incentive compatibility" by Holmstrom and Myerson (1983). Ex post equilibrium is increasingly studied in game theory (see Kalai (2002)) and is often used in mechanism design as a more robust solution concept (Cremer and McLean (1985), Dasgupta and Maskin (2000), Perry and Reny (2002), Bergemann and Valimaki (2002)).

Definition 3 (Interim equilibrium)
A pure strategy profile $s = (s_1, ..., s_I)$ is an interim equilibrium of the game $(\mathcal{T}, \mathcal{M})$ if

$$\sum_{t_{-i} \in T_{-i}} u_i \left(g\left(s\left(t\right)\right), \widehat{\theta}\left(t\right) \right) \widehat{\pi}_i\left(t_i\right)\left[t_{-i}\right] \geq \sum_{t_{-i} \in T_{-i}} u_i \left(g\left(\left(m_i, s_{-i}\left(\theta_{-i}\right)\right)\right), \widehat{\theta}\left(t\right) \right) \widehat{\pi}_i\left(t_i\right)\left[t_{-i}\right]$$

for all i, t_i and m_i.

Definition 4 (Interim Implementation)
Social choice function f is interim implementable on type space $\mathcal{T}$ if there exists a mechanism $\mathcal{M}$ such that every (pure strategy) equilibrium s of the game $(\mathcal{T}, \mathcal{M})$ satisfies

$$g\left(s\left(t\right)\right) = f\left(\widehat{\theta}\left(t\right)\right)$$

for all t.

4 Maskin Monotonicity and Ex Post Monotonicity

The existing literature on complete information and Bayesian implementation identifies "monotonicity" conditions that are necessary and "almost sufficient" for implemention. It is useful to introduce our notion of ex post monotonicity by comparing it with Maskin monotonicity.

4.1 Maskin Monotonicity

Maskin (1999) introduced a celebrated monotonicity notion for the complete information environment which constitutes a necessary and almost sufficient condition for complete information implementation.

Definition 5 (Maskin monotonicity)
Social choice function f is (Maskin) monotone, if

$$u_i\left(f\left(\theta'\right), \theta'\right) \geq u_i\left(y, \theta'\right) \Rightarrow u_i\left(f\left(\theta'\right), \theta\right) \geq u_i\left(y, \theta\right)$$

for all i and y, then

$$f\left(\theta'\right) = f\left(\theta\right).$$

"In words, monotonicity requires that if alternative x is f optimal with respect to some profile of preferences and the profile is then altered so that, in each individual's ordering a does not fall below any alternative that it was not below before, then x remains f optimal with respect to the new profile." (Maskin (1999)). Maskin monotonicity is necessary for complete information implementation and, when there are at least three agents and no veto power holds, also sufficient.

To motivate the monotonicity notions of this paper, it is useful to re-write this statement. First, we can give the equivalent contrapositive statement: if $f\left(\theta\right) \neq f\left(\theta'\right)$, then there exists i and y such that

$$u_i\left(f\left(\theta'\right), \theta'\right) \geq u_i\left(y, \theta'\right)$$

and

$$u_i\left(y, \theta\right) > u_i\left(f\left(\theta'\right), \theta\right).$$

Also, it is useful to think of the agents in a complete information setting engaging in a "deception" where they misreport the true type profile in a coordinated way. Write

$$\alpha : \Theta \to \Theta$$

for the common deception strategy. Now Maskin monotonicity requires that for every deception with $f \circ \alpha \neq f$, then there exists i, θ, and y such that

$$u_i(y, \theta) > u_i(f(\alpha(\theta)), \theta), \tag{2}$$

while

$$u_i(f(\alpha(\theta)), \alpha(\theta)) \geq u_i(y, \alpha(\theta)). \tag{3}$$

This alternative statement suggests a rather intuitive description why monotonicity is a necessary condition for implementation. Suppose that f is complete information implementable. Then if the agents were to deceive the designer by misreporting $\alpha(\theta)$ rather than reporting truthfully θ and if the deception $\alpha(\theta)$ would lead to a different allocation, i.e. $f(\alpha(\theta)) \neq f(\theta)$, then the designer should be able to fend off the deception. This requires that there is some agent i and profile θ such that the designer can offer agent i a reward y for denouncing the deception $\alpha(\theta)$ by the agents if the true type profile is θ. Yet, at the same time, the designer has be aware that the reward could be used in the wrong circumstances, namely when the true payoff type profile is $\alpha(\theta)$ and it is indeed reported to be $\alpha(\theta)$. The first strict inequality (2) then guarantees the existence of a whistle-blower, whereas the second weak inequality (3) guarantees incentive compatible behavior by the whistle-blower. Both these features will re-appear in all the monotonicity conditions studied in this paper.

4.2 Ex Post Monotonicity

With incomplete information, a *deception* - i.e., a non truth-telling strategy in the direct mechanism a *deception*, is a collection $\alpha = (\alpha_1, ..., \alpha_I)$, each $\alpha_i : \Theta_i \to \Theta_i$ and

$$\alpha(\theta) = (\alpha_1(\theta_1), ..., \alpha_I(\theta_I)).$$

In a direct revelation game α_i would indicate i's reported type as a function of his true type. For a direct revelation mechanism, if agents report the deception α rather than truthfully, then the resulting social outcome is given by $f(\alpha(\theta))$ rather than $f(\theta)$. We write $f \circ \alpha(\theta) \equiv f(\alpha(\theta))$. For any profile of payoff types of agents other than i, we write $Y_i^*(\theta_{-i})$ for the set of allocations that make agent i worse off than under the social choice function at all of his payoff types. So

$$Y_i^*(\theta_{-i}) \equiv \left\{ y : u_i\left(f\left(\theta_i', \theta_{-i}\right), \left(\theta_i', \theta_{-i}\right)\right) \geq u_i\left(y, \left(\theta_i', \theta_{-i}\right)\right), \ \forall \theta_i' \in \Theta_i. \right\}. \tag{4}$$

Definition 6 (Ex-post monotonicity)
Social choice function f satisfies ex post monotonicity (EM) if for every deception α with $f \neq f \circ \alpha$, there exists i, θ and $y \in Y_i^*(\alpha_{-i}(\theta_{-i}))$ such that

$$u_i(y, \theta) > u_i(f(\alpha(\theta)), \theta). \tag{5}$$

At first glace, ex post monotonicity looks like a stronger requirement than Maskin monotonicity, since the whistle-blowing constraint for Maskin monotonicity (2) stays the same, while the single incentive compatibility constraint (3) is replaced by the requirement that $y \in Y_i^*(\alpha_{-i}(\theta_{-i}))$, which implies a family of constraints,

$$u_i\left(f\left(\theta_i', \alpha_{-i}(\theta_{-i})\right), \left(\theta_i', \alpha_{-i}(\theta_{-i})\right)\right) \geq u_i\left(y, \left(\theta_i', \alpha_{-i}(\theta_{-i})\right)\right) \ \forall \theta_i' \in \Theta_i. \tag{6}$$

But because of the coordination built into the complete information deceptions, it becomes harder to find a reward y for Maskin monotonicity than for ex post monotonicity. In Section 6, we will describe an example that is Maskin monotonic and not ex post monotonic; and another example that is ex post monotonic but not Maskin monotonic.

5 Ex Post Implementation

We present necessary and sufficient conditions for a social choice function f to be ex-post implementable in the payoff type space. Our results extend the work of Maskin (1999) for complete information implementation and Jackson (1991) on Bayesian implementation (i.e., interim implementation on a fixed type space) to the notion of ex post equilibrium.

If we were just interested in partially implementing f - i.e., constructing a mechanism with an ex post equilibrium achieving f - then by the revelation principle we could restrict attention to direct mechanisms and a necessary and sufficient condition is the following ex post incentive compatibility condition.

Definition 7 (Ex Post Incentive Compatibility)
Social choice function f is ex post incentive compatible (EPIC) if

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

for all i, θ and θ_i'.

Ex post incentive and monotonicity conditions are necessary conditions for ex post implementation.

Theorem 1 (Necessity)
If f is ex post implementable, then it satisfies $(EPIC)$ and (EM).

Proof. Let (M,g) implement f with equilibrium strategies $s_i:\Theta_i\to M_i$. Consider any $i,\theta_i'\in\Theta_i$. Since s is an equilibrium,

$$u_i\left(g\left(s\left(\theta\right)\right),\theta\right) \geq u_i\left(g\left(s_i\left(\theta_i'\right),s_{-i}\left(\theta_{-i}\right)\right),\theta\right)$$

for all $\theta\in\Theta$. Noting that $g\left(s_i\left(\theta_i'\right),s_{-i}\left(\theta_{-i}\right)\right)=f\left(\theta_i',\theta_{-i}\right)$ establishes $(EPIC)$.
Suppose that for some deception α, $f\neq f\circ\alpha$. It must be that $s\circ\alpha$ is not an equilibrium at some $\theta\in\Theta$. Therefore there exists i and $m_i\in M_i$ such that we have

$$u_i\left(g\left(m_i,s_{-i}\left(\alpha_{-i}\left(\theta_{-i}\right)\right)\right),\theta\right) > u_i\left(g\left(s\left(\alpha\left(\theta\right)\right)\right),\theta\right)$$

Let $y\triangleq g\left(m_i,s_{-i}\left(\alpha_{-i}\left(\theta_{-i}\right)\right)\right)$. Then, from above,

$$u_i\left(y,\theta\right) > u_i\left(f\left(\alpha\left(\theta\right)\right),\theta\right).$$

But since s is an equilibrium it follows that

$$\begin{aligned} u_i\left(f\left(\theta_i',\alpha_{-i}\left(\theta_{-i}\right)\right),\left(\theta_i',\alpha_{-i}\left(\theta_{-i}\right)\right)\right) &= u_i\left(g\left(s\left(\theta_i',\alpha_{-i}\left(\theta_{-i}\right)\right)\right),\left(\theta_i',\alpha_{-i}\left(\theta_{-i}\right)\right)\right) \\ &\geq u_i\left(g\left(m_i,s_{-i}\left(\alpha_{-i}\left(\theta_{-i}\right)\right)\right),\left(\theta_i',\alpha_{-i}\left(\theta_{-i}\right)\right)\right) \\ &= u_i\left(y,\left(\theta_i',\alpha_{-i}\left(\theta_{-i}\right)\right)\right),\forall\theta_i'\in\Theta_i. \end{aligned}$$

This establishes that $y\in Y_i^*\left(\theta_{-i}\right)$. ■

We proceed by showing that in a wide class of environments, to be referred to as economic environments, ex post incentive and monotonicity condition are also sufficient conditions for ex post implementation.

Definition 8 (Economic environment)
An environment is economic at state $\theta \in \Theta$ if, for every allocation $a \in Y$, there exist $i \neq j$ and allocations x and y respectively such that

$$u_i(x, \theta) > u_i(a, \theta)$$

and

$$u_j(y, \theta) > u_j(a, \theta).$$

An environoment is economic if it is economic at every state.

We shall prove the sufficiency of the ex post monotonicity condition by using the following augmented mechanism. It is similar to mechanisms used to establish sufficiency in the complete information implementation literature (e.g., Maskin (1999)). Each agent sends a message of the form $m_i = (\theta_i, z_i, y_i)$, where $\theta_i \in \Theta_i$, z_i is a non-negative integer and $y_i \in Y$. The mechanism is described by three rules.

1. If $z_i = 0$ for all i, then $g(m) = f(\theta)$.
2. If $z_j = 1$ and $z_i = 0$ for all $i \neq j$, then outcome y_j is chosen if $y_j \in Y_j^*(\theta_{-j})$; otherwise outcome $f(\theta)$ is chosen.
3. In all other cases, $y_{\widetilde{j}(z)}$ is chosen, where $\widetilde{j}(z)$ is the agent i with the highest value of z_i (and, in the event of a tie, the lowest label).

A strategy profile in this game is a collection $s = (s_1, ..., s_I)$, with $s_i : \Theta_i \to M_i$ and we write

$$s_i(\theta) = \left(s_i^1(\theta), s_i^2(\theta), s_i^3(\theta)\right) \in \Theta_i \times \mathbb{Z}_+ \times Y;$$

and $s^k(\theta) = \left(s_i^k(\theta)\right)_{i=1}^I$. We shall refer to this mechanism as the *augmented mechanism.*

Theorem 2 (Economic Environment)
If $I \geq 3$ and f satisfies ex post incentive compatibility and ex post monotonicity and the environment is economic, then f is ex post implementable.

Proof. The proposition is proved in three steps, using the above mechanism.

Step 1. There is an ex post equilibrium s with $g(s(\theta)) = f(\theta)$ for all θ. Any strategy profile s of the following form is an ex post equilibrium:

$$s_i(\theta_i) = (\theta_i, 0, \cdot).$$

Suppose agent i thinks that his opponents are types θ_{-i} and deviates to a message of the form

$$s_i(\theta_i) = \left(\theta_i', z_i, y_i\right);$$

if either $z_i = 0$ or $z_i > 0$ but $y_i \notin Y_i^*(\theta_{-i})$, then the payoff gain is

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

which is non-positive by (EPIC); if $z_i = 1$ and $y_i \in Y_i^*(\theta_{-i})$, then the payoff gain is

$$u_i(y_i, (\theta_i, \theta_{-i})) - u_i\left(f(\theta_i, \theta_{-i}), f(\theta_i, \theta_{-i})\right),$$

which is non-positive by the definition of $Y_i^*(\theta_{-i})$.

Step 2. In any ex post equilibrium, $s_i^2(\theta_i) = 0$ for all i and θ_i. Suppose that rule 2 or rule 3 applies to the message profile sent at payoff type profile θ, so that there exists i such that $s_i^2(\theta_i) = 1$. Given the strategies of the other agents, any agent $j \neq i$ of type θ_j who thought his opponents were types θ_{-j} could send any message of the form

$$(\cdot, z_j, y_j)$$

and obtain utility $u_j(y_j, \theta)$. Thus we must have $u_j(g(s(\theta)), \theta) \geq u_j(a, \theta)$ for all a and all $j \neq i$. This contradicts the economic environment assumption.

Step 3. In any ex post equilibrium with $s_i^2(\theta_i) = 0$ for all i and θ_i, $f \circ s^1 = f$. Suppose that $f \circ s^1 \neq f$. By (EM), there exists i, θ and $y \in Y_i^*\left(s_{-i}^1(\theta_{-i})\right)$ such that

$$u_i(y, \theta) > u_i\left(f\left(s^1(\theta)\right), \theta\right).$$

Now suppose that type θ_i of agent i believes that his opponents are of type θ_{-i} and sends message $m_i = (\cdot, 1, y)$, while other agents send their equilibrium messages, then from the definition of $g(\cdot)$:

$$g(m_i, s_{-i}(\theta_{-i})) = y,$$

so that

$$\begin{aligned} u_i(g(m_i, s_{-i}(\theta_{-i})), \theta) &= u_i(y, \theta) \\ &> u_i\left(f\left(s^1(\theta)\right), \theta\right) \\ &= u_i(g(s(\theta)), \theta), \end{aligned}$$

and this completes the proof of sufficiency. ■

The economic environment condition was used to show that in the augmented mechanism in equilibrium, the integer reports z_i all have to say $z_i = 0$, or else any agent j could profitably change his report z_i and obtain a more desirable allocation to $f(\cdot)$, where the economic environment guaranteed the existence of agent j with a preferred allocation.

We now proceed to establish sufficient conditions for ex post implementation outside of economic environments. We begin by establishing an implication of non-economic environments.

Lemma 1 *The environment is non-economic at θ if and only if there exists j and $b \in Y$ such that $u_i(b, \theta) \geq u_i(a, \theta)$ for all $a \in A$ and $i \neq j$.*

Proof. The environment is non-economic (by definition) if and only if there exists an allocation b, such that if $u_j(y, \theta) > u_j(b, \theta)$ for some j, $y \in Y$, then there does not exist $i \neq j$ and $a \in Y$ such that $u_i(a, \theta) > u_i(b, \theta)$. Thus $u_i(b, \theta) \geq u_i(a, \theta)$ for all $a \in Y$ and $i \neq j$. ■

The ex post analogue of Jackson's "no veto hypothesis" is simply the requirement that the state be non-economic.

Definition 9 (No Veto Power)
Social choice function f satisfies no veto power at θ if $u_i(b, \theta) \geq u_i(a, \theta)$ for all $a \in Y$ and all $i \neq j$ implies that $f(\theta) = b$.

Definition 10 (Ex Post Monotonicity No Veto (EMNV))
A social choice function f satisfies ex post monotonicity no veto if the following is true. Fix any deception α and sets $\Phi_i \subset \Theta_i$ (write $\Phi = \times_{i=1}^{I} \Phi_i$). Suppose that the environment in non-economic at each $\theta \notin \Phi$. Suppose also that either $f(\alpha(\theta)) \neq f(\theta)$ for some $\theta \in \Phi$ or the no veto power property fails for some $\theta \notin \Phi$. Then there exists i, $\theta \in \Phi$ and $y \in Y_i^*(\alpha_{-i}(\theta_{-i}))$ such that

$$u_i(y, \theta) > u_i(f(\alpha(\theta)), \theta).$$

EPMV is almost equivalent to requiring ex post monotonicity and no veto power everywhere. More precisely, we have:

1. If ex post monotonicity holds and no veto power holds at every type profile, then EMNV holds.
2. If EPMV holds, then (1) ex post monotonicity holds and (2) if the environment is non-economic whenever $\theta_i = \theta_i^*$, then no veto power holds whenever $\theta_i = \theta_i^*$. To see (1), set $\Phi_i = \Theta_i$ for all i; to see (2), set α to be the truth-telling deception and, for some i, $\Phi_i = \Theta_i \backslash \{\theta_i^*\}$ and $\Phi_j = \Theta_j$ for all $j \neq i$.

Thus in an economic environment, EMNV is equivalent to ex post monotonicity.

Theorem 3 (Sufficiency)
For $I \geq 3$, f satisfies $(EPIC)$ and $(EMNV)$, then it is ex post implementable.

Proof. We use the same mechanism as before. The argument that there exists an ex post equilibrium s with $g(s(\theta)) = f(\theta)$ for all θ is the same as before. Now we establish three claims that hold for all equilibria. Let

$$\Phi_i = \{\theta_i : s_i(\theta_i) = (\cdot, 0, \cdot)\}$$

Claim 1. In any ex post equilibrium, for each $\theta \notin \Phi$, (a) there exists i such that $u_j(g(s(\theta)), \theta) \geq u_j(a, \theta)$ for all a and $j \neq i$; and thus (b) the environment is non-economic at θ.

First, observe that for each $\theta \notin \Phi$, there exists i such that $s_i^2(\theta_i) > 0$. Given the strategies of the other agents, any agent $j \neq i$ who thought his opponents were types θ_{-j} could send any message of the form

$$(\cdot, z_j, y_j)$$

and obtain utility $u_j(y_j, \theta)$. Thus we must have $u_j(g(s(\theta)), \theta) \geq u_j(a, \theta)$ for all a and $j \neq i$; thus the environment is non-economic for all $\theta \notin \Phi$.

Claim 2. In any ex post equilibrium, for all $\theta \in \Phi$,

$$u_i\left(f\left(s^1(\theta)\right), \theta\right) \geq u_i(y, \theta)$$

for all $y \in Y_i^*\left(s_{-i}^1(\theta_{-i})\right)$. Suppose that $y \in Y_i^*\left(s_{-i}^1(\theta_{-i})\right)$ and that type θ_i of agent i believes that his opponents are of type θ_{-i} and sends message $m_i = (\cdot, z_i, y)$, while other agents send their equilibrium messages. Now

$$g(m_i, s_{-i}(\theta_{-i})) = y;$$

so ex post equilibrium requires that

$$\begin{aligned} u_i(g(s(\theta)), \theta) &= u_i\left(f\left(s^1(\theta)\right), \theta\right) \\ &\geq u_i(g(m_i, s_{-i}(\theta_{-i})), \theta) \\ &= u_i(y, \theta). \end{aligned}$$

Claim 3. If EPMV is satisfied, then Claim 1 and 2 imply that $g(s(\theta)) = f(\theta)$ for all θ.

Fix any equilibrium. Claim 1(b) establishes that the environment is non-economic at all $\theta \in \Phi$. Suppose $g(s(\theta)) \neq f(\theta)$ for some $\theta \in \Phi$. Now EPMV implies that there exists i, $\theta \in \Phi$ and $y \in Y_i^*\left(s_{-i}^1(\theta_{-i})\right)$ such that $u_i(y, \theta) > u_i\left(f\left(s^1(\theta)\right), \theta\right)$, contradicting Claim 2. Suppose $g(s(\theta)) \neq f(\theta)$ for some $\theta \notin \Phi$. By claim 1(a), there exists i such that $u_j(g(s(\theta)), \theta) \geq u_j(a, \theta)$ for all a and $j \neq i$. This establishes that no veto power fails at θ. So again EPMV implies that there exists i, $\theta \in \Phi$ and $y \in Y_i^*(\alpha_{-i}(\theta_{-i}))$ such that $u_i(y, \theta) > u_i(f(\alpha(\theta)), \theta)$, contradicting Claim 2. ■

The structure of the proof is similar to Jackson (1991). The mechanism used to prove sufficiency is simpler as we require the strategies to be in an ex-post rather than an interim equilibrium. The entire argument is more compact due to the simplifying assumption of a social choice function rather than social choice set.

6 Interim Monotonicity and Robust Monotonicity

6.1 Interim Monotonicity

A deception for a type space $\mathcal{T}$ is a collection $\alpha = (\alpha_1, ..., \alpha_I)$, with

$$\alpha_i : T_i \to T_i.$$

Write $\alpha(t) = (\alpha_i(t_i))_{i=1}^I$; let $f \circ \widehat{\theta} : T \to A$ and $f \circ \widehat{\theta} \circ \alpha : T \to A$ be defined by

$$\begin{aligned} f \circ \widehat{\theta}(t) &= f\left(\widehat{\theta}(t)\right) \\ \text{and } f \circ \widehat{\theta} \circ \alpha(t) &= f\left(\widehat{\theta}(\alpha(t))\right) \end{aligned}$$

for all t.

Definition 11 (Interim Monotonicity)
Social choice function f satisfies interim monotonicity on type space $\mathcal{T}$ if, for every deception α with $f \circ \widehat{\theta} \circ \alpha \neq f \circ \widehat{\theta}$, there exists i, t_i and $y : T \to Y$ such that

$$\sum_{t_{-i} \in T_{-i}} u_i\left(y(\alpha(t)), \widehat{\theta}(t)\right) \widehat{\pi}_i(t_i)[t_{-i}] > \sum_{t_{-i} \in T_{-i}} u_i\left(f\left(\widehat{\theta}(\alpha(t))\right), \widehat{\theta}(t)\right) \widehat{\pi}_i(t_i)[t_{-i}], \tag{7}$$

and

$$\begin{aligned} &\sum_{t_{-i} \in T_{-i}} u_i\left(f\left(\widehat{\theta}(t_i', t_{-i})\right), \widehat{\theta}(t_i', t_{-i})\right) \widehat{\pi}_i(t_i')[t_{-i}] \\ \geq &\sum_{t_{-i} \in T_{-i}} u_i\left(y(\alpha_i(t_i), t_{-i}), \widehat{\theta}(t_i', t_{-i})\right) \widehat{\pi}_i(t_i')[t_{-i}], \ \forall t_i' \in T_i. \end{aligned} \tag{8}$$

Conditions like this are known as Bayesian monotonicity in the literature. We use the term interim monotonicity both because we are interested in the case when there is no common prior and to highlight the comparison with ex post monotonicity. Postlewaite and Schmeidler (1986) showed that such an interim monotonicity condition is necessary and sufficient for full implementation in an exchange economy with nonexclusive information and at least three agents. Palfrey and Srivastava (1989) provide separate necessary and sufficient conditions for interim implementation when there is exclusive information. Jackson (1991) showed that interim monotonicity is necessary and sufficient for interim implementation in economic environments and that a slightly strengthened property (Bayesian monotonicity no veto) is sufficient.

6.2 Robust Monotonicity

We will be interested in another new monotonicity notion that is equivalent to interim monotonicity on all type spaces. In defining robust monotonicity, we therefore formalize a deception as a point-to-set mapping. A deception is a collection $\boldsymbol{\beta} = (\beta_1, ..., \beta_I)$ with $\beta_i : \Theta_i \to 2^{\Theta_i}$ and $\theta_i \in \beta_i(\theta_i)$. The interpretation is that $\beta_i(\theta_i)$ is the collection of correct or incorrect reports that payoff type θ_i might send. A deception is *acceptable* if $\theta' \in \boldsymbol{\beta}(\theta) \Rightarrow f(\theta') = f(\theta)$. A deception is *unacceptable* if it is not acceptable. We write

$$\beta_i^{-1}(\theta_i') \equiv \left\{\theta_i : \theta_i' \in \beta_i(\theta_i)\right\}$$

and

$$\beta_{-i}^{-1}(\theta_{-i}') \equiv \underset{j \neq i}{\times} \beta_j^{-1}(\theta_j').$$

Thus $\beta_{-i}^{-1}(\theta_{-i}')$ is the collection of θ_{-i} who might report themselves to be θ_{-i}' under deception β.

Definition 12 (Robust Monotonicity)
Social choice function f satisfies robust monotonicity if for every unacceptable deception β, there exist i, θ_i, $\theta_i' \in \beta_i(\theta_i)$ such that, for all $\theta_{-i}' \in \Theta_{-i}$ and $\psi_i \in \Delta\left(\beta_{-i}^{-1}\left(\theta_{-i}'\right)\right)$, there exists $y \in Y_i^*\left(\theta_{-i}'\right)$ such that

$$\sum_{\theta_{-i}\in\Theta_{-i}} \psi_i(\theta_{-i})\, u_i(y,(\theta_i,\theta_{-i})) > \sum_{\theta_{-i}\in\Theta_{-i}} \psi_i(\theta_{-i})\, u_i\left(f\left(\theta_i',\theta_{-i}'\right),(\theta_i,\theta_{-i})\right). \tag{9}$$

Note that the allocation y is allowed to depend on the misreport θ_{-i}' and the distribution ψ_i.

The notion of robust monotonicity shares many features with the ex post monotonicity condition. Like ex post monotonicity, robust monotonicity refers only to payoff types and does not refer to priors or posteriors over payoff types nor does it refer to any general type spaces. Robust monotonicity also requires that the ex post incentive compatibility requirement $y \in Y_i^*\left(\theta_{-i}'\right)$ be satisfied. But the whistle-blower inequality (9) is a stronger version of the ex post requirement.

6.3 Comparing Monotonicity Properties: Results

In this Subsection, we establish the relation between various monotonicity notions. We first show that robust monotonicity is equivalent to interim monotonicity on all type spaces.

Theorem 4
Social choice function f satisfies robust monotonicity if and only if it satisfies interim monotonicity on every type space.

Proof. ($\Leftarrow$). We first prove that interim monotonicity on every type space implies robust monotonicity. It is convenient to work with the following contrapositive statement of robust monotonicity. Thus for all i, θ_i, $\theta_i' \in \beta_i(\theta_i)$, there exists a payoff profile $\theta_{-i}' \in \Theta_{-i}$, to be denoted by:

$$\zeta_i\left(\theta_i,\theta_i'\right) \triangleq \left(\zeta_{ij}\left(\theta_i,\theta_i'\right)\right)_{j\neq i} \in \Theta_{-i}$$

and a conditional probability distribution $\psi_i\left(\cdot|\theta_i,\theta_i'\right) \in \Delta\left(\beta_{-i}^{-1}\left(\zeta_i\left(\theta_i,\theta_i'\right)\right)\right)$ such that

$$u_i\left(f\left(\widetilde{\theta}_i,\zeta_i\left(\theta_i,\theta_i'\right)\right),\left(\widetilde{\theta}_i,\zeta_i\left(\theta_i,\theta_i'\right)\right)\right) \geq u_i\left(y,\left(\widetilde{\theta}_i,\zeta_i\left(\theta_i,\theta_i'\right)\right)\right), \tag{10}$$

for all $\widetilde{\theta}_i$ implies

$$\begin{aligned} &\sum_{\left\{\theta_{-i}\in\Theta_{-i}:\zeta_i\left(\theta_i,\theta_i'\right)\in\beta_{-i}(\theta_{-i})\right\}} \psi_i\left(\theta_{-i}|\theta_i,\theta_i'\right) u_i\left(f\left(\theta_i',\zeta_i\left(\theta_i,\theta_i'\right)\right),(\theta_i,\theta_{-i})\right) \\ \geq &\sum_{\left\{\theta_{-i}\in\Theta_{-i}:\zeta_i\left(\theta_i,\theta_i'\right)\in\beta_{-i}(\theta_{-i})\right\}} \psi_i\left(\theta_{-i}|\theta_i,\theta_i'\right) u_i(y,(\theta_i,\theta_{-i})). \end{aligned} \tag{11}$$

Now we construct a type space based on the deception $\boldsymbol{\beta}$ such that if the social choice function satisfies interim monotonicity on this type space, then $\boldsymbol{\beta}$ must be acceptable.

First, agent i has a set of "deception" types T_i^1 which are isomorphic to

$$\Psi_i = \left\{\left(\theta_i,\theta_i'\right) : \theta_i \in \Theta_i \text{ and } \theta_i' \in \beta_i(\theta_i)\right\}$$

and for simplicity we identify every type $t_i \in T_i^1$ simply by such a pair of payoff types $\left(\theta_i,\theta_i'\right)$, or $T_i^1 \triangleq \Psi_i$. The type $\left(\theta_i,\theta_i'\right)$ has payoff type θ_i and assigns probability $\psi_i\left(\theta_{-i}|\theta_i,\theta_i'\right)$ to the event that each agent j is type $\left(\theta_j,\zeta_{ij}\left(\theta_i,\theta_i'\right)\right)$.

Second, agent i has a set of "pseudo-complete information types" T_i^2, which are isomorphic to Θ, and for simplicity, again let $T_i^2 = \Theta_i$. The type corresponding to θ has payoff type θ_i and he is convinced that each other agent j is type θ.

More formally, we have

$$T_i = T_i^1 \cup T_i^2.$$

If $t_i \in T_i^1$ and $t_i = \left(\theta_i, \theta_i'\right)$, then

$$\widehat{\theta}_i\left(t_i\right) = \theta_i$$

and

$$\widehat{\pi}_i\left(t_i\right)\left[t_{-i}\right] = \begin{cases} \psi_i\left(\theta_{-i}|\theta_i, \theta_i'\right), \text{ if } t_j = \left(\theta_j, \zeta_{ij}\left(\theta_i, \theta_i'\right)\right) \text{ for each } j \neq i \\ 0, \text{ otherwise;} \end{cases}$$

if $t_i \in T_i^2$ and $t_i = \theta$, then

$$\widehat{\theta}_i\left(t_i\right) = \theta_i, \tag{12}$$

and

$$\widehat{\pi}_i\left(t_i\right)\left[t_{-i}\right] = \begin{cases} 1, \text{ if } t_j = \left(\theta_j, \theta_j\right) \text{ for each } j \neq i \\ 0, \text{ otherwise.} \end{cases} \tag{13}$$

Now we prove the proposition, by showing that interim monotonicity on this type space implies the deception β we started with must be acceptable. Consider the deception α_i on the constructed type space where each type $\left(\theta_i, \theta_i'\right)$ reports himself to be type $\left(\theta_i', \theta_i'\right)$, and all other types report their types truthfully. Thus:

$$\alpha_i\left(t_i\right) = \begin{cases} \left(\theta_i', \theta_i'\right), \text{ if } t_i = \left(\theta_i, \theta_i'\right) \\ t_i, \text{ otherwise} \end{cases}.$$

Notice that type $t_i = (\theta_i, \theta_i)$ reports his type truthfully under this deception α_i for all i. Now we apply the interim monotonicity condition as presented in Definition 11 to this deception. For any type $t_i \in T_i^2$, the deception α_i changes neither his action nor his beliefs about his opponents' reporting behavior. Thus he cannot be the critical type t_i in the definition who "reports the deception". More formally, for any type $t_i = \theta \in T_i^2$, the interim monotonicity conditions reduce to, after using (12) and (13):

$$u_i\left(y\left(\theta\right), \theta\right) > u_i\left(f\left(\theta\right), \theta\right)$$

and for all $t_i' = \theta' \in T_i^2$, we would have

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

which clearly leads to a contradiction for $t_i' = \theta$. Thus there must exist i, $t_i \in T_i^1$ and $y : T \to Y$ such that (7) and (8) hold. Letting $\widehat{t}_i = \left(\theta_i, \theta_i'\right)$, (7) becomes:

$$\begin{aligned} & \sum_{\left\{\theta_{-i} \in \Theta_{-i} : \zeta_i\left(\theta_i, \theta_i'\right) \in \beta_{-i}\left(\theta_{-i}\right)\right\}} \psi_i\left(\theta_{-i}|\theta_i, \theta_i'\right) u_i\left(y\left(\left(\theta_i', \theta_i'\right), \left(\zeta_{ij}\left(\theta_i, \theta_i'\right), \zeta_{ij}\left(\theta_i, \theta_i'\right)\right)_{j \neq i}\right), \left(\theta_i, \theta_{-i}\right)\right) \\ > & \\ & \sum_{\left\{\theta_{-i} \in \Theta_{-i} : \zeta_i\left(\theta_i, \theta_i'\right) \in \beta_{-i}\left(\theta_{-i}\right)\right\}} \psi_i\left(\theta_{-i}|\theta_i, \theta_i'\right) u_i\left(f\left(\theta_i', \zeta_i\left(\theta_i, \theta_i'\right)\right), \left(\theta_i, \theta_{-i}\right)\right). \end{aligned} \tag{14}$$

In the special case of the pseudo complete information types with $t_i' = \left(\widetilde{\theta}_i, \zeta_i\left(\theta_i, \theta_i'\right)\right)$, the interim incentive compatibility condition (8) becomes

$$\begin{aligned} & u_i\left(f\left(\widetilde{\theta}_i, \zeta_i\left(\theta_i, \theta_i'\right)\right), \left(\widetilde{\theta}_i, \zeta_i\left(\theta_i, \theta_i'\right)\right)\right) \\ \geq & \\ & u_i\left(y\left(\left(\theta_i', \theta_i'\right), \left(\zeta_{ij}\left(\theta_i, \theta_i'\right), \zeta_{ij}\left(\theta_i, \theta_i'\right)\right)_{j \neq i}\right), \left(\widetilde{\theta}_i, \zeta_i\left(\theta_i, \theta_i'\right)\right)\right), \ \forall \widetilde{\theta}_i. \end{aligned} \tag{15}$$

But now (10), (11) and (15) implies that (14) fails. Thus interim monotonicity on this type space requires that

$$f\left(\widehat{\theta}\left(t\right)\right) = f\left(\widehat{\theta}\left(\alpha\left(t\right)\right)\right) \text{ for all } t.$$

This requires $\boldsymbol{\beta}$ is acceptable. This completes the proof of robust monotonicity.

($\Rightarrow$) Suppose f satisfies robust monotonicity. Fix any type space $\mathcal{T}$ and any deception α with $f\left(\widehat{\theta}\left(t\right)\right) \neq f\left(\widehat{\theta}\left(\alpha\left(t\right)\right)\right)$ for some t. Define $\boldsymbol{\beta}$ by:

$$\beta_i\left(\theta_i\right) = \left\{\theta_i' : \exists t_i \text{ such that } \widehat{\theta}_i\left(t_i\right) = \theta_i \text{ and } \widehat{\theta}_i\left(\alpha_i\left(t_i\right)\right) = \theta_i'\right\}.$$

For every θ_i, $\beta_i\left(\theta_i\right)$ is the collection of payoff types θ_i' which will be reported by some type t_i when he is using the deception α_i and has a true payoff type θ_i. Deception $\boldsymbol{\beta}$ is unacceptable, so by robust monotonicity, there exist i, θ_i, $\theta_i' \in \beta_i\left(\theta_i\right)$ such that, for all $\theta_{-i}' \in \Theta_{-i}$ and for all ψ_i with

$$\psi_i \in \Delta\left(\left\{\theta_{-i} \in \Theta_{-i} : \theta_{-i}' \in \beta_{-i}\left(\theta_{-i}\right)\right\}\right),$$

there exists $y\left(\theta_{-i}', \psi_i\right)$ such that

$$\begin{aligned} & \sum_{\left\{\theta_{-i}\in\Theta_{-i}:\theta_{-i}'\in\beta_{-i}(\theta_{-i})\right\}} \psi_i\left(\theta_{-i}\right) u_i\left(y\left(\theta_{-i}', \psi_i\right), \left(\theta_i, \theta_{-i}\right)\right) \\ > & \sum_{\left\{\theta_{-i}\in\Theta_{-i}:\theta_{-i}'\in\beta_{-i}(\theta_{-i})\right\}} \psi_i\left(\theta_{-i}\right) u_i\left(f\left(\theta_i', \theta_{-i}'\right), \left(\theta_i, \theta_{-i}\right)\right) \end{aligned} \tag{16}$$

and

$$u_i\left(f\left(\widetilde{\theta}_i, \theta_{-i}'\right), \left(\widetilde{\theta}_i, \theta_{-i}'\right)\right) \geq u_i\left(y\left(\theta_{-i}', \psi_i\right), \left(\widetilde{\theta}_i, \theta_{-i}'\right)\right), \tag{17}$$

for all $\widetilde{\theta}_i$. We emphasize that the distribution ψ_i only generates positive probabilities over $\theta_{-i} \in \Theta_{-i}$ which could lead to a deception θ_{-i}' for some types $t_{-i} \in T_{-i}$. Thus in the following we omit the set specification $\left\{\theta_{-i} \in \Theta_{-i} : \theta_{-i}' \in \beta_{-i}\left(\theta_{-i}\right)\right\}$ in the summation whenever we take expectations with respect to $\psi_i\left(\theta_{-i}\right)$ as profiles θ_{-i}'' with $\theta_{-i}' \notin \beta_{-i}\left(\theta_{-i}''\right)$ receive probability zero anyhow. Now choose any t_i such that $\widehat{\theta}_i\left(t_i\right) = \theta_i$ and $\widehat{\theta}_i\left(\alpha_i\left(t_i\right)\right) = \theta_i'$. Let

$$\xi_i\left(\theta_{-i}'\right) \triangleq \sum_{\left\{t_{-i}\in T_{-i}:\widehat{\theta}_{-i}(\alpha_{-i}(t_{-i}))=\theta_{-i}'\right\}} \widehat{\pi}_i\left(t_i\right)\left[t_{-i}\right] \tag{18}$$

and

$$\psi_i\left(\theta_{-i}|\theta_{-i}'\right) \triangleq \frac{\sum_{\left\{t_{-i}\in T_{-i}:\widehat{\theta}_{-i}(t_{-i})=\theta_{-i} \text{ and } \widehat{\theta}_{-i}(\alpha_{-i}(t_{-i}))=\theta_{-i}'\right\}} \widehat{\pi}_i\left(t_i\right)\left[t_{-i}\right]}{\sum_{\left\{t_{-i}\in T_{-i}:\widehat{\theta}_{-i}(\alpha_{-i}(t_{-i}))=\theta_{-i}'\right\}} \widehat{\pi}_i\left(t_i\right)\left[t_{-i}\right]}. \tag{19}$$

For a given type space T and type t_i, $\xi_i\left(\theta_{-i}'\right)$ is the probability that agent i attaches to a payoff type report θ_{-i}' given the deception α_{-i}. Consequently, $\psi_i\left(\theta_{-i}|\theta_{-i}'\right)$ is the conditional probability that the true payoff type profile is θ_{-i} if the announced type profile is θ_{-i}'.

We construct a reward function $y\left(t\right)$ on the type space T by setting:

$$y\left(\alpha_i\left(t_i\right), t_{-i}\right) \triangleq y\left(\widehat{\theta}_{-i}\left(t_{-i}\right), \psi_i\left(\cdot \left| \widehat{\theta}_{-i}\left(t_{-i}\right)\right.\right)\right). \tag{20}$$

Using the probabilities distributions defined in (18) and (19), and the reward function defined in (20) we have the following equalities useful to establish the interim reward inequality:

$$\begin{aligned} & \sum_{t_{-i}\in T_{-i}} u_i\left(y\left(\alpha\left(t\right)\right), \widehat{\theta}\left(t\right)\right) \widehat{\pi}_i\left(t_i\right)\left[t_{-i}\right] \\ = & \sum_{\theta_{-i}'\in\Theta_{-i}} \sum_{\theta_{-i}\in\Theta_{-i}} u_i\left(y\left(\theta_{-i}', \psi_i\left(\cdot|\theta_{-i}'\right)\right), \theta\right) \psi_i\left(\theta_{-i}|\theta_{-i}'\right) \xi_i\left(\theta_{-i}'\right) \end{aligned} \tag{21}$$

and

$$\begin{aligned} &\sum_{t_{-i}\in T_{-i}} u_i\left(f\left(\widehat{\theta}\left(\alpha\left(t\right)\right)\right),\widehat{\theta}\left(t\right)\right)\widehat{\pi}_i\left(t_i\right)\left[t_{-i}\right] \\ =\ &\sum_{\theta'_{-i}\in\Theta_{-i}}\sum_{\theta_{-i}\in\Theta_{-i}} u_i\left(f\left(\theta'\right),\theta\right)\psi_i\left(\theta_{-i}|\theta'_{-i}\right)\xi_i\left(\theta'_{-i}\right). \end{aligned} \tag{22}$$

As the inequality (16) holds for every θ'_{-i}, we can infer from (16) that

$$\begin{aligned} &\sum_{\theta'_{-i}\in\Theta_{-i}}\sum_{\theta_{-i}\in\Theta_{-i}} u_i\left(y\left(\theta'_{-i},\psi_i\left(\cdot\left|\theta'_{-i}\right.\right)\right),\theta\right)\psi_i\left(\theta_{-i}|\theta'_{-i}\right)\xi_i\left(\theta'_{-i}\right) \\ >\ &\sum_{\theta'_{-i}\in\Theta_{-i}}\sum_{\theta_{-i}\in\Theta_{-i}} u_i\left(f\left(\theta'\right),\theta\right)\psi_i\left(\theta_{-i}|\theta'_{-i}\right)\xi_i\left(\theta'_{-i}\right) \end{aligned}$$

holds when we take the expectation with respect to $\xi_i\left(\theta'_{-i}\right)$. By appealing to the equalities (21) and (22), we establish that:

$$\begin{aligned} &\sum_{t_{-i}\in T_{-i}} u_i\left(y\left(\alpha\left(t\right)\right),\widehat{\theta}\left(t\right)\right)\widehat{\pi}_i\left(t_i\right)\left[t_{-i}\right] \\ >\ &\sum_{t_{-i}\in T_{-i}} u_i\left(f\left(\widehat{\theta}\left(\alpha\left(t\right)\right)\right),\widehat{\theta}\left(t\right)\right)\widehat{\pi}_i\left(t_i\right)\left[t_{-i}\right]. \end{aligned} \tag{23}$$

Using again the probabilities distributions defined in (18) and (19), the reward function defined in (20), we have the following equalities useful to establish the interim incentive inequalities:

$$\begin{aligned} &\sum_{t_{-i}\in T_{-i}} u_i\left(f\left(\widehat{\theta}\left(t'_i,t_{-i}\right)\right),\widehat{\theta}\left(t'_i,t_{-i}\right)\right)\widehat{\pi}_i\left(t'_i\right)\left[t_{-i}\right] \\ =\ &\sum_{\theta'_{-i}\in\Theta_{-i}}\sum_{\theta_{-i}\in\Theta_{-i}} u_i\left(f\left(\widehat{\theta}_i\left(t'_i\right),\theta_{-i}\right),\left(\widehat{\theta}_i\left(t'_i\right),\theta_{-i}\right)\right)\psi_i\left(\theta_{-i}|\theta'_{-i}\right)\xi_i\left(\theta'_{-i}\right) \end{aligned} \tag{24}$$

and

$$\begin{aligned} &\sum_{t_{-i}\in T_{-i}} u_i\left(y\left(\alpha_i\left(t_i\right),t_{-i}\right),\widehat{\theta}\left(t'_i,t_{-i}\right)\right)\widehat{\pi}_i\left(t'_i\right)\left[t_{-i}\right] \\ =\ &\sum_{\theta'_{-i}\in\Theta_{-i}}\sum_{\theta_{-i}\in\Theta_{-i}} u_i\left(y\left(\theta_{-i},\psi_i\left(\cdot\left|\theta_{-i}\right.\right)\right),\left(\widehat{\theta}_i\left(t'_i\right),\theta_{-i}\right)\right)\psi_i\left(\theta_{-i}|\theta'_{-i}\right)\xi_i\left(\theta'_{-i}\right),\ \forall t'_i. \end{aligned} \tag{25}$$

By appealing the ex post incentive inequalities of robust monotonicity, (17), we know that

$$u_i\left(f\left(\widehat{\theta}_i\left(t'_i\right),\theta_{-i}\right),\left(\widehat{\theta}_i\left(t'_i\right),\theta_{-i}\right)\right)\geq u_i\left(y\left(\theta'_{-i},\psi_i\left(\cdot\left|\theta_{-i}\right.\right)\right),\left(\widehat{\theta}_i\left(t'_i\right),\theta_{-i}\right)\right), \tag{26}$$

for all t'_i. The inequalities (26) then remain valid when we take expectations with respect to the conditional and marginal distributions $\psi_i\left(\theta_{-i}|\theta'_{-i}\right)$ and $\xi_i\left(\theta'_{-i}\right)$ respectively. By using the equalities (24) and (25) we can then establish the interim incentive compatibility conditions:

$$\begin{aligned} &\sum_{t_{-i}\in T_{-i}} u_i\left(f\left(\widehat{\theta}\left(t'_i,t_{-i}\right)\right),\widehat{\theta}\left(t'_i,t_{-i}\right)\right)\widehat{\pi}_i\left(t'_i\right)\left[t_{-i}\right] \\ \geq\ &\sum_{t_{-i}\in T_{-i}} u_i\left(y\left(\alpha_i\left(t_i\right),t_{-i}\right),\widehat{\theta}\left(t'_i,t_{-i}\right)\right)\widehat{\pi}_i\left(t'_i\right)\left[t_{-i}\right],\ \forall t'_i. \end{aligned} \tag{27}$$

But by (23) and (27), we have confirmed interim monotonicity on this type space. ■

The proof may appear rather intricate in its details. We next give a brief outline of the basic steps to show that interim implies robust monotonicity. We start with an arbitrary deception β which satisfies the inequalities (10) and (11) and, crucially, do not insist on β being acceptable. For the given deception β, we then create a type space, consisting of two components for every agent i. The first component for agent i is created by the set of pairs of payoff types $\left(\theta_i, \theta_i'\right)$, where the first entry is the true payoff type and the second entry is a feasible deception (under β), or $\theta_i' \in \beta_i(\theta_i)$. For this reason, we refer to these types as "deception types." For every such pair $\left(\theta_i, \theta_i'\right)$ there exists one particular payoff profile θ_{-i}' which is "salient" for agent i of type $\left(\theta_i, \theta_i'\right)$, as the deception β satisfies (10) and (11). Under the deception β, this payoff profile could have been reported by all true payoff profiles which are in the support of ψ_i. Consequently, the belief component of type $\left(\theta_i, \theta_i'\right)$ is given by simply adopting $\psi_i\left(\cdot \left| \theta_i, \theta_i'\right.\right)$. The second component are "pseudo complete information types", described by $t_i = \theta \in \Theta$, which have a probability one belief that the true payoff profile is given by θ and that all other agents report the deception type (θ_j, θ_j), and hence the "pseudo" in the labelling.

Given this type space T_i, we then consider a particular deception $\alpha_i : T_i \to T_i$. The deception α_i is localized around the "deception types" and the "pseudo complete information types" report thruthfully. The deception α_i consists of agent i always reporting his deception type rather than his true type, or $\alpha_i\left(\theta_i, \theta_i'\right) = \left(\theta_i', \theta_i'\right)$. We then verify whether f is interim monotone under α. The existence of the pseudo complete information types θ forces the interim incentive compatibility conditions to reduce to ex post incentive compatibility conditions. This guarantees the hypothesis in the robust monotonicity notion, namely inequality (10), and thus leads to the conclusion in form of the inequalities (11). But then we obtain a contradiction to the reward condition of interim monotonicity, unless the hypothesis for the interim monotonicity condition, namely $f \neq f \circ \alpha$, is not satisfied, i.e. $f = f \circ \alpha$ holds, but of course this implies that β is acceptable.

For the second part of the proof we use the full strength of robust monotonicity to establish interim monotonicity. We start out with a deception α on an arbitrary type space $\mathcal{T}$ such that $f \circ \alpha \neq f$. We then extract from given type t_i and associated belief type $\pi_i(t_i)[t_{-i}]$ a conditional distribution over payoff types $\xi_i(t_i)[\theta_{-i}]$. For this conditional distribution, we can then construct a reward by the robust monotonicity hypothesis, which we then employ for construct a reward allocation offer to induce type t_i to denounce the deception α.

Theorem 5
If f satisfies interim monotonicity on the complete information type space, then it satisfies Maskin monotonicity.

Proof. The proof is by contrapositive. Suppose then that f is not Maskin monotone, and hence there exists $\widehat{\alpha} : \Theta^I \to \Theta^I$ such that for all i, θ, with $f(\widehat{\alpha}(\theta)) \neq f(\theta)$, and all h such that

$$u_i(h(\widehat{\alpha}(\theta)), \theta) > u_i(f(\widehat{\alpha}(\theta)), \theta),$$

we have

$$u_i(f(\widehat{\alpha}(\theta)), \widehat{\alpha}(\theta)) < u_i(h(\widehat{\alpha}(\theta)), \widehat{\alpha}(\theta)).$$

Consider then the complete information type space $T_i = \Theta$. For every i, let $\alpha_i = \widehat{\alpha}$. To obtain the contradiction, let us then suppose that there exists i and t_i such that

$$\sum_{t_{-i} \in T_{-i}} u_i(h(\alpha(t)), t)\, \widehat{\pi}_i(t_i)[t_{-i}] > \sum_{t_{-i} \in T_{-i}} u_i(f(\alpha(t)), t)\, \widehat{\pi}_i(t_i)[t_{-i}] \tag{28}$$

while

$$\sum_{t_{-i} \in T_{-i}} u_i(f(t_i', t_{-i}), (t_i', t_{-i}))\, \widehat{\pi}_i(t_i')[t_{-i}] \geq \sum_{t_{-i} \in T_{-i}} u_i(h(\alpha_i(t_i), t_{-i}), t)\, \widehat{\pi}_i(t_i')[t_{-i}], \ \forall t_i' \neq t_i. \tag{29}$$

With the complete information type space and the symmetric deception strategy, the inequalities (28) and (29) reduce to

$$u_i\left(h\left(\widehat{\alpha}\left(\theta\right)\right),\theta\right) > u_i\left(f\left(\widehat{\alpha}\left(\theta\right)\right),\theta\right) \tag{30}$$

and

$$u_i\left(f\left(\theta'\right),\theta'\right) \geq u_i\left(h\left(\widehat{\alpha}\left(\theta\right),\theta',...,\theta'\right),\theta'\right),\ \forall\theta' \neq \theta, \tag{31}$$

but naturally there exists $\theta' = \widehat{\alpha}\left(\theta\right)$, and for this profile, the above inequality reads

$$u_i\left(f\left(\widehat{\alpha}\left(\theta\right)\right),\widehat{\alpha}\left(\theta\right)\right) \geq u_i\left(h\left(\widehat{\alpha}\left(\theta\right)\right),\widehat{\alpha}\left(\theta\right)\right),\ \theta' = \widehat{\alpha}\left(\theta\right),$$

which leads to the desired contradiction with Maskin monotonicity. ■

Theorem 6
If f satisfies robust monotonicity, then it satisfies ex post monotonicity.

Proof. Let α be an ex post deception with $f \neq f \circ \alpha$. Let β be a robust deception with $\beta_i\left(\theta_i\right) = \left\{\theta_i\right\} \cup \left\{\alpha_i\left(\theta_i\right)\right\}$. By the definition of robust monotonicity, there exists i, θ_i, $\theta'_i \in \beta_i\left(\theta_i\right)$ such that, for all $\theta'_{-i} \in \Theta_{-i}$ and $\psi_i \in \Delta\left(\beta_{-i}^{-1}\left(\theta'_{-i}\right)\right)$, there exists $y \in Y_i^*\left(\theta'_{-i}\right)$ such that

$$\sum_{\theta_{-i}\in\Theta_{-i}} \psi_i\left(\theta_{-i}\right) u_i\left(y,\left(\theta_i,\theta_{-i}\right)\right) > \sum_{\theta_{-i}\in\Theta_{-i}} \psi_i\left(\theta_{-i}\right) u_i\left(f\left(\theta'_i,\theta'_{-i}\right),\left(\theta_i,\theta_{-i}\right)\right).$$

By the construction of β, we must have $\theta'_i = \alpha_i\left(\theta_i\right)$. Thus there exists i, θ_i, $\theta'_i = \alpha_i\left(\theta_i\right)$, $\theta'_{-i} \in \Theta_{-i}$ with $\theta'_{-i} = \alpha_{-i}\left(\theta_{-i}\right)$ and $y \in Y_i^*\left(\theta'_{-i}\right)$ such that

$$u_i\left(y,\left(\theta_i,\theta_{-i}\right)\right) > u_i\left(f\left(\theta'_i,\theta'_{-i}\right),\left(\theta_i,\theta_{-i}\right)\right).$$

But this is ex post monotonicity. ■

While Maskin monotonicity is implied by interim monotonicity on complete information prior type spaces, we do not have an argument implying ex post monotonicity or robust monotonicity using type spaces that have a common prior, full support or common support. Because the strict inequalities in the definition of interim monotonicity give rise to a non-compact set, it is not clear that such an argument is possible. The following example shows how it is possible to have interim monotonicity satisfied for every type space with a sequence of full support priors, but fail in the limit.

6.4 Comparing Monotonicity Notions: Examples

6.4.1 Example B

The example satisfies Maskin monotonicity and interim monotonicity for all common priors over the payoff type space. Yet it fails to satisfy ex post monotonicity, and thus robust monotonicity.

There are three agents, $i = 1, 2, 3$ and each agent has a binary payoff type space $\theta_i \in \Theta_i = \{0, 1\}$. The entire payoff type space is given by $\Theta = \times_{i=1}^{3}\Theta_i$. For simplicity of the example, the allocation space is identical to the payoff type space, or $A = \Theta$ and the social choice function $f : \Theta \to A$ is given by the identity mapping $f\left(\theta\right) = \theta$ for all $\theta \in \Theta$. The payoffs of the agents satisfy symmetry across allocations a and payoff types: $u_i\left(a,\theta\right) = u_i\left(\theta,a\right)$ for all i, $a \in A$ and $\theta \in \Theta$ and invariance with respect to symmetric permutations, for all $i, a \in A, \theta \in \Theta$ and $\sigma : \Theta \to \Theta$, we have:

$$u_i\left(a,\theta\right) = u_i\left(\sigma\left(a\right),\sigma\left(\theta\right)\right).$$

The payoff matrices below represent the payoffs of the agents for allocation $a = (0, 0, 0)$ at all possible type profiles $\theta \in \Theta$. Agent 1 is the row player, agent 2 the column player and agent 3 the matrix player:

$(0,0,0)$			
	0	0	1
	0	$1,1,1$	$1+\varepsilon,0,1+\delta$
	1	$0,1+\delta,1+\varepsilon$	$0,0,0$

$(0,0,0)$			
	1	0	1
	0	$1+\delta,1+\varepsilon,0$	$0,0,0$
	1	$0,0,0$	$1,1,1$

The payoffs generated by the remaining allocations can be generated by the symmetry assumptions from the above matrices and hence are omitted. We assume that $0 < \varepsilon < \delta \ll 1$. The parameters ε and δ are assumed to be distinct solely to guarantee that the environment is an economic environment for which ex post monotonicity is a necessary as well as sufficient condition. With a direct mechanism the game displays two symmetric pure strategy ex post equilibria. The first symmetric equilibrium is the truthtelling equilibrium, or

$$s_i(\theta_i) = \theta_i \text{ for all } i \text{ and } \theta_i,$$

whereas the second symmetric equilibrium is the misreporting equilibrium:

$$s_i(\theta_i) \neq \theta_i \text{ for all } i \text{ and } \theta_i.$$

Naturally, these ex post equilibria are also interim equilibria on any type space.

Ex Post Monotonicity Fails We first show that this example fails ex post monotonicity by showing that for the "complete" deception $\alpha_i(\theta_i) \neq \theta_i$ for all i and θ_i the social choice function does not satisfy ex post monotonicity. By symmetry, it is sufficient to consider agent 1 and true state $\theta = (0,0,0)$. The complete deception (all misreport) leads to the allocation $\alpha(0,0,0) = (1,1,1)$. The only allocations which would improve the utility of agent 1 are $a \in \{(0,1,0),(0,0,1)\}$ and we shall argue next that neither of these allocations satisfies the incentive compatibility conditions of the monotonicity condition. Consider first the reward $y = (0,1,0)$, for which ex post incentive compatibility would have to satisfy:

$$1 = u_1((1,1,1),(1,1,1)) \geq u_1((0,1,0),(1,1,1)) = 0$$

as well as

$$1 = u_1((0,1,1),(0,1,1)) \geq u_1((0,1,0),(0,1,1)) = 1+\delta,$$

but obviously the second inequality is violated. Similarly, we observe that for the reward $y = (0,0,1)$, the ex post incentive compatibility conditions are:

$$1 = u_1((1,1,1),(1,1,1)) \geq u_1((0,0,1),(1,1,1)) = 0$$

as well as

$$1 = u_1((0,1,1),(0,1,1)) \geq u_1((0,0,1),(0,1,1)) = 1+\varepsilon,$$

and again the second inequality is violated. Thus we conclude that we can not find an allocation which acts as a reward, yet leads to an incentive compatible denouncement strategy.

Maskin Monotonicity Holds With respect to the "complete" deception: $\alpha_i(\theta_i) \neq \theta_i$ for all i and θ_i, the above discussion of ex post monotonicity already allows us to conclude that Maskin monotonicity is satisfied. The violation of the ex post incentive compatibility condition for either reward $y \in \{(0,1,0),(0,0,1)\}$ occurred at $\theta = (0,1,1)$, but not at the deception $\alpha(0,0,0) = (1,1,1)$ which is the only profile to be verified with Maskin monotonicity. For all other deceptions, it suffices to observe that at most two agents benefit from the deception $f(\alpha(\theta))$ relative to the social choice $f(\theta)$ and hence there is always a third agent who can be rewarded by simply offering him the allocation $y = f(\theta)$ at θ, which also guarantees the incentive compatibility of the reward.

Interim Monotonicity Holds on all Payoff Type Spaces We start by considering the "complete" deception: $\alpha_i(\theta_i) \neq \theta_i$ for all i and θ_i and then extend the argument to all deceptions. We first suggest a reward rule $y : \Theta \to A$ which will work for agent 1 at $\theta_1 = 0$ provided that $p((0,0)|0) > 0$ and $p((1,1)|0) \leq \frac{1}{1+\varepsilon}$. We offer the following contingent reward to agent 1:

$$y = \begin{cases} (0,0,1) & \text{if} \quad \theta = (1,1,1) \\ f & \text{if} \quad \theta \neq (1,1,1) \end{cases} \tag{32}$$

The reward condition at $\theta_1 = 0$ then reduces to, after eliminating terms on both sides of the inequality by using (32):

$$u_1((0,0,1),(0,0,0))\, p((0,0)|0) > u_1((1,1,1),(0,0,0))\, p((0,0)|0) \tag{33}$$

and the interim incentive compatibility conditions for $\theta_1 = 0$ is, after inserting the corresponding utilities,

$$1 \geq (1+\varepsilon) \cdot p((1,1)|0) \tag{34}$$

and for $\theta_1 = 1$:

$$1 \geq 1 - p((1,1)|1) \tag{35}$$

We observe that (33) is satisfied by hypothesis of $p((0,0)|0) > 0$, inequality (34) by hypothesis of $p((1,1)|0) \leq \frac{1}{1+\varepsilon}$ and inequality (35) is always satisfied.

For the instance of $p((0,0)|0) > 0$ but $p((1,1)|0) > \frac{1}{1+\varepsilon}$, we can offer a modified reward rule:

$$y = \begin{cases} (0,1,0) & \text{if} \quad \theta = (1,0,0) \\ f & \text{if} \quad \theta \neq (1,0,0) \end{cases} \tag{36}$$

which differs from the reward rule (32) only by the type profile at which it offers a reward. With this modified rule can then write the reward condition as:

$$u_1((0,1,0),(0,1,1))\, p((1,1)|0) > u_1((1,0,0),(0,1,1))\, p((1,1)|0) \tag{37}$$

and the incentive compatibility conditions for $\theta_1 = 0$ again after insert the utilities,

$$1 \geq (1+\varepsilon) \cdot p((0,0)|0) \tag{38}$$

and for $\theta_1 = 1$:

$$1 \geq 1 - p((0,0)|0) \tag{39}$$

By the hypothesis of $p((1,1)|0) > \frac{1}{1+\varepsilon}$, it follows that (37) and (38) holds, and (39) is always satisfied. We can thus conclude that we can satisfy interim monotonicity for the "complete" deception for all priors.

Consider finally all deceptions which are not complete in the above sense. In this case, there exists at least some agent i and some state θ_i where he reports the truth. It is also true that every deception must involve at least two agents who misreport for some types. (Observe that otherwise, we could simple replace the deception by a single agent with the true state which would strictly improve the welfare of the agent in question.) But at any type profile θ at which exactly two agents misreport, the payoff for every agent is 0, whereas it is 1 if we were to choose the corresponding social choice $f(\theta)$, which then provides the reward and guarantees ex post incentive compatibility.

We would like to point out that all of the above arguments did not depend on a common prior nor did we need to make any full support assumption. The only necessary ingredient to demonstrate the success of interim implementation was the fact that every payoff type has exactly one belief type generate by the conditional belief, derived from a common prior or not.

6.4.2 Example C

There are three agents, $i = 1, 2, 3$ and each agent has a binary payoff type space $\Theta_i = \left\{\theta_i, \theta_i'\right\}$. The allocation space is given by $A = \{a, b, c, d, z_1, z_2, z_3\}$. The social choice function $f : \Theta \to A$ is given by:

θ_3	θ_2	θ_2'	θ_3'	θ_2	θ_2'
θ_1	a	b	θ_1	b	c
θ_1'	b	c	θ_1'	c	d

The payoffs of the agents are identical for every allocation which appears at least once in the social choice function. It therefore suffices to represent the payoff of agent 1 for each of these four allocations $\{a, b, c, d\}$

a :

θ_3	θ_2	θ_2'	θ_3'	θ_2	θ_2'
θ_1	1	0	θ_1	0	0
θ_1'	0	0	θ_1'	0	0

b :

θ_3	θ_2	θ_2'	θ_3'	θ_2	θ_2'
θ_1	-1	ε	θ_1	ε	-1
θ_1'	ε	-1	θ_1'	-1	-1

c :

θ_3	θ_2	θ_2'	θ_3'	θ_2	θ_2'
θ_1	-1	-1	θ_1	-1	ε
θ_1'	-1	ε	θ_1'	ε	-1

d :

θ_3	θ_2	θ_2'	θ_3'	θ_2	θ_2'
θ_1	-1	-1	θ_1	-1	-1
θ_1'	-1	-1	θ_1'	-1	ε

The allocation a is efficient if all agents are of type θ_i and d is efficient if all agents are of type θ_i'. In the remaining case the allocation b is efficient if a majority of agents is of type θ_i and the allocation c is efficient if a majority of agents is of type θ_i'. The difference between allocation a and b, c, d is that if a is efficient it has a strongly positive payoff $1 \gg \varepsilon > 0$ and if a is inefficient, then it has a 0 payoff, but not a strongly negative payoffs as the other allocations. For this reason, receiving the allocation a even if it is not efficient is not as damaging as receiving any other inefficient allocation.

The allocations z_1, z_2, z_3 are not called upon by the social choice function and they are merely introduced to turn the environment into an economic environment. We specify the payoffs as

$$u_i(\theta, z_i) = x, \quad \forall i, \forall \theta$$

and

$$u_i(\theta, z_j) = -x, \quad \forall i \neq j, \ \forall \theta$$

The allocation z_i is thus the most preferred alternative for agent i in all states and for this reason cannot be used as a reward as it would immediately violate the incentive constraints in the monotonicity condition.

In the game induced by the direct mechanism there exists only one ex post equilibrium, namely truthtelling, whereas depending on the priors over the payoff type space there may be many interim equilibria. We shall now briefly argue that the social choice function indeed satisfies ex post monotonicity and then display uniform and independent priors over the payoff types for which interim monotonicity fails.

Ex Post Monotonicity The social choice function is maximizes the sum of utilities at every type profile θ. Thus if a deception α generates a different social outcome at θ than $f(\theta)$, or $f(\alpha(\theta)) \neq f(\theta)$, then we can always offer the reward $y = f(\theta)$ following the report $\alpha(\theta)$ to anyone of the three agents. Since the social choice function is ex post incentive compatible and efficient we satisfy the reward as well as the incentive constraints. This establishes ex post monotonicity.

Maskin Monotonicity The same reward strategy to elicit the use of deceptions by the agents also establishes that the social choice function satisfies Maskin monotonicity. Yet if we change the payoffs for all the agents resulting from allocation a at $\theta = \left(\theta'_1, \theta'_2, \theta'_3\right)$ and increase it from 0 to $u_i\left(a, \left(\theta'_1, \theta'_2, \theta'_3\right)\right) = 2\varepsilon$, then f no longer satisfies Maskin monotonicity for the deception $\alpha\left(\theta'_1, \theta'_2, \theta'_3\right) = (\theta_1, \theta_2, \theta_3)$ as we cannot offer a suitable reward to elicit the denunciation. Yet, the social choice function f preserves ex post monotonicity in this modified environment as the incomplete information deception $\alpha_i\left(\theta'_i\right) = \theta_i$ for all i leads has type profiles, say $\left(\theta_1, \theta'_2, \theta'_3\right)$, where the misreports by agent 2 and agent 3 lead the social choice function to select either a or b when c is the efficient choice and indeed can be used as a reward to eliminate the possibility of deceptive equilibrium. Thus this example shows that a social choice function may satisfy ex post monotonicity, yet display or not display Maskin monotonicity.

Interim Monotonicity Finally consider the notion of interim monotonicity with a uniform prior over the payoff type space:

$$p(\theta) = \frac{1}{8}, \ \forall \theta.$$

For this type space we analyze the following "pooling" deception in which every agent always reports his type to be θ_i:

$$\alpha_i(\cdot) = \theta_i, \quad \forall i, \forall \theta_i,$$

Under this deception, the social choice function recommends to select allocation a for all true payoff type profiles. As the designer attempts to identify a reward allocation $y : \Theta \to A$, he faces the problem that all types report identically θ_i, and he has to offer a single allocation regardless of the true type profile. Thus he is necessarily forced to select an allocation, different from a, at payoff type profiles where it is not efficient. With the given payoffs this will lead to substantial utility losses whereas the allocation a, even if it is not efficient, only leads to a small payoff loss. With the uniform prior, the best possible reward structure relative to the equilibrium utility is to offer c to an agent i of type θ'_i, yet when we evaluate the reward inequality:

$$\sum_{\theta_{-i} \in \Theta_{-i}} u_i\left(y\left(\alpha\left(\theta\right)\right), \theta\right) p\left(\theta_{-i} \left| \theta_i \right.\right) > \sum_{\theta_{-i} \in \Theta_{-i}} u_i\left(f\left(\alpha\left(\theta\right)\right), \theta\right) p\left(\theta_{-i} \left| \theta_i \right.\right)$$

we obtain

$$\varepsilon\left(\frac{1}{4} + \frac{1}{4}\right) + (-1)\left(\frac{1}{4} + \frac{1}{4}\right) > 0$$

which is clearly violated for small ε and hence interim monotonicity will be violated for a large sets of priors over the payoff type space.

7 Implementation on All Type Spaces

Proposition 1 *If f is interim implementable on every type space $\mathcal{T}$, then f satisfies (EPIC) and robust monotonicity.*

The necessity of EPIC is proved in our companion paper, Bergemann and Morris (2003).

To prove the necessity of robust monotonicity, it is enough to show the following lemma.

Proposition 2 *If f is interim implementable on type space $\mathcal{T}$, then f satisfies interim monotonicity on type space $\mathcal{T}$.*

This argument is standard from the Bayesian implementation literature and dates back to Postlewaite and Schmeidler (1986). For completeness, we report a proof. One subtlety is that such results are usually stated under the assumption of common support, with implementation required only on that support. We do not make this assumption and we emphasize that our definition requires implementation at every type profile whether or not any agent thinks it occurs with positive probability.

Proof. Let (M, g) implement f with equilibrium strategies $s_i : T_i \to M_i$. Suppose that for some deception α, $f \circ \widehat{\theta} \neq f \circ \widehat{\theta} \circ \alpha$. It must be that $s \circ \alpha$ is not an equilibrium. Therefore there exists i, $t_i \in T_i$ and $m_i \in M_i$ such that

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

Let $y \triangleq g \left(m_i, s_{-i} \left(\alpha_{-i} \left(t_{-i} \right) \right) \right)$. Then, from above,

$$u_i \left(y, \theta \right) > u_i \left(f \left(\alpha \left(\theta \right) \right), \theta \right).$$

But since s is an equilibrium it follows that

$$\begin{aligned} u_i \left(f \left(\theta_i', \alpha_{-i} \left(\theta_{-i} \right) \right), \left(\theta_i', \alpha_{-i} \left(\theta_{-i} \right) \right) \right) &= u_i \left(g \left(s \left(\theta_i', \alpha_{-i} \left(\theta_{-i} \right) \right) \right), \left(\theta_i', \alpha_{-i} \left(\theta_{-i} \right) \right) \right) \\ &\geq u_i \left(g \left(m_i, s_{-i} \left(\alpha_{-i} \left(\theta_{-i} \right) \right) \right), \left(\theta_i', \alpha_{-i} \left(\theta_{-i} \right) \right) \right) \\ &= u_i \left(y, \left(\theta_i', \alpha_{-i} \left(\theta_{-i} \right) \right) \right), \forall \theta_i' \in \Theta_i. \end{aligned}$$

This establishes that $y \in Y_i^* \left(\theta_{-i} \right)$. ■

Definition 13 (Robustly Economic Environment)
An environment is robustly economic at θ if for any $y^* \in Y$, there exist i and j, $i \neq j$, and allocations a and b such that

$$\begin{aligned} u_i \left(a, \theta \right) &> u_i \left(y^*, \theta \right) \\ \text{and } u_i \left(a, \left(\theta_i, \theta_{-i}' \right) \right) &\geq u_i \left(y, \left(\theta_i, \theta_{-i}' \right) \right) \end{aligned}$$

for all $y \in Y$ and $\theta_{-i}' \in \Theta_{-i}$; and

$$\begin{aligned} u_j \left(b, \theta \right) &> u_j \left(y^*, \theta \right) \\ u_j \left(b, \left(\theta_j, \theta_{-j}' \right) \right) &\geq u_j \left(y, \left(\theta_j, \theta_{-j}' \right) \right) \end{aligned}$$

for all $y \in Y$ and $\theta_{-j}' \in \Theta_{-j}$. An environment is robustly economic if it is robustly economic at all θ.

Evidently, a robustly economic environment is also an ex post economic environment as the former shares the strict inequality with the later condition. Yet, the former is more stringent requirement through the addition of the weak inequalities which are necessitated by the robust implementation. As different types t_i of agent i may share the same payoff type $\widehat{\theta}_i \left(t_i \right)$, but pursue different deception strategies, the weak inequalities have to hold for the allocations $a, b \in Y$ against all allocations $y \in Y$ rather than a selection $y : \Theta \to Y$ as in the notion of an interim economic environment.

Proposition 3 (Sufficiency of Robust Monotonicity)
In robustly economic environments with $I \geq 3$, if f satisfies robust monotonicity and (EPIC), then there exists a mechanism that implements f on all full support type spaces.

The following mechanism $\mathcal{M}^R$ will be employed in the proof. Each agent reports a message $m_i = (\theta_i, z_i, \gamma_i, y_i)$, where $\theta_i \in \Theta_i$; z_i is a non-negative integer, $\gamma_i : \Theta_{-i} \to A$ is a mapping from payoff type profiles to outcomes, satisfying the property that $\gamma_i(\theta_{-i}) \in Y_i^*(\theta_{-i})$ for all θ_{-i}; and $y_i \in Y$. Outcomes are determined by the following rules:

1. If $z_i = 0$ for all i, then $g(m) = f(\theta)$.
2. If there exists j such that $z_i = 0$ for all $i \neq j$ and $z_j \geq 1$, then $g(m) = \gamma_j(\theta_{-j})$.
3. In all other cases, $g(m) = y_{\widetilde{j}(z)}$, where $\widetilde{j}(z)$ is uniquely determined by the following rules:
   (a) $z_{\widetilde{j}(z)} \geq z_i$ for all i;
   (b) if $z_i = z_{\widetilde{j}(z)}$, then $i \geq \widetilde{j}(z)$.

Proof. Fix any common support type space $\mathcal{T}$ (with common support T^*).

STEP 1. There is an equilibrium where every type t_i of agent i always sends a message of the form $\left(\widehat{\theta}_i(t_i), 0, \cdot, \cdot\right)$. This strategy profile implements f.

No agent has an incentive to deviate. By choosing a message of the form $\left(\theta_i', 0, \cdot, \cdot\right)$, his payoff gain (if his opponents have payoff type profile θ_{-i}) is

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

which is less than or equal to 0 by EPIC. By choosing a message of the form $(\theta_i, z_i, \gamma_i, \cdot)$ with $z_i \geq 1$, his payoff gain (if his opponents have payoff type profile θ_{-i}) is

$$u_i\left(\gamma_i(\theta_{-i}), (\theta_i, \theta_{-i})\right) - u_i\left(f(\theta_i, \theta_{-i}), (\theta_i, \theta_{-i})\right)$$

which is less than or equal to 0 by the requirement that $\gamma_i(\theta_{-i}) \in Y_i^*(\theta_{-i})$.

STEP 2. In any equilibrium, all types of each agent sends a message of the form $(\cdot, 0, \cdot, \cdot)$.

We argue by contradiction. First suppose that there exists j and $t_j \in T_j$ with $s_j(t_j) = (\cdot, z_j, \cdot, \cdot)$ for some $z_j \geq 1$. Pick any t_{-j} such that $(t_j, t_{-j}) \in T^*$. Now by the robustly economic environment, there exists an agent $i \neq j$ and an allocation a such that

$$u_i\left(a, \widehat{\theta}(t)\right) > u_i\left(g(s(t)), \widehat{\theta}(t)\right)$$

and

$$u_i\left(a, \widehat{\theta}\left(t_i, t_{-i}'\right)\right) \geq u_i\left(g\left(s\left(t_i, t_{-i}'\right)\right), \widehat{\theta}\left(t_i, t_{-i}'\right)\right) \text{ for all } t_{-i}'.$$

But then agent i could strictly increase his utility by setting $s_i(t_i) = (\cdot, z_i, \cdot, a)$, with z_i higher than the integer chosen by any other type in equilibrium.

STEP 3. In any equilibrium where all types of each agent send a message of the form $(\cdot, 0, \cdot, \cdot)$, social choice function f is implemented.

Again, we argue by contradiction. Let

$$\beta_i(\theta_i) = \{\theta_i\} \cup \left\{\theta_i' \in \Theta_i : \exists t_i \text{ with } \widehat{\theta}_i(t_i) = \theta_i \text{ and } s_i(t_i) = \left(\theta_i', 0, \cdot, \cdot\right)\right\}.$$

If f is not implemented, we must have that β is not acceptable. By robust monotonicity, there exists i, θ_i and $\theta_i' \in \beta_i(\theta_i)$ such that for every θ_{-i}' and $\psi_i \in \Delta\left(\beta_{-i}^{-1}\left(\theta_{-i}'\right)\right)$, there exists $y\left(\theta_{-i}', \psi_i\right) \in Y_i\left(\theta_{-i}'\right)$ such that

$$\sum_{\theta_{-i} \in \Theta_{-i}} \psi_i(\theta_{-i})\, u_i\left(y\left(\theta_{-i}', \psi_i\right), (\theta_i, \theta_{-i})\right) > \sum_{\theta_{-i} \in \Theta_{-i}} \psi_i(\theta_{-i})\, u_i\left(f\left(\theta_i', \theta_{-i}'\right), (\theta_i, \theta_{-i})\right).$$

Now pick any t_i such that $\widehat{\theta}_i(t_i) = \theta_i$ and $s_i(t_i) = \left(\theta_i', 0, \cdot, \cdot\right)$. Let

$$\psi_i\left(\theta_{-i}|\theta_{-i}'\right) = \frac{\sum_{\left\{t_{-i}\in T_{-i}:\widehat{\theta}_j(t_j)=\theta_j \text{ and } s_j(t_j)=\left(\theta_j',0,\cdot,\cdot\right),\forall j\neq i\right\}} \widehat{\pi}_i[t_i](t_{-i})}{\sum_{\left\{t_{-i}\in T_{-i}:\ s_j(t_j)=\left(\theta_j',0,\cdot,\cdot\right),\forall\ j\neq i\right\}} \widehat{\pi}_i[t_i](t_{-i})}.$$

If type t_i follows his proposed strategy, he obtains utility

$$\begin{aligned} &\sum_{t_{-i}\in T_{-i}} \widehat{\pi}_i[t_i](t_{-i})\, u_i\left(f\left(s^1(t)\right), \widehat{\theta}(t)\right) \qquad (40)\\ =\ &\sum_{\left\{t_{-i}:\ s_j(t_j)=\left(\theta_j',0,\cdot,\cdot\right),\forall j\neq i\right\}} \widehat{\pi}_i[t_i](t_{-i}) \sum_{\theta_{-i}\in\Theta_{-i}} \psi_i\left(\theta_{-i}|\theta_{-i}'\right) u_i\left(f\left(\theta_i',\theta_{-i}'\right),(\theta_i,\theta_{-i})\right). \end{aligned}$$

If instead, he chooses $(\cdot, 1, \gamma_i, \cdot)$, where

$$\gamma_i(\theta_{-i}) = y\left(\theta_{-i}, \psi_i(\cdot\ |\theta_{-i})\right)$$

for each θ_{-i}, then he obtains utility

$$\begin{aligned} &\sum_{t_{-i}\in T_{-i}} \widehat{\pi}_i[t_i](t_{-i})\, u_i\left(\gamma_i\left(s_{-i}^1(t_{-i})\right), \widehat{\theta}(t)\right) \qquad (41)\\ =\ &\sum_{\left\{t_{-i}\in T_{-i}: s_j(t_j)=\left(\theta_j',0,\cdot,\cdot\right),\forall j\neq i\right\}} \widehat{\pi}_i[t_i](t_{-i}) \sum_{\theta_{-i}\in\Theta_{-i}} \psi_i\left(\theta_{-i}|\theta_{-i}'\right) u_i\left(y\left(\theta_{-i}', \psi_i(\cdot\ |\theta_{-i})\right),(\theta_i,\theta_{-i})\right). \end{aligned}$$

But now the whistle blower inequality (**??**) implies that (41) is strictly greater than (40). ■

7.1 Iterative Implementation

We now show that interim implementation on all type spaces (including non-common support type spaces) is equivalent to implementation in the strategy set surviving iterated deletion of never weak best responses.

We begin by setting the notation for iterated deletion of never weak best responses. For a fixed mechanism $\mathcal{M} = (M_1, ..., M_I, g)$, we define the set of surviving reports for agent i of payoff type θ_i after k rounds $\left\{M_i^k(\theta_i)\right\}_{i,\theta_i\in\Theta_i}$ recursively as follows. Let $M_i^0(\theta_i) = M_i$ and define recursively:

$$M_i^{k+1}(\theta_i) = \left\{ m_i \in M_i^k(\theta_i) : \left| \begin{array}{l} \text{there exists } \lambda_i \in \Delta\left(M_{-i}\times\Theta_{-i}\right) \text{ such that} \\ (1)\ \lambda_i(m_{-i},\theta_{-i}) > 0 \Rightarrow m_j \in M_j^{k-1}(\theta_j) \text{ for each } j\neq i \\ (2)\ \sum\limits_{m_{-i},\theta_{-i}} \lambda_i(m_{-i},\theta_{-i}) \left[\begin{array}{c} u_i\left(g\left(m_i,m_{-i}\right),(\theta_i,\theta_{-i})\right) \\ -u_i\left(g\left(m_i',m_{-i}\right),(\theta_i,\theta_{-i})\right) \end{array}\right] \geq 0 \text{ for all } m_i' \in M_i \end{array} \right. \right\}.$$

We write

$$M_i^*(\theta_i) = \bigcap_{k\geq 0} M_i^k(\theta_i) \quad \text{and} \quad M^\infty(\theta) = \left\{M_i^\infty(\theta_i)\right\}_{i=1}^I.$$

When the mechanism is finite, iterative deletion of never weak best responses is equivalent to iterated deletion of strictly dominated strategies; and a countable number of rounds of deletion is enough to converge to a fixed point. If there are an infinite number of messages, we might need to delete a transfinitely, as noted by Lipman (1994) in a complete information contect. In this case, we should understand the above definition as consisting of enough rounds of deletion to reach a fixed point.

Definition 14 *Social choice function f is iterative implementable if there exists a mechanism $\mathcal{M}$ such that*

$$m \in M^{\infty}(\theta) \Rightarrow g(m) = f(\theta).$$

We refer to iterative implementable rather than the more exhaustive implementable in strategies surviving iterated deletion of never weak best responses. We next present two examples to illustrate this definition. The first example augments the introductory example by two additional outcomes which are not called upon by the social choice function f. This example has the feature that the social choice function is iterative implementable, yet not dominant strategy implementable. We show iterative implementability by explicitly constructing the mechanism. The second example exactly reprises the introductory example and shows that even though there the social choice function f is ex post implementable, there does not exist a mechanism which would make f iterative implementable.

7.2 Example D

The introductory Example A had two agents, $i = 1, 2$ with binary payoff types: $\Theta_1 = \{\theta_1, \theta_1'\}$, $\Theta_2 = \{\theta_2, \theta_2'\}$. The only variation is in the allocation space $A = \{a, b, c, d, z_1, z_2\}$ which contains the additional elements z_1 and z_2. The social choice function is still given by:

f	θ_2	θ_2'
θ_1	a	b
θ_1'	c	d

and the payoffs of the agents remain identical for the original allocations $\{a, b, c, d\}$:

a	θ_2	θ_2'
θ_1	$3,3$	$0,0$
θ_1'	$0,0$	$1,1$

b	θ_2	θ_2'
θ_1	$0,0$	$3,3$
θ_1'	$1,1$	$0,0$

c	θ_2	θ_2'
θ_1	$0,0$	$1,1$
θ_1'	$3,3$	$0,0$

d	θ_2	θ_2'
θ_1	$1,1$	$0,0$
θ_1'	$0,0$	$3,3$

and for z_1 and z_2 are given by:

z_1	θ_2	θ_2'
θ_1	$2,2$	$2,0$
θ_1'	$2,2$	$2,0$

z_2	θ_2	θ_2'
θ_1	$2,0$	$2,2$
θ_1'	$2,0$	$2,2$

Consider the following augmented mechanism $g(\cdot)$ in which agent 1 can report besides his payoff type also a third message ϕ whereas agent 2 is again restricted to report his payoff type:

$g(\cdot)$	θ_2	θ_2'
θ_1	a	b
θ_1'	c	d
ϕ	y	z

The corresponding incomplete information game has the following payoffs:

	type	θ_2		θ_2'	
type	report	θ_2	θ_2'	θ_2	θ_2'
θ_1	θ_1	$3,3$	$0,0$	$0,0$	$3,3$
	θ_1'	$0,0$	$1,1$	$1,1$	$0,0$
	ϕ	$2,2$	$2,0$	$2,0$	$2,2$
θ_1'	θ_1	$0,0$	$1,1$	$1,1$	$0,0$
	θ_1'	$3,3$	$0,0$	$0,0$	$3,3$
	ϕ	$2,2$	$2,0$	$2,0$	$2,2$

If we perform iterated deletion of never weak best responses, then we arrive in four steps at a singleton for every type of every agent:

$$\begin{aligned}
M_1^0(\theta_1) &= \left\{\theta_1, \theta_1', \phi\right\}, M_1^0\left(\theta_1'\right) = \left\{\theta_1, \theta_1', \phi\right\}, M_2^0(\theta_2) = \left\{\theta_2, \theta_2'\right\}, M_2^0\left(\theta_2'\right) = \left\{\theta_2, \theta_2'\right\} \\
M_1^1(\theta_1) &= \{\theta_1, \phi\}, M_1^1\left(\theta_1'\right) = \left\{\theta_1', \phi\right\}, M_2^1(\theta_2) = \left\{\theta_2, \theta_2'\right\}, M_2^1\left(\theta_2'\right) = \left\{\theta_2, \theta_2'\right\} \\
M_1^2(\theta_1) &= \{\theta_1, \phi\}, M_1^2\left(\theta_1'\right) = \left\{\theta_1', \phi\right\}, M_2^2(\theta_2) = \{\theta_2\}, M_2^2\left(\theta_2'\right) = \left\{\theta_2'\right\} \\
M_1^3(\theta_1) &= \{\theta_1\}, M_1^3\left(\theta_1'\right) = \left\{\theta_1'\right\}, M_2^3(\theta_2) = \{\theta_2\}, M_2^3\left(\theta_2'\right) = \left\{\theta_2'\right\}
\end{aligned}$$

7.3 Example A Revisited

We now return to the original example and simply omit the allocations z_1 and z_2. Here we will prove that the social choice function is not iterative implementable. We argue by contradiction. Thus suppose that there is a finite mechanism $\mathcal{M}$ such that

$$m \in M^\infty(\theta) \Rightarrow g(m) = f(\theta).$$

Let

$$M_i^*(\theta_i) = \{m_i : g(m_i, m_j) = f(\theta_i, \theta_j) \text{ for some } m_j, \theta_j\}.$$

By induction, $M_i^*(\theta_i) \subseteq M_i^k(\theta_i)$ for all k. Suppose that this is true for k. Then for any $m_i \in M_i^*(\theta_i) \subseteq M_i^k(\theta_i)$, there exists $m_j \in M_j^*(\theta_j) \subseteq M_j^k(\theta_j)$ such that $g(m_i, m_j) = f(\theta_i, \theta_j)$. So $m_i \in M_i^{k+1}(\theta_i)$.

Thus we must have that $(m_1, m_2) \in M_1^*(\theta_1) \times M_2^*(\theta_2)$ implies $g(m_1, m_2) = f(\theta_1, \theta_2)$. Let $m_i^*(\cdot)$ be any selection from $M_i^*(\cdot)$. Now let k^* be the lowest k such that, for some i,

$$m_i^*\left(\theta_i'\right) \notin M_i^k(\theta_i).$$

Without loss of generality, let $i = 1$. Note $m_2^*\left(\theta_2'\right) \in M_2^{k-1}(\theta_2)$ by assumption. If agent 1 was type θ_1 and was sure his opponent were type θ_2 and choosing action $m_2^*\left(\theta_2'\right)$, we know that he could guarantee himself a payoff of 1 by choosing $m_1^*\left(\theta_1'\right)$. Since $m_1^*\left(\theta_1'\right)$ is deleted for type θ_1 at round k, we know that there exists $m_1' \in M_1$ such that

$$g_1\left(m_1', m_2^*\left(\theta_2'\right)\right) > 1$$

and thus there exists m_1' such that $g_1\left(m_1', m_2^*\left(\theta_2'\right)\right) = f(\theta_1, \theta_2)$. This implies that $m_2^*\left(\theta_2'\right) \in M_2^*(\theta_2)$, a contradiction.

Both examples use the fact that the social choice function always selects an outcome that is strictly Pareto-optimal and - paradoxically - it this feature which inhibits iterative implementation in the current example.[5] Borgers (1995) proves the impossibility of complete information implementation of non-dictatorial social choice functions in iteratively undominated strategies when the set

[5] In this context, it is worthwhile to observe that the social choice function f in Example A satisfies robust monotonicity. Example A

We return to Example A of Section 2 and verify that is satisfies robust monotonocity. To see why, first observe that $Y_i^*(\theta_{-i}) = Y$ for all i and θ_{-i}. So it is enough to show that for any $\theta_i \neq \theta_i'$, $\theta_j \neq \theta_j'$, $\widehat{\theta}_j$ and $\psi \in [0,1]$, there exists $y \in Y$ such that

$$\psi u_i(y, (\theta_i, \theta_j)) + (1-\psi) u_i(y, (\theta_i, \theta_j')) > \psi u_i\left(f\left(\theta_i', \widehat{\theta}_j\right), (\theta_i, \theta_j)\right) + (1-\psi) u_i\left(f\left(\theta_i', \widehat{\theta}_j\right), (\theta_i, \theta_j')\right).$$

But this is true, since if $\psi > \frac{1}{3}$, we can set $y = f(\theta_i, \theta_j)$, and if $\psi < \frac{2}{3}$, we can set $y = f\left(\theta_i, \theta_j'\right)$.

Since robust monotonicity is satisfied, Maskin monotonicity, ex post monotonicity and interim monotonicity on every type space are also satisfied.

Yet, as the environment is distinctly non-economic, monotonicity is only a necessary but not sufficient condition for interim implementation. More precisely, in this example any attempt to create a reward allocation has to rely on the use of the efficient allocations, and this necessarily creates mutliple equilibria, not all of them implement the social choice function f.

of feasible preference profiles includes such unanimous preference profiles and the argument here is reminiscent of Borgers' argument.

7.4 Characterization

We will use the following straightforward lemma.

Lemma 2 *For each $m_i \in M_i^\infty(\theta_i)$, there exists $\lambda_i \in \Delta(\Theta_{-i} \times M_{-i})$ such that:*

1. *$\lambda_i(\theta_{-i}, m_{-i}) = 0$ if $m_j \notin M_j^\infty(\theta_j)$ for some $j \neq i$;*
2. *$m_i \in \arg\max_{m_i'} \sum_{\theta_{-i}, m_{-i}} \lambda_i(\theta_{-i}, m_{-i}) u_i(g(m_i', m_{-i}), (\theta_i, \theta_{-i}))$.*

Definition 15 *Social choice function f is uniformly implementable if there exists a mechanism $\mathcal{M}$ such that for every finite type space $\mathcal{T}$, every (pure strategy) interim equilibrium s of the game $(\mathcal{T}, \mathcal{M})$ satisfies*

$$g(s(t)) = f\left(\widehat{\theta}(t)\right).$$

We then use the characterization of iterative implementation provided by Lemma 2 to relate iterative implementation and implementation on all type spaces.

Theorem 7 *There is a mechanism which implements social choice function f on all type spaces if and only if f is iteratively implementable.*

Proof. First, suppose that f is iterative implementable. Fix any type space $\mathcal{T}$. Choose a mechanism $\mathcal{M}$ that iterative implements f. Fix any equilibrium s of the game $(\mathcal{T}, \mathcal{M})$ and let

$$\widehat{M}_i(\theta_i) = \left\{m_i : s_i(t_i) = m_i \text{ and } \widehat{\theta}_i(t_i) = \theta_i\right\}.$$

By induction, $\widehat{M}_i(\theta_i) \subseteq M_i^k(\theta_i)$ for all k, and thus $\widehat{M}_i(\theta_i) \subseteq M_i^\infty(\theta_i)$. Now $g(s(t)) = f\left(\widehat{\theta}(t)\right)$.

Now suppose that f is not iterative implementable. Then for any mechanism $\mathcal{M}$, there exists m^* such that $m^* \in M^\infty(\theta^*)$ but $g(m^*) \neq f(\theta^*)$. Recall from Lemma 2 that for each $m_i \in M_i^\infty(\theta_i)$, there exists $\lambda_i(\cdot|m_i) \in \Delta(\Theta_{-i} \times M_{-i})$ such that:

1. $\lambda_i(\theta_{-i}, m_{-i}) = 0$ if $m_j \notin M_j^\infty(\theta_j)$ for some $j \neq i$;
2. $m_i \in \arg\max_{m_i'} \sum_{\theta_{-i}, m_{-i}} \lambda_i(\theta_{-i}, m_{-i}) u_i(g(m_i', m_{-i}), (\theta_i, \theta_{-i}))$.

Now we construct a type space where

$$\begin{aligned} T_i &= \{(\theta_i, m_i) \in \Theta_i \times M_i : m_i \in M_i^\infty(\theta_i)\} \\ \widehat{\theta}_i((\theta_i, m_i)) &= \theta_i \\ \widehat{\pi}_i((\theta_i, m_i))\left[(\theta_j, m_j)_{j \neq i}\right] &= \lambda_i(\theta_{-i}, m_{-i} | m_i). \end{aligned}$$

By construction, there is an equilibrium s of the game $(\mathcal{T}, \mathcal{M})$ with

$$s_i((\theta_i, m_i)) = m_i.$$

But now $g(s(\theta^*, m^*)) = g(m^*) \neq f(\theta^*)$, while $\widehat{\theta}(\theta^*, m^*) = \theta^*$. ■

This argument is a straightforward application of a more general game theoretic argument. Brandenburger and Dekel (1987) showed that the following result. Fix a complete information game and a type space. Since there is complete information, all types are identical in terms of payoffs, but may differ in their beliefs over others' types. Ask which actions may be played in a Bayesian Nash equilibrium of this rather degenerate incomplete information game on any type space (including those where agents' beliefs are not derived from a common prior). This is equivalent to asking which actions may be played in a subjective correlated equilibrium of the underlying complete information game. Brandenburger and Dekel show that the answer is the set of all actions which survive iterated deletion of strictly dominated strategies.

This result can be extended to an incomplete information setting as follows. Let each agent i have one of a finite set of payoff types, Θ_i. Fix an incomplete information payoff function, where agents' payoffs depend on the profile of actions chosen and the profile of payoff types. Take any rich type space of the form we defined in Section 3.2, where an agent's type includes a description of his payoff type and his beliefs about others' types. Ask which actions might be played by a given payoff type in any equilibrium of the resulting game, for any type space. The answer is the set of actions that survive iterated deletion of strictly dominated actions, where an action is dominated for a payoff type if there is a mixed strategy that gives a strictly higher payoff for every action/payoff type profile of the remaining players that has not yet been deleted. Proposition 7 is direct application of this result. Battigalli and Siniscalchi (2003) have reported incomplete information generalizations of the Brandenburger and Dekel (1987) that can incorporate the argument here as a special case. Arguments in Lipman (1994) can be used to show the extension to infinite actions.

8 Private Values and Dominant Strategies

We conclude this section by noting the connection between robust monotonicity and dominant strategies.

Definition 16 *Social choice function f satisfies strict dominant strategies incentive compatibility if for all i, θ, θ' with $\theta'_i \neq \theta_i$,*

$$u_i\left(f\left(\theta_i, \theta'_{-i}\right), \theta\right) > u_i\left(f\left(\theta'_i, \theta'_{-i}\right), \theta\right).$$

Definition 17 *Social choice function f satisfies dominant strategies incentive compatibility if for all i, θ, θ',*

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

Intermediate between these notions, we have conditions where strict inequalities are required only at some subset of deviations.

Definition 18 *Social choice function f satisfies selective dominant strategy incentive compatibility (SDI) if for every deception α with $f \neq f \circ \alpha$, there exists i and θ such that*

$$u_i\left(f\left(\theta_i, \alpha_{-i}\left(\theta_{-i}\right)\right), \theta\right) > u_i\left(f\left(\alpha_i\left(\theta_i\right), \alpha_{-i}\left(\theta_{-i}\right)\right), \theta\right).$$

Definition 19 *Social choice function f satisfies selective dominant strategies incentive compatibility ($SD2$) if f satisfies dominant incentive compatibility and, for all unacceptable deceptions β, there exists i, θ_i and $\theta'_i \in \beta_i(\theta_i)$ such that*

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

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

First, we observe that SD1 and SD2 imply ex post monotonicity and robust monotonicity, respectively.

Lemma 3 *If social choice function f satisfies SD1 and EPIC, then f satisfies ex post monotonicity.*

PROOF. Fix any deception α with $f \neq f \circ \alpha$. By SD1, there exists i and θ such that

$$u_i\left(f\left(\theta_i, \alpha_{-i}\left(\theta_{-i}\right)\right), \theta\right) > u_i\left(f\left(\alpha_i\left(\theta_i\right), \alpha_{-i}\left(\theta_{-i}\right)\right), \theta\right).$$

Setting $y \equiv f\left(\theta_i, \alpha_{-i}\left(\theta_{-i}\right)\right)$, we have

$$u_i\left(y, \theta\right) > u_i\left(f\left(\alpha_i\left(\theta_i\right), \alpha_{-i}\left(\theta_{-i}\right)\right), \theta\right).$$

But by EPIC,

$$\begin{aligned} u_i\left(f\left(\theta_i', \alpha_{-i}\left(\theta_{-i}\right)\right), \left(\theta_i', \alpha_{-i}\left(\theta_{-i}\right)\right)\right) &\geq u_i\left(f\left(\theta_i, \alpha_{-i}\left(\theta_{-i}\right)\right), \left(\theta_i', \alpha_{-i}\left(\theta_{-i}\right)\right)\right) \\ &= u_i\left(y, \left(\theta_i', \alpha_{-i}\left(\theta_{-i}\right)\right)\right) \end{aligned}$$

for all θ_i'. So $y \in Y_i^*\left(\alpha_{-i}\left(\theta_{-i}\right)\right)$.

Lemma 4 *If social choice function f satisfies SD2 and EPIC, then f satisfies robust monotonicity.*

PROOF. Fix any unacceptable deception β. If f satisfies SD2, there exist i, θ_i and $\theta_i' \in \beta_i\left(\theta_i\right)$ with

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

for all θ_{-i} and $\theta_{-i}' \in \beta_{-i}\left(\theta_{-i}\right)$. Setting $y = f\left(\theta_i, \theta_{-i}'\right)$, we have

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

for all θ_{-i} and $\theta_{-i}' \in \beta_{-i}\left(\theta_{-i}\right)$. Now by EPIC,

$$\begin{aligned} u_i\left(f\left(\widetilde{\theta}_i, \theta_{-i}'\right), \left(\widetilde{\theta}_i, \theta_{-i}'\right)\right) &\geq u_i\left(f\left(\theta_i, \theta_{-i}'\right), \left(\widetilde{\theta}_i, \theta_{-i}'\right)\right) \\ &= , u_i\left(y, \left(\widetilde{\theta}_i, \theta_{-i}'\right)\right) \end{aligned}$$

for all $\left(\widetilde{\theta}_i, \theta_{-i}'\right) \in \Theta$.

We also show that under the private value assumption, SD1 and SD2 are implied by ex post monotonicity and robust monotonicity, respectively.

Definition 20 *The social choice environment satisfies private values if*

$$u_i\left(y, \left(\theta_i, \theta_{-i}\right)\right) = \widehat{u}_i\left(y, \theta_i\right)$$

for all i, y, θ_i and θ_{-i}.

Lemma 5 *In a private values environment, if f satisfies ex post monotonicity, then f satisfies SD1.*

Proof. By ex post monotonicity, for every deception α with $f \neq f \circ \alpha$, there exists i, θ and $y \in Y_i^*\left(\alpha_{-i}\left(\theta_{-i}\right)\right)$ such that

$$u_i\left(y, \theta_i\right) > u_i\left(f\left(\alpha\left(\theta\right)\right), \theta_i\right)$$

and

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

for all θ_i'. Thus

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

■

Clearly, strict dominant implies selective dominant which in turn implies dominant strategies incentive compatibility. The relationship between selective dominant strategies incentive compatibility and robust monotonicity is established next.

Lemma 6 *In a private values environment, if f satisfies robust monotonicity, then f satisfies SDs.*

Proof. By robust monotonicity condition, for every unacceptable deception β, there exist i, θ_i, $\theta_i' \in \beta_i\left(\theta_i\right)$, such that for every θ_{-i}' and $\psi_i \in \Delta\left(\beta_{-i}^{-1}\left(\theta_{-i}'\right)\right)$, there exists $y\left(\theta_{-i}',\psi_i\right)$ with

$$\sum_{\theta_{-i}} \psi_i\left(\theta_{-i}\right)\widehat{u}_i\left(y\left(\theta_{-i}',\psi_i\right),\theta_i\right) > \sum_{\theta_{-i}} \psi_i\left(\theta_{-i}\right)\widehat{u}_i\left(f\left(\theta_i',\theta_{-i}'\right),\theta_i\right).$$

and

$$\widehat{u}_i\left(f\left(\widetilde{\theta}_i,\theta_{-i}'\right),\widetilde{\theta}_i\right) \geq \widehat{u}_i\left(y\left(\theta_{-i}',\psi_i\right),\widetilde{\theta}_i\right)$$

for all $\widetilde{\theta}_i$. Thus

$$\widehat{u}_i\left(f\left(\theta_i,\theta_{-i}'\right),\theta_i\right) \geq \widehat{u}_i\left(y\left(\theta_{-i}',\psi_i\right),\theta_i\right) > \widehat{u}_i\left(f\left(\theta_i',\theta_{-i}'\right),\theta_i\right)$$

for all θ_{-i}'. ■

9 Discussion

To simplify the presentation and to facilitate comparisons with the existing literature, we maintained a number of standard (perhaps unfortunately) assumptions from the literature: infinite actions games were allowed, type spaces were finite, and only pure strategy equilibria were allowed. We briefly discussed how the results would vary if we relaxed those assumptions.

9.1 Infinite Action Games

We used "integer games" to ensure that actions we added to the direct in the augmented mechanism were never played in equilibrium. As is standard in the literature (e.g., Jackson (1991)), it would be straightforward to replace the integer game with finite action "modulo games." This would change our results in two ways. First, it would imply that our result showing the sufficiency of robust monotonicity for interim implementation on all type spaces would require a different mechanism for every finite type space. Our existing result used a single mechanism for all type spaces. Second, in this case the pure strategy would have bite for the interim implementation on all type spaces problem.

9.2 Finite Type Spaces

For simplicity, we restricted attention to finite type spaces. We use the discrete type space to have an unambiguous notion of common support.

9.3 The Pure Strategy Restriction

For the ex post implementation question, the pure strategy restriction has bite. For interim implementation on all type spaces, if pure strategy implementation was possible but mixed strategy implmentation was not, we could always add types to purify the bad mixed strategy equilibria.

9.4 Conclusion

This paper examined the robustness of the classical implementation problem. We formalized robustness by requiring that the implementation problem remains solvable as we gradually relax common knowledge among the agents and the designer. The weakening of common knowledge was achieved by considering large type spaces in which the private information of the individual agents becomes more prominent.

Motivated by the recent literature on mechanism design with interdependent valuations which focuses on the notion of ex post equilibrium we presented initially necessary and sufficient conditions for ex post implementation. We then proceeded to relate interim implementation on large type spaces to ex post and complete information implementation. The obtained results point to the essential role of type spaces and the representation of private information in the implementation problem. While interim implementation on *all common prior type spaces* implies ex post and complete information implementation, the implication fails to hold if we were to consider only *all common prior payoff type spaces*, wherein the canonical model of the mechanism design literature resides. Moreover, and in contrast to our earlier results on truthful implementation (Bergemann and Morris (2003)) ex post implementation does not imply interim implementation even when we consider only common prior payoff type spaces. The analysis thus suggests that the ex post equilibrium notion may not capture robustness and concerns about detail free solutions as well for implementation as it does for truthful implementation problems.

The robustness results are all derived for general environment and exact implementation. It remains an open question whether more detailed relationships between these notions arise in specific environments such as single crossing or supermodular environments. Likewise it would be interesting to purse to the robustness analysis for virtual rather than exact implementation.

References

[1] Abreu, D. and H. Matsushima. 1992. "Virtual Implementation in Iterative Undominated Strategies: Complete Information." *Econometrica* 60: 993-1008.

[2] Battigalli, P. 1999. "Rationalizability in Incomplete Information Games."

[3] Battigalli, P. and M. Siniscalchi. 2003. "Rationalization and Incomplete Information", *Advances in Theoretical Economics* Vol. 3: No. 1, Article 3. `http://www.bepress.com/bejte/advances/vol3/iss1/art3`

[4] Bergemann, D. and S. Morris. 2001. "Robust Mechanism Design." early draft at `http://www.econ.yale.edu/sm326/rmd-nov2001.pdf`

[5] Bergemann, D. and S. Morris. 2003. "Robust Mechanism Design." Cowles Foundation Discussion Paper No. 1421. http://ssrn.com/abstract=412497.

[6] Bergemann, D. and S. Morris. 2004. "Notes on Complete Information Implementation with Rich Type Spaces."

[7] Bergemann, D. and J. Valimaki. 2002. "Information Acquisition and Mechanism Design." *Econometrica* 70: 1007-1033.

[8] Bernheim, D. 1984. "Rationalizable Strategic Behavior." *Econometrica* 52: 1007-1028.

[9] Borgers, T. 1995. "A Note on Implementation and Strong Dominance." *Social Choice, Welfare and Ethics*, W. Barnett, H. Moulin, M. Salles, Schofield, eds. Cambridge University Press.

[10] Brandenburger, A. and E. Dekel. 1987. "Rationalizability and Correlated Equilibria." *Econometrica* 55: 1391-1402.

[11] Brandenburger, A. and E. Dekel. 1993. "Hierarchies of Beliefs and Common Knowledge." *Journal of Economic Theory* 59: 189-198.

[12] Brandenburger, A. and A. Friedenberg. 2002. "Common Assumption of Rationality in Games."

[13] Chung, K.-S. and J. Ely. 2003. "Implementation with Near-Complete Information." *Econometrica* 71: 857-871.

[14] Cremer, J. and R. McLean. 1985. "Optimal Selling Strategies Under Uncertainty for a Discriminating Monopolist when Demands are Interdependent." *Econometrica* 53: 345-361.

[15] Cremer, J. and R. McLean. 1988. "Full Extraction of the Surplus in Bayesian and Dominant Strategy Auctions." *Econometrica* 56: 1247-1258.

[16] Dasgupta, P. and E. Maskin. 2000. "Efficient Auctions." *Quarterly Journal of Economics* 115: 341-388.

[17] Harsanyi, J. 1967/68. "Games with Incomplete Information Played by Bayesian agents." *Management Science* 14, 159-182, 320-334, 486-502.

[18] Heifetz, A. and Z. Neeman. 2003. "On the Generic Impossibility of Full Surplus Extraction in Mechanism Design"

[19] Heifetz, A. and D. Samet. 1888. "Topology-Free Typology of Beliefs." *Journal of Economic Theory* 82, 324-341.

[20] Holmstrom, B. and R. Myerson. 1983. "Efficient and Durable Decision Rules with Incomplete Information." *Econometrica* 51: 1799-1819.

[21] Jackson, M. 1991. "Bayesian Implementation." *Econometrica 59: 461-477.*

[22] Jackson, M. 1992. "Implementation in Undominated Strategies: A Look at Bounded Mechanisms." *Review of Economic Studies* 59, 757-775.

[23] Jehiel, P. and B. Moldovanu. 2001. "Efficient Design with Interdependent Valuations." *Econometrica* 65: 1237-1259.

[24] Kajii, A. and S. Morris. 1997. "The Robustness of Equilibria to Incomplete Information." *Econometrica* 65: 1283-1309.

[25] Kalai, E. 2002. "Large Robust Games." Northwestern University.

[26] Lipman, B. 1994. A Note on the Implications of Common Knowledge of Rationality." *Games and Economic Behavior* 6, 114-129.

[27] Maskin, E. 1999. "Nash Equilibrium and Welfare Optimality." *Review of Economic Studies 66: 23-38.*

[28] Maskin, E. and T. Sjostrom. 2001. "Implementation Theory." To appear in *Handbook of Social Choice and Welfare*, edited by K. Arrow, A. Sen and K. Suzumura.

[29] McLean, R. and A. Postlewaite. 2001. "Efficient Auction Mechanisms with Multidimensional Signals."

[30] Mertens, J.-F. and S. Zamir. 1985. "Formulation of Bayesian Analysis for Games of Incomplete Information." *International Journal of Game Theory* 14: 1-29.

[31] Mookerjee, D. and S. Reichelstein. 1989. "Implementation Via Augmented Revelation Mechanisms." *Review of Economic Studies* 57: 453-475.

[32] Morris, S. 2002. "Typical Types." Available at `http://www.econ.yale.edu/~sm326/typical.pdf`.

[33] Neeman, Z. 2001. "The Relevance of Private Information in Mechanism Design."

[34] Palfrey, T. and S. Srivastava. 1989. Implementation with Incomplete Information in Exchange Economies." *Econometrica* 57: 115-134.

[35] Pearce, D. 1984. "Rationalizable Strategic Behavior." *Econometrica* 52: 1007-1029.

[36] Perry, M. and P. Reny. 2002. "An Ex Post Efficient Auction." *Econometrica* 70: 1199-1212.

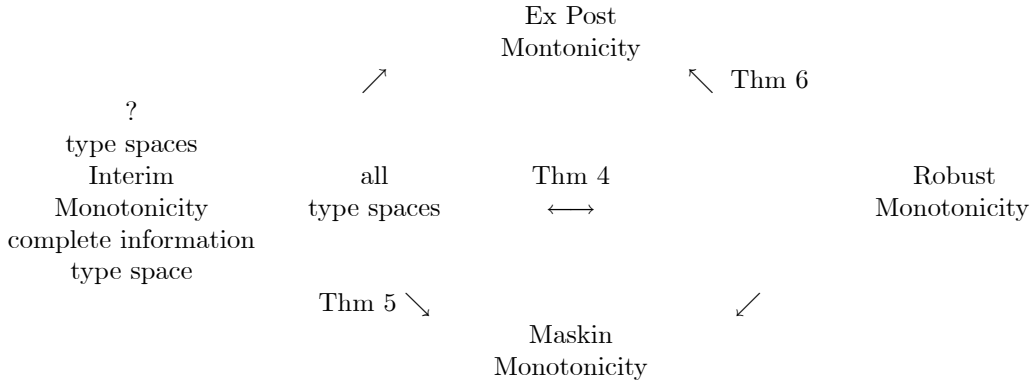
[37] Postlewaite, A. and D. Schmeidler. 1986. "Implementation in Differential Information Economies." *Journal of Economic Theory* 39: 14-33.

[38] Serrano, R. and R. Vohra. 2002. "A Characterization of Virtual Bayesian Implementation."

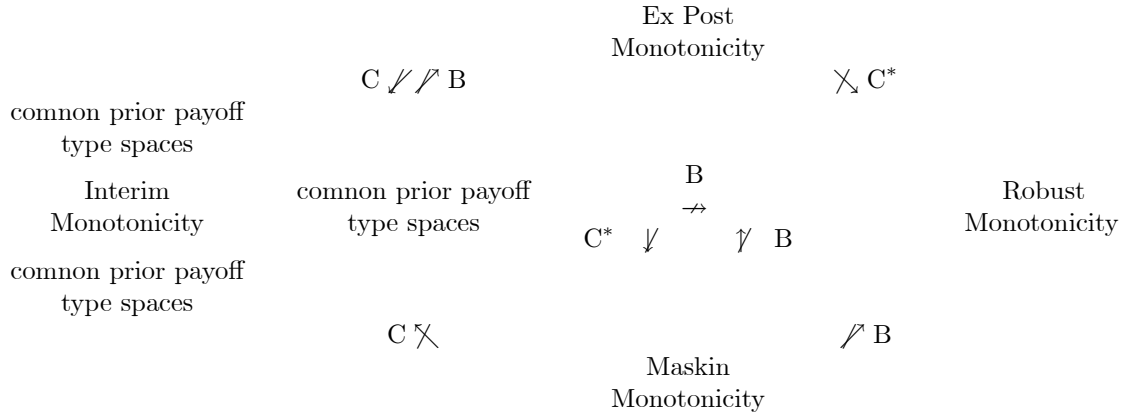
[39] Wilson, R. 1987. "Game-Theoretic Analyses of Trading Processes." In Advances in Economic Theory: Fifth World Congress, ed. Truman Bewley. Cambridge: Cambridge University Press chapter 2, pp. 33-70.

Overview over Results

1. Positive Results Regarding Monotonicity:



2. Negative Results Regarding Monotonicity:



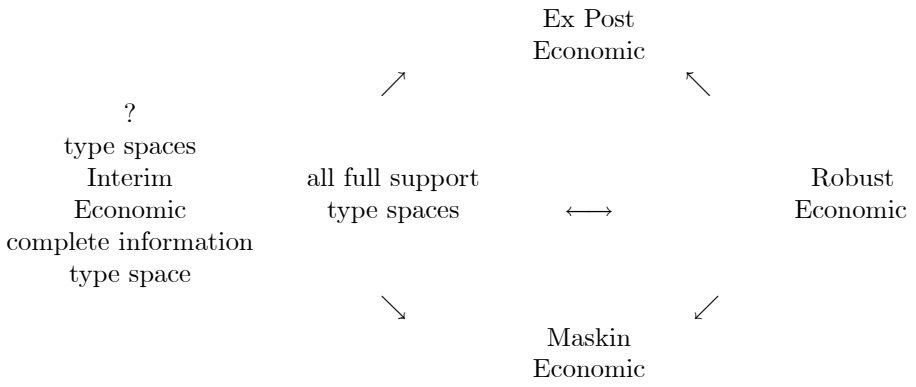
We refer to C* as the modification of Example C suggested in that very section.

3. Open and Possible Results Regarding Monotonicity.

- Does interim monotonicity on a subset of all type spaces imply ex post monotonicity. By the equivalence between robust and interim monotonicity on all type spaces, we know that interim monotonicity on all type spaces implies ex post monotonicity, but a stronger implication could be possible, but it would have to be weaker than common prior payoff type spaces.
- Ex Post Monotonicity should fail to imply Robust Monotonicity. The answer is yes, the only question is whether C already violates robust monotonicity or whether only the modification C^* violates robust monotonicity.
- Does interim monotonicity on all common prior payoff type spaces fail to imply maskin monotonicity

4. Open and Possible Results Regarding Economic Environment.

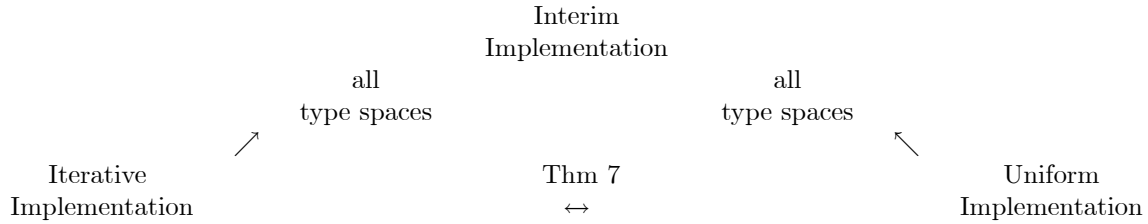
We could reasonably expect that the implications for economic environment are identical to the ones derived for monotonicity. We would then start to establish sufficient condition for implementation for a specific class of environments. Yet this has still to be established. Similarly, we might ask whether the results hold at least partially also for the more elaborate No Veto Monotonicity Hypothesis.



- the additional example shows that robust econonomic and robust monotonicity is only a sufficient condition for full support type spaces and thus we would hope that full suport for interim would be necessary and sufficient as well
- this raises the question whether there is reasonable strengthening of robust economic so that the equivalence holds for all type spaces, not only for full support type spaces. Do we think that the failure lies in robust economic or robust monotonicity.
- once we have sufficient conditions for robust economic, we can then think whether this is enough to bring close and how close to iterative implementation.
- to be integrated....note on robust and single mechanism...notes on compelete information....
- what can be said about the auction world of Maskin and Dasgupta, can they be implemented, to take on the question of Kfir Eliaz.

4. Implementation

We finally can relate iterative and uniform implementation on all type spaces. We may think of interim implementation on all type spaces as robust implementation.



It then remains an open question as to whether interim implementation on all type spaces in turn implies either iterative or uniform implementation.