Robust Implementation: The Role of Large Type Spaces*

Dirk Bergemann† Stephen Morris‡

First Version: March 2003
This Version: December 2003

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.

These results follow the implementation literature in focussing on pure strategy equilibria and allowing a finite mechanism to be chosen after the finite type space is chosen. We say that there is uniform implementation if there exists a finite mechanism that fully implements a social choice function for every finite type space that could be constructed for a fixed set of payoff types (under this concept there can be no gap between pure and mixed strategy implementation). We show that uniform implementation is equivalent to an ex post version of dominance solvability and show by example that uniform implementation may be impossible even though implementation is possible on every type space.

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 and Mike Riordan. We would like to thank seminar audiences at Columbia University, 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).

All the results reported thus far maintain two standard assumptions from the implementation literature. There is a restriction to pure strategies and the finite mechanism was chosen after the type space was fixed. We next examine the uniform implementation problem. Suppose that we fix the environment consisting of finite payoff types, outcomes, agent payoff functions and social choice function. We ask if there exists a finite mechanism such that every equilibrium of every possible

type space implies that the social choice function is realized. In other words, the finite mechanism is chosen after the finite type space is fixed. It does not matter for uniform implementability whether there is a pure strategy restriction. We show that uniform implementation 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. We show by example that this requirement is strictly stronger than robust monotonicity and thus uniform implementation is strictly stronger than interim implementation on every type space.

This last result about uniform implementation 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 reports our analysis of the ex post implementation problem. Section 5 reports on the connection between our ex post monotonicity and earlier notions of monotonicity. Section 6 introduces our notion of robust monotonicity and shows that it is equivalent to interim monotonicity on all type space. Section 7 reports results on uniform implementability. Section 8 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, $\Theta_1 = \left\{\theta_1^1, \theta_1^2\right\}$ and $\Theta_2 = \left\{\theta_2^1, \theta_2^2\right\}$. There are four possible social outcomes, $A = \{a, b, c, d\}$. The payoffs of the two agents are given by:

a	θ_2^1	θ_2^2
θ_1^1	$3,3$	$0,0$
θ_1^2	$0,0$	$1,1$

b	θ_2^1	θ_2^2
θ_1^1	$0,0$	$3,3$
θ_1^2	$1,1$	$0,0$

c	θ_2^1	θ_2^2
θ_1^1	$0,0$	$1,1$
θ_1^2	$3,3$	$0,0$

d	θ_2^1	θ_2^2
θ_1^1	$1,1$	$0,0$
θ_1^2	$0,0$	$3,3$

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

	θ_2^1	θ_2^2
θ_1^1	a	b
θ_1^2	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? First, observe that the social choice function is ex post incentive compatible. Thus if the planner simply invites the agents to announce their payoff types, they will have an incentive to tell the truth as long as they expect others to do so. Thus truth telling is an ex post equilibrium of the game where agents' types are just their payoff types. It is also an interim (Bayesian) equilibrium of the game played on a richer type space.

However, this game also has another ex post equilibrium where each type of each agent always misreports his type. This is also thus an interim equilibrium on any richer type space. However, it is easy to augment the simple mechanism to one where all (pure strategy) ex post equilibria yield desirable outcomes. 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_2^1	θ_2^2
(θ_1^1, truth)	a	b
(θ_1^2, truth)	c	d
(θ_1^1, lie)	b	a
(θ_1^2, lie)	d	c

(1)

What are the (pure strategy) ex post equilibria of this game? In any ex post equilibrium, type θ_2^1 of agent 2 must announce θ_2^1 or θ_2^2. If type θ_2^1 of agent 2 announces θ_2^1, then type θ_1^1 of agent 1 must announce (θ_1^1, truth) and type θ_1^2 of agent 1 must announce (θ_1^2, truth); so type θ_2^2 of agent 2 must announce θ_2^2. On the other hand, if type θ_2^1 of agent 2 announces θ_2^2, then type θ_1^1 of agent 1 must announce (θ_1^1, lie) and type θ_1^2 of agent 1 must announce (θ_1^2, lie); so type θ_2^2 of agent 2 must announce θ_2^1. Thus there are two possible ex post equilibria and both implement the social choice function.[1]

Thus for this example, we have shown the possibility of ex post implementation. Theorem 1 in Section 4 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.

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

	t_2^1	t_2^2	t_2^3	t_2^4	
t_1^1	$\frac{1}{8}(1-\varepsilon)$	$\frac{1}{8}(1-\varepsilon)$	$\frac{1}{8}\varepsilon$	$\frac{1}{8}\varepsilon$	θ_1^1
t_1^2	$\frac{1}{8}(1-\varepsilon)$	$\frac{1}{8}(1-\varepsilon)$	$\frac{1}{8}\varepsilon$	$\frac{1}{8}\varepsilon$	θ_1^2
t_1^3	$\frac{1}{8}\varepsilon$	$\frac{1}{8}\varepsilon$	$\frac{1}{8}(1-\varepsilon)$	$\frac{1}{8}(1-\varepsilon)$	θ_1^1
t_1^4	$\frac{1}{8}\varepsilon$	$\frac{1}{8}\varepsilon$	$\frac{1}{8}(1-\varepsilon)$	$\frac{1}{8}(1-\varepsilon)$	θ_1^2
	θ_2^1	θ_2^2	θ_2^1	θ_2^2	

[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."

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:

$$s_1(\cdot) = \begin{cases} (\theta_1^1, \text{truth}) & \text{if} \quad t_1^1 \\ (\theta_1^2, \text{truth}) & \text{if} \quad t_1^2 \\ (\theta_1^1, \text{lie}) & \text{if} \quad t_1^3 \\ (\theta_1^2, \text{lie}) & \text{if} \quad t_1^4 \end{cases}$$

and

$$s_2(\cdot) = \begin{cases} \theta_2^1 & \text{if} \quad t_2^1 \\ \theta_2^2 & \text{if} \quad t_2^2 \\ \theta_2^2 & \text{if} \quad t_2^3 \\ \theta_2^1 & \text{if} \quad t_2^4 \end{cases}$$

To see why this is an equilibrium, note that if $\varepsilon = 0$, then we have disjoint type spaces consisting of types $\left(t_1^1, t_1^2; t_2^1, t_2^2\right)$; and types $\left(t_1^3, t_1^4; t_2^3, t_2^4\right)$, respectively and the above type space reduces to:

	t_2^1	t_2^2	t_2^3	t_2^4	
t_1^1	$\frac{1}{8}$	$\frac{1}{8}$	0	0	θ_1^1
t_1^2	$\frac{1}{8}$	$\frac{1}{8}$	0	0	θ_1^2
t_1^3	0	0	$\frac{1}{8}$	$\frac{1}{8}$	θ_1^1
t_1^4	0	0	$\frac{1}{8}$	$\frac{1}{8}$	θ_1^2
	θ_2^1	θ_2^2	θ_2^1	θ_2^2	

In this new type space, the first disjoint type space $\left(t_1^1, t_1^2; t_2^1, t_2^2\right)$ play according to one ex post equilibrium of the augmented mechanism (1), whereas the second disjoint type space $\left(t_1^3, t_1^4; t_2^2, t_2^4\right)$ 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 may be easier to ex post implement than to interim implement. Later in the paper, we will give an example where it is possible to ex post implement on any type space but it is not possible to interim implement on some type space. We will also give an example where is possible to interim implement on any full support type space but not possible to ex post implement.[2]

It also turns out that in this example, there is no single finite mechanism that interim implements the social choice function on every type space. We return to this point in section 7.

3 The Implementation Problem

We consider 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 A. Each individual has utility function $u_i : A \times \Theta \to \mathbb{R}$. 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 $f : \Theta \to A$. If the true

[2] In this example, we could already sustain equilibria which do not implement the social choice function for some priors over the payoff types provided that we consider strategies which with partial lying (misreporting only for some, but not all types).

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, many of them with a richer set 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. 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:

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

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 finite set of messages available to agent i. Let $g(m)$ be the outcome if action profile m is chosen. Thus mechanisms that do not involve randomization contingent on the message profile. But randomization can be built into the outcome space A. Thus a mechanism is a collection

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

where $g : M \to A$. Note that finiteness is built into the definition of a mechanism.

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.[3]

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

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

[3] 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)).

As is standard in this literature, we restrict attention to pure strategy equilibria for most of the paper. The importance of this restriction is discussed in detail in section 7.

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

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

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.

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 Ex Post Implementation

We present necessary and sufficient conditions for a social choice function f to be ex-post equilibrium 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. We start with a brief review of the notion of 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 for all θ,θ' and

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

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 the no veto hypothesis holds, also sufficient.

For our purposes, it will be useful to state the above definition in its contrapositive form. In addition, it will be convenient to interpret θ' as the reported payoff profile of the agents when the true type profile is given by θ.

We call a non truth-telling strategy of the agents in the direct mechanism a *deception*. We recall that in the complete information environment, every agent i is informed about the entire vector of payoff types, θ. A successful deception by the agents relative to the principal (in the direct mechanism) therefore requires that they all report the same payoff profile. We thus focus on the case where all agents follows a common deception strategy:

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

The notion of a deception is meant to represent the possibility of multiple equilibria, in which agents do not necessarily report truthfully, but rather misreport systematically as represented by α.

Definition 6 (Maskin monotonicity)
Social choice function f is (Maskin) monotone, if either of the following equivalent conditions holds:

1. *If $f(\theta) \neq f(\theta')$, 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

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

2. *For all deceptions $\alpha : \Theta \to \Theta$, if $f \circ \alpha \neq f$, then there exists i, θ, and y such that*

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

while

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

The first version of the definition is simply the original definition in its contrapositive form. The equivalence between the first and the second version can be obtained by considering a pair θ and θ' with $f(\theta) \neq f(\theta')$ and define the deception α by setting $\theta' = \alpha(\theta)$.

The second version of the definition 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.

The ex post and interim monotonicity conditions to be presented shortly will all have these two components of "incentive compatible" "whistle-blower", but reflect the incomplete information environment and the different informational requirements of interim and ex post equilibrium.

4.2 Ex Post Monotonicity

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 Compatible)
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'.

However, there might exist multiple ex post equilibria in the direct mechanism. As before, we call a non truth-telling strategy in the direct mechanism a *deception*, with $\alpha = (\alpha_1, ..., \alpha_I)$, each $\alpha_i : \Theta_i \to \Theta_i$ and

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

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 $f\left(\alpha\left(\theta\right)\right)$ rather than $f\left(\theta\right)$. We write $f \circ \alpha\left(\theta\right) \equiv f\left(\alpha\left(\theta\right)\right)$.

Definition 8 (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 $y : \Theta \to A$ such that

$$u_i\left(y,\theta\right) > u_i\left(f\left(\alpha\left(\theta\right)\right),\theta\right), \tag{4}$$

and

$$u_i\left(f\left(\theta_i',\alpha_{-i}\left(\theta_{-i}\right)\right),\left(\theta_i',\alpha_{-i}\left(\theta_{-i}\right)\right)\right) \geq u_i\left(y,\left(\theta_i',\alpha_{-i}\left(\theta_{-i}\right)\right)\right),\ \forall \theta_i' \in \Theta_i. \tag{5}$$

The notions of Maskin and ex post monotonicity differ due to the informational assumptions inherent to each notion. In the complete information environment, every agent has complete information about the entire type profile. For this reason, the deception by every agent i is a mapping $\alpha_i : \Theta \to \Theta$. Moreover as it is a complete information environment, agent i cannot credibly issue a report distinct from the other agents in equilibrium. For this reason every agent has to report the same deception and it suffices to consider all common deceptions, dropping the index i completely. In the incomplete information environment, agent i has a private information about θ_i, his deception occurs only with respect to Θ_i and hence $\alpha_i : \Theta_i \to \Theta_i$.

The synchronicity in the complete information deception α and the asynchronicity in the incomplete information deception α_i affect the notions in two different ways. As the agents can synchronize their deception it is strictly harder to find a reward y with Maskin monotonicity than it is with ex post monotonicity. On the other hand, with complete information, there is no private information to agent i and it is strictly harder to satisfy the ex post incentive constraints. For this reason, neither one of the conditions implies the other, as will be demonstrated by Examples B and C.

For f to be ex-post implementable, it has to be that for every deception α, there exists an agent i and a type profile θ, such that at type profile θ agent i has a strict incentive to denounce the deception by choosing an alternative allocation.

However it has to be guaranteed that this particular agent i doesn't denounce the remaining agents when their announced type profile $\alpha_{-i}\left(\theta_{-i}\right)$ is in fact their true type profile. The later incentive constraint is represented by the second set of (weak) inequalities.

We next present an equivalent, but slightly more compact definition which we use in subsequent proofs. Let

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

The set $Y_i(\theta_{-i})$ comprises all allocations such that at every true profile $\left(\theta_i', \theta_{-i}\right)$, the social choice function $f\left(\theta_i', \theta_{-i}\right)$ weakly dominates every y in the set for every true type profile θ_i' of agent i. The following definition is equivalent to the above definition.

Definition 9 (Ex-post monotonicity)
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

$$u_i(y, \theta) > u_i(f(\alpha(\theta)), \theta). \tag{6}$$

In the second version of ex post monotonicity the set of incentive compatibility conditions is simply represented through the requirement that the allocation y is in the set $Y_i(\theta_{-i})$. We next demonstrate that 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,

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

for all $\theta \in \Theta$. Noting that $g\left(s_i\left(\theta_i'\right), s_{-i}(\theta_{-i})\right) = f\left(\theta_i', \theta_{-i}\right)$ 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

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

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

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

But since s is an equilibrium it follows that

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

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

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

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

and

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

The environment is said to be *non-economic* if it is not economic.

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)). The message space of each agent is given by:

$$M_i \equiv \Theta_i \times \{0,1\} \times A \times \{1,2,...,I\}$$

with typical element:

$$m_i \equiv (\theta_i, x_i, y_i, z_i)\,.$$

Thus a message profile is a vector

$$m \equiv (\theta_i, x_i, y_i, z_i)_{i=1}^{I}\,.$$

The mechanism is described by three rules.

1. If $x_i = 0$ for all i, then outcome $f(\theta)$ is chosen.
2. If $x_j = 1$ and $x_i = 0$ 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_i is chosen where the identity of agent i is determined by the following modulo construction:

$$i(z) = \left(\sum_{j=1}^{I} z_j\right)_{\mathrm{mod}\, I}\,.$$

More formally, we can describe the outcome function $g(\cdot)$:

$$g(\theta,x,y,z) = \begin{cases} f(\theta), & \text{if} \quad x_i = 0 \text{ for all } i \\ y_j, & \text{if} \quad x_j = 1,\ x_i = 0 \text{ for all } i \neq j \text{ and } y_j \in Y_j^*(\theta_{-j}) \\ f(\theta), & \text{if} \quad x_j = 1,\ x_i = 0 \text{ for all } i \neq j \text{ and } y_j \notin Y_j^*(\theta_{-j}) \\ y_{i(z)}, & \text{if} \quad \#\{i : x_i = 1\} \geq 2 \end{cases}\,. \tag{7}$$

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

$$s_i(\theta) = \left(s_i^1(\theta), s_i^2(\theta), s_i^3(\theta), s_i^4(\theta)\right) \in \Theta_i \times \{0,1\} \times A \times \{1,2,...,I\}\,;$$

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

For comparison, in the mechanism suggested by Maskin (1999) for complete information implementation, the message space of each agent is given by:

$$M_i = \mathcal{R} \times A \times \mathcal{N},$$

where $\mathcal{R}$ is the set of preference profiles and $\mathcal{N}$ the set of natural numbers. In the complete information environment, each agent knows the entire preference profile and therefore each agent can indicate by his suggestion of an allocation $a \in A$ whether the agents are reporting truthfully and choose $a = f(\mathcal{R})$ or whether the agents deceive and then report some $a \notin f(\mathcal{R})$. For this reason, the additional binary element $\{0,1\}$ in the message of each agent is not necessary in the complete information environment. In the mechanism offered by Maskin (1999), agents play an integer game, which we replace by a finite modulo game.

Theorem 2 (Economic Environment)
If $I \geq 3$ and f satisfies ex post incentive compatibility and ex post monotonicity and the economic condition (at all θ), 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:

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

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

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

if either $x_i = 0$ or $x_i = 1$ 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 $x_i = 1$ 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 $j \neq i$ who thought his opponents were types θ_{-j} could send any message of the form

$$\left(\cdot, 0, y_j, \left(j + \sum_{k \neq j} (I - z_k)\right)_{\bmod I}\right)$$

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 (EPM), 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 $m_i = (\cdot, 1, y, \cdot)$, while other agents send their equilibrium messages, then from the definition (7) 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 binary reports $x_i \in \{0, 1\}$ all have to say $x_i = 0$, or else any agent j could profitably change his binary report x_i and choose a modulo number z_i to 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 $b \in A$ 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, $y \in A$, then there does not exist $i \neq j$ and $a \in A$ such that $u_i(a,\theta) > u_i(b,\theta)$. Thus $u_i(b,\theta) \geq u_i(a,\theta)$ for all $a \in A$ 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 11 (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 $a \in A$ and all $i \neq j$ implies that $f(\theta) = b$.

Definition 12 (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

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

In the special case where $\Phi_i = \Theta_i$ for all i, the EMNV reduces to the ex post monotonicity condition. In the special case where $\Phi_i = \varnothing$ for all i, the EMNV requires *if* the environment is non-economic everywhere, then the no veto condition must hold everywhere. In the special case where α is the truth-telling deception and, for some i, $\Phi_i = \Theta_i \backslash \{\theta_i^*\}$ and $\Phi_j = \Theta_j$ for all $j \neq i$, then EMNV requires that if the environment is non-economic whenever $\theta_i = \theta_i^*$, then the environment satisfies no veto power whenever $\theta_i = \theta_i^*$. If ex post monotonicity holds and no veto power holds at every type profile, then EMNV is satisfied. Finally, observe that 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.

Claim 1. In any ex post equilibrium, the environment is non-economic for all $\theta \notin \Phi$. Let

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

First, observe that for each $\theta \notin \Phi$, there exists i such that $s_i^2(\theta_i) = 1$. 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

$$\left(\cdot, 0, y_j, \left(j + \sum_{k \neq j} z_k\right)_{\bmod I}\right)$$

and obtain utility $u_j(y_j,\theta)$. Thus we must have $u_j(g(m(\theta)),\theta) \geq u_j(a,\theta)$ for all a and all $j \neq i$. 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 $y \in Y_i\left(s^1_{-i}(\theta_{-i})\right)$. Suppose that $y \in Y_i\left(s^1_{-i}(\theta_{-i})\right)$ and that type θ_i of agent i believes that his opponents are of type θ_{-i} and sends message $m_i = (\cdot, 1, y, \cdot)$, 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. Note that Claim 1 establishes that the environment is non-economic at all $\theta \in \Phi$. Suppose $g(s(\theta)) \neq f(\theta)$ for some $\theta \in \Phi$. Since 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. Suppose $g(s(\theta)) \neq f(\theta)$ for some $\theta \notin \Phi$. 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.

A brief comparison between the proof strategy for the economic environment and the general monotonicity no veto condition elucidates the role of the no veto condition. In the economic environment, we proceed directly to eliminate message profiles $x \neq 0$ by using the hypothesis of an economic environment. With the general no veto monotonicity condition, we first split the type space θ into two complimentary subsets, Φ and $\Theta - \Phi$. On the subset Φ, the message profiles satisfy $x = 0$ and it is on the restricted domain Φ that we will eventually apply the monotonicity condition. On the remaining set $\Theta - \Phi$, we do not attempt to eliminate the message profile $x \neq 0$ as part of possible equilibrium strategy, but rather argue that if these message profiles are to be part of an equilibrium, then the implied social choice function $g(s)$ has to satisfy the no veto condition. The proof then uses the no veto condition to show that without the no veto condition, m, could not be part of an equilibrium. More precisely, if the no-veto condition fails at θ, then it follows that m at θ has to satisfy $\#\{x_i = 1\} > 1$. But if the no veto condition fails and $\#\{x_i = 1\} > 1$, then there are is at least one agent which would prefer z_j to $g(s)$ which he could force to be realize as $\#\{x_i = 1\} > 1$, and this contradicts the fact that s is an ex post equilibrium.

In either case, the condition of no veto or of an economic environment allows us to address type profiles θ with associated messages $\#\{x_i = 1\} > 1$. The hypothesis of ex post monotonicity by itself is sufficient to address all message profiles except $\{x_i = 1\} = 1$ yet $y_i \notin Y_i(\theta_{-i})$.The proofs show that the economic environment does not allow equilibrium messages $\#\{x_i = 1\} > 1$ whereas the no veto hypothesis forces the ex post monotonicity to operate on a smaller set of type profiles.

5 Comparing Monotonicity Conditions

In the previous section, we introduced the notions of Maskin and ex post monotonicity. We then identified ex post monotonicity as a necessary and (almost) sufficient condition for ex post implementation in incomplete information environments. In this section we begin to relate ex post and interim monotonicity conditions. In Subsection 5.1 we briefly state the interim monotonicity condition developed in the literature on Bayesian implementation. We then show that if the social choice function f satisfies interim monotonicity on all common prior type spaces, then it guarantees that f

satisfies ex post and Maskin monotonicity. We then illustrate the limitations of this result by means of two examples. In Subsection 5.2, we present Example B, in which interim monotonicity is satisfied for all common prior payoff type spaces, but ex post monotonicity fails. This is meant to illustrate the role (of the size) of the type space for the relation between ex post and interim monotonicity. In Subsection 5.3, Example C then shows that the converse of the result fails to hold. Namely, in the example ex post and Maskin monotonicity are satisfied yet interim monotonicity fails for a uniform prior over the payoff type space.

5.1 Interim Implementation

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

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

Definition 13 (Interim Monotonicity)
Social choice function f satisfies interim monotonicity on type space $\mathcal{T}$ if, for every deception α and $f \neq f \circ \alpha$, there exists i, t_i and $y : T \to A$ such that

$$\sum_{t_{-i} \in T_{-i}} u_i \left(y\left(\alpha\left(t\right)\right), \widehat{\theta}\left(t\right) \right) \widehat{\pi}_i\left(t_i\right)\left[t_{-i}\right] > \sum_{t_{-i} \in T_{-i}} u_i \left(f\left(\widehat{\theta}\left(\alpha\left(t\right)\right)\right), \widehat{\theta}\left(t\right) \right) \widehat{\pi}_i\left(t_i\right)\left[t_{-i}\right], \tag{8}$$

and

$$\begin{aligned} & \sum_{t_{-i} \in T_{-i}} u_i \left(f\left(\widehat{\theta}\left(t_i', t_{-i}\right)\right), \widehat{\theta}\left(t_i', t_{-i}\right) \right) \widehat{\pi}_i\left(t_i'\right)\left[t_{-i}\right] \\ \geq & \sum_{t_{-i} \in T_{-i}} u_i \left(y\left(\alpha_i\left(t_i\right), t_{-i}\right), \widehat{\theta}\left(t_i', t_{-i}\right) \right) \widehat{\pi}_i\left(t_i'\right)\left[t_{-i}\right], \ \forall t_i'. \end{aligned} \tag{9}$$

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.

Theorem 4
If f satisfies interim monotonicity on all common prior type spaces then

1. *it satisfies Maskin monotonicity;*
2. *it satisfies ex post monotonicity.*

Proof. (1.) 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 $f\left(\widehat{\alpha}\left(\theta\right)\right) \neq f\left(\theta\right)$, and all h such that

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

we have

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

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

$$\sum_{t_{-i} \in T_{-i}} u_i\left(h\left(\alpha\left(t\right)\right), t\right) p_i\left(t_{-i} \left| t_i \right.\right) > \sum_{t_{-i} \in T_{-i}} u_i\left(f\left(\alpha\left(t\right)\right), t\right) p_i\left(t_{-i} \left| t_i \right.\right) \tag{10}$$

while

$$\sum_{t_{-i}\in T_{-i}} u_i\left(f\left(t_i',t_{-i}\right),\left(t_i',t_{-i}\right)\right)p_i\left(t_{-i}\left|t_i'\right.\right) \geq \sum_{t_{-i}\in T_{-i}} u_i\left(h\left(\alpha_i\left(t_i\right),t_{-i}\right),t\right)p_i\left(t_{-i}\left|t_i'\right.\right),\ \forall t_i'\neq t_i. \quad (11)$$

With the complete information type space and the symmetric deception strategy, the inequalities (10) and (11) reduce to

$$u_i\left(h\left(\widehat{\alpha}\left(\theta\right)\right),\theta\right) > u_i\left(f\left(\widehat{\alpha}\left(\theta\right)\right),\theta\right) \quad (12)$$

and

$$u_i\left(f\left(\theta'\right),\theta'\right) \geq u_i\left(h\left(\widehat{\alpha}\left(\theta\right),\theta',...,\theta'\right),\theta'\right),\ \forall\theta'\neq\theta, \quad (13)$$

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.

(2.) We proceed by contrapositive. Suppose that f is not ex post monotone. Then there exists α such that for all i,θ and y if

$$u_i\left(y,\theta\right) > u_i\left(f\left(\alpha\left(\theta\right)\right),\theta\right) \quad (14)$$

then there exists θ_i' such that

$$u_i\left(f\left(\theta_i',\alpha_{-i}\left(\theta_{-i}\right)\right),\left(\theta_i',\alpha_{-i}\left(\theta_{-i}\right)\right)\right) < u_i\left(y,\left(\theta_i',\alpha_{-i}\left(\theta_{-i}\right)\right)\right). \quad (15)$$

We consider the following full support common prior type space: $T_i=T_i^1\cup T_i^2$ with

$$\pi\left(t\right)=\left\{\begin{array}{lll}\gamma, & \text{if} & t\in T^1\\ \varepsilon, & \text{if} & t\notin T^1,\ \exists i,\ t_{-i}\in T_{-i}^1\\ 0, & \text{if} & \text{otherwise}\end{array}\right| \quad (16)$$

with $\gamma\gg\varepsilon$. The first subset of the type space satisfies $T_i^1=\Theta_i$ with the following uniform belief property over the payoff types:

$$\widehat{\psi}_i\left(t_i\right)\left[\theta_{-i}\right]=\frac{1}{\#\Theta_{-i}}$$

and a bijection $\widehat{\theta}_i:T_i^1\rightarrow\Theta_i$. The second subset of the type space satisfies $T_i^2=\Theta$ with the belief property:

$$\widehat{\pi}_i\left(t_i\right)\left[t_{-i}\right]=\left\{\begin{array}{lll}1, & \text{if} & t_{-i}=\theta_{-i}\\ 0, & \text{if} & t_{-i}\neq\theta_{-i}\end{array}\right|$$

For this type space we then consider a deception $\beta_i:T_i\rightarrow T_i$ which replicates the deception α on which f failed to display ex post monotonicity

$$\forall t_i\in T_i^1:\qquad \beta_i\left(t_i\right)=\beta_i\left(\theta_i\right)=\alpha_i\left(\theta_i\right)$$

and

$$\forall t_i\in T_i^2:\qquad \beta_i\left(t_i\right)=\beta_i\left(\widehat{\theta}_i\left(t_i\right)\right)=\alpha_i\left(\theta_i\right).$$

The deception is a pooling deception in so far that all types with the same payoff type, irrespective of their belief type, choose the same deception. For this deception, we can then represent the interim monotonicity condition as follows. We start with the reward inequality. For all $t_i\in T_i^1$:

$$\sum_{\theta_{-i}\in\Theta_{-i}} u_i\left(y\left(\alpha\left(\theta\right)\right),\theta\right) > \sum_{\theta_{-i}\in\Theta_{-i}} u_i\left(f\left(\alpha\left(\theta\right)\right),\theta\right) \quad (17)$$

and for all $t_i \in T_i^2$:

$$u_i\left(y\left(\alpha\left(\theta\right)\right),\theta\right) > u_i\left(f\left(\alpha\left(\theta\right)\right),\theta\right) \tag{18}$$

The incentive inequalities are for $t_i \in T_i^1$

$$\sum_{\theta_{-i}\in\Theta_{-i}} u_i\left(f\left(\theta\right),\theta\right) \geq \sum_{\theta_{-i}\in\Theta_{-i}} u_i\left(y\left(\alpha_i\left(\theta_i\right),\theta_{-i}\right),\theta\right) \tag{19}$$

and for $t_i \in T_i^2$:

$$u_i\left(f\left(\left(\theta_i',\alpha_{-i}\left(\theta_{-i}\right)\right)\right),\left(\theta_i',\alpha_{-i}\left(\theta_{-i}\right)\right)\right) \geq u_i\left(y\left(\alpha\left(\theta\right)\right),\left(\theta_i',\alpha_{-i}\left(\theta_{-i}\right)\right)\right). \tag{20}$$

It is now immediate to verify that if ex post monotonicity is violated at α, interim monotonicity will be as well. Observe first that clearly, the reward cannot be offered for $t_i \in T_i^2$ without violating the corresponding incentive compatibility conditions (20). Suppose then that we seek to satisfy the reward inequality for some $t_i \in T_i^1$, then it follows that there must exist some $y\left(\alpha\left(\theta\right)\right)$ and θ where reward is provided, but by the hypothesis of failure of ex post monotonicity, we can then find a violation of (20). ■

While Maskin monotonicity is implied by interim monotonicity on common 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.

5.2 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. 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 payoff matrices below represent the payoffs of the agents in each true state $\{\theta\}$ when the allocation changes due to changes in its first, second or third entry. It is convenient to vary the allocation and holding the true state constant as this represents the ex post strategic opportunities for every agent given the true state of the world.

$\{000\}$		
	0	1
0	$1,1,1$	$1+\varepsilon,0,1+\delta$
1	$0,1+\delta,1+\varepsilon$	$0,0,0$

$\{000\}$		
	1	0	1
0	$1+\delta,1+\varepsilon,0$	$0,0,0$
1	$0,0,0$	$1,1,1$

The payoffs in the remaining states are simple permutations of the payoffs in state $\theta = 000$. The payoff state space of every agent is binary. Every permutation $\sigma_i : \Theta_i \to \Theta_i$ can therefore simply be thought of as an instruction to either keep the state or change the state of agent i. A permutation profile is $\sigma = (\sigma_1, \sigma_2, \sigma_3)$. The payoffs of the agents have the following symmetry property: for all i and all σ:

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

We list the payoffs in the remaining states for completeness:

$\{001\}$		
	0	1
0	$1+\delta,1+\varepsilon,0$	$0,0,0$
1	$0,0,0$	$1,1,1$

$\{001\}$		
	1	0	1
0	$1,1,1$	$1+\varepsilon,0,1+\delta$
1	$0,1+\delta,1+\varepsilon$	$0,0,0$

{010}			
	0	0	1
0	$1+\varepsilon, 0, 1+\delta$	$1, 1, 1$	
1	$0, 0, 0$	$0, 1+\delta, 1+\varepsilon$	

{010}			
	1	0	1
0	$0, 0, 0$	$1+\delta, 1+\varepsilon, 0$	
1	$1, 1, 1$	$0, 0, 0$	

{011}			
	0	0	1
0	$0, 0, 0$	$1+\delta, 1+\varepsilon, 0$	
1	$1, 1, 1$	$0, 0, 0$	

{011}			
	1	0	1
0	$1+\varepsilon, 0, 1+\delta$	$1, 1, 1$	
1	$0, 0, 0$	$0, 1+\delta, 1+\varepsilon$	

{100}			
	0	0	1
0	$0, 1+\delta, 1+\varepsilon$	$0, 0, 0$	
1	$1, 1, 1$	$1+\varepsilon, 0, 1+\delta$	

{100}			
	1	0	1
0	$0, 0, 0$	$1, 1, 1$	
1	$1+\delta, 1+\varepsilon, 0$	$0, 0, 0$	

{101}			
	0	0	1
0	$0, 0, 0$	$1, 1, 1$	
1	$1+\delta, 1+\varepsilon, 0$	$0, 0, 0$	

{101}			
	1	0	1
0	$0, 1+\delta, 1+\varepsilon$	$0, 0, 0$	
1	$1, 1, 1$	$1+\varepsilon, 0, 1+\delta$	

{110}			
	0	0	1
0	$0, 0, 0$	$0, 1+\delta, 1+\varepsilon$	
1	$1+\varepsilon, 0, 1+\delta$	$1, 1, 1$	

{110}			
	1	0	1
0	$1, 1, 1$	$0, 0, 0$	
1	$0, 0, 0$	$1+\delta, 1+\varepsilon, 0$	

{111}			
	0	0	1
0	$1, 1, 1$	$0, 0, 0$	
1	$0, 0, 0$	$1+\delta, 1+\varepsilon, 0$	

{111}			
	1	0	1
0	$0, 0, 0$	$0, 1+\delta, 1+\varepsilon$	
1	$1+\varepsilon, 0, 1+\delta$	$1, 1, 1$	

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 ex post 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.

5.2.1 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 $\theta = 000$. The complete deception leads to the allocation $\alpha(000) = 111$. The only allocations which would improve the utility of agent 1 are $a \in \{010, 001\}$ and we shall argue next that neither of these allocations satisfies the incentive compatibility conditions of the monotonicity condition. Consider first the reward $y = 010$, for which ex post incentive compatibility would have to satisfy:

$$1 = u_1(111, 111) \geq u_1(010, 111) = 0$$

as well as

$$1 = u_1(011, 011) \geq u_1(010, 011) = 1 + \delta,$$

but obviously the second inequality is violated. Similarly, we observe that for the reward $y = 001$, the ex post incentive compatibility conditions are:

$$1 = u_1(111, 111) \geq u_1(001, 111) = 0$$

as well as

$$1 = u_1(011, 011) \geq u_1(001, 011) = 1 + \varepsilon,$$

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.

5.2.2 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 $y \in \{010, 001\}$ occurred at $\theta = 011$, but not at the deception $\alpha(000) = 111$ 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.

5.2.3 Interim Monotonicity Holds for all Payoff Common Priors

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 $p(00\,|0) > 0$ and $p(11\,|0) \leq \frac{1}{1+\varepsilon}$. We offer the following contingent reward to agent 1:

$$y = \begin{cases} 001 & \text{if} \quad \theta = 111 \\ f & \text{if} \quad \theta \neq 111 \end{cases} \tag{21}$$

The reward condition at $\theta_1 = 0$ then reduces to, after eliminating terms on both sides of the inequality by using (21):

$$u_1(001, 000)\, p(00\,|0) > u_1(111, 000)\, p(00\,|0) \tag{22}$$

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

$$1 \geq (1 + \varepsilon) \cdot p(11\,|0) \tag{23}$$

and for $\theta_1 = 1$:

$$1 \geq 1 - p(11\,|1) \tag{24}$$

We observe that (22) is satisfied by hypothesis of $p(00\,|0) > 0$, inequality (23) by hypothesis of $p(11\,|0) \leq \frac{1}{1+\varepsilon}$ and inequality (24) is always satisfied.

For the instance of $p(00\,|0) > 0$ but $p(11\,|0) > \frac{1}{1+\varepsilon}$, we can offer a modified reward rule:

$$y = \begin{cases} 010 & \text{if} \quad \theta = 100 \\ f & \text{if} \quad \theta \neq 100 \end{cases} \tag{25}$$

which differs from the reward rule (21) only by the type profile at which it offers a reward. With this modified rule can then write the reward condition as:

$$u_1(010, 011)\, p(11\,|0) > u_1(100, 011)\, p(11\,|0) \tag{26}$$

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

$$1 \geq (1+\varepsilon) \cdot p\,(00\,|0\,) \tag{27}$$

and for $\theta_1 = 1$:

$$1 \geq 1 - p\,(00\,|0\,) \tag{28}$$

By the hypothesis of $p\,(11\,|0\,) > \frac{1}{1+\varepsilon}$, it follows that (26) and (27) holds, and (28) 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.

5.3 Example C

There are three agents, $i = 1, 2, 3$ and each agent has a binary payoff type space $\Theta_i = \left\{\theta_i^1, \theta_i^2\right\}$. 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_3^1	θ_2^1	θ_2^2		θ_3^2	θ_2^1	θ_2^2
θ_1^1	a	b		θ_1^1	b	c
θ_1^2	b	c		θ_1^2	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 $\{a, b, c, d\}$

	θ_3^1	θ_2^1	θ_2^2	θ_3^2	θ_2^1	θ_2^2
a :	θ_1^1	1	0	θ_1^1	0	0
	θ_1^2	0	0	θ_1^2	0	0
	θ_3^1	θ_2^1	θ_2^2	θ_3^2	θ_2^1	θ_2^2
b :	θ_1^1	-1	ε	θ_1^1	ε	-1
	θ_1^2	ε	-1	θ_1^2	-1	-1
	θ_3^1	θ_2^1	θ_2^2	θ_3^2	θ_2^1	θ_2^2
c :	θ_1^1	-1	-1	θ_1^1	-1	ε
	θ_1^2	-1	ε	θ_1^2	ε	-1
	θ_3^1	θ_2^1	θ_2^2	θ_3^2	θ_2^1	θ_2^2
d :	θ_1^1	-1	-1	θ_1^1	-1	-1
	θ_1^2	-1	-1	θ_1^2	-1	ε

The allocation a is efficient if all agents are of type θ_i^1 and d is efficient if all agents are of type θ_i^2. In the remaining case the allocation b is efficient if a majority of agents is of type θ_i^1 and the allocation c is efficient if a majority of agents is of type θ_i^2. 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 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.

5.3.1 Ex Post Monotonicity

The social choice function is efficient 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.

5.3.2 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 $\theta = \theta_1^2\theta_2^2\theta_3^2$ and increase it from 0 to $u_i\left(a, \theta_1^2\theta_2^2\theta_3^2\right) = 2\varepsilon$, then f no longer satisfies Maskin monotonicity for the deception $\alpha\left(\theta_1^2\theta_2^2\theta_3^2\right) = \theta_1^1\theta_2^1\theta_3^1$ 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 $\alpha_i\left(\theta_i^2\right) = \theta_i^1$ for all i leads has type profiles, say $\theta_1^1\theta_2^2\theta_3^2$, 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 environment, yet display or not display Maskin monotonicity.

5.3.3 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^1:

$$\alpha_i(\cdot) = \theta_i^1, \quad \forall i, \forall \theta_i,$$

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^1, 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^2, 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.

6 Robust Monotonicity

The results presented in the previous section show that interim monotonicity on all common prior type spaces imply ex post and Maskin monotonicity. Yet, Example C showed that ex post monotonicity does not even imply interim monotonicity on all common prior payoff type spaces. We therefore propose a stronger and novel monotonicity notion, to be called robust monotonicity, which is necessary and sufficient for interim monotonicity on all type spaces. If the designer does not know the true type space, i.e. the agent's beliefs and higher order beliefs about other agents' types, then he might want to find a mechanism that works for every type space. We show that robust monotonicity achieves this objective and that it is strictly stronger than both Maskin and ex post monotonicity.

6.1 Definition

The new requirement precisely addresses the gap that appeared between ex post and interim monotonicity in Example C. The failure of interim monotonicity in Example C came from a pooling deception in which different payoff types generated the same deceptive report and the designer failed to find a reward which would work across the different true type profiles, even though for every particular profile, such a reward existed by ex post monotonicity. The strengthening of the reward inequality then asks that we can find a reward in response to a report θ'_{-i} which could have been generated by an arbitrary distribution over true types θ_{-i} rather than a fixed given type profile θ_{-i}.

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 $\theta_i\in\beta_i\left(\theta_i\right)$. The interpretation is that $\beta_i\left(\theta_i\right)$ is the collection of correct or incorrect reports that payoff type θ_i might send. A deception is *acceptable* if $\theta'\in\boldsymbol{\beta}\left(\theta\right)\Rightarrow f\left(\theta'\right)=f\left(\theta\right)$. A deception is *unacceptable* if it is not acceptable.

Definition 14 (Robust Monotonicity)
Social choice function f satisfies robust monotonicity if one of the following equivalent conditions holds:

1. *For every unacceptable deception β, 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, where*

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

there exists y such that

$$\sum_{\theta_{-i}\in\Theta_{-i}} \psi_i(\theta_{-i})\, u_i(y,(\theta_i,\theta_{-i})) > \sum_{\theta_{-i}\in\Theta_{-i}} \psi_i(\theta_{-i})\, u_i\left(f\left(\theta'_i,\theta'_{-i}\right),(\theta_i,\theta_{-i})\right) \tag{29}$$

and

$$u_i\left(f\left(\widetilde{\theta}_i,\theta'_{-i}\right),\left(\widetilde{\theta}_i,\theta'_{-i}\right)\right) \geq u_i\left(y,\left(\widetilde{\theta}_i,\theta'_{-i}\right)\right), \tag{30}$$

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

2. *Fix a deception $\boldsymbol{\beta}$. If for all i, θ_i,$\theta'_i \in \beta_i(\theta_i)$, there exist θ'_{-i} and ψ_i with*

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

such that

$$u_i\left(f\left(\widetilde{\theta}_i,\theta'_{-i}\right),\left(\widetilde{\theta}_i,\theta'_{-i}\right)\right) \geq u_i\left(y,\left(\widetilde{\theta}_i,\theta'_{-i}\right)\right), \text{ for all } \widetilde{\theta}_i, \tag{31}$$

implies

$$\sum_{\theta_{-i}\in\Theta_{-i}} \psi_i(\theta_{-i})\, u_i\left(f\left(\theta'\right),(\theta_i,\theta_{-i})\right) \geq \sum_{\theta_{-i}\in\Theta_{-i}} \psi_i(\theta_{-i})\, u_i(y,(\theta_i,\theta_{-i}))\,; \tag{32}$$

then $\boldsymbol{\beta}$ is acceptable.

The second version of the definition is simply the contrapositive statement of the first version. We use the first version to show that robust monotonicity implies interim monotonicity, and the second version to show that interim monotonicity on all type spaces implies robust monotonicity.

The notion of robust monotonicity shares many features with the ex post monotonicity condition. We shall use the first version of Definition 14 for comparison here. As the notion of 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 shares the ex post incentive compatibility conditions (30) with ex post monotonicity. The reward inequality (29) on the other hand is a stronger version of the ex post notion. We first observe that the reward inequality takes expectation over payoff types θ_{-i} which could lead to a deception profile θ'_{-i} for a given deception β, which is incorporated in the restriction that ψ_i satisfies:

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

The second observation is that we evaluate the social choice function f and the reward function y only at the reported payoff type profile θ'_{-i}. This leads to apparent discrepancy with respect to the interim notion, as the interim notion takes expectations over all possible deceptions generated by the entire set of payoff type profiles. However, we can easily bring the reward inequality of the robust notion closer to the interim notion by defining y to be equal to the social choice function for all reported profiles different than θ'_{-i}.

Finally we would like to emphasize that the allocation y is allowed to depend on the misreport θ'_{-i} and the distribution ψ_i.

6.2 Equivalence

Next we establish the equivalence between robust monotonicity and interim monotonicity on all type spaces.

Theorem 5
If f satisfies interim monotonicity on every type space, then f satisfies robust monotonicity.

Proof. Fix a deception β. Suppose that the premise of the definition of robust monotonicity (in its second version) holds. 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 ψ_i over true payoff profiles θ_{-i} which can generate $\zeta_i\left(\theta_i, \theta_i'\right)$ under β_{-i}:

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

such that

$$u_i\left(f\left(\widetilde{\theta}_i, \zeta_i\left(\theta_i, \theta_i'\right)\right), \left(\widetilde{\theta}_i, \zeta_i\left(\theta_i, \theta_i'\right)\right)\right) \geq u_i\left(y, \left(\widetilde{\theta}_i, \zeta_i\left(\theta_i, \theta_i'\right)\right)\right), \tag{33}$$

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

$$\begin{aligned} & \sum_{\left\{\theta_{-i} \in \Theta_{-i} : \zeta_i\left(\theta_i, \theta_i'\right) \in \beta_{-i}\left(\theta_{-i}\right)\right\}} \psi_i\left(\theta_{-i}|\theta_i, \theta_i'\right) u_i\left(f\left(\theta_i', \zeta_i\left(\theta_i, \theta_i'\right)\right), \left(\theta_i, \theta_{-i}\right)\right) \qquad (34) \\ \geq & \sum_{\left\{\theta_{-i} \in \Theta_{-i} : \zeta_i\left(\theta_i, \theta_i'\right) \in \beta_{-i}\left(\theta_{-i}\right)\right\}} \psi_i\left(\theta_{-i}|\theta_i, \theta_i'\right) u_i\left(y, \left(\theta_i, \theta_{-i}\right)\right). \end{aligned}$$

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

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

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

$$\widehat{\theta}_i\left(t_i\right) = \theta_i, \tag{35}$$

and

$$\widehat{\pi}_i\left(t_i\right)\left[t_{-i}\right] = \begin{cases} 1, \text{ if } t_j = \left(\theta_j, \theta_j\right) \text{ for each } j \neq i \\ 0, \text{ otherwise.} \end{cases} \tag{36}$$

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 13 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 (35) and (36):

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

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 $y : T \to A$ such that (8) and (9) hold. Letting $\widehat{t}_i = \left(\theta_i, \theta_i'\right)$, (8) becomes:

$$\begin{aligned} &\sum_{\left\{\theta_{-i}\in\Theta_{-i}:\zeta_i\left(\theta_i,\theta_i'\right)\in\beta_{-i}(\theta_{-i})\right\}} \psi_i\left(\theta_{-i}|\theta_i,\theta_i'\right) u_i\left(y\left(\left(\theta_i',\theta_i'\right),\left(\zeta_{ij}\left(\theta_i,\theta_i'\right),\zeta_{ij}\left(\theta_i,\theta_i'\right)\right)_{j\neq i}\right),\left(\theta_i,\theta_{-i}\right)\right) \\ > \\ &\sum_{\left\{\theta_{-i}\in\Theta_{-i}:\zeta_i\left(\theta_i,\theta_i'\right)\in\beta_{-i}(\theta_{-i})\right\}} \psi_i\left(\theta_{-i}|\theta_i,\theta_i'\right) u_i\left(f\left(\theta_i',\zeta_i\left(\theta_i,\theta_i'\right)\right),\left(\theta_i,\theta_{-i}\right)\right). \end{aligned} \tag{37}$$

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 (9) becomes

$$\begin{aligned} &u_i\left(f\left(\widetilde{\theta}_i,\zeta_i\left(\theta_i,\theta_i'\right)\right),\left(\widetilde{\theta}_i,\zeta_i\left(\theta_i,\theta_i'\right)\right)\right) \\ \geq \\ &u_i\left(y\left(\left(\theta_i',\theta_i'\right),\left(\zeta_{ij}\left(\theta_i,\theta_i'\right),\zeta_{ij}\left(\theta_i,\theta_i'\right)\right)_{j\neq i}\right),\left(\widetilde{\theta}_i,\zeta_i\left(\theta_i,\theta_i'\right)\right)\right),\ \forall\widetilde{\theta}_i. \end{aligned} \tag{38}$$

But now (33), (34) and (38) implies that (37) 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. ■

The proof may appear rather intricate in its details and we give a brief outline of the basic steps next. We start with an arbitrary deception β which satisfies the inequalities (31) and (32) 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 $\theta_i' \in \beta_i\left(\theta_i\right)$. 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 (31) and (32). 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 (31), and thus leads to the conclusion in form of the inequalities (32). 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.

Theorem 6
If f satisfies robust monotonicity, then f satisfies interim monotonicity on every type space.

Proof. 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

$$\begin{aligned}&\sum_{\left\{\theta_{-i}\in\Theta_{-i}:\theta_{-i}'\in\beta_{-i}\left(\theta_{-i}\right)\right\}}\psi_i\left(\theta_{-i}\right)u_i\left(y\left(\theta_{-i}',\psi_i\right),\left(\theta_i,\theta_{-i}\right)\right)\\ >&\sum_{\left\{\theta_{-i}\in\Theta_{-i}:\theta_{-i}'\in\beta_{-i}\left(\theta_{-i}\right)\right\}}\psi_i\left(\theta_{-i}\right)u_i\left(f\left(\theta_i',\theta_{-i}'\right),\left(\theta_i,\theta_{-i}\right)\right)\end{aligned}\tag{39}$$

and

$$u_i\left(f\left(\widetilde{\theta}_i,\theta_{-i}'\right),\left(\widetilde{\theta}_i,\theta_{-i}'\right)\right)\geq u_i\left(y\left(\theta_{-i}',\psi_i\right),\left(\widetilde{\theta}_i,\theta_{-i}'\right)\right),\tag{40}$$

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

$$\xi_i\left(\theta_{-i}'\right)\triangleq\sum_{\left\{t_{-i}\in T_{-i}:\widehat{\theta}_{-i}\left(\alpha_{-i}\left(t_{-i}\right)\right)=\theta_{-i}'\right\}}\widehat{\pi}_i\left(t_i\right)\left[t_{-i}\right]\tag{41}$$

and

$$\psi_i\left(\theta_{-i}|\theta_{-i}'\right)\triangleq\frac{\sum_{\left\{t_{-i}\in T_{-i}:\widehat{\theta}_{-i}\left(t_{-i}\right)=\theta_{-i}\text{ and }\widehat{\theta}_{-i}\left(\alpha_{-i}\left(t_{-i}\right)\right)=\theta_{-i}'\right\}}\widehat{\pi}_i\left(t_i\right)\left[t_{-i}\right]}{\sum_{\left\{t_{-i}\in T_{-i}:\widehat{\theta}_{-i}\left(\alpha_{-i}\left(t_{-i}\right)\right)=\theta_{-i}'\right\}}\widehat{\pi}_i\left(t_i\right)\left[t_{-i}\right]}.\tag{42}$$

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 $y(t)$ on the type space T by setting:

$$y(\alpha_i(t_i), t_{-i}) \triangleq y\left(\widehat{\theta}_{-i}(t_{-i}), \psi_i\left(\cdot \left| \widehat{\theta}_{-i}(t_{-i})\right.\right)\right). \tag{43}$$

Using the probabilities distributions defined in (41) and (42), and the reward function defined in (43) we have the following equalities useful to establish the interim reward inequality:

$$\begin{aligned} &\sum_{t_{-i}\in T_{-i}} u_i\left(y(\alpha(t)), \widehat{\theta}(t)\right)\widehat{\pi}_i(t_i)[t_{-i}] \\ = &\sum_{\theta'_{-i}\in\Theta_{-i}}\sum_{\theta_{-i}\in\Theta_{-i}} u_i\left(y\left(\theta'_{-i}, \psi_i\left(\cdot|\theta'_{-i}\right)\right), \theta\right)\psi_i\left(\theta_{-i}|\theta'_{-i}\right)\xi_i\left(\theta'_{-i}\right) \end{aligned} \tag{44}$$

and

$$\begin{aligned} &\sum_{t_{-i}\in T_{-i}} u_i\left(f\left(\widehat{\theta}(\alpha(t))\right), \widehat{\theta}(t)\right)\widehat{\pi}_i(t_i)[t_{-i}] \\ = &\sum_{\theta'_{-i}\in\Theta_{-i}}\sum_{\theta_{-i}\in\Theta_{-i}} u_i\left(f\left(\theta'\right), \theta\right)\psi_i\left(\theta_{-i}|\theta'_{-i}\right)\xi_i\left(\theta'_{-i}\right). \end{aligned} \tag{45}$$

As the inequality (39) holds for every θ'_{-i}, we can infer from (39) 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 (44) and (45), we establish that:

$$\begin{aligned} &\sum_{t_{-i}\in T_{-i}} u_i\left(y(\alpha(t)), \widehat{\theta}(t)\right)\widehat{\pi}_i(t_i)[t_{-i}] \\ > &\sum_{t_{-i}\in T_{-i}} u_i\left(f\left(\widehat{\theta}(\alpha(t))\right), \widehat{\theta}(t)\right)\widehat{\pi}_i(t_i)[t_{-i}]. \end{aligned} \tag{46}$$

Using again the probabilities distributions defined in (41) and (42), the reward function defined in (43), we have the following equalities useful to establish the interim incentive inequalities:

$$\begin{aligned} &\sum_{t_{-i}\in T_{-i}} u_i\left(f\left(\widehat{\theta}(t'_i, t_{-i})\right), \widehat{\theta}(t'_i, t_{-i})\right)\widehat{\pi}_i(t'_i)[t_{-i}] \\ = &\sum_{\theta'_{-i}\in\Theta_{-i}}\sum_{\theta_{-i}\in\Theta_{-i}} u_i\left(f\left(\widehat{\theta}_i(t'_i), \theta_{-i}\right), \left(\widehat{\theta}_i(t'_i), \theta_{-i}\right)\right)\psi_i\left(\theta_{-i}|\theta'_{-i}\right)\xi_i\left(\theta'_{-i}\right) \end{aligned} \tag{47}$$

and

$$\begin{aligned} &\sum_{t_{-i}\in T_{-i}} u_i\left(y(\alpha_i(t_i), t_{-i}), \widehat{\theta}(t'_i, t_{-i})\right)\widehat{\pi}_i(t'_i)[t_{-i}] \\ = &\sum_{\theta'_{-i}\in\Theta_{-i}}\sum_{\theta_{-i}\in\Theta_{-i}} u_i\left(y(\theta_{-i}, \psi_i(\cdot|\theta_{-i})), \left(\widehat{\theta}_i(t'_i), \theta_{-i}\right)\right)\psi_i\left(\theta_{-i}|\theta'_{-i}\right)\xi_i\left(\theta'_{-i}\right), \ \forall t'_i. \end{aligned} \tag{48}$$

By appealing the ex post incentive inequalities of robust monotonicity, (40), we know that

$$u_i\left(f\left(\widehat{\theta}_i\left(t_i'\right),\theta_{-i}\right),\left(\widehat{\theta}_i\left(t_i'\right),\theta_{-i}\right)\right)\geq u_i\left(y\left(\theta_{-i}',\psi_i\left(\cdot\left|\theta_{-i}\right.\right)\right),\left(\widehat{\theta}_i\left(t_i'\right),\theta_{-i}\right)\right), \tag{49}$$

for all t_i'. The inequalities (49) 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 (47) and (48) we can then establish the interim incentive compatibility conditions:

$$\begin{aligned}&\sum_{t_{-i}\in T_{-i}}u_i\left(f\left(\widehat{\theta}\left(t_i',t_{-i}\right)\right),\widehat{\theta}\left(t_i',t_{-i}\right)\right)\widehat{\pi}_i\left(t_i'\right)\left[t_{-i}\right] \\ \geq\ &\sum_{t_{-i}\in T_{-i}}u_i\left(y\left(\alpha_i\left(t_i\right),t_{-i}\right),\widehat{\theta}\left(t_i',t_{-i}\right)\right)\widehat{\pi}_i\left(t_i'\right)\left[t_{-i}\right],\ \forall t_i'.\end{aligned} \tag{50}$$

But by (46) and (50), we have confirmed interim monotonicity on this type space. ■

The proof of the above theorem uses 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 $\pi_i\left(t_i\right)\left[t_{-i}\right]$ a conditional distribution over payoff types $\xi_i\left(t_i\right)\left[\theta_{-i}\right]$. 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.

The following property is a simple implication of robust monotonicity.

Definition 15 (Pairwise Robust Monotonicity)
If $f\left(\theta\right)\neq f\left(\theta'\right)$, then there exist i and y such that

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

and

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

The pairwise notion shares the requirement that a reward can be offered for every pair θ,θ' of payoff profiles, where θ is true payoff profile and θ' is the deception, but then requires that the ex post incentive constraints are satisfied for all possible misreports regarding payoff types of agent i. The reward inequality is thus identical to the one imposed by Maskin monotonicity (see Definition 6.1) but the incentive constraints are extended to the private information of agent i, as opposed to the complete information assumption inherent to Maskin monotonicity. It is easy to verify that Example C satisfies pairwise robust monotonicity, but not robust monotonicity. The next result relates robust, pairwise robust, ex post and Maskin monotonicity.

Proposition 1 (Pairwise Robust Monotonicity)

1. *If f satisfies robust monotonicity, then f satisfies pairwise robust monotonicity;*
2. *If f satisfies pairwise robust monotonicity, then f satisfies Maskin monotonicity;*
3. *If f satisfies pairwise robust monotonicity, then f satisfies ex post monotonicity.*

Proof. (1.) Suppose that $f\left(\theta\right)\neq f\left(\theta'\right)$ and consider the unacceptable deception where

$$\beta_i\left(\widetilde{\theta}_i\right)=\begin{cases}\left\{\widetilde{\theta}_i\right\},\text{ if }\widetilde{\theta}_i\neq\theta_i\\ \left\{\theta_i,\theta_i'\right\},\text{ if }\widetilde{\theta}_i=\theta_i\end{cases}.$$

By the second definition of robust monotonicity, there exists i such that for all $\theta'_{-i} \in \Theta_{-i}$ and $\psi_i \in \Delta\left(\left\{\theta_{-i} \in \Theta_{-i} : \theta'_{-i} \in \beta_{-i}(\theta_{-i})\right\}\right)$, there exists y such that

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

and

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

for all $\widetilde{\theta}_i$. Letting ψ_i put mass 1 on θ_{-i}, we have

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

and

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

(2.) Restricting $\widetilde{\theta}_i$ to be equal to θ'_i in the pairwise robust monotonicity condition, we get the second definition of Maskin monotonicity.

(3.) Fix any ex post deception α with $f(\theta) \neq f(\alpha(\theta))$ for some θ. Letting $\theta' = \alpha(\theta)$ in the definition of pairwise robust monotonicity, we have that there exist i and y such that

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

and

$$u_i\left(f\left(\widetilde{\theta}_i, \alpha_{-i}(\theta_{-i})\right), \left(\widetilde{\theta}_i, \alpha_{-i}(\theta_{-i})\right)\right) \geq u_i\left(y, \left(\widetilde{\theta}_i, \alpha_{-i}(\theta_{-i})\right)\right), \ \forall \widetilde{\theta}_i.$$

But this is just the second definition of ex post monotonicity. ■

6.3 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',*

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

Definition 18 *Social choice function f satisfies selective dominant strategies incentive compatibility 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*

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

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

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.

Proposition 2 (Selective Dominance)

1. *If social choice function f satisfies selective dominance incentive compatibility then f satisfies robust monotonicity.*
2. *If social choice function f satisfies private values and robust monotonicity then f satisfies selective dominance incentive compatibility.*

Proof. (1.) If f satisfies selective dominance, then f satisfies robust monotonicity, since for any β and any $\theta_i' \in \beta_i(\theta_i)$ with $\theta_i' \neq \theta_i$, we will have

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

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

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

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

(2.) The social choice environment satisfies *private values* if

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

for all i, y, θ_i and θ_{-i}. If there are private values, then the robust monotonicity condition implies that for every unacceptable deception β, there exist i, θ_i, $\theta_i' \in \beta_i(\theta_i)$ and $y : \Theta_{-i} \to A$ such that

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

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

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

for all $\left(\widetilde{\theta}_i, \theta_{-i}'\right) \in \Theta$. Setting $\widetilde{\theta}_i = \theta_i$ in the latter condition, we have

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

Thus for every unacceptable deception β, there exist i, θ_i, $\theta_i' \in \beta_i(\theta_i)$ such that

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

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

7 Uniform Implementation

The notion of interim implementation allowed us to specify a distinct mechanism for every type space. Combined with restriction to finite mechanisms, this is restrictive. It is then natural to ask whether we can specify a single mechanism which interim implements the social choice function for all (finite) type spaces. We refer to this as *uniform* implementation. In this section we establish the relation between uniform implementation and implementation in strategies surviving iterated deletion of strictly dominated strategies, where we refer to the later as *iterative* implementation.

7.1 Iterative Implementation

We begin by setting the notation for iterated deletion of strictly dominated strategies. 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:

$$M_i^{k+1}(\theta_i) = \left\{ m_i \in M_i^k(\theta_i) \;\middle|\; \begin{array}{l} \nexists \mu_i \in \Delta(M_i) \text{ s.th.} \\ \sum_{m_i'} \mu_i(m_i') u_i(g(m_i', m_{-i}), (\theta_i, \theta_{-i})) > u_i(g(m_i, m_{-i}), (\theta_i, \theta_{-i})) \\ \forall \theta_{-i} \in \Theta_{-i} \text{ and } \forall m_{-i} \in M_{-i}^k(\theta_{-i}) \end{array} \right\}.$$

We write

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

Definition 19 *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 strict dominated strategies. 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: $\Theta_1 = \left\{\theta_1^1, \theta_1^2\right\}$, $\Theta_2 = \left\{\theta_2^1, \theta_2^2\right\}$. 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_2^1	θ_2^2
θ_1^1	a	b
θ_1^2	c	d

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

a	θ_2^1	θ_2^2
θ_1^1	$3, 3$	$0, 0$
θ_1^2	$0, 0$	$1, 1$

b	θ_2^1	θ_2^2
θ_1^1	$0, 0$	$3, 3$
θ_1^2	$1, 1$	$0, 0$

c	θ_2^1	θ_2^2
θ_1^1	$0, 0$	$1, 1$
θ_1^2	$3, 3$	$0, 0$

d	θ_2^1	θ_2^2
θ_1^1	$1, 1$	$0, 0$
θ_1^2	$0, 0$	$3, 3$

and for z_1 and z_2 are given by:

z_1	θ_2^1	θ_2^2
θ_1^1	$2, 2$	$2, 0$
θ_1^2	$2, 2$	$2, 0$

z_2	θ_2^1	θ_2^2
θ_1^1	$2, 0$	$2, 2$
θ_1^2	$2, 0$	$2, 2$

Consider the following augmented mechanism 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:

	θ_2^1	θ_2^2
θ_1^1	a	b
θ_1^2	c	d
ϕ	y	z

The corresponding incomplete information game has the following payoffs:

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

If we perform iterated deletion of ex post dominated strategies, then we arrive in four steps at a singleton for every type of every agent:

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

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 $g(m_i, m_j) = f(\theta_i, \theta_j)$. Thus there does not exist $\mu_i \in \Delta(M_i)$ such that

$$\sum_{m_i'} \mu_i(m_i') u_i(g(m_i', m_j), (\theta_i, \theta_j)) > u_i(g(m_i, m_j), (\theta_i, \theta_j)) = 3.$$

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,

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

Without loss of generality, let $i = 1$. Note $m_2^*(\theta_2') \in M_2^{k-1}(\theta_2)$ by assumption. If agent 1 was type θ_1 and was sure his opponent were type θ_2 and choosing action $m_2^*(\theta_2')$, we know that he could guarantee himself a payoff of 1 by choosing $m_1^*(\theta_1')$. Since $m_1^*(\theta_1')$ is deleted for type θ_1 at round k, we know that there exists $\mu_1 \in \Delta(M_1)$ such that

$$\sum_{m_1'} \mu_1(m_1') g_1(m_1', m_2^*(\theta_2')) > 1$$

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.[4] Borgers (1995) proves the impossibility of complete information implementation of non-dictatorial social choice functions in iteratively undominated strategies when the set of feasible preference profiles includes such unanimous preference profiles and the argument here is reminiscent of Borgers' argument.

It will be useful to report an alternative characterization of $M^\infty(\theta)$ based on the classical duality argument relating iterated deletion of strictly dominated strategies to rationalizability (Pearce (1984)).

$$M_i^{k+1}(\theta_i) = \left\{ m_i \in M_i^k(\theta_i) \;\middle|\; \begin{array}{c} \exists \lambda_i^k \in \Lambda_i^k \text{ s.t.} \\ m_i \in \arg\max_{m_i'} \sum_{\theta_{-i}, m_{-i}} \lambda_i^k(\theta_{-i}, m_{-i})\, u_i\left(g\left(m_i', m_{-i}\right), (\theta_i, \theta_{-i})\right) \end{array} \right\},$$

where

$$\Lambda_i^k = \left\{ \lambda_i^k \in \Delta(\Theta_{-i} \times M_{-i}) \;\middle|\; \lambda_i^k(\theta_{-i}, m_{-i}) = 0 \text{ if } m_j \notin M_j^k(\theta_j) \text{ for some } j \neq i \right\}.$$

The equivalence of this definition follows immediately from the well known equivalence between an action being strictly dominated and being never a weak best response. Now the mechanism is finite, we know that there exists K such that $M_i^k(\theta_i) = M_i^\infty(\theta_i)$ for all $k \geq K$. Thus we have the following crucial lemma concerning iterative implementation.

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. *$m_i \in \arg\max_{m_i'} \sum_{\theta_{-i}, m_{-i}} \lambda_i(\theta_{-i}, m_{-i})\, u_i\left(g\left(m_i', m_{-i}\right), (\theta_i, \theta_{-i})\right)$.*

7.4 Characterization

With this background material, we can proceed to define and analyze uniform implementation.

Definition 20 *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 alternative characterization of iterative implementation provided by Lemma 2 to relate uniform and iterative implementation.

Theorem 7 *Social choice function f is uniformly implementable if and only if it is iterative implementable.*

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

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.

7.5 The Not-So-Universal Type Space

We do not know how big is the gap between "implementation on all type spaces" (for which robust monotonicity is a necessary and - in economic environments with full support type spaces - a sufficient

condition) and "uniform implementability" (for which iterative implementation is necessary and sufficient). But in this section, we speculate on where the gap comes from.

The traditional approach in the incomplete information implementation literature - which we followed in Sections 4 through 6 - is to fix a type space and ask if we can construct a mechanism such that every pure strategy equilibrium delivers the social desirable outcomes at every state. The pure strategy restriction is an ad hoc restriction that is made for reasons of tractability and has been much criticized (see, for example, Jackson (1992) and Abreu and Matsushima (1992)).

The fixed type space is rarely questioned. We conjecture that this is because of the following "folk" argument that fixing the type space is without loss of generality. If there was uncertainty about the true type space, we could always incorporate that uncertainty into a larger type space and implement on that larger type space. If there was still not common knowledge of the type space, we could construct a yet bigger type space. We know this process constructing larger and larger type space is guaranteed to terminate because of the universal type space construction of Harsanyi (1967/68) and Mertens and Zamir (1985).

The pure strategy restriction only has bite, however, if the type space is fixed. If we are allowed to add payoff irrelevant types to purify mixed strategies, then mixed strategy implementation becomes necessary for pure strategy purification. Thus the ad hoc pure strategy results depend on the fixed type space assumption. But we will argue the folk argument in support of fixing the type space is wrong.

To make this argument, we first report how the standard construction of the universal type space would work in this setting. The only non-standard feature of this construction is that we want to assume that it is common knowledge that each agent knows his own payoff type. We will build this feature into the construction (Neeman (2001) and Heifetz and Neeman (2003) report similar constructions). Agent i's 0-th level type is his payoff type $t_i^0 = \theta_i \in \Theta_i$. Let $T_i^0 \equiv \Theta_i$ be agent i's set of 0-th level types. agent i's 1st level type must specify his payoff type and his belief about other agents' 0th level types. Thus $t_i^1 \in T_i^1 \equiv \Theta_i \times \Delta\left(T_{-i}^0\right)$. agent i's 2nd level type must specify his payoff type and his belief about other agents' 1st level types. Thus $t_i^2 \in T_i^2 \equiv \Theta_i \times \Delta\left(T_{-i}^1\right)$. Iterating this construction, we have $t_i^k \in T_i^k \equiv \Theta_i \times \Delta\left(T_{-i}^{k-1}\right)$, and we obtain an infinite hierarchy of beliefs $\left(t_i^0, t_i^1, t_i^2, ...\right)$. We want to require that high level types, which intuitively contain more information than lower level types, are consistent with lower levels. Formally, an infinite hierarchy is *coherent* if all higher level types have the same payoff type as lower level types and if the projection of their beliefs over other agents' types onto lower level type spaces is consistent with lower level types' beliefs. Now if we impose some topological structure on the belief spaces, we can let agent i's possible types, T_i, be the set of all infinite hierarchies of beliefs. The crucial property of such type spaces is that the set of types, constructed as infinite hierarchies, can be identified with pairs of payoff types and beliefs types, so that, for each i, there exists a homeomorphism $h_i : T_i \to \Theta_i \times \Delta\left(T_{-i}\right)$. For example, Brandenburger and Dekel (1993) show that if we topologize the belief spaces with a complete separable metric, then this follows from Kolmogorov's Existence Theorem. Now letting $\widehat{\theta}_i$ be the projection of h_i onto Θ_i and letting $\widehat{\pi}_i$ be the projection of h_i onto $\Delta\left(T_{-i}\right)$, this canonical universal type space fits (apart from the infinite set of types) into the language for type spaces described in section 3.2.

Now we could ask if it is possible to pure (or mixed) strategy implement social choice function $f : \Theta \to A$ on this universal type space. Since this type space is infinite, we would presumably want to allow for infinite mechanisms, perhaps with messages at least as rich as the already rich type space. While the exact answer to the question of whether f could be implemented on the universal type space might be sensitive to assumptions such as the richness of the type space, the question could certainly be precisely posed (we haven't done it and would be interested to see it done!).

However, even if f could be implemented on the universal type space, it would not prove that f could be implemented on all type spaces. It is a misunderstanding of what was proved in the universal type space construction to believe that this is so. To illustrate this point, consider the

following type space:

$$T_1 = \{t_1, t_1'\},$$
$$T_2 = \{t_2, t_2'\},$$

with a single payoff type for each agent:

$$\widehat{\theta}_1(t_1) = \widehat{\theta}_1(t_1') = \theta_1,$$
$$\widehat{\theta}_2(t_2) = \widehat{\theta}_2(t_2') = \theta_2,$$

and the following associated belief types

$$\widehat{\pi}_1(t_1)[t_2] = \frac{2}{3},$$
$$\widehat{\pi}_1(t_1')[t_2] = \frac{1}{3},$$
$$\widehat{\pi}_2(t_2)[t_1] = \frac{2}{3},$$
$$\widehat{\pi}_2(t_2')[t_1] = \frac{1}{3}.$$

Since all types have the same payoff type, the infinite hierarchy of beliefs is degenerate. I.e., each type of agent i is sure that he has payoff type θ_i, he is sure that his opponent j has payoff type θ_j, and so on. However, because of the opportunities for correlation, rational strategic behavior on this type space may be very different from the type space where each agent has only a single possible type.

Similarly, suppose we constructed a mechanism that implemented f on the universal type space, with message spaces even richer than the (large) set of types. Then we could always add some extra payoff irrelevant types to reflect players' strategic uncertainty about how others were playing the game. If the implementation of the social choice function uses the equilibrium assumption (rather than just iterated deletion), then we can add types in a way that undermines the implementation of the social choice function. This is a general lesson from the literature on the epistemic foundations of game theoretic solution concepts (e.g., Brandenburger and Dekel (1987)) and it was a general lesson from Proposition 7.

8 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.

We establish equivalence between ex post implementation and interim implementation on all type spaces provided that the social choice function satisfies the new notion of robust monotonicity. We finally extend the line of argument and ask when a given mechanism can interim implement the social choice function for every (finite) type spaces and relate uniform implementation to the notion of iterated deletion of strictly dominated strategies.

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.

9 Appendix: Full Support Uniform Implementation

The arguments for our characterization of uniform implementation clearly rely on allowing for type spaces where beliefs do not have full support and, in particular, agents' beliefs' supports may be inconsistent. In this appendix, we discuss how important this issue might be.

First, note that by imposing a full support assumption, we will clearly move from a characterization corresponding to iterated deletion of strictly dominated strategies to one which allows deletion of at least some weakly dominated strategies. On the one hand, this suggests that we would be able to obtain much more permissive results, since we know that there is sometimes much to be gained in the full implementation literature by allowing the deletion of weakly dominated strategies (give refs). On the other hand, we are interested in *robust* implementation. We are maintaining the assumption that there is common knowledge of the set of possible payoff types, even as we allow for a rich set of higher order belief or payoff irrelevant types. Clearly, if there is behavior that can be supported on non-full support type spaces then, by adding payoff perturbations to the payoff types, we could support this behavior on full support type spaces. While this type of small payoff perturbation is not formally part of our model, it would be easy to add on in a way that destroys the prior of weak dominance. For example, Chung and Ely (2003) give an argument that small uncertainty about payoff types in a complete information setting destroys the ability of weak dominance arguments to dispense with Maskin monotonicity.

However, it is nonetheless natural to ask the attempt to characterize the following natural full support uniform implementation concept:

Definition 21 *Social choice function f is full support uniformly implementable if there exists a mechanism $\mathcal{M}$ such that for every full support finite type space $\mathcal{T}$, every (pure strategy) interim equilibrium s of the game $(\mathcal{T}, \mathcal{M})$ satisfies*

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

Note that as in the case of uniform implementation, the pure strategy restriction is without loss of generality here: if mixed strategy implementation was impossible on some full support type space, we could construct another larger full support type space where pure strategy implementation was impossible.

Here, we are able to present a characterization although we do not know much about its properties. We first give an alternative fixed point characterization of ex post rationalizable implementation that extends more easily to the full support case.

9.1 Fixed Point Characterization of Uniform Implementation

Fix a mechanism $\mathcal{M} = (M_1, ..., M_I, g)$. Let $S_i : \Theta_i \to 2^{M_i} / \varnothing$ and $S = (S_1, ..., S_I)$.

Definition 22 *S satisfies the ex post best response property if for all i, θ_i and $m_i \in S_i(\theta_i)$, there exists $\lambda_i \in \Delta(\Theta_{-i} \times M_{-i})$ such that:*

1. $\lambda_i(\theta_{-i}, m_{-i}) > 0 \Rightarrow m_j \in S_j(\theta_j)$ *for all* $j \neq i$;
2. $m_i \in \arg\max_{m_i'} \sum_{\theta_{-i}, m_{-i}} \lambda_i(\theta_{-i}, m_{-i})\, u_i(g(m_i', m_{-i}), (\theta_i, \theta_{-i}))$.

By standard arguments, we know that the set of strategies surviving iterated deletion satisfies the best response property, and every S satisfying the best response property is contained in the set of strategies surviving interated deletion, i.e., if S satisfies the best response property, then

$$S_i(\theta_i) \subseteq M_i^{\infty}(\theta_i)$$

for all i and θ_i. So f is iteratively implementable (and thus uniformly implementable) if and only if for every S satisfy the best response property,

$$m \in S(\theta) \Rightarrow g(m) = f(\theta).$$

We will generalize this characterization of uniform implementability to the full support case.

9.2 Fixed Point Characterization of Full Support Uniform Implementation

Definition 23 *S satisfies the ex post full support best response property if for all i, θ_i and $m_i \in S_i(\theta_i)$, there exists $\lambda_i \in \Delta(\Theta_{-i} \times M_{-i})$ such that:*

1. *$\lambda_i(\theta_{-i}, m_{-i}) > 0$ if and only if $m_j \in S_j(\theta_j)$ for all $j \neq i$;*
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 24 *Social choice function f is full support iteratively (FSI) implementable if for every S satisfying the ex post full support best response property, $m \in S(\theta) \Rightarrow g(m) = f(\theta)$.*

Proposition 3 *Social choice function f is full support uniformly implementable if and only if f is FSI implementable.*

Proof. First, suppose that f is FSI implementable. Then $m \in S(\theta) \Rightarrow g(m) = f(\theta)$ for every S satisfying the ex post full support best response property. Fix any full support type space $\mathcal{T}$. Choose a mechanism $\mathcal{M}$ that FSI implements f. Fix any equilibrium s of the game $(\mathcal{T}, \mathcal{M})$ and let

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

By construction, S satisfies the ex post full support best response property. Thus $g(s(t)) = f\left(\widehat{\theta}(t)\right)$.

Now suppose that f is not FSI implementable. Then for any mechanism $\mathcal{M}$, there exists S satisfying (1) the ex post full support best response property and (2) $g(m^*) \neq f(\theta^*)$ for some θ^* and $m^* \in S(\theta^*)$. Recall from Definition 23 that for each $m_i \in S_i(\theta_i)$, 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 and only if $m_j \in S_j(\theta_j)$ for all $j \neq i$;
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 construct the type space where

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

By construction, there this is a full support type space and 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 $\widehat{\theta}(\theta^*, m^*) = \theta^*$. ■

To understand the significance of the full support best response property, consider the special case where each agent has only one type. In this case, we are looking at a complete information solution concept that refines the set of rationalizable actions. Sets of actions satisfying the full best response property have the additional property that each action for each player is admissible *with respect to the strategies of other players in that set.*

This is related to the self-admissible sets of Brandenburger and Friedenberg (2003). They show that an action is consistent with "common assumption of rationality" if and only if it is included in some self-admissible set. They impose the additional requirement that strategies that are inadmissible with respect to the whole strategy set cannot be included. And they have the additional requirement that strategies whose payoffs are a convex combinations of other strategies must also be included in a self-admissible set. In any case, the set of actions that are included in some self-admissible set include all strategies surviving iterated deletion of weakly dominated strategies and is included in the set of strategies surviving iterated deletion of strictly dominated strategies. The complete information analogue of our solution concept would be similarly situated.[5]

[5] We are grateful to Amanda Friedenberg for explaining this connection.

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 J. Valimaki. 2002. "Information Acquisition and Mechanism Design." *Econometrica* 70: 1007-1033.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

[26] 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.

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

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

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

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

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

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

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

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

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

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

[37] 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.