STRATEGIC DISTINGUISHABILITY
WITH AN APPLICATION TO ROBUST VIRTUAL IMPLEMENTATION

By

Dirk Bergemann and Stephen Morris

June 2007

COWLES FOUNDATION DISCUSSION PAPER NO. 1609

Strategic Distinguishability

with an application to Robust Virtual Implementation[*]

Dirk Bergemann[†]   Stephen Morris[‡]

First Version: March 2006
This Version: May 2007

Abstract

In a general interdependent preference environment, we characterize when two payoff types can be distinguished by their rationalizable strategic choices without any prior knowledge of their beliefs and higher order beliefs. We show that two types are *strategically distinguishable* if and only if they satisfy a separability condition. The separability condition for each agent essentially requires that there is not too much interdependence in preferences across agents.

A social choice function - mapping payoff type profiles to outcomes - can be *robustly virtually implemented* if there exists a mechanism such that every equilibrium on every type space achieves an outcome arbitrarily close to the social choice function: this definition is equivalent to requiring virtual implementation in iterated deletion of strategies that are strictly dominated for all beliefs. The social choice function is *robustly measurable* if strategically indistinguishable types receive the same allocation. We show that ex post incentive compatibility and robust measurability are necessary and sufficient for robust virtual implementation.

KEYWORDS: Mechanism Design, Virtual Implementation, Robust Implementation, Rationalizability, Ex-Post Incentive Compatibility.

JEL CLASSIFICATION: C79, D82

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

[‡]Department of Economics, Princeton University, Prospect Street, Princeton NJ 08544, smorris@princeton.edu.

1

1 Introduction

Preferences are assumed to be interdependent for informational or psychological reasons in many areas of economics. But there has been little attempt to identify what are the observable implications of such preferences. A classic and well developed "revealed preference" theory underlies economists' way of understanding individual choice. An analogous *strategic* revealed preference understanding of interdependent preferences is required. This paper proposes an approach to this question.

Fix an interdependent preferences environment, with a finite set of agents, each with a finite set of possible *payoff types*, with expected utility preferences over lotteries depending on the whole profile of types. Say that two payoff types of an agent are *strategically distinguishable* if they have disjoint rationalizable strategic choices in some finite game for all possible beliefs and higher order beliefs about others' types. Thus a pair of payoff types are strategically *indistinguishable* if in every game, there exists some action which each type might rationally choose given some beliefs and higher order beliefs. We are able to provide an exact and insightful characterization of strategic distinguishability. If we have sets of types, Ψ_1 and Ψ_2, of agents 1 and 2, respectively, we say that Ψ_2 separates Ψ_1 if knowing agent 1's preferences and knowing that agent 1 is sure that agent 2's type is in Ψ_2, we can rule out at least one type of agent 1. Now consider an iterative process where we start, for each agent, with all subsets of his type set and - at each round - delete subsets of actions that are separated by every remaining subset of types of his opponents. A pair of types are said to be *pairwise inseparable* if the set consisting of that pair of types survives this process. We show that two types are strategically indistinguishable if and only if they are pairwise inseparable.

If there are private values and every type is value distinguished, then every pair of types will be pairwise separable and thus strategically distinguishable. Thus strategic indistinguishability arises when the degree of interdependence in preferences is large. We can illustrate this with a simple example. Suppose that agent i's payoff type is $\theta_i \in [0,1]$ and agent i's valuation of a private good is $\theta_i + \gamma \sum_{j \neq i} \theta_j$. Each agent has quasilinear utility, i.e., his utility from money is linear and additive. We show all distinct pairs of types are strategically distinguishable if $|\gamma| < \frac{1}{I-1}$ where I is the number of agents. All pairs of types are strategically indistinguishable if $|\gamma| \geq \frac{1}{I-1}$.

Two rational payoff types are strategically indistinguishable if they *might* choose the same action (in any game). We will show that this strategic revealed preference relation on payoff types is key to the implementation problem, when one cannot allow for the possibility of two distinct types behaving the same in every mechanism. But say that two payoff types are *strategically equivalent* if the sets of actions they might choose are the *same* (in any game). In other words, two types are strategically equivalent if they have the same set of rationalizable actions in every game. We contrast strategic distinguishability with strategic equivalence and note that strategic equivalence generates a much finer partition on agents' types; for example, in the linear example of the previous paragraph, no distinct types are strategically equivalent. If two rational payoff types are strategically equivalent, it is not possible that they will behave differently in any game. Strategic equivalence is the relevant strategic revealed preference notion if one is interested in identifying the finest behaviorally relevant description of agents' interdependent types. This is the question studied by Gul and Pesendorfer (2005), who pioneered the study of the revealed preference implications

of interdependent preferences. While they do not explicitly allow for uncertainty or incorporate strategic choices, the strategic equivalence question in our framework is closely related to their exercise. In particular, to facilitate the comparison, we show that a version of their validity condition is sufficient to ensure that distinct types are not strategically equivalent.

As well as its intrinsic interest, strategic distinguishability is key in characterizing when *robust virtual implementation* is possible. Suppose that a social planner would like to design a mechanism that will induce self-interested agents to make strategic choices that will lead to the selection of socially desirable outcomes. A *social choice function* specifies the social desired outcomes as a function of unobserved payoff types of the agents. The planner would like to be sure that outcomes specified by the social choice function arise with probability arbitrarily close to 1: thus she requires *virtual* implementation; she would like *every* possible equilibrium to virtually implement the social choice function: thus she requires *full* implementation; and she would like every equilibrium to virtually implement the social choice function whatever the agents' beliefs and higher order beliefs about others' types; thus she requires *robust* implementation. In this paper, we provide a characterization of when *robust virtual implementation* is possible in a general interdependent preference environment.

One necessary condition for robust virtual implementation will be *ex post incentive compatibility*: under the social choice function, each agent must have an incentive to truthfully report his type if others' report their types truthfully, whatever their types. Ex post incentive compatibility is sufficient to ensure the existence of desirable equilibria, but, as the existing incomplete information implementation literature has emphasized, further restrictions on the social choice function are required to rule out other, undesirable, equilibria. If a mechanism is to fully implement a social choice function, it must be that two types who are treated differently by the social choice function are guaranteed to behave differently in the implementing mechanism. If two types are guaranteed to behave differently in the implementing mechanism, then - under our definition outlined above - they are strategically distinguishable. Thus a second necessary condition for robust virtual implementation will be *robust measurability*: strategically indistinguishable types are treated the same by the social choice function. We show that ex post incentive compatibility and robust measurability are also sufficient for robust virtual implementation (under an economic assumption).

Our characterization result for strategic distinguishability (theorem 1) comes in two parts. If two types of an agent are pairwise inseparable, then they belong to a set of types which are not separable by a profile of sets of types of that agent's opponents. The set of types of each opponent in that profile is then not separable by a profile of sets of types of that opponent's opponents. And there is a continuing chain of inseparable sets in the chain. We prove that pairwise inseparable types are strategically indistinguishable (proposition 1) by induction, showing that in any mechanism at any round in the iterated deletion of messages that are never best responses and for every set of types in the chain of inseparable type sets, there is a common action which is played. The inseparability property ensures that we can always construct beliefs for each type that make the same message a best response.

To show the converse result (proposition 2), we construct a single, finite *maximally revealing mechanism* with the property that all pairwise separable types have disjoint sets of rationalizable actions. The construction exploits the linearity of expected utility preferences and duality theory. Whenever a set of types

of one agent is separated by a profile of sets of types of other agents, we are able to construct a finite set of lotteries such that knowing the first agent's preference over those lotteries will always rule out at least one of his types. We can take the union over all such finite sets constructed for each profile of type sets where the separability property holds. We then construct a finite "test set" of lotteries such that knowing an agent's most preferred outcome in that test set implicitly reveals his ranking of outcomes in all the original sets. Finally, we consider a mechanism where each agent gets to pick a lottery with some positive probability, then guesses which lotteries others chose and gets to pick another lottery, with small probability, contingent on other agents making the choice he conjectured, and so on. With a large, but finite, number of rounds this mechanism will eventually lead pairwise separable types to make distinct choices.

Our proof of the sufficiency of ex post incentive compatibility and robust measurability (corollary 1) for robust virtual implementation builds on an ingenious construction used by Abreu and Matsushima (1992b) to establish an extremely permissive result for complete information virtual implementation; in Abreu and Matsushima (1992c), they adapted the argument to a standard Bayesian virtual implementation problem; we in turn adapt the argument to our robust virtual implementation problem.

While our sufficiency argument for robust virtual implementation builds on Abreu and Matsushima (1992c), the interpretation of our results ends up being rather different. Abreu and Matsushima (1992c) characterized virtual implementation in a standard Bayesian environment, where there was common knowledge of a common prior over a fixed set of types, using the solution concept of iterated deletion of strictly dominated strategies and restricting attention to well-behaved (finite) mechanisms. Bayesian incentive compatibility of the social choice function is a necessary condition: a standard compactness argument shows that the weakening to *virtual* implementation does not weaken the incentive compatibility requirement. In addition, they showed that a *measurability* condition was necessary. Put each agent's types into equivalence classes that have the same preferences over outcomes - unconditional on other agents' types. Having distinguished some types by their unconditional preferences, we can then further refine agents' types, by distinguishing types with different preferences conditional on other agents' types in the first round. We can continue this process of refining agents' types based on preferences conditional on other agents' types revealed so far. The social choice function is *Abreu-Matsushima measurable* if it is measurable with respect to the limit of this iterative refinement. This seems to be a weak restriction that is generically satisfied.[1] They show that Bayesian incentive compatibility and Abreu-Matsushima measurability are sufficient as well as necessary for virtual implementation in iterated deletion of strictly dominated strategies.

Robust virtual implementation is equivalent to requiring that there is a single mechanism that implements a social choice function, for all possible type spaces that could be constructed for the environment with fixed payoff types and utility functions for the agents. It is instructive to see how to get from Abreu and Matsushima (1992c) to the robust virtual implementation application in this paper.

Observe that Abreu and Matsushima (1992c)'s solution concept naturally uses agents' given beliefs about others' types in their solution concept: when strategies are deleted, it is because they are strictly dominated

---

[1] Abreu and Matsushima (1992c) and Serrano and Vohra (2005) note that a simple sufficient condition for all social choice functions to be A-M measurable is *type diversity*: every type has distinct preferences over lotteries unconditional on others' types.

conditional on their beliefs. We want implementation for all possible beliefs; we therefore establish our results under an incomplete information version of rationalizability that does not make use of any beliefs over others' types; it is equivalent to iteratively deleting strategies that are *ex post strictly dominated*, i.e., strictly dominated for all possible beliefs over others' types. We work with this solution concept throughout the paper. However, results from the epistemic foundations of game theory establish that an action is rationalizable in this sense for a payoff type if and only if it could be played in an equilibrium on some type space with beliefs and higher order beliefs, by a type with that payoff type (Brandenburger and Dekel (1987) and Battigalli and Siniscalchi (2003)). Thus a bonus of our "robust" analysis is that the distinction between equilibrium and rationalizability (or iterated deletion of strictly dominated strategies) becomes moot.

Now ex post incentive compatibility is the robust analogue of Bayesian incentive compatibility and robust measurability is the robust analogue of Abreu-Matsushima measurability. Abreu and Matsushima (1992c) could reasonably argue that - in a standard Bayesian setting - their measurability condition is a weak technical requirement.[2] As a result, the "bottom line" of the virtual implementation literature has been that full implementation, i.e., getting rid of undesirable equilibria, does not impose any substantive constraints beyond incentive compatibility, i.e., the existence of desirable equilibria. By requiring the more demanding, but more plausible, robust formulation of incomplete information, we end up with a condition that is substantive, imposing significantly more structure in interdependent value environments than incentive compatibility, easily interpretable and - via the relation to strategic distinguishability - of independent conceptual interest.

This paper adds to a recent literature on robust mechanism design that provides one operationalization of the so-called "Wilson doctrine"[3] that progress in practical mechanism design will come from relaxing the implicit common knowledge assumption in the formulation of mechanism design problems.[4] Neeman (2004) highlighted the fact that full surplus extraction with correlated type results (Myerson (1981) and Cremer and McLean (1985)) rely on the implicit assumption that there is common knowledge of a mapping from beliefs to payoff types of all agents (a "beliefs determine preferences" property). This (counterintuitive) assumption is implied by the "generic" choice of a common prior on a fixed type space where distinct types are assumed to have different preferences. The apparent weakness of the Abreu-Matsushima measurability condition (and the fact that it is satisfied for generic priors) relies on the same property. We believe that by relaxing this unnatural implicit assumption, we get a better insight into the nature of the extra requirement for full implementation over and above incentive compatibility conditions.

It is possible to interpret our result as rather negative: ex post incentive compatibility is already a very strong condition, as emphasized by the recent work of Jehiel, Moldovanu, Meyer-Ter-Vehn, and Zame (2006); robust measurability adds the further substantive restriction that there not be too much interdependence of preferences; and, in any case, the mechanism that we use to robustly virtually implement social choice functions is complicated to describe and presumably hard to play. However, we can show that in one large and interesting class of economic environments with interdependent preferences, robust virtual implementation

---

[2] Although Serrano and Vohra (2001) describe an economic example where all individually rational and Bayesian incentive compatible social choice functions fail Abreu-Matsushima measurability because types have identical conditional preferences.

[3] Wilson (1987) contains a statement of what Eric Maskin has dubbed the "Wilson doctrine".

[4] Neeman (2004), Bergemann and Morris (2005c), Heifetz and Neeman (2006), Chung and Ely (2007).

is not only possible but is possible in the direct mechanism where agents simply report their payoff types. Say that an environment has *aggregator single crossing* preferences if the profile of agents' types can be aggregated into a single number and preferences are single crossing with respect to that number. Efficient social choice functions satisfying ex post incentive compatibility often exist in such environments. Bergemann and Morris (2005a) showed that in such an environment, exact robust implementation is possible if the social choice function satisfies strict ex post incentive compatibility and a contraction property. In this paper, we observe that the contraction property is equivalent to robust measurability, so that - under the weak condition that there exists some strictly ex post incentive compatible social choice function - whenever robust virtual implementation is possible, it is possible in the direct mechanism.

The remainder of the paper is organized as follows. Section 2 introduces the environment and the solution concept. Section 3 illustrates the notion of separability in the context of a single private good with interdependent preferences. Section 4 defines and characterizes strategic distinguishability, constructing the maximally revealing mechanism to show the equivalence between strategic distinguishability to pairwise separability. Section 5 reports our results on robust virtual implementation. Section 6 contains discussion of the stronger notion strategic equivalence (discussed above), the epistemic foundations for the solution concept, weak rather than strict dominance, postive results in the direct mechanisms and the relation to exact implementation results. Section 7 concludes.

2 Setting

2.1 Environment

There is a finite set of agents $1, ..., I$ and each agent i has finite set of possible payoff types:

$$\Theta_i = \left\{ \theta_i^1, ..., \theta_i^l, ..., \theta_i^L \right\}.$$

We assume without loss of generality that the cardinality of each set Θ_i is equal to L for all i. The finite set X of pure outcomes is given by

$$X = \left\{ x_1, ..., x_n, ..., x_N \right\}.$$

The lottery space over the set of outcome is $Y = \Delta(X)$. A lottery y is an N dimensional vector $y = (y_1, ..., y_n, ..., y_N)$ with

$$y_n \geq 0, \quad \sum_{n=1}^{N} y_n = 1.$$

Each agent has a von Neumann Morgenstern expected utility function $u_i : Y \times \Theta \to \mathbb{R}$ with

$$u_i(y, \theta) = \sum_{n=1}^{N} u_i(x_n, \theta) y_n.$$

We will abuse notation by writing x for the lottery putting probability 1 on outcome x and X for the set of degenerate lotteries.

It is often convenient to work with underlying preferences over lotteries rather than any of their representations. We write $\mathcal{R}$ for the collection of expected utility preference relations on Y. We will write $R_{\theta_i, \lambda_i} \in \mathcal{R}$ for the preference relation of agent i if his payoff type is θ_i and he has belief $\lambda_i \in \Delta(\Theta_{-i})$ about the types of others:

$$\forall y, y' \in Y: \qquad y R_{\theta_i, \lambda_i} y' \quad \Leftrightarrow \quad \sum_{\theta_{-i} \in \Theta_{-i}} \lambda_i (\theta_{-i}) u_i (y, (\theta_i, \theta_{-i})) \geq \sum_{\theta_{-i} \in \Theta_{-i}} \lambda_i (\theta_{-i}) u_i (y', (\theta_i, \theta_{-i}));$$

and we write P_{θ_i, λ_i} for the strict preference relation corresponding to R_{θ_i, λ_i}.

We make a non-degeneracy assumption on preferences: every agent i, whatever his type $\theta_i \in \Theta_i$ and beliefs $\lambda_i \in \Delta(\Theta_{-i})$, has a strict preference over some pair of outcomes:

Assumption 1 (Non-Degeneracy)
For each i, $\theta_i \in \Theta_i$ and $\lambda_i \in \Delta(\Theta_{-i})$, there exist $x, x' \in X$ such that $x P_{\theta_i, \lambda_i} x'$.

We maintain this assumption throughout the paper.[5] We denote by $\overline{y}$ the *central lottery* which puts equal probability on each of the pure outcomes. Now non-degeneracy implies that every agent i, whatever his type θ_i and beliefs $\lambda_i \in \Delta(\Theta_{-i})$, strictly prefers some pure outcome x to $\overline{y}$; and compactness implies that those strict preferences are uniformly strict:

Lemma 1 *There exists $c > 0$ such that, for each i, $\theta_i \in \Theta_i$ and $\lambda_i \in \Delta(\Theta_{-i})$, there exists $x \in X$ such that*

$$\sum_{\theta_{-i} \in \Theta_{-i}} \lambda_i (\theta_{-i}) u_i (x, (\theta_i, \theta_{-i})) > \sum_{\theta_{-i} \in \Theta_{-i}} \lambda_i (\theta_{-i}) u_i (\overline{y}, (\theta_i, \theta_{-i})) + c.$$

The lemma is proved in appendix 8.1 and we will use c in our later constructions. We will also exploit the existence of an upper bound on payoff differences C which follows immediately from the finiteness of pure outcomes and states:

Lemma 2 *There exists $C > 0$ such that*

$$|u_i (y, \theta) - u_i (y', \theta)| \leq C,$$

for all i, y, y', θ.

2.2 Mechanisms and Solution Concept

A mechanism $\mathcal{M}$ is a collection $((M_i)_{i=1}^I, g)$ where each M_i is finite and $g : M \to Y$. We denote a belief of agent i over the product of payoff type and message spaces of the other agents by $\mu_i \in \Delta(\Theta_{-i} \times M_{-i})$. We consider the process of iteratively eliminating never best responses, without making assumptions on agents' beliefs about others' payoff types.

The set of messages surviving the $k-$th level of elimination for type θ_i in mechanism $\mathcal{M}$ are iteratively defined by

$$S_i^{\mathcal{M}, 0} (\theta_i) = M_i$$

---

[5] Our results can be extended to allow for non-degeneracy, see appendix 8.6 for details.

and, for each $k = 0, 1, ..$

$$S_i^{\mathcal{M},k+1}(\theta_i) = \left\{ m_i \in S_i^{\mathcal{M},k}(\theta_i) \left| \begin{array}{l} \exists\ \mu_i \in \Delta\left(\Theta_{-i} \times M_{-i}\right) \text{ s.t.:} \\[2ex] (1)\ \mu_i\left(\theta_{-i}, m_{-i}\right) > 0 \Rightarrow m_{-i} \in S_{-i}^{\mathcal{M},k}\left(\theta_{-i}\right) \\[2ex] (2)\ m_i \in \underset{m_i'}{\arg\max} \sum_{\theta_{-i}, m_{-i}} \mu_i\left(\theta_{-i}, m_{-i}\right) u_i\left(g\left(m_i', m_{-i}\right), \left(\theta_i, \theta_{-i}\right)\right) \end{array} \right. \right\};$$

we let

$$S_i^{\mathcal{M}}(\theta_i) = \bigcap_{k \geq 0} S_i^{\mathcal{M},k}(\theta_i).$$

We refer to $S_i^{\mathcal{M}}(\theta_i)$ as the *rationalizable* messages of type θ_i of agent i in mechanism $\mathcal{M}$. This incomplete information version of rationalizability was studied in Battigalli (1998) and Battigalli and Siniscalchi (2003). A standard and well known duality argument implies that this solution concept is equivalent to iterated deletion of ex post strictly dominated strategies.

$S_i^{\mathcal{M}}(\theta_i)$ is the set of messages that type θ_i might send consistent with knowing that his payoff type is θ_i, common knowledge of rationality and the set of possible payoff types of the other players, but no restrictions on his beliefs and higher order beliefs about other types. Equivalently, it is the set of messages that might be played in any equilibrium on any type space by a type of player i with payoff type θ_i and any possible beliefs and higher order beliefs about others' payoff types. In section 6.2, we report a formal argument confirming this interpretation. In the body of the paper, we work directly with this solution concept.

2.3   Separability

We will be interested in the set of preferences that an agent might have if his payoff type is θ_i and he knows that the type θ_j of each opponent j belongs to some subset Ψ_j of his possible types Θ_j. Thus writing $\Psi_{-i} = \{\Psi_j\}_{j \neq i}$ for a profile of subsets of i's opponents, we define

$$\mathcal{R}_i\left(\theta_i, \Psi_{-i}\right) = \left\{R \in \mathcal{R}\,\middle|\,R = R_{\theta_i, \lambda_i} \text{ for some } \lambda_i \in \Delta\left(\Psi_{-i}\right)\right\}.$$

Now suppose we observed i's preferences over lotteries and knew that i assigned probability 1 to his opponents' type profile θ_{-i} being an element of Ψ_{-i}, what would we be able to deduce about i's type? We will say that Ψ_{-i} *separates* Ψ_i if - whatever those realized preferences - we could rule out at least one possible type of i.

Definition 1 (Separation)
Type set profile Ψ_{-i} separates Ψ_i if

$$\underset{\theta_i \in \Psi_i}{\cap}\ \mathcal{R}_i\left(\theta_i, \Psi_{-i}\right) = \varnothing.$$

We will be interested in a process by which we iteratively delete type sets of each agent that are separated by some type set profile of his opponents. Thus writing Ξ_i^k for the kth level inseparable sets of player i, we have:

$$\Xi_i^0 = 2^{\Theta_i}, \tag{1}$$

and

$$\Xi_i^{k+1} = \left\{ \Psi_i \in \Xi_i^k \mid \Psi_{-i} \text{ does not separate } \Psi_i, \text{ for some } \Psi_{-i} \in \Xi_{-i}^k \right\}, \tag{2}$$

and a (finite) limit type set profile is defined by:

$$\Xi_i^* = \bigcap_{k \geq 0} \Xi_i^k. \tag{3}$$

Finally, we say that a pair of types are pairwise inseparable if they cannot be iteratively separated in this way:

Definition 2 (Pairwise Inseparability)
Types θ_i and θ_i' are pairwise inseparable (written $\theta_i \sim \theta_i'$) if $\left\{ \theta_i, \theta_i' \right\} \in \Xi_i^*$.

Note that the relation $\sim$ is reflexive and symmetric by construction, but it is not necessarily transitive. The following "fixed point" characterization of pairwise inseparability will be useful in the analysis that follows. Let $\Xi = (\Xi_i)_{i=1}^I \in \times_{i=1}^I 2^{\Theta_i}$ be a profile of type sets for each agent.

Definition 3 (Mutual Inseparability)
Ξ is mutually inseparable if, for each i and $\Psi_i \in \Xi_i$, there exists $\Psi_{-i} \in \Xi_{-i}$ such that Ψ_{-i} does not separate Ψ_i.

Lemma 3 *Types θ_i and θ_i' are pairwise inseparable if and only if there exists mutually inseparable $\Xi = (\Xi_i)_{i=1}^I$ and $\Psi_i \in \Xi_i$ with $\left\{ \theta_i, \theta_i' \right\} \subseteq \Psi_i$.*

if. Suppose there exists $\widehat{\Xi} = \left(\widehat{\Xi}_i \right)_{i=1}^I$ and $\Psi_i \in \widehat{\Xi}_i$ with $\left\{ \theta_i, \theta_i' \right\} \subseteq \Psi_i$. We claim that

$$\left\{ \Psi_i \, \middle| \, \Psi_i \subseteq \Psi_i' \text{ and } \Psi_i' \in \widehat{\Xi}_i \text{ for some } \Psi_i' \right\} \subseteq \Xi_i^k$$

for each $k = 0, 1, \dots$. The claim holds for $k = 0$ by definition. Suppose the claim holds for arbitrary k and suppose that $\Psi_i \subseteq \Psi_i'$ and $\Psi_i' \in \widehat{\Xi}_i$. Because $\widehat{\Xi}$ is mutually inseparable, there exists $\Psi_{-i} \in \widehat{\Xi}_{-i} \subseteq \Xi_{-i}^k$ such that Ψ_{-i} does not separate Ψ_i'. By the definition of separation, since $\Psi_i \subseteq \Psi_i'$, Ψ_{-i} does not separate Ψ_i. So $\Psi_i \in \Xi_i^{k+1}$ and

$$\left\{ \theta_i, \theta_i' \right\} \subseteq \Psi_i \in \Xi_i^* = \bigcap_{k \geq 0} \Xi_i^k.$$

[only if] Observe that $\Xi_i^{k+1} \subseteq \Xi_i^k$ for each $k = 0, 1, \dots$ by construction. Thus $(\Xi_i^*)_{i=1}^I$ is mutually inseparable. Thus if $\theta_i \sim \theta_i'$, there exists mutually inseparable Ξ^* with $\left\{ \theta_i, \theta_i' \right\} \in \Xi_i^*$. ∎

3   An Environment with Interdependent Values for a Single Good

We consider a quasi-linear environment with a single good with interdependent values to illustrate the notion of separability. There are I agents and agent i's payoff type is $\theta_i \in [0, 1]$. If the type profile is θ, agent i's valuation of an object is given by:

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

with $\gamma \in \mathbb{R}_+$. The parameter γ measures the amount of interdependence in valuations: the case of private values is given by $\gamma = 0$ and the case of pure common values is $\gamma = 1$. The net utility of agent i depends on his probability y_i of receiving the object and the monetary transfer t_i:

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

We determine the conditions for separability of types in this preference environment.[6]

Type set profile Ψ_{-i} separates Ψ_i if, knowing i's preferences and knowing that he is sure that others' type profile is Ψ_{-i}, we can always rule out some θ_i. In this example, because the utility function $u_i(\cdot)$ is linear in the monetary transfer for all types and all agents, separability must come from different valuations of the object. For given type set profile Ψ_{-i} of all but i, we can identify the set of possible (expected) valuations of agent i with type θ_i by writing:

$$\begin{aligned} V_i(\theta_i, \Psi_{-i}) &= \left\{ v_i \in \mathbb{R}_+ \;\middle|\; \exists \lambda_i \in \Delta(\Psi_{-i}) \text{ s.t. } v_i = \theta_i + \gamma \sum_{\theta_{-i} \in \Psi_{-i}} \lambda_i(\theta_{-i}) \sum_{j \neq i} \theta_j \right\} \\ &= \left[\theta_i + \gamma \sum_{j \neq i} \min \Psi_j, \; \theta_i + \gamma \sum_{j \neq i} \max \Psi_j \right]. \end{aligned} \tag{4}$$

Now Ψ_{-i} separates Ψ_i if and only if

$$\bigcap_{\theta_i \in \Psi_i} V_i(\theta_i, \Psi_{-i}) = \varnothing.$$

This is equivalent to requiring that

$$V_i(\max \Psi_i, \Psi_{-i}) \cap V_i(\min \Psi_i, \Psi_{-i}) = \varnothing.$$

By (4), this will hold if and only if

$$\max \Psi_i + \gamma \sum_{j \neq i} \min \Psi_j > \min \Psi_i + \gamma \sum_{j \neq i} \max \Psi_j.$$

We can rewrite the inequality as

$$\max \Psi_i - \min \Psi_i > \gamma \sum_{j \neq i} (\max \Psi_j - \min \Psi_j).$$

Thus Ψ_{-i} separates Ψ_i if and only if the difference between the smallest and the largest element in the set Ψ_i is larger than the weighted sum of the differences of the smallest and the largest element in the remaining sets Ψ_j for all $j \neq i$. Conversely, Ψ_{-i} does not separate Ψ_i if the above inequality is reversed, i.e.,

$$\max \Psi_i - \min \Psi_i \leq \gamma \sum_{j \neq i} (\max \Psi_j - \min \Psi_j). \tag{5}$$

---

[6] The example has a continuum of types and a continuum of deterministic monetary allocations. In contrast, the general model is defined for a finite number of types and pure outcomes. We could rewrite the example and the corresponding results without loss in the finite setting. With a finite model, integer problems would need to be taken into account in deriving the inequalities to make sure that the process of elimination proceeds. In particular, the exact value of the critical threshold for interdependence, to be determined below, would depend on the size of the grid. Naturally, as the grid becomes finer, the critical thresholds converge to the ones of the continuous example here.

Now we can identify the kth level inseparable sets, described in (1)-(3), for our example. We have

$$\Xi_i^0 = 2^{[0,1]}$$

and, by (5),

$$\Xi_i^{k+1} = \left\{ \Psi_i \in \Xi_i^k \,\bigg|\, \max \Psi_i - \min \Psi_i \leq \gamma \sum_{j \neq i} \max_{\Psi_j \in \Xi_j^k} (\max \Psi_j - \min \Psi_j) \right\},$$

Now by induction, we have that

$$\Xi_i^{k+1} = \left\{ \Psi_i \,\bigg|\, \max \Psi_i - \min \Psi_i \leq (\gamma (I-1))^k \right\}.$$

Thus if $\gamma (I-1) < 1$, Ξ_i^* consists of singletons, $\Xi_i^* = (\{\theta_i\})_{\theta_i \in [0,1]}$, while if $\gamma (I-1) \geq 1$, Ξ_i^* consists of all subsets, $\Xi_i^* = 2^{[0,1]}$.

Thus if $\gamma < \frac{1}{I-1}$, so that interdependence is not too large, every distinct pair of types are pairwise separable. If $\gamma \geq \frac{1}{I-1}$, every pair of types are pairwise inseparable. We note that the linear structure of the valuations $v_i(\cdot)$ leads to the strong converse result. But the example illustrates the general principle that pairwise separability corresponds to not too much interdependence.[7]

Our later results will show that if $\gamma \geq \frac{1}{I-1}$, no social choice function (except for a constant one) is robustly virtually implementable; but if $\gamma < \frac{1}{I-1}$, any ex post incentive compatible allocation can be robustly virtually implemented. One can construct generalized VCG payments such that efficient allocation is ex post incentive compatible in this environment if $\gamma \leq 1$ (Cremer and McLean (1988), Dasgupta and Maskin (2000)). Thus the efficient allocation is robustly virtually implementable if and only if $\gamma < \frac{1}{I-1}$. We return to this example, after describing our results for general environments, in section 6.4.[8]

4   Strategic Distinguishability

4.1   Definition

Two payoff types are strategically distinguishable if there exists a mechanism where the rationalizable actions of those payoff types are disjoint; thus they are strategically indistinguishable if they have a rationalizable action in common in every mechanism.

Definition 4 (Strategically Indistinguishable)
Types θ_i and θ_i' are strategically indistinguishable if $S^{\mathcal{M}}(\theta_i) \cap S^{\mathcal{M}}(\theta_i') \neq \varnothing$ for every $\mathcal{M}$.

We have two reasons for being interested in characterizing strategic distinguishability.

---

[7] This observation can be straightforwardly extended to $\gamma < 0$, i.e., negative interdependence in preferences; now if $|\gamma| < \frac{1}{I-1}$, all distinct pairs of types are pairwise separable; if $|\gamma| \geq \frac{1}{I-1}$, all pairs of types are pairwise inseparable.

[8] In fact, robust virtual implementation is possible in the direct mechanism. Chung and Ely (2001) first identified this condition as sufficient for (exact) implementation of the efficient outcome in iterated deletion of weakly dominated strategies. We discuss the relation in section 6.3.

First, assumptions of interdependence of preferences for informational or psychological reasons are prevalent in many areas of economics, but there has been little attempt to identify what are the observable implications of such preferences. A classic and well developed "revealed preference" theory underlies economists' way of understanding individual choice. An analogous strategic revealed preference approach to understanding interdependent preferences is required. We believe that our characterization of "strategic distinguishability" delivers some clean insights about strategic revealed preferences and may be a useful component of a more general approach to the question.

Second, our characterization of strategic distinguishability turns out to be the key step in our characterization of robust virtual implementation, described in the next section.

We say that two types are strategically indistinguishable if they do not have any rationalizable messages in common. A more demanding relation between types would be to require that they have the same rationalizable actions: thus types θ_i and θ_i' are *strategically equivalent* if $S^{\mathcal{M}}(\theta_i) = S^{\mathcal{M}}(\theta_i')$ for every $\mathcal{M}$. Two types are strategically indistinguishable if they might behave the same. They are strategically equivalent if nothing they might rationally do could distinguish them. We briefly discuss the alternative, much stronger notion of strategic equivalence, in section 6.1.

As we noted in the introduction, Gul and Pesendorfer (2005) pioneered the study of the revealed preference implications of interdependent preferences. We follow them in characterizing when there are distinct observational implications of two interdependent preference types, and our characterization, like theirs, is based on iterated statements about conditional preferences. We do not follow them in constructing a canonical universal type space based on our notion of distinct observational implications, although this would be an interesting exercise. There are a number of ways in which our frameworks differ: we explicitly incorporate uncertainty with expected utility preferences about others' types; and our notion of revealed (interdependent) preferences is based on rationalizable strategic behavior, while theirs is based on sincere announcements. Nonetheless, we note in section 6.1 that a property closely related to their validity condition is sufficient to ensure that distinct types are not strategically equivalent.

4.2   Main Result

The main result of this paper is a characterization of strategic indistinguishability.

Theorem 1 (Equivalence)
Types θ_i and θ_i' are strategically indistinguishable if and only if they are pairwise inseparable.

This result will be proved in two parts. First, proposition 1 shows that under any finite mechanism, if θ_i and θ_i' are pairwise inseparable, then the intersection of the set of rationalizable messages for θ_i and θ_i' will always be non-empty. This observation follows easily from our definitions.

Proposition 1

If θ_i and θ_i' are pairwise inseparable ($\theta_i \sim \theta_i'$), then $S_i^{\mathcal{M}}(\theta_i) \cap S_i^{\mathcal{M}}(\theta_i') \neq \varnothing$ in any mechanism $\mathcal{M}$.

Proof. By lemma 3, if $\theta_i \sim \theta_i'$, there exists mutually inseparable Ξ with $\{\theta_i, \theta_i'\} \subseteq \Psi_i^* \in \Xi_i$.

Now fix any mechanism $\mathcal{M}$. We will show, by induction on k, that for each k, i and $\Psi_i \in \Xi_i$, there exists $m_i^k(\Psi_i) \in M_i$ such that $m_i^k(\Psi_i) \in S_i^{\mathcal{M},k}(\widetilde{\theta}_i)$ for each $\widetilde{\theta}_i \in \Psi_i$. This is true by definition for $k = 0$. Suppose that it is true for k. Now fix any i and $\Psi_i \in \Xi_i$. Since Ξ is mutually inseparable, there exists $\Psi_{-i} \in \Xi_{-i}$, R and, for each $\widetilde{\theta}_i \in \Psi_i$, $\lambda_i^{\widetilde{\theta}_i} \in \Delta(\Psi_{-i})$ such that $R_{\widetilde{\theta}_i, \lambda_i^{\widetilde{\theta}_i}} = R$. Now let $m_i^{k+1}(\Psi_i)$ be any optimal message of agent i when he believes that his opponents will sent message profile $m_{-i}^k(\Psi_{-i})$ with probability 1 and has beliefs $\lambda_i^{\widetilde{\theta}_i}$ about the type profile of his opponents, i.e.,

$$m_i^{k+1}(\Psi_i) \in \arg\max_{m_i'} \sum_{\theta_{-i}} \lambda_i^{\widetilde{\theta}_i}(\theta_{-i}) u_i\left(g\left(m_i', m_{-i}^k(\Psi_{-i})\right), \left(\widetilde{\theta}_i, \theta_{-i}\right)\right).$$

By construction, $m_i^{k+1}(\Psi_i) \in S_i^{\mathcal{M},k+1}(\widetilde{\theta}_i)$ for all $\widetilde{\theta}_i \in \Psi_i$.

By the finiteness of the mechanism, there exists K such that $S_i^{\mathcal{M},k}(\widetilde{\theta}_i) = S_i^{\mathcal{M}}(\widetilde{\theta}_i)$ for all i, $\widetilde{\theta}_i$ and $k \geq K$. Thus for each $\Psi_i \in \Xi_i$, there exists $m_i(\Psi_i) \in M_i$ such that $m_i(\Psi_i) \in S_i^{\mathcal{M}}(\widetilde{\theta}_i)$ for each $\widetilde{\theta}_i \in \Psi_i^*$. Thus there exists $m_i \in S_i^{\mathcal{M}}(\theta_i) \cap S_i^{\mathcal{M}}(\theta_i')$. ∎

The second part of the theorem's proof is the converse result.

Proposition 2 (Existence of Maximally Revealing Mechanism)

There exists $\mathcal{M}^*$ such that $\theta_i \nsim \theta_i' \Rightarrow S_i^{\mathcal{M}^*}(\theta_i) \cap S_i^{\mathcal{M}^*}(\theta_i') = \varnothing$.

Propositions 1 and 2 immediately imply theorem 1. Proposition 2 is proved by the explicit construction of a mechanism which will lead every pair of distinguishable types to choose different messages. We refer to the specific mechanism as the "maximally revealing mechanism", and spend the rest of this section describing its construction and finding its properties.

4.3   The Maximally Revealing Mechanism:

We will construct a mechanism that will work for any environment. In the canonical mechanism, each agent is given K simultaneous opportunities to select a preferred allocation from a given "test set" of allocations. For each opportunity k to select a preferred allocation, with $k = 1, ..., K$, the agent is asked to report a profile of possible choices by the remaining agents in the opportunities preceding the k-th opportunity. If the report of the agent at opportunity k matches the choices of the other agents in the opportunities below k, then he will be given the right to choose a preferred allocation. On the other hand, if his report fails to replicate the choices of the other agents in the opportunities before k, then the designer will simply select the central lottery $\bar{y}$. While the mechanism is entirely static, it requires each agent to make a series of choices, each one contingent on the choices of the other agents. In particular, by asking the agent at opportunity k to match his report with the choices of the other agents at the opportunities before k, we introduce an inductive structure into the series of choices by each agent. We therefore refer to the k-th opportunity as the k-th stage or k-th step of the mechanism even though the mechanism itself is entirely static.

The central aspect of the inductive structure of the choice mechanism is that it allows us to analyze the behavior of the agent in the mechanism in terms of the iterative elimination of dominated strategies. The precise construction of the choice mechanism is based on two central concepts, the notion of a test set and the notion of an augmentation of a given mechanism. The test set will give each agent a finite set of choices and the choice behavior by the agent allows us to distinguish between different types of the agent. The construction of the set of test allocations relies on a few critical implications of our notion of separation. In turn, the notion of an augmentation permits us to show that we can always construct a more informative mechanism on the basis of a given mechanism.

4.3.1   Test Allocations

The canonical mechanism will ask each agent to make a series of binary choices between the central lottery $\bar{y}$ and a specific lottery y from the test set. If the test set is to be successful in eliciting the private information from agent i, then the test set should contain a sufficient number of allocations such that for every type θ_i and every belief λ_i of agent i there exists some allocation y that is strictly preferred to the central lottery $\bar{y}$.

Lemma 4 (Duality)
Type set profile Ψ_{-i} separates Ψ_i if and only if there exists $\widetilde{y} : \Psi_i \to Y$ such that

$$\sum_{\theta_i \in \Psi_i} (\widetilde{y}(\theta_i) - \overline{y}) = 0, \tag{6}$$

and

$$\widetilde{y}(\theta_i) \, P_{\theta_i, \lambda_i} \, \overline{y}, \tag{7}$$

for all $\theta_i \in \Psi_i$ and all $\lambda_i \in \Delta(\Psi_{-i})$.

This result says that for each $\theta_i \in \Psi_i$, we can identify a direction in lottery space, $\widetilde{y}(\theta_i) - \overline{y}$, that agent i likes whatever his beliefs about Ψ_{-i}, such that the sum of those changes add up to zero. The proof of the lemma appears in appendix 8.1. It follows from the following duality result in Samet (1998):

Proposition 3 (Samet (1998))
Let $V_1, ..., V_L$ be closed, convex, subsets of the N-dimensional simplex Δ^N. These sets have an empty intersection if and only if there exist $z_1, ..., z_L \in \mathbb{R}^N$ such that

$$\sum_{l=1}^{L} z_l = 0,$$

and

$$v \cdot z_l > 0, \text{ for each } l = 1, ..., L \text{ and } v \in V_l.$$

To obtain some intuition for this result, note that it was introduced in Samet (1998) in order to provide a simple proof of the observation that asymmetrically informed agents will trade against each other if and only if they do not share a common prior, i.e., their posterior beliefs could not have been derived by updating a

common prior.[9] Suppose that there are N states and L agents. Each agent l observes one of a collection of signals about the true state. Each signal leads him to have a posterior $v \in \Delta^N$ over the states. Let V_l be the convex hull of his set of possible posteriors. Notice that V_l represents the set of prior beliefs he might have held over the state space before observing his signal. Thus posterior beliefs are consistent with a common prior if and only if the intersection of the V_l sets is non-empty. Now consider a multilateral bet specifying that if state n was realized, agent l will receive payment z_{ln} where the total payments sum to zero:

$$\sum_{l=1}^{L} z_{ln} = 0 \text{ for all } n.$$

Writing $z_l = (z_{ln})_{n=1}^{N}$, we then have

$$\sum_{l=1}^{L} z_l = 0.$$

There exists such a bet where every agent has a strictly positive expected value from accepting the bet conditional on every signal if $v \cdot z_l > 0$, for each $l = 1, ..., L$ and $v \in V_l$.

We now use lemma 4 to show how, if Ψ_{-i} separates Ψ_i, we can construct a *finite* set of lotteries $\widetilde{Y}_i(\Psi_i, \Psi_{-i}) \subseteq Y$ such that knowing that agent i knows that his opponent's type is in Ψ_{-i} and knowing his preferences on $\widetilde{Y}_i(\Psi_i, \Psi_{-i})$ will always be enough to rule out at least one type in Ψ_i for agent i.

Lemma 5 *If Ψ_{-i} separates Ψ_i, then there exists a finite set $\widetilde{Y}_i(\Psi_i, \Psi_{-i}) \subseteq Y$, such that for each $\theta_i \in \Psi_i$ and $\lambda_i \in \Delta(\Psi_{-i})$, there exists $y \in \widetilde{Y}_i(\Psi_i, \Psi_{-i})$ such that*

$$\overline{y} P_{\theta_i, \lambda_i} y, \tag{8}$$

and for some $\theta_i' \in \Psi_i$,

$$y P_{\theta_i', \lambda_i'} \overline{y}, \tag{9}$$

for all $\lambda_i' \in \Delta(\Psi_{-i})$.

Proof. By lemma 4, there exists $\widetilde{y} : \Psi_i \to Y$ such that

$$\sum_{\theta_i \in \Psi_i} (\widetilde{y}(\theta_i) - \overline{y}) = 0,$$

and

$$\widetilde{y}(\theta_i) P_{\theta_i, \lambda_i} \overline{y} \quad \text{for all } \theta_i \in \Psi_i \text{ and } \lambda_i \in \Delta(\Psi_{-i}).$$

Let $\widetilde{Y}_i(\Psi_i, \Psi_{-i}) = \{\widetilde{y}(\theta_i)\}_{\theta_i \in \Psi_i}$. Fix $\theta_i \in \Psi_i$ and $\lambda_i \in \Delta(\Psi_{-i})$. Write $\widetilde{Y}_i(\Psi_i, \Psi_{-i}) = \{y^1, ..., y^K\}$, with $y^1 = \widetilde{y}(\theta_i)$. Let $\overline{y}^0 = \overline{y}$ and

$$\overline{y}^l = \overline{y} + \varepsilon \sum_{\kappa=1}^{l} (y^\kappa - \overline{y}),$$

with $\varepsilon > 0$ chosen sufficiently small such that $\overline{y}^l \in Y$ for all $l = 1, ..., K$. We know $\overline{y}^1 P_{\theta_i, \lambda_i} \overline{y}^0$. Suppose $\overline{y}^{l+1} R_{\theta_i, \lambda_i} \overline{y}^l$ for all $l = 1, ..., K-1$. By transitivity, this would imply that:

$$\overline{y}^K P_{\theta_i, \lambda_i} \overline{y}^0.$$

---

[9] This converse to the no trade theorem was originally proved by Morris (1994), by a more indirect duality argument.

But $\overline{y}^K = \overline{y}^0$, so we have a contradiction. We conclude that, for some $l = 1, ..., K - 1$, $\overline{y}^l \ P_{\theta_i, \lambda_i} \ \overline{y}^{l+1}$. This implies that there exists θ_i' such that

$$\overline{y} \ P_{\theta_i, \lambda_i} \ y \left(\theta_i' \right).$$

Since

$$y \left(\theta_i' \right) \ P_{\theta_i', \lambda_i'} \ \overline{y} \ \text{ for all } \ \lambda_i' \in \Delta \left(\Psi_{-i} \right),$$

the inequalities (8) and (9) are established. ∎

Now we will construct a large enough finite set of lotteries (the "test set") such that knowing just an agent's most preferred outcome on the test set will always reveal enough information about his preferences to separate out a type, if it is possible to do so.

For any set of lotteries $\widehat{Y} \subseteq Y$, let $B_i^{\widehat{Y}} (\theta_i, \lambda_i)$ be agent i's most preferred lotteries in the set $\widehat{Y}$ if he has payoff type θ_i and (with a minor abuse of notation) let $B_i^{\widehat{Y}} (\theta_i, \Psi_{-i})$ be agent i's possible most preferred lotteries if he has payoff type θ_i and assigns probability 1 to his opponents having types in Ψ_{-i}, so that

$$B_i^{\widehat{Y}} (\theta_i, \lambda_i) = \left\{ y \in \widehat{Y} \ \middle| \ y \ R_{\theta_i, \lambda_i} \ y' \text{ for all } y' \in \widehat{Y} \right\},$$

and by extension:

$$B_i^{\widehat{Y}} (\theta_i, \Psi_{-i}) = \bigcup_{\lambda_i \in \Delta(\Psi_{-i})} B_i^{\widehat{Y}} (\theta_i, \lambda_i).$$

Proposition 4 (Existence of Finite Test Set)

There exists a finite test set $Y^* \subseteq Y$ such that:

1. *for each i, θ_i and $\lambda_i \in \Delta (\Theta_{-i})$, $B_i^{Y^*} (\theta_i, \lambda_i) \neq Y^*$;*

2. *for each i, Ψ_i and Ψ_{-i}, if Ψ_{-i} separates Ψ_i, then for each $\theta_i \in \Psi_i$ and $\lambda_i \in \Delta (\Psi_{-i})$, there exists $\theta_i' \in \Psi_i$ such that*
$$B_i^{Y^*} (\theta_i, \lambda_i) \cap B_i^{Y^*} (\theta_i', \Psi_{-i}) = \varnothing.$$

Our proof is constructive. We first construct a set $\widetilde{Y}$ consisting of the degenerate lotteries X and the $\widetilde{Y}_i (\Psi_i, \Psi_{-i})$ sets constructed in lemma 5, for every triple (i, Ψ_i, Ψ_{-i}) with Ψ_{-i} separating Ψ_i. Knowing an agent's ranking of each element of $\widetilde{Y}$ relative to the central lottery $\overline{y}$ would reveal all the information we need to extract. In order to extract this information in a single choice, we let the agent pick $f : \widetilde{Y} \to \{0, 1\}$. For each $y \in \widetilde{Y}$, y is chosen with probability $1/\widetilde{Y}$ if $f(y) = 1$, otherwise the central lottery $\overline{y}$ is chosen. We let Y^* be the set of all such lotteries. Now observing an agent's most preferred outcome in Y^* reveals his binary preference between $\overline{y}$ and each element of $\widetilde{Y}$. Since $\widetilde{Y}$ contains each $\widetilde{Y}_i (\Psi_i, \Psi_{-i})$, this will ensure part (2). Since $\widetilde{Y}$ contains degenerate lotteries, the agents will have strict preferences ensuring part (1).

Proof. Let

$$\widetilde{Y} = X \ \cup \bigcup_{\{(i,\Psi_i,\Psi_{-i})|\Psi_{-i} \text{ separates } \Psi_i\}} \widetilde{Y}_i\left(\Psi_i, \Psi_{-i}\right).$$

Now for any $f : \widetilde{Y} \rightarrow \{0,1\}$, let y_f be the lottery obtained by picking an element $y \in \widetilde{Y}$ with uniform probability and then choosing lottery y if $f(y) = 1$ and $\overline{y}$ if $f(y) = 0$. Thus we define:

$$y_f \equiv \overline{y} + \frac{1}{\#\widetilde{Y}} \sum_{y \in \widetilde{Y}} f(y)(y - \overline{y}).$$

Let Y^* be the set of such lotteries, i.e.,

$$Y^* = \left\{ y \in Y \left| \exists f : \widetilde{Y} \rightarrow \{0,1\} \text{ such that } y = y_f \right. \right\}.$$

To prove part (1) of the proposition, fix any $\theta_i \in \Theta_i$ and $\lambda_i \in \Delta\left(\Theta_{-i}\right)$. By lemma 1, there exists $x \in X \subseteq \widetilde{Y}$ such that $xP_{\theta_i,\lambda_i}\overline{y}$; now let

$$f^0(y) = 0, \text{ for all } y \in \widetilde{Y},$$

and

$$f^*(y) = \begin{cases} 0, \text{ if } y \neq x \\ 1, \text{ if } y = x \end{cases}$$

So we can write:

$$y_{f^0} = \overline{y}, \quad y_{f^*} = \overline{y} + \frac{1}{\#\widetilde{Y}}(x - \overline{y})$$

and so $y_{f^0} \notin B_i^{Y^*}\left(\theta_i, \lambda_i\right)$.

To prove part (2) of the proposition, suppose that Ψ_{-i} separates Ψ_i. Fix $\theta_i \in \Psi_i$ and $\lambda_i \in \Delta\left(\Psi_{-i}\right)$. By lemma 5, there exists $y \in \widetilde{Y}_i\left(\Psi_i, \Psi_{-i}\right)$ and $\theta'_i \in \Psi_i$ such that $\overline{y} \ P_{\theta_i,\lambda_i} \ y$ and $y \ P_{\theta'_i,\lambda'_i} \ \overline{y}$ for all $\lambda'_i \in \Delta\left(\Psi_{-i}\right)$. So

$$y_f \in B_i^{Y^*}\left(\theta_i, \lambda_i\right) \Rightarrow f(y) = 0,$$

while

$$y_f \in B_i^{Y^*}\left(\theta'_i, \Psi_i\right) \Rightarrow f(y) = 1,$$

and so

$$B_i^{Y^*}\left(\theta_i, \lambda_i\right) \cap B_i^{Y^*}\left(\theta'_i, \Psi_i\right) = \varnothing,$$

which establishes the result. ∎

4.3.2 Uniformly Worse Responses and Augmentation

In this subsection, we report two results that we will use in our analysis. The routine proofs are reported in Appendix 8.1. First, we note that for any fixed finite mechanism $\mathcal{M}$, when we iteratively delete messages that are not best responses, they are uniformly worse responses, i.e., there exists $\eta_{\mathcal{M}} > 0$ such that each of those deleted messages is not even an $\eta_{\mathcal{M}}$-best response.

Lemma 6 (Uniformly Worse Responses)

For any mechanism $\mathcal{M}$, there exists $\eta_{\mathcal{M}} > 0$ such that if $m_i \in S_i^{\mathcal{M},k}(\theta_i)$, $m_i \notin S_i^{\mathcal{M},k+1}(\theta_i)$ and $\mu_i \in \Delta(\Theta_{-i} \times M_{-i})$ satisfies

$$\mu_i(\theta_{-i}, m_{-i}) > 0 \Rightarrow m_j \in S_j^{\mathcal{M},k}(\theta_j) \text{ for each } j \neq i$$

then there exists $\overline{m}_i$ such that

$$\sum_{\theta_{-i}, m_{-i}} \mu_i(\theta_{-i}, m_{-i}) u_i(g^*(\overline{m}_i, m_{-i}), (\theta_i, \theta_{-i})) > \sum_{\theta_{-i}, m_{-i}} \mu_i(\theta_{-i}, m_{-i}) u_i(g^*(m_i, m_{-i}), (\theta_i, \theta_{-i})) + \eta_{\mathcal{M}}.$$

Second, we use the uniform lower bound in stating a key result about "augmenting" mechanisms. We use this "augmentation lemma" in the construction of both the maximally revealing mechanism (in this section) and the canonical mechanism for robust virtual implementation (in the next section). For each player i, fix finite message sets M_i^0 and M_i^1 and let $M_i = M_i^0 \times M_i^1$. Fix $g^0 : M^0 \to Y$, $g^1 : M^1 \to Y$ and $g^+ : M \to Y$.

Lemma 7 (Augmentation)

Fix $\pi^0, \pi^1, \pi^+ \geq 0$, let $g : M \to Y$ be defined by

$$g(m) = \pi^0 g^0(m^0) + \pi^1 g^1(m^1) + \pi^+ g^+(m),$$

and consider the mechanism

$$\mathcal{M}^0 = \left(\left(M_i^0\right)_{i=1}^I, g^0\right),$$

and the augmented mechanism

$$\mathcal{M} = \left(\left(M_i\right)_{i=1}^I, g\right).$$

If $\pi^+ C \leq \pi^0 \eta_{\mathcal{M}^0}$, then

$$\left(m_i^0, m_i^1\right) \in S_i^{\mathcal{M}}(\theta_i) \Rightarrow m_i^0 \in S_i^{\mathcal{M}^0}(\theta_i).$$

The lemma states that if the weight put on the original payoff function g^0 in the augmented mechanism (π^0) is much larger than the weight put on any other component of the mechanism where m^0 effects the allocation (π^+), then any rationalizable message in the augmented mechanism must entail sending a message m_i^0 that was rationalizable in the original mechanism.

We are now prepared to provide an explicit construction of the maximally revealing mechanism.

4.3.3 Construction of the Maximally Revealing Mechanism

The maximally revealing mechanism offers each agent i a series of K opportunities to select a preferred allocation from the test set Y^*. The set of messages for each agent in the maximally revealing mechanism is defined as follows. Let $M_i^0 = \{\overline{m}_i^0\}$ and inductively define

$$M_i^{k+1} = M_i^k \times M_{-i}^k \times Y^*.$$

Thus $M_i^0 = \{\overline{m}_i^0\}$, $M_i^1 = \{\overline{m}_i^0\} \times M_{-i}^0 \times Y^*$, $M_i^2 = \{\overline{m}_i^0\} \times M_{-i}^0 \times Y^* \times M_{-i}^1 \times Y^*$, and so on. The message m_i^{k+1} of agent i in stage $k+1$ thus reiterates his message from step k and reports a message profile of the

remaining agents in the preceding stage k. Due to the inductive structure of the messages, we can write a typical element m_i^k of M_i^k as a list of the form

$$m_i^k = \left\{ m_i^0, r_i^1, y_i^1, r_i^2, y_i^2, ..., r_i^k, y_i^k \right\},$$

with $m_i^0 = \overline{m}_i^0$ and each $r_i^k \in M_{-i}^{k-1}$ and each $y_i^k \in Y^*$. The entry r_i^k constitutes the report of agent i regarding the message of the other agents in the previous round $k-1$. The message set of agent i is then given by M_i^K.

The outcome function in the revealing mechanism is given by

$$g^{K,\varepsilon}\left(m\right) = \overline{y} + \frac{1 - \varepsilon^K}{1 - \varepsilon} \frac{1}{I} \left(\sum_{k=1}^{K} \varepsilon^{k-1} \sum_{i=1}^{I} \mathbb{I}\left(r_i^k, m_{-i}^{k-1}\right) \left(y_i^k - \overline{y}\right) \right),$$

for some $\varepsilon > 0$ and where $\mathbb{I}$ is the indicator function,

$$\mathbb{I}\left(r_i^k, m_{-i}^{k-1}\right) = \left\{ \begin{array}{l} 1, \text{ if } r_i^k = m_{-i}^{k-1} \\ 0, \text{ otherwise} \end{array} \right. .$$

For a given $\varepsilon > 0$ and positive integer K, we refer to the (K, ε) revealing mechanism as

$$\mathcal{M}^{K,\varepsilon} = \left(M^K, g^{K,\varepsilon} \right). \tag{10}$$

In words, the mechanism has K stages. In each stage k, an agent is asked to announce a stage $k-1$ message profile of messages he thinks his opponents might have sent and - with positive probability - gets to pick a lottery from Y^*. Lotteries from early rounds are much more likely to be chosen than lotteries from later rounds. We can now analyze how the series of messages can iteratively and interactively identify the types of each agent.

4.3.4 Properties of the Maximally Revealing Mechanism

For small $\varepsilon > 0$, an agent's choice of message at the kth round will be independent of what messages he thinks others will send at round k and higher and thus also independent of K, the total number of rounds of messages that will be sent. Our strategy for characterizing rationalizable messages is to first find the set $\overline{\Theta}_i^1\left(m_i^1\right)$ of types of player i who could possibly send first round message m_i^1. Since we will ignore later rounds, this will be independent of ε and K. Taking these sets as given, we will then find the set $\overline{\Theta}_i^2\left(m_i^2\right)$ of types of player i who could possibly send second round message m_i^2. And so on. We will end up with an inductive characterization of the set $\overline{\Theta}_i^k\left(m_i^k\right)$ of types of player i who could possibly send kth round message m_i^k. We will then appeal to lemma 7 to verify that we can ignore later rounds for sufficiently small $\varepsilon > 0$. Finally, we appeal to proposition 4 to verify that - for sufficiently small $\varepsilon > 0$ and sufficiently large K - the richness of the test set ensures that any pair of mutually separable types are sending distinct messages in the (K, ε) revealing mechanism.

Heuristic We initialize

$$\overline{\Theta}_i^0 \left(\overline{m}_i^0 \right) = \Theta_i,$$

and inductively define $\overline{\Theta}_i^k \left(m_i^k \right)$ as follows:

$$\overline{\Theta}_i^k \left(m_i^k \right) = \overline{\Theta}_i^k \left(\left(m_i^{k-1}, r_i^k, y_i^k \right) \right) = \left\{ \theta_i \left| \begin{array}{l} \text{(i)} \ \ \theta_i \in \overline{\Theta}_i^{k-1} \left(m_i^{k-1} \right); \\ \text{(ii)} \ \ \overline{\Theta}_{-i}^{k-1} \left(r_i^k \right) \neq \varnothing; \text{ and} \\ \text{(iii)} \ \ y_i^k \in B_i^{Y^*} \left(\theta_i, \overline{\Theta}_{-i}^{k-1} \left(r_i^k \right) \right). \end{array} \right. \right\}, \tag{11}$$

The set $\overline{\Theta}_i^k \left(m_i^k \right)$ identifies the set of types of agent i for whom the message $m_i^k = \left(m_i^{k-1}, r_i^k, y_i^k \right)$ could be a "best response" in stage k, given that the messages in the previous rounds encoded a "best response" in the test set Y^*. The analysis of the limit behavior of $\overline{\Theta}_i^k \left(m_i^k \right)$ is heuristic in the sense that the inductive process assumes the properties (ii) and (iii) in (11). In particular, it is simply assumed that agent i in round k announces a past message profile of the remaining agents which could have been sent by some type profile of the other agents, and it is simply assumed that agent i will select an allocation which is a best response to some belief in stage k.

Limit Argument We now show that these choices are indeed the result of iteratively elimination of strictly dominated strategies. More precisely, we verify that $\overline{\Theta}_i^k \left(m_i^k \right)$ is an upper bound on the set of types who could send kth round message m_i^k in any $\mathcal{M}^{k,\varepsilon}$ for sufficiently small ε.

Lemma 8 (Limit)
For each k, there exists $\overline{\varepsilon} > 0$ such that

$$\left\{ \theta_i \in \Theta_i \left| m_i^k \in S^{\mathcal{M}^{k,\varepsilon}} \left(\theta_i \right) \right. \right\} \subseteq \overline{\Theta}_i^k \left(m_i^k \right).$$

for all $\varepsilon \leq \overline{\varepsilon}$ and $m_i^k \in M_i^k$.

Proof. By induction. The claim of the lemma holds for $k = 0$, since

$$\left\{ \theta_i \in \Theta_i \left| m_i^0 \in S^{\mathcal{M}^{0,\varepsilon}} \left(\theta_i \right) \right. \right\} = \Theta_i = \overline{\Theta}_i^0 \left(m_i^0 \right).$$

Now suppose that the claim holds for k. Thus there exists $\overline{\varepsilon}_k > 0$, such that

$$\left\{ \theta_i \in \Theta_i \left| m_i^k \in S^{\mathcal{M}^{k,\varepsilon}} \left(\theta_i \right) \right. \right\} \subseteq \overline{\Theta}_i^k \left(m_i^k \right) \text{ for all } \varepsilon \leq \overline{\varepsilon}_k \text{ and } m_i^k \in M_i^k.$$

Now observe that $\mathcal{M}^{k+1,\varepsilon}$ is an augmentation of $\mathcal{M}^{k,\varepsilon}$ and thus - by lemma 7 - there exists $\overline{\varepsilon}_{k+1} \in (0, \overline{\varepsilon}_k]$, such that for all $\varepsilon \leq \overline{\varepsilon}_{k+1}$,

$$m_i^{k+1} = \left(m_i^k, r_i^{k+1}, y_i^{k+1} \right) \in \ S^{\mathcal{M}^{k+1,\varepsilon}} \left(\theta_i \right) \Rightarrow m_i^k \in \ S^{\mathcal{M}^{k,\varepsilon}} \left(\theta_i \right). \tag{12}$$

Now by the inductive hypothesis, we also have

$$\theta_i \in \overline{\Theta}_i^k \left(m_i^k \right). \tag{13}$$

$m_i^{k+1} = S^{\mathcal{M}^{k+1,\varepsilon}}(\theta_i)$ also implies there must exist $\mu_i \in \Delta\left(\Theta_{-i} \times M_{-i}^{k+1}\right)$ such that (1):

$$\mu_i\left(\theta_{-i}, m_{-i}^{k+1}\right) > 0 \quad \Rightarrow \quad m_j^{k+1} \in S^{\mathcal{M}^{k+1,\varepsilon}}(\theta_j) \text{ for each } j \neq i$$

and (2):

$$m_i^{k+1} \in \underset{\overline{m}_i^{k+1} \in M_i^{k+1}}{\arg\max} \sum_{\theta_{-i}, m_{-i}^{k+1}} \mu_i\left(\theta_{-i}, m_{-i}^{k+1}\right)\left[u_i\left(g^{k+1,\varepsilon}\left(\overline{m}_i^{k+1}, m_{-i}^{k+1}\right), (\theta_i, \theta_{-i})\right)\right].$$

But note that $\left(r_i^{k+1}, y_i^{k+1}\right)$ - the last components of m_i^{k+1} - effect only one additively separable component of the above expression. In particular, $\left(r_i^{k+1}, y_i^{k+1}\right)$ must maximize:

$$\sum_{\theta_{-i}, m_{-i}^{k+1}} \mu_i\left(\theta_{-i}, m_{-i}^{k+1}\right) \mathbb{I}\left(r_i^{k+1}, m_{-i}^k\right)\left(u_i\left(y_i^{k+1}, (\theta_i, \theta_{-i})\right) - u_i\left(\overline{y}, (\theta_i, \theta_{-i})\right)\right) \tag{14}$$

which we can rewrite as

$$\sum_{\theta_{-i}} \sum_{\left\{m_{-i}^{k+1} \mid m_{-i}^k = r_i^{k+1}\right\}} \mu_i\left(\theta_{-i}, m_{-i}^{k+1}\right)\left(u_i\left(y_i^{k+1}, (\theta_i, \theta_{-i})\right) - u_i\left(\overline{y}, (\theta_i, \theta_{-i})\right)\right).$$

The later expression is zero if

$$\mu_i\left(r_i^{k+1}\right) \equiv \sum_{\theta_{-i}} \sum_{\left\{m_{-i}^{k+1} \mid m_{-i}^k = r_i^{k+1}\right\}} \mu_i\left(\theta_{-i}, m_{-i}^{k+1}\right) = 0.$$

But if $\mu_i\left(r_i^{k+1}\right) > 0$ and $y_i^{k+1} \in B_i^{Y^*}(\theta_i, \lambda_i)$, where

$$\lambda_i\left(\theta_{-i}\right) = \frac{\displaystyle\sum_{\left\{m_{-i}^{k+1} \mid m_{-i}^k = r_i^{k+1}\right\}} \mu_i\left(\theta_{-i}, m_{-i}^{k+1}\right)}{\displaystyle\sum_{\theta'_{-i}} \sum_{\left\{m_{-i}^{k+1} \mid m_{-i}^k = r_i^{k+1}\right\}} \mu_i\left(\theta'_{-i}, m_{-i}^{k+1}\right)},$$

then (14) must be strictly positive, by the first part of proposition 4. Thus we must have $\left(r_i^{k+1}, y_i^{k+1}\right)$ chosen such that $\mu_i\left(r_i^{k+1}\right) > 0$ and $y_i^{k+1} \in B_i^{Y^*}(\theta_i, \lambda_i)$. Now $\mu_i\left(r_i^{k+1}\right) > 0$, (12) and the inductive hypothesis imply that

$$\overline{\Theta}_{-i}^k\left(r_i^{k+1}\right) \neq \varnothing; \tag{15}$$

and

$$\lambda_i \in \Delta\left(\overline{\Theta}_{-i}^k\left(r_i^{k+1}\right)\right) \text{ and } y_i^{k+1} \in B_i^{Y^*}(\theta_i, \lambda_i). \tag{16}$$

Now (13), (15) and (16) together imply that, for any $m_i^{k+1} \in S^{\mathcal{M}^{k+1,\varepsilon}}(\theta_i)$, $\theta_i \in \overline{\Theta}_i^{k+1}\left(m_i^{k+1}\right)$. $\blacksquare$

We next show that the sets $\overline{\Theta}_i^k$ are closely related to kth level inseparable sets Ξ_i^k, as defined earlier in (1)-(3).

Lemma 9 *For all i and k, $\overline{\Theta}_i^k\left(m_i^k\right) \in \Xi_i^k$ for all $m_i^k \in M_i^k$.*

Proof. By induction. The claim is true for $k = 0$ by definition. Suppose $\overline{\Theta}_{-i}^{k-1}\left(m_{-i}^{k-1}\right) \in \Xi_{-i}^{k-1}$ for all $m_{-i}^{k-1} \in M_{-i}^{k-1}$. Now fix any $m_i^k = \left(m_i^{k-1}, r_i^k, y_i^k\right) \in M_i^k$ and let $\Psi_i = \overline{\Theta}_i^k\left(m_i^k\right)$ and let $\Psi_{-i} = \overline{\Theta}_{-i}^{k-1}\left(r_i^k\right)$. By proposition 4 part (1), every type has some strict preference over Y^* and thus will set r_i^k equal to some m_{-i}^{k-1} he assigns positive probability to. By our inductive assumption, $\Psi_{-i} \in \Xi_{-i}^{k-1}$. Now suppose Ψ_{-i} separates Ψ_i and fix $\theta_i \in \Psi_i$. By proposition 4 part (2), there exists $\theta'_i \in \Psi_i$ such that $y_i^k \notin B_i^{Y^*}(\theta'_i, \Psi_{-i})$. Thus $\theta'_i \notin \overline{\Theta}_i^k\left(m_i^k\right)$, a contradiction. We conclude that Ψ_{-i} does not separate Ψ_i. $\blacksquare$

Conclusion of proof of proposition 2   The proof of proposition 2 can now be completed. For a fixed environment we can choose K such that $\Xi^K = \Xi^*$. By lemmas 8 and 9, there exists $\varepsilon > 0$ such that

$$\left\{ \theta_i \in \Theta_i \,\middle|\, m_i^K \in S^{\mathcal{M}^{K,\varepsilon}}(\theta_i) \right\} \subseteq \overline{\Theta}_i^K \left(m_i^K \right) \in \Xi_i^K = \Xi_i^*,$$

for all $m_i^K \in M_i^K$. So

$$m_i^K \in S^{\mathcal{M}^{K,\varepsilon}}(\theta_i) \cap S^{\mathcal{M}^{K,\varepsilon}}(\theta_i') \Rightarrow \{\theta_i, \theta_i'\} \in \Xi_i^*.$$

Thus the mechanism $\mathcal{M}^{K,\varepsilon}$ satisfies the property of proposition 2.

5   Robust Virtual Implementation

In this section, we use the notions of strategic distinguishability and the maximally revealing mechanism to establish necessary and sufficient conditions for robust virtual implementation. Virtual implementation of a social choice function requires a mechanism such that the desired outcomes are realized with probability arbitrarily close to 1 (see Abreu and Matsushima (1992b) and Abreu and Matsushima (1992c)). Robust implementation requires implementation of a social choice function depending on agents' "payoff types" independent of their beliefs and higher order beliefs about others' payoff types (see Bergemann and Morris (2005a) and Bergemann and Morris (2005b)). Our definition of robust virtual implementation is the natural one incorporating both these notions.

5.1   Definitions

Write $\|y - y'\|$ for the Euclidean distance between a pair of lotteries y and y', i.e.,

$$\|y - y'\| = \sqrt{\sum_{x \in X} \left(y(x) - y'(x) \right)^2}.$$

Definition 5 (Robust ε-Implementation)
The mechanism $\mathcal{M}$ robustly ε-implements the social choice function f if

$$m \in S^{\mathcal{M}}(\theta) \Rightarrow \|g(m) - f(\theta)\| \leq \varepsilon.$$

f is robustly ε-implementable if there exists a mechanism $\mathcal{M}$ that robustly ε-implements f.

We can now define the notion of robust virtual implementation.

Definition 6 (Robust Virtual Implementation)
Social choice function f is robustly virtually implementable if, for every $\varepsilon > 0$, f is robustly ε-implementable.

The relevant incentive compatibility condition required for our robust problem is ex post incentive compatibility.

Definition 7 (EPIC)

Social choice function f satisfies ex post incentive compatibility (EPIC) if, for all i, θ_i, θ_{-i} and θ_i':

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

"Robust measurability" requires that if θ_i is strategically indistinguishable from θ_i', then the social choice function must treat the two types the same. This condition is the robust analogue of the measurability condition in Abreu and Matsushima (1992c).

Definition 8 (Robust Measurability)

Social choice function f satisfies robust measurability if $\theta_i \sim \theta_i' \Rightarrow f\left(\theta_i,\theta_{-i}\right) = f\left(\theta_i',\theta_{-i}\right)$ for all θ_{-i}.

5.2 Necessity

It is well known from the literature on virtual Bayesian implementation (e.g., Abreu and Matsushima (1992c)) that the relaxation to virtual implementation does not relax incentive compatibility conditions by a standard compactness argument.[10]

Theorem 2 (Necessity)

If f is robustly virtually implementable, then f satisfies ex post incentive compatibility and robust measurability.

Proof. We first establish ex post incentive compatibility. Fix any mechanism $\mathcal{M}$ that robustly ε-implements f. Fix θ_{-i} and $m_{-i} \in S_{-i}^{\mathcal{M}}\left(\theta_{-i}\right)$. For any $m_i' \in S_i^{\mathcal{M}}\left(\theta_i'\right)$, virtual implementation requires

$$\left\|g\left(m_i',m_{-i}\right) - f\left(\theta_i',\theta_{-i}\right)\right\| \leq \varepsilon. \tag{17}$$

Now suppose that player i is type θ_i and is convinced that his opponent is type θ_{-i} sending message m_{-i}. Let m_i be any message which is a best response to that belief. Then $m_i \in S_i^{\mathcal{M}}\left(\theta_i\right)$, implying that

$$\left\|g\left(m_i,m_{-i}\right) - f\left(\theta_i,\theta_{-i}\right)\right\| \leq \varepsilon. \tag{18}$$

In particular, by the best response property of m_i:

$$u_i\left(g\left(m_i,m_{-i}\right),\left(\theta_i,\theta_{-i}\right)\right) \geq u_i\left(g\left(m_i',m_{-i}\right),\left(\theta_i,\theta_{-i}\right)\right). \tag{19}$$

Now (17) and lemma 2 imply

$$\left|u_i\left(g\left(m_i',m_{-i}\right),\left(\theta_i,\theta_{-i}\right)\right) - u_i\left(f\left(\theta_i',\theta_{-i}\right),\left(\theta_i,\theta_{-i}\right)\right)\right| \leq \varepsilon C, \tag{20}$$

and (18) and lemma 2 imply

$$\left|u_i\left(g\left(m_i,m_{-i}\right),\left(\theta_i,\theta_{-i}\right)\right) - u_i\left(f\left(\theta_i,\theta_{-i}\right),\left(\theta_i,\theta_{-i}\right)\right)\right| \leq \varepsilon C. \tag{21}$$

---

[10] Dasgupta, Hammond, and Maskin (1979) and Ledyard (1979) argued in a private value environment that dominant strategy incentive compatibility was implied by Bayesian incentive compatibility for all priors on a fixed type space. In the case of a social choice function, this argument - generalized to interdependent values - shows the necessity of ex post incentive compatibility (see Bergemann and Morris (2005c)).

Now combining (19), (20) and (21), we obtain

$$u_i \left(f \left(\theta_i, \theta_{-i} \right), \left(\theta_i, \theta_{-i} \right) \right) \geq u_i \left(f \left(\theta'_i, \theta_{-i} \right), \left(\theta_i, \theta_{-i} \right) \right) - 2\varepsilon C.$$

But virtual implementation implies that this holds for all $\varepsilon > 0$, so we have

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

and this establishes EPIC as necessary condition.

Next we establish robust measurability. Suppose that f is robustly virtually implementable. Fix any $\varepsilon > 0$. Since f is robustly virtually implementable, there exists a mechanism $\mathcal{M}^\varepsilon$ such that

$$m \in S^{\mathcal{M}^\varepsilon} \left(\theta \right) \Rightarrow \left\| g \left(m \right) - f \left(\theta \right) \right\| \leq \varepsilon.$$

Now fix any θ_{-i} and $m^\varepsilon_{-i} \in S^{\mathcal{M}^\varepsilon}_{-i} \left(\theta_{-i} \right)$. Also fix any $\theta_i \sim \theta'_i$, so by proposition 1, there exists $m^\varepsilon_i \in S^{\mathcal{M}^\varepsilon}_i \left(\theta_i \right) \cap S^{\mathcal{M}^\varepsilon}_i \left(\theta'_i \right)$. Now $\left\| g \left(m^\varepsilon_i, m^\varepsilon_{-i} \right) - f \left(\theta_i, \theta_{-i} \right) \right\| \leq \varepsilon$ and $\left\| g \left(m^\varepsilon_i, m^\varepsilon_{-i} \right) - f \left(\theta'_i, \theta_{-i} \right) \right\| \leq \varepsilon$. Thus $\left\| f \left(\theta_i, \theta_{-i} \right) - f \left(\theta'_i, \theta_{-i} \right) \right\| \leq 2\varepsilon$. This is true for each $\varepsilon > 0$, so $f \left(\theta_i, \theta_{-i} \right) = f \left(\theta'_i, \theta_{-i} \right)$. ∎

5.3 Sufficiency

We first describe the construction of a *canonical mechanism* that will be used to establish sufficiency. Our construction follows the logic of Abreu and Matsushima (1992c), which in turn builds on Abreu and Matsushima (1992b). In the mechanism we construct, each agent simultaneously announces (i) a message in the maximally revealing mechanism described above; (ii) L announcements of his payoff type. With probability close to $\frac{1}{L}$, the outcome is chosen according the agents' lth announcement of their payoff types in part (ii) of their messages. But with small probability, the outcome is chosen according to the maximally revealing mechanism and their part (i) messages. The mechanism then checks to see which agents were the "first" to "lie", in the sense that his lth report of his type is not consistent with the message he sent in the maximally revealing mechanism and no other agent sent an inconsistent message in an "earlier" report. If an agent is not one of the first to lie, then the agent is rewarded. For this part of the mechanism, we need an economic property.

Definition 9 (Economic Property)

The uniform economic property is satisfied if there exist a profile of lotteries, $(z_i)_{i=1}^I$, such that, for each i and θ, $u_i \left(z_i, \theta \right) > u_i \left(\overline{y}, \theta \right)$ and $u_j \left(\overline{y}, \theta \right) \geq u_j \left(z_i, \theta \right)$ for all $j \neq i$.

Under the uniform economic property, there will exist a constant c_0 such that

$$u_i \left(z_i, \theta \right) > u_i \left(\overline{y}, \theta \right) + c_0 \tag{22}$$

for all i and θ.

In the canonical mechanism, part (i) announcements for the maximally revealing mechanism are made as if the maximally revealing mechanism was being played as a stand alone mechanism (since the probability of rewards can be chosen sufficiently small). An agent will never allow himself to be one of the first to lie:

sending a message that ensures that he is not the first to lie (given his beliefs about others' strategies) will always strictly improve on his expected payoff, since if others are telling the truth, truth-telling is a weak best response by ex post incentive compatibility, and if they are lying, for sufficiently large L, the reward will outweigh the cost of not lying in one round of the mechanism.

We write $\mathcal{M}^* = \left((M_i^*)_{i=1}^I, g^* \right)$ for the maximally revealing mechanism. We use three numbers in defining the canonical mechanism: c_0 is the uniform lower bound on an agent's utility gain from having his uniformly preferred lottery rather than the central lottery; recall from lemma 2 that C is an upper bound on payoff differences in the environment; recall from lemma 6 whenever a message is deleted in the iterated deletion process for the maximally revealing mechanism $\mathcal{M}^*$, it is not even an $\eta_{\mathcal{M}^*}$-best response to any conjecture. We will use these three numbers c_0, C and $\eta_{\mathcal{M}^*}$, together with the number of players I, to define two further numbers δ and L that will be used in the construction of the canonical mechanism. Choose $\delta > 0$ such that

$$\delta < \frac{\eta_{\mathcal{M}^*}}{C}, \tag{23}$$

and an integer L such that

$$L > \frac{IC}{\delta^2 c_0}. \tag{24}$$

Now the message space of the canonical mechanism is

$$M_i = M_i^* \times \overbrace{\Theta_i \times .... \times \Theta_i}^{L \text{ times}} = M_i^* \times \Theta_i^L.$$

Thus a typical message will be written as $m_i = \left(m_i^0, m_i^1, ..., m_i^L \right)$, with $m_i^0 \in M_i^*$; $m_i^l \in \Theta_i$ for each $l = 1, ..., L$. The idea is that an agent is "supposed" to truthfully report his payoff type in each round $l = 1, ..., L$ and will receive a small punishment if he is one of the "first" to report a type that is not consistent with his 0th message. The small individual rewards and punishments are provided by

$$r_i(m) = \begin{cases} \overline{y}, \text{ if} & \exists k \in \{1, ..., L\} \text{ s.t. } m_i^0 \notin S_i^{\mathcal{M}^*}\left(m_i^k \right) , \\ & \text{and } m_j^0 \in S_j^{\mathcal{M}^*}\left(m_j^l \right) \ \forall j \neq i \text{ and } l = 1, ..., k-1; \\ \\ z_i, \text{ otherwise.} \end{cases}$$

(In slight abuse of notation, we use $r_i(m)$ here to denote rewards whereas we used r_i^k earlier in Subsection 4.3.4.) Now the outcome function of the canonical mechanism is:

$$g(m) = \left(1 - \delta - \delta^2 \right) \frac{1}{L} \sum_{l=1}^L f\left(m^l \right) + \delta g^*\left(m^0 \right) + \frac{\delta^2}{I} \sum_{i=1}^I r_i(m).$$

The mechanism $g(m)$ has three components. The first component, which carries the largest probability, is the social choice function f itself. The appropriate allocation $f\left(m^l \right)$ will be selected by L replicas, each one of which is chosen with a small probability $1/L$. The second component is the maximally revealing mechanism outcome function g^* which receives a smaller weight of δ. The third and final component, $r_i(m)$, represents a small reward or punishment. It is designed to give each agent an incentive to replicate in strip l the report issued in the previous strips. It provides a small "punishment" $(\overline{y})$ if player i is the first to

report in the message component, m_i^l, a type inconsistent with previous reports, otherwise $r_i(m)$ provides the small "reward" (z_i).

Theorem 3 *Under the uniform economic property, if f satisfies EPIC and robust measurability, then the canonical mechanism $\delta(1+\delta)$ robustly virtually implements f.*

This immediately implies the sufficiency part of our characterization of robust virtual implementation, since we can choose δ arbitrarily close to 0 in the canonical mechanism.

Corollary 1 (Sufficiency) *Under the uniform economic property, if f satisfies EPIC and robust measurability, then f is robustly virtually implementable.*

Proof. To prove the theorem, it is enough to establish that, for each i, $m_i = \left(m_i^0, m_i^1, ..., m_i^L\right) \in S_i^{\mathcal{M}}(\theta_i)$ implies that (1) $m_i^0 \in S_i^{\mathcal{M}^*}(\theta_i)$ and (2) $m_i^0 \in S_i^{\mathcal{M}^*}\left(m_i^l\right)$ for each $l = 1, ..., L$. To see why, observe that $m_i^0 \in S_i^{\mathcal{M}^*}(\theta_i) \cap S_i^{\mathcal{M}^*}\left(m_i^l\right)$ implies θ_i is strategically indistinguishable from m_i^l, which implies, by robust measurability, that $f\left(m_i^l, m_{-i}^l\right) = f\left(\theta_i, m_{-i}^l\right)$. Since this holds for each i, we have $f\left(m^l\right) = f(\theta)$. Since this is true for each l, we have that the mechanism selects $f(\theta)$ with probability at least $1 - \delta - \delta^2$.

Claim (1) above - that $\left(m_i^0, m_i^1, ..., m_i^L\right) \in S_i^{\mathcal{M}}(\theta_i) \Rightarrow m_i^0 \in S_i^{\mathcal{M}^*}(\theta_i)$ - follows from lemma 7 and inequality (23), since m^0 influences the outcome only through weight δ on $g^*\left(m^0\right)$ and weight δ^2 on $\frac{1}{I}\sum_{i=1}^{I} r_i(m)$.

We will now establish claim (2) above - that $\left(m_i^0, m_i^1, ..., m_i^L\right) \in S_i^{\mathcal{M}}(\theta_i) \Rightarrow m_i^0 \in S_i^{\mathcal{M}^*}\left(m_i^l\right)$ for all i and $l = 1, ..., L$.

Suppose this claim were false. Then there must exist i and $m_i = \left(m_i^0, m_i^1, ..., m_i^L\right) \in S_i^{\mathcal{M}}(\theta_i)$ with $m_i^0 \notin S_i^{\mathcal{M}^*}\left(m_i^{l^*}\right)$ for some $l^* \in \{1, ..., L\}$ and $m_j^0 \in S_j^{\mathcal{M}^*}\left(m_j^l\right)$ for all j and $1 \leq l < l^*$. Now fix any conjecture $\mu_i \in \Delta(\Theta_{-i} \times M_{-i})$ with $\mu_i(\theta_{-i}, m_{-i}) > 0 \Rightarrow m_j \in S_j^{\mathcal{M}}(\theta_j)$ for all $j \neq i$. Consider two cases. First, suppose that

$$\mu_i(\theta_{-i}, m_{-i}) > 0 \implies m_j^0 \in S_j^{\mathcal{M}^*}\left(m_j^l\right) \text{ for all } j \neq i \text{ and } l = 1, ..., L. \tag{25}$$

In this case, sending the message

$$\overline{m}_i = (m_i^0, \overbrace{\theta_i, \theta_i, ..., \theta_i}^{L \text{ times}})$$

instead of m_i will strictly increase i's utility: since he is certain that each agent is reporting a type that is strategically indistinguishable in each of the L strips, EPIC and robust monotonicity ensure that his utility will not decrease from truthtelling in the L strips; his utility will be unchanged in the maximally revealing mechanism; and his utility will be strictly increased in the punishment component. Secondly, suppose that (25) fails. In this case, we can define

$$\widehat{l} = \min\left\{l \in \{1, ..., L\} : \exists (\theta_{-i}, m_{-i}) \text{ with } \mu_i(\theta_{-i}, m_{-i}) > 0 \text{ and } m_j^0 \notin S_j^{\mathcal{M}^*}\left(m_j^l\right) \text{ for some } j \neq i\right\}.$$

Now sending the message

$$\overline{m}_i = (m_i^0, \overbrace{\theta_i, \theta_i, ..., \theta_i}^{\widehat{l} \text{ times}}, m_i^{\widehat{l}+1}, ..., m_i^L)$$

instead of m_i will strictly increase i's utility: since he is certain that each agent is reporting a type that is strategically indistinguishable in each of the first $\widehat{l} - 1$ strips, EPIC and robust monotonicity ensure that his utility will not decrease from truthtelling in the first $\widehat{l} - 1$ strips; his utility will be unchanged in the maximally revealing mechanism; if it turns out that $m_j^0 \in S_j^{\mathcal{M}^*}\left(m_j^{\widehat{l}}\right)$ for some $j \neq i$, then i's utility will also not be reduced in the $\widehat{l}$−th strip or in the punishment component; but if it turns out that $m_j^0 \notin S_j^{\mathcal{M}^*}\left(m_j^{\widehat{l}}\right)$ for all $j \neq i$, then i's utility will be reduced in the $\widehat{l}$−th strip by at most $\left(1 - \delta - \delta^2\right)\frac{1}{L}C$ and will increase in the punishment component by at least $\frac{\delta^2}{I}c_0$. The latter exceeds the former by (24).

We conclude that for no conjecture is m_i a best response, contradicting our original assumption. This proves our second claim. ∎

While the basic construction of this proof follows Abreu and Matsushima (1992c), there are some complications that arise in our robust formulation. The messages sent in the maximally revealing mechanism do not partition an agent's types. Rather, for each set of types that survives the iterated deletion of sets that can always be separated, there is a message that may be sent by all types in that set. So we say that message m_i^l is consistent with m_i^0 if message m_i^0 is one that might be sent by $m_j^0 \in S_i^{\mathcal{M}^*}\left(m_i^l\right)$.

The economic property can weakened along the lines of assumption 2 in Abreu and Matsushima (1992c). It would be enough to have that the economic property holds for any type set profile Ψ in the inseparable type set Ξ^*, i.e. for each set profile $\Psi = (\Psi_i)_{i=1}^I \in \Xi^*$, there exists $(z_i)_{i=1}^I$, such that, for each i and $\theta \in \times_{i=1}^I \Psi_j$, $u_i(z_i, \theta) > u_i(\overline{y}, \theta)$ and $u_j(\overline{y}, \theta) \geq u_j(z_i, \theta)$ for all $j \neq i$.

6 Discussion

6.1 Strategic Equivalence

We briefly discuss the relation between strategic distinguishability and a finer natural relation on payoff types, strategic equivalence.[11]

Definition 10 (Strategic Equivalence)
Types θ_i and θ_i' are strategically equivalent if $S_i^{\mathcal{M}}(\theta_i) = S_i^{\mathcal{M}}(\theta_i')$ for every $\mathcal{M}$.

In words, the types θ_i and θ_i' are strategically equivalent if the set of rationalizable actions of θ_i and θ_i' agree in every mechanism. In contrast, the types θ_i and θ_i' were said to be strategically indistinguishable if the intersection of their rationalizable actions is nonempty for every mechanism. The following strengthening of pairwise inseparability is sufficient for strategic equivalence. Consider a sequence of profiles of partitions of types of each agent, $\mathcal{P}^k = \left(\mathcal{P}_1^k, ..., \mathcal{P}_I^k\right)$, defined as follows. First, $\mathcal{P}_i^0 = \{\Theta_i\}$ for all i. Second, writing $P_i^k(\theta_i)$ for the unique element of $\mathcal{P}_i^k$ containing θ_i, let

$$P_i^{k+1}(\theta_i) = \left\{\theta_i' | \mathcal{R}_i\left(\theta_i', \Psi_{-i}\right) = \mathcal{R}_i(\theta_i, \Psi_{-i}) \text{ for each } \Psi_{-i} \in \mathcal{P}_{-i}^k\right\}$$

for each i, θ_i and $k = 0, 1, ...$ .Finally,

$$P_i^*(\theta_i) = \lim_{k \to \infty} P_i^k(\theta_i).$$

---

[11] We are grateful to Faruk Gul and Wolfgang Pesendorfer for discussions about this connection.

One can easily verify that in the single good example of section 3, P_i^* is the finest partition for any value of γ. Thus strategic equivalence is a much more discriminating criterion of distinguishability than pairwise separability. The requirement that each $P_i^*(\theta_i)$ is a singleton is essentially the validity property of Gul and Pesendorfer (2005), translated into our setting. The differences are that they do not explicitly incorporate uncertainty, so they have a more general outcome space and their analogue of $\mathcal{R}_i(\theta_i, \Psi_{-i})$ corresponds to all preferences that you might have given certain beliefs about others' types; and their analogue has player i's types equivalent at round k if his conjecture about all agents' preferences are the same, not just player i.

Now we have the following result:

Proposition 5 *If $\theta_i' \in P_i^*(\theta_i)$, then θ_i' and θ_i are strategically equivalent.*

The proof of proposition 5 is in appendix 8.5. We conjecture that a converse to the proposition is true: if θ_i' and θ_i are strategically equivalent, then $\theta_i' \in P_i^*(\theta_i)$. This could be shown by constructing, for each $\theta_i' \notin P_i^*(\theta_i)$, a mechanism $\mathcal{M}$ with message $m_i \in M_i$ with $m_i \in S_i^{\mathcal{M}}(\theta_i)$ and $m_i' \notin S_i^{\mathcal{M}}(\theta_i')$. Such a construction would rely on constructing mechanism that reveals agents' beliefs and higher order beliefs about others types, along the lines of Dekel, Fudenberg, and Morris (2006).

6.2 Rationalizability and All Equilibria on All Type Spaces

Our analysis took as given the solution concept of incomplete information rationalizability for our environment. Thus we assumed that if the agents' true payoff type profile was

$$\theta = (\theta_1, ..., \theta_I),$$

they might send any message profile

$$m \equiv (m_1, ..., m_I) \in \underset{i=1}{\overset{I}{\times}} S_i^{\mathcal{M}}(\theta_i) \equiv S^{\mathcal{M}}(\theta).$$

Our motivation for employing this solution concept is that we did not want to make any assumption about agents' beliefs and higher order beliefs about other agents' payoff types. In fact, suppose one constructed a "type space" $\mathcal{T}$ specifying for each agent a set of possible epistemic types, and, for each epistemic type, a description of his (known) payoff type and his beliefs about others' epistemic types. By standard universal type space arguments, we can incorporate any beliefs and higher order beliefs about others' payoff types in such a type space. Now the type space $\mathcal{T}$ and a mechanism $\mathcal{M}$ together define a standard incomplete information game. The set of messages that can be sent by *any* type of agent i with payoff type θ_i in *any* Bayesian Nash equilibrium of the game $(\mathcal{T}, \mathcal{M})$ for *any* type space $\mathcal{T}$ is equal to $S_i^{\mathcal{M}}(\theta_i)$. This result is the straightforward incomplete information extension of the classic epistemic foundations result of Brandenburger and Dekel (1987), showing that the set of actions that can be played in the subjective correlated equilibria of a complete information game equals the set of actions that survive iterated deletion of strictly dominated actions in that game. Battigalli and Siniscalchi (2003) reported the incomplete information version of this result as Propositions 4.2 and 4.3. For completeness, we formally state and prove this result in appendix 8.2.

This observation means that the gap between the solution concepts of pure strategy Bayesian Nash equilibrium (Serrano and Vohra (2001), Serrano and Vohra (2005)) and iterated deletion of (interim) strictly

dominated strategies (Abreu and Matsushima (1992c)) in incomplete information virtual implementation disappears in our robust approach.[12] We consider this to be an attraction of our approach. The intuition is that the extra bite obtained by the assumption of equilibrium is lost without complementary strong assumptions on beliefs and higher order beliefs for the implementation problem.

6.3 Iterated Deletion of Weakly Dominated Strategies

Our incomplete information rationalizability solution concept is equivalent to iterated deletion of strictly dominated strategies. What would happen if we looked at iterated deletion of weakly dominated strategies instead? In other words, we let $W_i^{\mathcal{M},0}(\theta_i) = M_i$,

$$
W_i^{\mathcal{M},k+1}(\theta_i) = \left\{ m_i \in W_i^{\mathcal{M},k}(\theta_i) \left| \begin{array}{l} \exists\, \mu_i \in \Delta_{++}(\Theta_{-i} \times M_{-i}) \text{ s.t.:} \\[2ex] (1)\ \mu_i(\theta_{-i}, m_{-i}) > 0 \Rightarrow m_{-i} \in W_{-i}^{\mathcal{M},k}(\theta_{-i}) \\[2ex] (2)\ m_i \in \arg\max_{m_i'} \sum_{\theta_{-i}, m_{-i}} \mu_i(\theta_{-i}, m_{-i})\, u_i\left(g\left(m_i', m_{-i}\right), (\theta_i, \theta_{-i})\right) \end{array} \right. \right\};
$$

and

$$
W_i^{\mathcal{M}}(\theta_i) = \bigcap_{k \geq 0} W_i^{\mathcal{M},k}(\theta_i).
$$

It is easy to see that our "negative" results would go through unchanged. If two types are pairwise inseparable $(\theta_i \sim \theta_i')$ then the argument of proposition 1 - unchanged - implies that they will have iteratively weakly undominated actions in common in every mechanism, or

$$
W_i^{\mathcal{M}}(\theta_i) \cap W_i^{\mathcal{M}}(\theta_i') \neq \varnothing \text{ for all } \mathcal{M}.
$$

Thus robust measurability is a necessary condition for implementation (virtual or exact) of any social choice function in iterated deletion of weakly dominated strategies in a finite (or compact) mechanism: the argument of proposition 2 will go through unchanged in this case.

Abreu and Matsushima (1994) show their argument for *virtual* complete information implementation in iterated deletion of *strictly* dominated strategies can be adapted to show the possibility of *exact* complete information implementation in iterated deletion of *weakly* dominated strategies, with some extra restrictions on the environment. It is a reasonable conjecture that this extension could be adapted to the standard incomplete information implementation setting of Abreu and Matsushima (1992c) and our robust incomplete information setting. However, we have not attempted this extension.

Chung and Ely (2001) have shown that in an auction environment with interdependent valuations as in section 3, the efficient outcome can be implemented in the direct mechanism under iterated deletion of

---

[12] Abreu and Matsushima (1992c) showed that their measurability condition was necessary for virtual implementation in mixed strategy Bayesian Nash equilibrium restricting attention to well-behaved mechanisms. But it remains an open question whether the measurability condition is necessary for virtual implementation in pure strategy Bayesian Nash equilibrium restricting attention to well-behaved mechanisms (see Serrano and Vohra (2005)).

weakly dominated strategies (i.e., the solution concept described above) under the assumption that $\gamma < \frac{1}{I-1}$. Our results supply a strong converse: if $\gamma \geq \frac{1}{I-1}$, it is not possible to implement (exactly or virtually) any non-trivial social choice function in iterated deletion of weakly dominated strategies in any finite (or compact) mechanism, direct or indirect.[13]

6.4 Implementation in a Direct Mechanism

We restricted attention in this paper to finite mechanisms. Thus the mechanisms here do not include any of the pathological features of "integer games" that play an important role in the full implementation literature and have been much criticized (see, e.g., Jackson (1992)). Nonetheless, the mechanisms in this paper are complex. The canonical mechanism for robust virtual implementation inherits the complexity of the mechanism of Abreu and Matsushima (1992c), on which it builds. Our maximally revealing mechanism generating strategic distinguishability is no simpler. While the mechanisms are theoretically kosher, it has been argued that their complexity and the logic of the iteration deletion in the mechanism might make them hard to use in practise. For example, Glazer and Rosenthal (1992) have made this argument about the mechanism used by Abreu and Matsushima (1992b) for complete information virtual implementation (see Abreu and Matsushima (1992a) for a response and Sefton and Yavas (1996) for later experiments inspired by the mechanism).

By requiring robustness to agents' beliefs and higher order beliefs, we reduce the amount of common knowledge about the environment that can be used by the planner in designing a mechanism. This will make it harder to achieve positive results (and our robust measurability condition is rather strong in applications). But one motivation for studying robust implementation is that we hope that robustness considerations will endogenously lead to simpler mechanisms when positive results can be achieved. By adapting results from our earlier work on exact robust implementation in direct mechanisms (Bergemann and Morris (2005a)), we can report that, in at least one broad class of economic environments of interest, whenever robust virtual implementation is possible according to corollary 1, it is possible in a direct mechanism where agents simply report their payoff types.

This result can be nicely illustrated in the environment with interdependent valuations for a single good of section 3. Recall that if $\gamma \geq \frac{1}{I-1}$, all pairs of types are pairwise inseparable, so - by this paper's theorem 1 - all pairs of types are strategically indistinguishable, and - by this paper's theorem 2 - robust virtual implementation of any non-trivial social choice function is impossible in any mechanism. However, it turns out if $\gamma < \frac{1}{I-1}$, not only does there exist a finite mechanism that robustly virtually implements the efficient allocation, there is in fact a *direct* mechanism - where each agent's message space is his set of payoff types - that does so. To see why, first observe that there is a well-known simple mechanism that allocates the object efficiently as a function of agents' reports of their types. Each agent makes a report b_i about his payoff type θ_i. The object is awarded to the highest bidder who must pay the "pivotal" value

$$\max_{j \neq i}\{b_j\} + \gamma \sum_{j \neq i} b_j. \tag{26}$$

---

[13]Our results are stated for a lottery space over finite outcomes, but the extension to any compact space and compact mechanisms is straightforward.

Truth-telling is ex post incentive compatible in this mechanism; i.e., if you are sure that others will bid truthfully, you have an incentive to bid truthfully whatever you think that others will bid. It is straightforward to modify this direct revelation mechanism to one that virtually allocates the object efficiently with strict ex post incentive compatibility. With probability $1 - \varepsilon$, allocate the object to the highest bidder; but with probability ε, there is a random allocation rule where one of the agents is chosen with probability $\frac{1}{I}$ and he is given the object with probability b_i (independently of others' bids). If bidder i receives the object as the highest bidder, he must pay the pivotal value (26); if he receives the object under the random allocation rule, he must pay $\frac{1}{2}b_i + \gamma \sum_{j \neq i} b_j$. Truth-telling is strictly ex post incentive compatible in this mechanism and the object is allocated efficiently with probability at least $1 - \varepsilon$. Bergemann and Morris (2005a) establish that - if $\gamma < \frac{1}{I-1}$ - truth-telling is the unique rationalizable message in this mechanism.[14]

This observation generalizes to an economically intuitive class of environments. Preferences satisfy *aggregator single crossing* (ASC) if each agent i's preferences at type profile θ belong to a single crossing class parameterized by $h_i(\theta)$, where $h_i : \Theta \to \mathbb{R}$ is a monotonic aggregator. Bergemann and Morris (2005a) established that exact robust implementation by a compact mechanism is possible if and only if the social choice function satisfies strict ex post incentive compatibility and a *contraction property* on the aggregator functions $h = (h_1, ..., h_I)$. In appendix 8.4, we show that under the ASC assumption, robust measurability is always satisfied under the contraction property. If preferences display an easily satisfied *strict ex post preferences* (SEP) condition and *a* social choice function satisfies EPIC, then there exists a nearby social choice function that satisfies strict EPIC. So under ASC and SEP, if f satisfies EPIC and robust measurability, f is robustly virtually implementable in a *direct* mechanism. Formal statements and proofs of these results appear in appendix 8.4.

6.5   Exact Implementation and Integer Games

The first papers on incomplete information implementation focussed on exact implementation. Postlewaite and Schmeidler (1986) and Jackson (1991) identified a Bayesian monotonicity condition which (together with Bayesian incentive compatibility) was necessary and (under weak economic conditions) sufficient for exact implementation in Bayesian Nash equilibrium. Bergemann and Morris (2005b) provide a robust analogue of this result, showing that ex post incentive compatibility and a *robust monotonicity* condition are necessary and - under weak economic conditions - sufficient for exact robust implementation. All these papers follow a tradition in the implementation literature of allowing very badly behaved mechanisms, like integer games, in proving their general results. In this paper, we follow Abreu and Matsushima (1992c) in restricting attention to finite - and thus well-behaved - mechanisms. We briefly discuss the relation between these results in this section: a more complete and formal discussion in contained in appendix 8.3.

Robust measurability and robust monotonicity turn out to be equivalent in the important class of aggregator single crossing preferences (see appendix 8.4). However, in general, one can show by example that

---

[14]Chung and Ely (2001) earlier noted that the efficient outcome was the only one surviving iterated deletion of weakly dominated strategies in the original fully efficient auction without the modification to generate strict EPIC. We discuss the relation in section 6.3.

robust measurability neither implies nor is implied by robust monotonicity. Thus requiring only virtual implementation is sometimes a strict relaxation; and allowing badly-behaved mechanisms is sometimes a strict relaxation. We do not have a characterization of when exact robust implementation by a well behaved mechanism is possible (just as analogous characterizations do not exist for complete information and classical Bayesian implementation). We know only that robust measurability, robust monotonicity and strict ex post incentive compatibility will all be necessary.

We restrict attention in our analysis to social choice functions rather than social choice correspondences. Bergemann and Morris (2005c) considered the problem of *partially* robustly implementing a social choice correspondence, i.e., ensuring that whatever players' beliefs and higher order beliefs about others' types, there is *an* equilibrium leading to outcomes contained in the social choice correspondence. In the special case where the social choice correspondence is a function (and more generally in a class of separable environments), this is possible only if the function (or a selection from the correspondence in separable environments) is ex post incentive compatible. But in the general case, we do not have a satisfactory characterization of when partial robust implementation is possible. For this reason, we have not even attempted a characterization of (full) robust implementation of social choice correspondences.

7   Conclusion

In an environment with interdependent preferences we introduced a notion of strategic distinguishability by saying that two payoff types of an agent can be distinguished if they have disjoint rationalizable actions in some finite game for all possible beliefs and higher order beliefs about others' types. Conversely, a pair of payoff types are strategically *indistinguishable* if in every game, there exists some action which each type might rationally choose given some beliefs and higher order beliefs. We provided an exact and insightful characterization of strategic distinguishability.

The notion of strategic distinguishability is related to the idea of incentive compatibility in the context of information revelation in a mechanism. The difference between distinguishability and incentive compatibility arises from the two central features of strategic distinguishability. First, we say that two payoff types can be strategically distinguished if there exists *some* mechanism and hence *some* outcome function for which the types have disjoint rationalizable actions. In contrast, the analysis of incentive compatibility is typically concerned with a specific mechanism and hence a specific outcome function. Second, strategic distinguishability requires that the two payoff types display disjoint rationalizable actions for *all possible* beliefs and higher order beliefs. In contrast, the analysis of incentive compatibility is typically concerned with a fixed and common prior belief of the agents.

Despite this distinct perspective suggested by the notion of strategic distinguishability, we then showed that strategic distinguishability plays an important and natural role in the robust version of virtual implementation. By virtual implementation of a social choice function f, we require that a given social choice function is only realized with probability $1 - \varepsilon$ for every $\varepsilon > 0$. The link between strategic distinguishability and virtual implementation is established by the remaining ε probability. Here we are allowed to select an arbitrary outcome function, and in particular an outcome function which can identify strategically distin-

guishable types. Consequently, we show that a social choice function can be virtually implemented for all possible beliefs and higher order beliefs, i.e. it is robustly virtually implementable if and only if the social choice function is measurable which respect to strategically distinguishable types.

8   Appendix

8.1   Omitted Proofs

Proof of lemma 1.   Suppose that

$$\sum_{\theta_{-i}\in\Theta_{-i}} \lambda_i\left(\theta_{-i}\right) u_i\left(\overline{y},\left(\theta_i,\theta_{-i}\right)\right) \geq \sum_{\theta_{-i}\in\Theta_{-i}} \lambda_i\left(\theta_{-i}\right) u_i\left(x,\left(\theta_i,\theta_{-i}\right)\right) \tag{27}$$

for all $x \in X$. If

$$\sum_{\theta_{-i}\in\Theta_{-i}} \lambda_i\left(\theta_{-i}\right) u_i\left(\overline{y},\left(\theta_i,\theta_{-i}\right)\right) > \sum_{\theta_{-i}\in\Theta_{-i}} \lambda_i\left(\theta_{-i}\right) u_i\left(x',\left(\theta_i,\theta_{-i}\right)\right) \tag{28}$$

for some $x' \in X$, we could conclude, that

$$\sum_{\theta_{-i}\in\Theta_{-i}} \lambda_i\left(\theta_{-i}\right) u_i\left(\overline{y},\left(\theta_i,\theta_{-i}\right)\right) > \frac{1}{N}\sum_{x\in X}\sum_{\theta_{-i}\in\Theta_{-i}} \lambda_i\left(\theta_{-i}\right) u_i\left(x,\left(\theta_i,\theta_{-i}\right)\right) = \sum_{\theta_{-i}\in\Theta_{-i}} \lambda_i\left(\theta_{-i}\right) u_i\left(\overline{y},\left(\theta_i,\theta_{-i}\right)\right),$$

a contradiction. So (27) implies

$$\sum_{\theta_{-i}\in\Theta_{-i}} \lambda_i\left(\theta_{-i}\right) u_i\left(\overline{y},\left(\theta_i,\theta_{-i}\right)\right) = \sum_{\theta_{-i}\in\Theta_{-i}} \lambda_i\left(\theta_{-i}\right) u_i\left(x,\left(\theta_i,\theta_{-i}\right)\right) \tag{29}$$

for all $x \in X$. But (29) implies that R_{θ_i,λ_i} is indifferent between all pure outcomes and thus all lotteries. This contradicts assumption 1 on non=degeneracy. We conclude that the non-degeneracy assumption implies that (27) fails for all i, i.e., that for all i, $\theta_i \in \Theta_i$ and $\lambda_i \in \Delta\left(\Theta_{-i}\right)$, there exists $x \in X$ such that

$$\sum_{\theta_{-i}\in\Theta_{-i}} \lambda_i\left(\theta_{-i}\right) u_i\left(x,\left(\theta_i,\theta_{-i}\right)\right) > \sum_{\theta_{-i}\in\Theta_{-i}} \lambda_i\left(\theta_{-i}\right) u_i\left(\overline{y},\left(\theta_i,\theta_{-i}\right)\right). \tag{30}$$

Now suppose that the conclusion of the lemma fails, so that for all $\varepsilon > 0$, there exists i, $\theta_i \in \Theta_i$ and $\lambda_i \in \Delta\left(\Theta_{-i}\right)$ such that

$$\sum_{\theta_{-i}\in\Theta_{-i}} \lambda_i\left(\theta_{-i}\right) u_i\left(x,\left(\theta_i,\theta_{-i}\right)\right) \leq \sum_{\theta_{-i}\in\Theta_{-i}} \lambda_i\left(\theta_{-i}\right) u_i\left(\overline{y},\left(\theta_i,\theta_{-i}\right)\right) + \varepsilon$$

Thus there exists i and $\theta_i \in \Theta_i$ such that for each $\varepsilon > 0$, there exists $\lambda_i^\varepsilon \in \Delta\left(\Theta_{-i}\right)$ such that

$$\sum_{\theta_{-i}\in\Theta_{-i}} \lambda_i^\varepsilon\left(\theta_{-i}\right) u_i\left(x,\left(\theta_i,\theta_{-i}\right)\right) \leq \sum_{\theta_{-i}\in\Theta_{-i}} \lambda_i^\varepsilon\left(\theta_{-i}\right) u_i\left(\overline{y},\left(\theta_i,\theta_{-i}\right)\right) + \varepsilon$$

for all $x \in X$. The sequence λ_i^ε has a convergent subsequence with limit λ_i^* and

$$\sum_{\theta_{-i}\in\Theta_{-i}} \lambda_i^*\left(\theta_{-i}\right) u_i\left(x,\left(\theta_i,\theta_{-i}\right)\right) \leq \sum_{\theta_{-i}\in\Theta_{-i}} \lambda_i^*\left(\theta_{-i}\right) u_i\left(\overline{y},\left(\theta_i,\theta_{-i}\right)\right)$$

for all $x \in X$. This contradicts (30).   ∎

Proof of lemma 4.   By definition, type set profile Ψ_{-i} separates Ψ_i if, for every $R \in \mathcal{R}$, there exists $\theta_i \in \Psi_i$ such that $R_{\theta_i,\lambda_i} \neq R$ for every $\lambda_i \in \Delta\left(\Psi_{-i}\right)$. Write

$$X = \{x_1, x_2, ..., x_N\}, \quad \Theta_i = \left\{\theta_i^1, \theta_i^2, ..., \theta_i^L\right\}, \quad \text{and } \Theta_{-i} = \left\{\theta_{-i}^1, \theta_{-i}^2, ..., \theta_{-i}^M\right\}, \text{ with } M = L^{I-1}.$$

The vector

$$v_{lm} = \left(u_i \left(x^n, \left(\theta_i^l, \theta_{-i}^m \right) \right) \right)_{n=1}^N,$$

is an element of $\mathbb{R}^N$. Without loss of generality (since expected utility preferences can be represented by any affine transformation), we can assume that each v_{lm} is an element of the N dimensional simplex Δ^N. Now $(v_{lm})_{m=1}^M$ is a collection of M elements of Δ^N, and the set of preferences

$$\left\{ R_{\theta_i^l, \lambda_i} : \lambda_i \in \Delta \left(\Psi_{-i} \right) \right\},$$

are represented by the convex hull of $(v_{lm})_{m=1}^M$, which we write as

$$V_l = \mathrm{conv} \left((v_{lm})_{m=1}^M \right) \subseteq \Delta^N.$$

Thus Ψ_{-i} separates Ψ_i exactly if

$$\bigcap_{l=1}^L V_l = \varnothing.$$

By proposition 3, this is true if and only if there exist $z_1, ..., z_L \in \mathbb{R}^N$ such that

$$\sum_{l=1}^L z_l = 0, \tag{31}$$

and

$$\sum_{n=1}^N v_n z_{ln} > 0, \tag{32}$$

for each l and $v \in V_l$. But if $(z_l)_{l=1}^L$ satisfy (31) and (32), we may choose $\varepsilon > 0$ sufficiently small such that

$$\widetilde{y} \left(\theta_i^l \right) = \overline{y} + \varepsilon z_l \in Y \quad \text{for each } l,$$

and we have established (6) and (7).

Conversely, if (6) and (7) hold and we set $z_l = \widetilde{y} \left(\theta_i^l \right) - \overline{y}$ for $l = 1, ..., L$, then $(z_l)_{l=1}^L$ satisfy (31) and (32). ∎

Proof of lemma 6. Fix any $m_i \notin S_i^{\mathcal{M}} (\theta_i)$. Then there exists k such that $m_i \in S_i^{\mathcal{M}, k} (\theta_i)$ and $m_i \notin S_i^{\mathcal{M}, k+1} (\theta_i)$. Consider

$$\Delta_i^k = \left\{ \mu_i \in \Delta \left(\Theta_{-i} \times M_{-i} \right) \Big| \mu_i \left(\theta_{-i} \times m_{-i} \right) > 0 \ \Rightarrow \ m_{-i} \in S_{-i}^{\mathcal{M}, k} \left(\theta_{-i} \right) \text{ for each } j \neq i \right\}.$$

For all $\mu_i \in \Delta_i^k$, there exists $\overline{m}_i$ such that

$$\sum_{\theta_{-i}, m_{-i}} \mu_i \left(\theta_{-i}, m_{-i} \right) u_i \left(g \left(\overline{m}_i, m_{-i} \right), \left(\theta_i, \theta_{-i} \right) \right) > \sum_{\theta_{-i}, m_{-i}} \mu_i \left(\theta_{-i}, m_{-i} \right) u_i \left(g \left(m_i, m_{-i} \right), \left(\theta_i, \theta_{-i} \right) \right).$$

By compactness of Δ_i^k, there exists $\overline{\varepsilon}_i (m_i) > 0$ such that for all $\mu_i \in \Delta_i^k$ there exists $\overline{m}_i$ such that

$$\sum_{\theta_{-i}, m_{-i}} \mu_i \left(\theta_{-i}, m_{-i} \right) u_i \left(g \left(\overline{m}_i, m_{-i} \right), \left(\theta_i, \theta_{-i} \right) \right) > \sum_{\theta_{-i}, m_{-i}} \mu_i \left(\theta_{-i}, m_{-i} \right) u_i \left(g \left(m_i, m_{-i} \right), \left(\theta_i, \theta_{-i} \right) \right) + \overline{\varepsilon}_i (m_i).$$

Now let

$$\eta_{\mathcal{M}} = \min_{i, \; \theta_i \text{ and } m_i \notin S_i^{\mathcal{M}}(\theta_i)} \overline{\varepsilon}_i(m_i).$$

∎

Proof of lemma 7. Suppose $\pi^+ C \leq \pi^0 \eta_{\mathcal{M}^0}$. We will argue, by induction on k, that

$$\left(m_i^0, m_i^1\right) \in S_i^{\mathcal{M},k}(\theta_i) \Rightarrow m_i^0 \in S_i^{\mathcal{M}^0,k}(\theta_i)$$

for all $k \geq 0$. This is true by definition for $k = 0$; suppose that it is true for k. Now suppose that $m_i^0 \notin S_i^{\mathcal{M}^0,k+1}(\theta_i)$ but $\left(m_i^0, m_i^1\right) \in S_i^{\mathcal{M},k+1}(\theta_i)$ and so $\left(m_i^0, m_i^1\right) \in S_i^{\mathcal{M},k}(\theta_i)$ and - by the inductive hypothesis - $m_i^0 \in S_i^{\mathcal{M}^0,k}(\theta_i)$. Now fix any $\mu_i \in \Delta\left(\Theta_{-i} \times M_{-i}\right)$ satisfying

$$\mu_i\left(\theta_{-i}, \left(m_j^0, m_j^1\right)_{j \neq i}\right) > 0 \Rightarrow \left(m_j^0, m_j^1\right)_{j \neq i} \in S_{-i}^{\mathcal{M},k}(\theta_{-i}) \Rightarrow m_{-i}^0 \in S_{-i}^{\mathcal{M}^0,k}(\theta_{-i}).$$

Let

$$\overline{\mu}_i\left(\theta_{-i}, m_{-i}^0\right) = \sum_{\left(m_j^1\right)_{j \neq i} \in M_{-i}^1} \mu_i\left(\theta_{-i}, \left(m_j^0, m_j^1\right)_{j \neq i}\right).$$

By lemma 6, there exists $\overline{m}_i^0$ such that:

$$\sum_{\theta_{-i}, m_{-i}^0} \overline{\mu}_i\left(\theta_{-i}, m_{-i}^0\right) u_i\left(g^0\left(\overline{m}_i^0, m_{-i}^0\right), (\theta_i, \theta_{-i})\right) - \sum_{\theta_{-i}, m_{-i}^0} \overline{\mu}_i\left(\theta_{-i}, m_{-i}^0\right) u_i\left(g^0\left(m_i^0, m_{-i}^0\right), (\theta_i, \theta_{-i})\right) > \eta_{\mathcal{M}^0}.$$

Thus

$$\sum_{\theta_{-i}, m_{-i}} \mu_i\left(\theta_{-i}, m_{-i}\right) u_i\left(g\left(\left(\overline{m}_i^0, m_i^1\right), m_{-i}\right), (\theta_i, \theta_{-i})\right) - \sum_{\theta_{-i}, m_{-i}} \mu_i\left(\theta_{-i}, m_{-i}\right) u_i\left(g\left(\left(m_i^0, m_i^1\right), m_{-i}\right), (\theta_i, \theta_{-i})\right)$$

$$> \quad \pi^0 \eta_{\mathcal{M}^0} - \pi^+ C$$

$$\geq \quad 0.$$

This contradicts our premise that $\left(m_i^0, m_i^1\right) \in S_i^{\mathcal{M},k+1}(\theta_i)$. We conclude that $\left(m_i^0, m_i^1\right) \in S_i^{\mathcal{M},k+1}(\theta_i) \Rightarrow m_i^0 \in S_i^{\mathcal{M}^0,k+1}(\theta_i)$. ∎

8.2 Formal Statement and Proof of Epistemic Result

A type space $\mathcal{T}$ is defined by $\mathcal{T} = \left(T_i, \widehat{\theta}_i, \widehat{\pi}_i\right)_{i=1}^I$, where T_i is a countable set, $\widehat{\theta}_i : T_i \rightarrow \Theta_i$ and $\widehat{\pi}_i : T_i \rightarrow \Delta\left(T_{-i}\right)$.[15] A type space $\mathcal{T}$ and a mechanism $\mathcal{M} = (M, g)$ define a game where a behavioral strategy for agent i is a mapping $\sigma_i : T_i \rightarrow \Delta\left(M_i\right)$ and the interim equilibrium payoff of type t_i of agent i if he sends message m_i and other agents follow strategies σ_{-i} is

$$U_i\left(m_i, \sigma_i, t_i\right) \equiv \sum_{t_{-i} \in T_{-i}} \widehat{\pi}_i(t_i)[t_{-i}] \sum_{m_{-i} \in M_{-i}} \left(\prod_{j \neq i} \sigma_j(t_j)[m_j]\right) u_i\left(g\left(m_i', m_{-i}\right), \left(\widehat{\theta}_i(t_i), \widehat{\theta}_{-i}(t_{-i})\right)\right).$$

---

[15] The arguments extend to uncountable type spaces. We focus on countable type spaces for notational convenience.

Strategy profile $\sigma = (\sigma_i)_{i=1}^I$ is an *interim equilibrium* of $(\mathcal{T}, \mathcal{M})$ if

$$\sigma_i(t_i)[m_i] > 0 \Rightarrow m_i \in \underset{m_i'}{\arg\max} \, U_i(m_i', \sigma_i, t_i)$$

for all i, $t_i \in T_i$ and $m_i \in M_i$. This is the standard definition of Bayesian Nash equilibrium, expressed in terms of agents interim incentives to choose an optimal action. Without a common prior on the type space, equilibrium cannot be given an ex ante formulation.

Proposition 6 $m_i \in S_i^{\mathcal{M}}(\theta_i)$ *if and only if there exists (i) a type space $\mathcal{T} = \left(T_i, \widehat{\theta}_i, \widehat{\pi}_i\right)_{i=1}^I$; (ii) an interim equilibrium σ of $(\mathcal{T}, \mathcal{M})$ and (iii) a type $t_i \in T_i$ with $\widehat{\theta}_i(t_i) = \theta_i$ and $\sigma_i(t_i)[m_i] > 0$.*

Proof. Fix (i) a type space $\mathcal{T} = \left(T_i, \widehat{\theta}_i, \widehat{\pi}_i\right)_{i=1}^I$; (ii) an interim equilibrium σ of $(\mathcal{T}, \mathcal{M})$ and (iii) a type $t_i^* \in T_i$ with (a) $\widehat{\theta}_i(t_i^*) = \theta_i^*$; and (b) $\sigma_i(t_i^*)[m_i^*] > 0$. Let

$$S_i^*(\theta_i) = \left\{ m_i \in M_i | \exists t_i \in T_i \text{ s.t. } \sigma_i(t_i)[m_i] > 0 \text{ and } \widehat{\theta}_i(t_i) = \theta_i \right\}.$$

Now for each i, θ_i in the range of $\widehat{\theta}_i$ and $m_i \in S_i^*(\theta_i)$, let

$$\lambda_i^{\theta_i, m_i}(\theta_{-i}, m_{-i}) = \sum_{\left\{ t_{-i} \in T_{-i} \,:\, \widehat{\theta}_{-i}(t_{-i}) = \theta_{-i} \right\}} \widehat{\pi}_i(t_i)[t_{-i}] \sum_{m_{-i} \in M_{-i}} \left(\prod_{j \neq i} \sigma_j(t_j)[m_j] \right). \tag{33}$$

Now because σ is an equilibrium,

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

Now we show by induction on k that $S_i^*(\theta_i) \subseteq S_i^{\mathcal{M}, k}(\theta_i)$ for all i, θ_i and k. This is true for $k = 0$ by definition. Suppose that it is true for k. Now $\lambda_i^{\theta_i, m_i}(\theta_{-i}, m_{-i}) > 0$ implies that $m_{-i} \in S_{-i}^*(\theta_{-i})$, by construction, which implies $m_{-i} \in S_{-i}^{\mathcal{M}, k}(\theta_{-i})$ by the inductive hypothesis. Together with (33), this establishes $m_i \in S_i^{\mathcal{M}, k+1}(\theta_i)$. This proves the induction. Now $m_i^* \in S_i^*(\theta_i^*) \subseteq S_i^{\mathcal{M}}(\theta_i^*)$, proving the "if" claim of the proposition.

Conversely, suppose that $m_i^* \in S_i^{\mathcal{M}}(\theta_i^*)$. Observe that for each i, θ_i and $m_i \in S_i^{\mathcal{M}}(\theta_i)$, there exists $\lambda_i^{\theta_i, m_i} \in \Delta(\Theta_{-i} \times M_{-i})$ such that:

(a) $\lambda_i^{\theta_i, m_i}(\theta_{-i}, m_{-i}) > 0 \Rightarrow m_{-i} \in S_{-i}^{\mathcal{M}}(\theta_{-i})$
(b) $m_i \in \arg\max_{m_i'} \sum_{\theta_{-i}, m_{-i}} \lambda_i^{\theta_i, m_i}(\theta_{-i}, m_{-i}) u_i(g(m_i', m_{-i}), (\theta_i, \theta_{-i}))$

Now construct (i) a type space $\mathcal{T}$ with

$$T_i = \left\{ (\theta_i, m_i) \in \Theta_i \times M_i \, | \, m_i \in S_i^{\mathcal{M}}(\theta_i) \right\},$$
$$\widehat{\theta}_i((\theta_i, m_i)) = \theta_i, \text{ and}$$
$$\widehat{\pi}_i((\theta_i, m_i)) \left[(\theta_j, m_j)_{j \neq i} \right] = \lambda_i^{\theta_i, m_i}(\theta_{-i}, m_{-i});$$

((a) above ensures that this is well-defined) and (ii) a strategy profile σ with

$$\sigma_i((\theta_i, m_i))[m_i'] = \begin{cases} 1, & \text{if } m_i' = m_i \\ 0, & \text{otherwise} \end{cases}.$$

Now (b) ensures that σ is an equilibrium and by construction $t_i^* = (\theta_i^*, m_i^*) \in T_i$ with $\widehat{\theta}_i(t_i^*) = \theta_i$; and $\sigma_i(t_i^*)[m_i^*] > 0$. This establishes the "only if" part of the proposition. $\blacksquare$

This straightforward incomplete information generalization of Brandenburger and Dekel (1987) was first stated to our knowledge by Battigalli (1998). A more general version of this result is reported as propositions 4.2 and 4.3 in Battigalli and Siniscalchi (2003), where they allow for additional restrictions ("Δ") on beliefs. The above result is implied by Battigalli and Siniscalchi (2003) when Δ is empty.

8.3   Exact Implementation

In this section, we formally put the contribution of this paper in the context of the larger implementation literature.

The classical complete information literature identified a monotonicity condition ("Maskin monotonicity") on social choice correspondences that is necessary and - under a weak economic condition - sufficient for full implementation in Nash equilibrium. These arguments were generalized to a standard incomplete information environment by Postlewaite and Schmeidler (1986) and Jackson (1991) who showed that Bayesian incentive compatibility and a "Bayesian monotonicity" condition were necessary and - again, under some weak economic assumption - sufficient for full implementation. The sufficiency parts of all these results have been widely criticized for relying on badly behaved mechanisms, i.e., integer games, with pathological features, i.e., the non-existence of best responses to some conjectures (see, e.g., Jackson (1992)).

In a complete information environment, Abreu and Matsushima (1992b) showed that if

1. the outcome space is a lottery space over finite outcomes and

2. the implementation notion is weakened to require only virtual implementation; but

3. the solution concept is strengthened to require implementation in iterated deletion of strictly dominated strategies and

4. mechanisms are restricted to be "well-behaved,"[16]

then essentially all social choice functions could be implemented. Abreu and Matsushima (1992c) generalized these results to incomplete information, showing that Bayesian incentive compatibility and a measurability condition ("AM measurability") were necessary and sufficient for virtual (Bayesian) implementation in iterated deletion of (interim) strictly dominated strategies. Notice that (1) is a domain restriction; (2) is a restriction that makes it easier to obtain a positive result; (3) and (4) are restrictions that make it harder to obtain a positive result.

A natural question to ask is what happens with different combinations of assumptions (2), (3) and (4). Abreu and Matsushima (1992c) showed that Bayesian incentive compatibility and AM measurability are necessary conditions for virtual implementation in Bayesian Nash equilibrium in well-behaved mechanisms. Serrano and Vohra (2005) describe a "virtual monotonicity" condition - a weakening of the Bayesian

---

[16] Their results are true whether one restricts attention to finite mechanisms or allow more general compact mechanisms where best responses are well defined for all conjectures.

monotonicity condition of Postlewaite and Schmeidler (1986) and Jackson (1991) - which, together with Bayesian incentive compatibility, is necessary and sufficient for virtual implementation in Bayesian Nash equilibrium using perhaps badly behaved mechanisms. Virtual monotonicity must therefore be a weakening of AM measurability. Example 2 in Serrano and Vohra (2001) exhibits an environment where all non-constant social choice functions fail AM measurability fails but all social choice functions satisfy virtual monotonicity and many satisfy Bayesian monotonicity. On the other hand, the social choice function allocating a single object efficiently under private values will fail Bayesian monotonicity (any efficient allocation mechanism will allow undesirable equilibria) but will satisfy AM measurability. Thus Bayesian monotonicity neither implies nor is implied by AM measurability.

We derived the robust analogue of Jackson (1991) and the exact Bayesian implementation literature in Bergemann and Morris (2005b). The environment was as in this paper, except that the outcome space Y was not restricted to be a lottery space. We showed that ex post incentive compatibility and a robust monotonicity condition were necessary and - under a weak economic condition - sufficient for full implementation on all possible type spaces. As in the classic implementation literature, we did not restrict attention to "well-behaved" mechanisms. To define the robust monotonicity condition, let $\beta_i : \Theta_i \to 2^{\Theta_i}$ with $\theta_i \in \beta_i(\theta_i)$; the interpretation is that $\beta_i(\theta_i)$ describes the set of types that type θ_i might report himself to be. A *deception profile* β is given by $\beta = (\beta_1, ..., \beta_I)$. A deception β is *acceptable* if agents' reports will always lead to the right outcome if they are interpreted truthfully, i.e., $\theta'_i \in \beta_i(\theta_i)$ for each $i \Rightarrow f(\theta') = f(\theta)$. Now:

Definition 11 (Robust Monotonicity)
Social choice function f satisfies robust monotonicity if for every unacceptable deception β, there exist i, θ_i, $\theta'_i \in \beta_i(\theta_i)$ such that, for all $\theta'_{-i} \in \Theta_{-i}$, there exists y such that

$$u_i(y, (\theta_i, \theta_{-i})) > u_i(f(\theta'_i, \theta'_{-i}), (\theta_i, \theta_{-i})) \tag{34}$$

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

$$u_i(f(\theta''_i, \theta'_{-i}), (\theta''_i, \theta'_{-i})) \geq u_i(y, (\theta''_i, \theta'_{-i})) \tag{35}$$

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

Under the other necessary condition of ex post incentive compatibility, this condition - like the robust measurability condition of this paper - has the interpretation that it requires that there be not too much interdependence in preferences. To see why, consider the extreme case of private values, when we remove the dependence of utility on others' types. In this case, ex post incentive compatibility becomes dominant strategies incentive compatibility,

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

for all i, θ_i, θ'_i and θ_{-i}; and robust monotonicity reduces to the requirement that for every unacceptable deception β, there exist i, θ_i, $\theta'_i \in \beta_i(\theta_i)$ such that

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

for all $\theta'_{-i} \in \Theta_{-i}$. The latter requirement is simply a requirement that ex post incentive constraints be strict often enough. The robust monotonicity requirement adds little to incentive compatibility conditions under private values. As values become more interdependent, the set of rewards y that can be offered in definition 11 are reduced, and the condition becomes stronger.

In section 8.4, we show that in a large class of economic significant environments, robust monotonicity and robust measurability have an identical "not too much interdependence" characterization. However, in the remainder of this section we document that robust measurability neither implies nor is implied by robust monotonicity. This observation parallels our summary observation above about the standard Bayesian implementation literature: AM measurability neither implies nor is implied by Bayesian monotonicity.

8.3.1 Example 1: Robust Measurability holds while Robust Monotonicity fails

Consider an environment with two agents, a and b. The payoff type space of each agent i is given by $\Theta_i = \{\theta_i^1, \theta_i^2, \theta_i^3\}$. The allocation space is given by $X = \{x_1, x_2, x_3, x_4\}$. Below we display the payoff of agent a. The payoffs for agent b are symmetric.

x_1	θ_b^1	θ_b^2	θ_b^3
θ_a^1	1	1	1
θ_a^2	1	1	0
θ_a^3	0	0	1

x_2	θ_b^1	θ_b^2	θ_b^3
θ_a^1	0	0	1
θ_a^2	1	1	1
θ_a^3	1	1	0

x_3	θ_b^1	θ_b^2	θ_b^3
θ_a^1	1	1	0
θ_a^2	0	0	1
θ_a^3	1	1	1

x_4	θ_b^1	θ_b^2	θ_b^3
θ_a^1	$\frac{2}{3}+\varepsilon$	$\frac{2}{3}-\varepsilon$	$\frac{2}{3}$
θ_a^2	$\frac{2}{3}+\varepsilon$	$\frac{2}{3}-\varepsilon$	$\frac{2}{3}$
θ_a^3	$\frac{2}{3}+\varepsilon$	$\frac{2}{3}-\varepsilon$	$\frac{2}{3}$

$$(36)$$

With the payoff matrix given by (36), it is easy to characterize pairwise separability. Observe that Ψ_j separates Ψ_i whenever $\#\Psi_i \geq \#\Psi_j$ and $\#\Psi_i > 1$ but not if $\#\Psi_i < \#\Psi_j$ or $\#\Psi_i = 1$. Thus while the 0th level inseparable sets of agent i, Ξ_i^0, will consist of all sets of types, the 1st level inseparable sets of agent i, Ξ_i^1, will consist of all sets of types with cardinality at most 2, and the 2nd level inseparable sets of agent i, Ξ_i^2, will consist exactly of all singletons. Thus all pairs of distinct types are pairwise separable and thus strategically distinguishable, and so any social choice function will satisfy robust measurability.

We will show that robust monotonicity fails for some f. In fact, we will show that a weaker property implied by robust monotonicity in the context of lotteries fails:

Definition 12 (Directional Robust Monotonicity)
Social choice function f satisfies directional robust monotonicity if for every unacceptable deception β, there exist i, θ_i, $\theta'_i \in \beta_i(\theta_i)$ such that, for all $\theta'_{-i} \in \Theta_{-i}$, there exists y such that

$$u_i(y, (\theta_i, \theta_{-i})) > u_i(\overline{y}, (\theta_i, \theta_{-i})) \tag{37}$$

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

$$u_i(\overline{y}, (\theta'_i, \theta'_{-i})) \geq u_i(y, (\theta'_i, \theta'_{-i})). \tag{38}$$

Consider the deception with $\beta_i(\theta_i) = \Theta_i$ for all i and θ_i. By symmetry of the payoffs and the deceptions across states, it suffices to consider a single type profile $\theta_i = \theta_a^1$ and $\theta'_i = \theta_a^2$ and likewise $\theta'_{-i} = \theta_b^2$. We

require that there exists y such that - by (37) -

$$
\begin{aligned}
u_a\left(y,\left(\theta_a^1,\theta_b^1\right)\right) &> u_a\left(\overline{y},\left(\theta_a^1,\theta_b^1\right)\right), \\
u_a\left(y,\left(\theta_a^1,\theta_b^2\right)\right) &> u_a\left(\overline{y},\left(\theta_a^1,\theta_b^2\right)\right), \\
u_a\left(y,\left(\theta_a^1,\theta_b^3\right)\right) &> u_a\left(\overline{y},\left(\theta_a^1,\theta_b^3\right)\right),
\end{aligned}
\tag{39}
$$

and

$$
u_a\left(\overline{y},\left(\theta_a^2,\theta_b^2\right)\right) \geq u_a\left(y,\left(\theta_a^2,\theta_b^2\right)\right).
\tag{40}
$$

We rewrite the inequalities (39)-(40) explicitly in terms of payoffs and probabilities with $y = (y_1,y_2,y_3,y_4) = (y_1,y_2,y_3,1-y_1-y_2-y_3)$ and $\overline{y} = \left(\frac{1}{4},\frac{1}{4},\frac{1}{4},\frac{1}{4}\right)$.

$$
\begin{aligned}
y_1 + y_3 + \left(\frac{2}{3}+\varepsilon\right)(1-y_1-y_2-y_3) &> \frac{1}{4}+\frac{1}{4}+\frac{1}{4}\left(\frac{2}{3}+\varepsilon\right), \\
y_1 + y_3 + \left(\frac{2}{3}-\varepsilon\right)(1-y_1-y_2-y_3) &> \frac{1}{4}+\frac{1}{4}+\frac{1}{4}\left(\frac{2}{3}-\varepsilon\right), \\
y_1 + y_2 + \left(\frac{2}{3}\right)(1-y_1-y_2-y_3) &> \frac{1}{4}+\frac{1}{4}+\frac{1}{4}\frac{2}{3},
\end{aligned}
\tag{41}
$$

and

$$
\frac{1}{4}+\frac{1}{4}+\frac{1}{4}\left(\frac{2}{3}-\varepsilon\right) \geq y_1 + y_2 + \left(\frac{2}{3}-\varepsilon\right)(1-y_1-y_2-y_3).
\tag{42}
$$

But now we have a contradiction to the condition of directional robust monotonicity as (41) and (42) jointly cannot be satisfied.

8.3.2 Example 2: Robust Measurability fails but Robust Monotonicity holds

There are two agents 1 and 2 and each agent i has two types, θ_i and θ_i'. There are six pure outcomes, $X = \{x_1,x_2,x_3,x_4,x_5,x_6\}$. The state dependent utility of agents 1 and 2 are depicted in the following tables:

x_1	θ_2	θ_2'
θ_1	1,1	1,1
θ_1'	1,1	1,1

x_2	θ_2	θ_2'
θ_1	0,0	0,0
θ_1'	0,0	0,0

x_3	θ_2	θ_2'
θ_1	2,2	$-2,0$
θ_1'	0,0	0,2

x_4	θ_2	θ_2'
θ_1	$-2,0$	2,2
θ_1'	0,2	0,0

x_5	θ_2	θ_2'
θ_1	0,0	0,2
θ_1'	2,2	$-2,0$

x_6	θ_2	θ_2'
θ_1	0,2	0,0
θ_1'	$-2,0$	2,2

Suppose agent 1 assigns equal probability to each type of agent 2. Then - whatever his type - his expected utility from allocations $(x_1,x_2,x_3,x_4,x_5,x_6)$ are $(1,0,0,0,0,0)$. Thus $\{\theta_2,\theta_2'\}$ does not separate $\{\theta_1,\theta_1'\}$. A symmetric argument establishes that $\{\theta_1,\theta_1'\}$ does not separate $\{\theta_2,\theta_2'\}$. We conclude that no pair of types of each agent are pairwise separable and only constant social choice functions satisfy robust measurability.

But consider the following (Pareto-efficient) social choice function:

f	θ_2	θ_2'
θ_1	x_3	x_4
θ_1'	x_5	x_6

Strict ex post incentive compatibility and robust monotonicity both hold. To verify the latter, observe that consider any deception with $\beta_1\left(\tilde{\theta}_1\right) = \Theta_1$ for some $\tilde{\theta}_1$. Without loss of generality, assume that $\beta_1\left(\theta_1\right) = \Theta_1$. Type θ_1 reporting θ_1' can be offered outcome a to report the deception. Now consider any deception with $\beta_1\left(\tilde{\theta}_1\right) = \left\{\tilde{\theta}_1\right\}$ for each $\tilde{\theta}_1$ and $\beta_2\left(\tilde{\theta}_2\right) = \Theta_2$ for some $\tilde{\theta}_2$. Without loss of generality, assume that $\beta_2\left(\theta_2\right) = \Theta_2$. Type θ_2 reporting θ_2' receiving report θ_1' can be offered outcome $f\left(\theta_1', \theta_2\right)$ to report the deception.

Results in Bergemann and Morris (2005b) show that robust implementation by a perhaps badly behaved mechanism must be possible. We can illustrate this result with the following mechanism. Agent 1 sends a message $m_1 \in M_1 = \left\{\theta_1, \theta_1'\right\} \cup \{1, 2, 3, ...\}$, agent 2 sends a message $m_2 \in M_2 = \left\{\theta_2, \theta_2'\right\}$. If $m_1 \in \left\{\theta_1, \theta_1'\right\}$, then $g\left(m_1, m_2\right) = f\left(m_1, m_2\right)$; if $m_1 \in \{1, 2, 3, ...\}$, then $g\left(m_1, m_2\right)$ is the lottery putting probability $\frac{1}{m_1}$ on x_2 and probability $1 - \frac{1}{m_1}$ on x_1. Now truth-telling survives iterated deletion of never best responses. Also, (i) sending message 2 (expected payoff: $\frac{1}{2}$) is always a better response for agent 1 than mis-reporting his type (payoff: 0) and (ii) choosing $m_1 + 1$ (expected payoff: $\frac{m_1}{m_1+1}$) is always a better response for agent 1 than sending message m_1 (payoff: $\frac{m_1-1}{m_1}$). So player 1 must tell the truth. Truth-telling is then the only best response for agent 2.

8.4 Aggregator Single Crossing Preferences

In this subsection, we consider the implications of the results in this paper for a class of economic environments introduced in Bergemann and Morris (2005a).

The main structural assumption in this subsection is the existence of a monotonic aggregator $h_i\left(\theta\right)$ for each i which aggregates the types of the agents. The aggregator $h_i\left(\theta\right)$ serves as a sufficient statistic of the entire type profile $\theta = \left(\theta_1, ..., \theta_I\right)$ in determination of the preference of agent i.

Definition 13 (Aggregator Single Crossing Preferences)
Preferences are Aggregator Single Crossing (ASC) if, for each i, there exists $h_i : \Theta \to \mathbb{R}$ and $v_i : Y \times \mathbb{R} \to \mathbb{R}$ such that

1. *h_i is strictly increasing in θ_i and increasing in θ_{-i};*

2. *$v_i\left(y, z\right)$ is continuous, an expected utility functional on Y, for each $z \in \mathbb{R}$, and strict single crossing, i.e.,*
$$v_i\left(y, z'\right) = v_i\left(y', z'\right) \Rightarrow \text{sign}\left(v_i\left(y, z\right) - v_i\left(y', z\right)\right) = -\text{sign}\left(v_i\left(y, z''\right) - v_i\left(y', z''\right)\right)$$
   for all $z > z' > z''$;

3. *for any θ_i and $\lambda_i \in \Delta\left(\Theta_{-i}\right)$, there exists $z \in \left[h_i\left(\theta_i, \min \text{supp}\left(\lambda_i\right)\right), h_i\left(\theta_i, \max \text{supp}\left(\lambda_i\right)\right)\right]$ such that $v\left(\cdot, z\right)$ represents R_{θ_i, λ_i}.*

We maintain this assumption on preferences throughout this section. Bergemann and Morris (2005a) introduced this class of preferences in a strictly more general environment than that in this paper: Y was an arbitrary compact space, each Θ_i was a compact space, each h_i was continuous in θ, each $h_i\left(\theta_i, \cdot\right)$ did not

need to be increasing in θ_{-i} and v_i was a continuous function (not necessarily an expected utility functional). Bergemann and Morris (2005a) showed that the following contraction property was equivalent to the robust monotonicity of all social choice functions. We write β^* for the "truth-telling" deception, i.e., $\beta_i^*\left(\theta_i\right) = \{\theta_i\}$ for each i and θ_i.

Definition 14 (Contraction Property)
The aggregator functions h satisfy the contraction property if, for all $\beta \neq \beta^*$, there exists i and $\theta_i' \in \beta_i\left(\theta_i\right)$ with $\theta_i' \neq \theta_i$, such that

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

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

So results in Bergemann and Morris (2005b) - described in section 8.3 - show that ex post incentive compatibility and the contraction property are necessary and - under a weak economic condition - sufficient for robust implementation in some (perhaps badly behaved) mechanism. Bergemann and Morris (2005a) consider well behaved (i.e. compact) mechanisms. Consider the following strict ex post incentive compatibility condition:

Definition 15 (Strict EPIC)
Social choice function f satisfies strict EPIC if for all i, θ_i and θ_{-i}:

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

if $f\left(\theta_i, \theta_{-i}'\right) \neq f\left(\theta_i', \theta_{-i}'\right)$ for some θ_{-i}'.

Bergemann and Morris (2005a) show that strict EPIC is necessary for robust implementation by a compact mechanism. Moreover, they show that strict EPIC and the contraction property are sufficient for robust implementation in the direct mechanism. With this background, we show what our results in this paper imply about ASC preferences.

We recall that Ξ^* is the limit of the kth level inseparable sets defined in (1) - (3). We say that there is *full separation* if each Ξ_i^* is the collection of singletons and thus every distinct pair of types of every agent is pairwise separable. We will show that the contraction property is equivalent to full separation.

Proposition 7 *The contraction property implies full separation.*

Proof. The proof is by contradiction. We suppose that the contraction property holds and full separation fails. By lemma 3, there exists mutually inseparable Ξ, and for some agent i there is a set $\Psi_i \in \Xi_i$ which is not a singleton. We now use this collection of sets Ξ to construct a deception as follows:

$$\beta_i\left(\theta_i\right) = \left\{\theta_i' \,\middle|\, \theta_i, \theta_i' \in \Psi_i \text{ for some } \Psi_i \in \Xi_i\right\}.$$

By construction, we have $\beta \neq \beta^*$. By the hypothesis of the contraction property, there exists i, θ_i and $\theta_i' \in \beta_i\left(\theta_i\right)$ such that for all θ_{-i} and all $\theta_{-i}' \in \beta_{-i}\left(\theta_{-i}\right)$, we have

$$\operatorname{sign}\left(\theta_i - \theta_i'\right) = \operatorname{sign}\left(h_i\left(\theta_i, \theta_{-i}\right) - h_i\left(\theta_i', \theta_{-i}'\right)\right). \tag{43}$$

By construction, we have $\theta_i, \theta_i' \in \Psi_i \in \Xi_i$. Now consider any $\Psi_{-i} \in \Xi_{-i}$. By construction of the deception β, we have

$$\Psi_{-i} = \beta_{-i}(\theta_{-i}) = \left\{ \theta_{-i}' : \theta_{-i} \in \beta_{-i}(\theta_{-i}') \right\}$$

for some $\theta_{-i} \in \Theta_{-i}$. We will show that Ψ_{-i} separates Ψ_i. Suppose without loss of generality that $\theta_i > \theta_i'$. Then it follows from (43) that for all $\theta_{-i}, \theta_{-i}' \in \Psi_{-i}$, we have $h_i(\theta_i, \theta_{-i}) > h_i(\theta_i', \theta_{-i}')$. In particular,

$$\min_{\theta_{-i} \in \Psi_{-i}} h_i(\theta_i, \theta_{-i}) > \max_{\theta_{-i}' \in \Psi_{-i}} h_i(\theta_i', \theta_{-i}'),$$

but that means that we can find a pair of allocations (y, y') such that we have a preference reversal for all types $\theta_{-i}, \theta_{-i}'$. Thus

$$\mathcal{R}_i(\theta_i, \Psi_{-i}) \cap \mathcal{R}_i(\theta_i', \Psi_{-i}) = \varnothing.$$

But now we have established that Ψ_{-i} separates Ψ_i for all $\Psi_{-i} \in \Xi_{-i}$. This contradicts our assumption that Ξ was mutually separable. ∎

Say that β is a *symmetric* deception if $\theta_i' \in \beta_i(\theta) \Rightarrow \theta_i \in \beta_i(\theta_i')$. To prove the converse result, we will exploit the following alternative characterization of the contraction property.

Lemma 10 *The contraction property holds if, for all symmetric $\beta \neq \beta^*$, there exists i and $\theta_i' \in \beta_i(\theta_i)$ with $\theta_i' \neq \theta_i$, such that*

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

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

Proposition 8 *Full separation implies the contraction property.*

Proof. Suppose full separation holds. Consider any $\beta \neq \beta^*$ and let

$$\Xi_i = \left\{ \theta_i' \,\middle|\, \theta_i' \in \beta_i(\theta_i) \text{ and } \theta_i' \leq \theta_i \text{ for some } \theta_i \in \Theta_i \right\} \cup \left\{ \theta_i' \,\middle|\, \theta_i' \in \beta_i(\theta_i) \text{ and } \theta_i' \geq \theta_i \text{ for some } \theta_i \in \Theta_i \right\}. \tag{44}$$

By full separation and lemma 3, there exists i and Ψ_i such that Ψ_{-i} separates Ψ_i for all $\Psi_{-i} \in \Xi_{-i}$. By the construction of Ξ, there exists θ_i^* such that either $\Psi_i = \left\{ \theta_i' \,\middle|\, \theta_i' \in \beta_i(\theta_i^*) \text{ and } \theta_i' \leq \theta_i^* \right\}$ or $\Psi_i = \left\{ \theta_i' \,\middle|\, \theta_i' \in \beta_i(\theta_i^*) \text{ and } \theta_i' \geq \theta_i^* \right\}$. Suppose without loss of generality that $\Psi_i = \left\{ \theta_i' \,\middle|\, \theta_i' \in \beta_i(\theta_i^*) \text{ and } \theta_i' \geq \theta_i^* \right\}$. Fix any $\Psi_{-i} \in \Xi_{-i}$. Since Ψ_{-i} separates Ψ_i, we have that

$$\bigcap_{\theta_i \in \Theta_i} \mathcal{R}_i(\theta_i, \Psi_{-i}) = \varnothing.$$

By the monotonicity of h_i, this implies that

$$h_i(\theta_i', \max \Psi_{-i}) < h_i(\theta_i, \min \Psi_{-i}).$$

By symmetry of β and construction of Ξ_{-i}, there exists $\theta_{-i} \in \Theta_{-i}$ such that

$$\theta_{-i} = \min \Psi_{-i}$$

and

$$\theta'_{-i} = \max \Psi_{-i} \in \beta_{-i} (\theta_{-i}).$$

So we have i, θ_i and $\theta'_i \in \beta_i (\theta_i)$ such that

$$\text{sign} (\theta_i - \theta'_i) = \text{sign} (h_i (\theta_i, \theta_{-i}) - h_i (\theta'_i, \theta'_{-i}))$$

for all θ_{-i} and $\theta'_{-i} \in \beta_{-i} (\theta_{-i})$. By lemma 10, we have established the contraction property. $\blacksquare$

Taken together, propositions 7 and 8 establish that - with ASC preferences - full separation is equivalent to the contraction property. Thus, for a social choice function that is everywhere sensitive to the state, robust measurability, robust monotonicity and the contraction property are all equivalent.

We now have that in the environment of this subsection, the results of this paper imply that robust virtual implementation is possible by some finite mechanism if and only if the contraction property holds and the social choice function satisfies ex post incentive compatibility. Bergemann and Morris (2005a) show that robust exact implementation is possible in the direct mechanism if and only if the contraction property holds and the social choice function satisfies strict ex post incentive compatibility.

Intuitively, it is easy to turn weak incentive constraints into strict incentive constraints if one can relax exact to virtual implementation. The following is an easy sufficient condition for doing this.

Definition 16 (Strict Ex Post Preferences)
Preferences are strict ex post (SEP) if there exists a *social choice function satisfying strict EPIC.*

Corollary 2 *If preferences are strict ex post and the social choice function satisfies EPIC and the contraction property, then the social choice function is robustly virtually implementable in the direct mechanism.*

Proof. If preferences are strict ex post, there exists a social choice function g which is strict EPIC. Now fix any social choice function f which satisfies EPIC and consider the social choice function $f^\varepsilon = \varepsilon g + (1 - \varepsilon) f$; f^ε satisfies strict EPIC. So, since the contraction property is satisfied, f^ε is (exactly) robustly implementable in the direct mechanism, by the main result in Bergemann and Morris (2005a). Thus f is robustly virtually implementable in the direct mechanism. $\blacksquare$

8.5 Strategic Equivalence

Proof of proposition 5. Fix any mechanism $\mathcal{M}$. We will show, by induction on k, that for each k, $\theta'_i \in P_i^k (\theta_i) \Rightarrow S_i^{\mathcal{M},k} (\theta'_i) = S_i^{\mathcal{M},k} (\theta_i)$. This is true by definition for $k = 0$. Suppose that it is true for k. Now fix any i, $\theta'_i \in P_i^{k+1} (\theta_i)$ and $m_i \in S_i^{\mathcal{M},k+1} (\theta_i)$. The latter property implies that there exists $\mu_i \in \Delta (\Theta_{-i} \times M_{-i})$ such that

$$\mu_i (\theta_{-i}, m_{-i}) > 0 \Rightarrow m_{-i} \in S_{-i}^{\mathcal{M},k} (\theta_{-i}) \tag{45}$$

and

$$m_i \in \underset{m'_i}{\arg\max} \sum_{\theta_{-i}, m_{-i}} \mu_i (\theta_{-i}, m_{-i}) u_i (g (m'_i, m_{-i}), (\theta_i, \theta_{-i})). \tag{46}$$

Let

$$\lambda_i\left(\theta_i\right) = \sum_{m_{-i} \in M_{-i}} \mu_i\left(\theta_{-i}, m_{-i}\right).$$

Now $\theta_i' \in P_i^{k+1}\left(\theta_i\right)$ implies that, for each $\Psi_{-i} \in \mathcal{P}_{-i}^k$, there exists $\lambda_{i,\Psi_{-i}}, \lambda_{i,\Psi_{-i}}' \in \Delta\left(\Psi_{-i}\right)$ such that $R_{\theta_i',\lambda_i'} = R_{\theta_i,\lambda_i}$. Thus there exist $\xi_i, \xi_i' \in \Delta\left(\mathcal{P}_{-i}^k\right)$ such that

$$\sum_{\Psi_{-i} \in \mathcal{P}_{-i}^k, \theta_{-i} \in \Theta_{-i},} \xi_i'\left(\Psi_{-i}\right) \lambda_{i,\Psi_{-i}}'\left(\theta_{-i}\right) u_i\left(\cdot, \left(\theta_i', \theta_{-i}\right)\right)$$

is an affine transformation of

$$\sum_{\Psi_{-i} \in \mathcal{P}_{-i}^k, \theta_{-i} \in \Theta_{-i},} \xi_i\left(\Psi_{-i}\right) \lambda_{i,\Psi_{-i}}\left(\theta_{-i}\right) u_i\left(\cdot, \left(\theta_i, \theta_{-i}\right)\right).$$

Now if we let

$$\mu_i'\left(\theta_{-i}, m_{-i}\right) = \begin{cases} \xi_i'\left(\Psi_{-i}\right) \lambda_{i,\Psi_{-i}}'\left(\theta_{-i}\right) \dfrac{\mu_i\left(\theta_{-i}, m_{-i}\right)}{\sum\limits_{\theta_{-i}'} \mu_i'\left(\theta_{-i}, m_{-i}\right)}, & \text{if } \sum\limits_{\theta_{-i}'} \mu_i'\left(\theta_{-i}, m_{-i}\right) > 0 \\ \\ 0, \text{ otherwise} \end{cases},$$

(45) implies

$$\mu_i'\left(\theta_{-i}, m_{-i}\right) > 0 \Rightarrow m_{-i} \in S_{-i}^{\mathcal{M},k}\left(\theta_{-i}\right);$$

and (46) and the above affine equivalence imply that

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

Thus $m_i \in S_i^{\mathcal{M},k+1}\left(\theta_i\right)$. ∎

8.6 Relaxing the Non-Degeneracy Assumption

We maintained assumption 1 on non-degeneracy throughout the paper. Our results can be extended to allow degenerate preferences. Write $\overline{R}$ to be the degenerate preference relation of complete indifference over all lotteries. Consider the following modified definition of kth level inseparable sets allowing the possibility of degenerate preferences:

$$\begin{aligned} \Xi_i^0 &= 2^{\Theta_i} \\ \Xi_i^{k+1} &= \left\{ \Psi_i \in \Xi_i^k \;\middle|\; \begin{array}{l} \exists \Psi_{-i} \in \Xi_{-i}^k \text{ and } R \in \mathcal{R} \setminus \{\overline{R}\} \text{ such that for all } \theta_i \in \Psi_i \\ \text{either (1) } \overline{R} \in \mathcal{R}_i\left(\theta_i, \Psi_{-i}'\right) \text{ for some } \Psi_{-i}' \in \Xi_{-i}^k \\ \text{or (2) } R \in \mathcal{R}_i\left(\theta_i, \Psi_{-i}\right) \end{array} \right\}, \quad (47) \\ \Xi_i^{**} &= \cap_{k \geq 1} \Xi_i^k \end{aligned}$$

Thus Ψ_i survives at the $(k+1)$th round only if there is $\Psi_{-i} \in \Xi_{-i}^k$ and $R \in \mathcal{R} \setminus \{\overline{R}\}$ such that, for each type of θ_i, observing his preferences cannot rule out the possibility that he thinks Ψ_{-i} is the set of possible

type profiles of his opponents and has preferences R. There are two ways this could happen. First, he may think some other set $\Psi'_{-i} \in \Xi^k_{-i}$ is the set of possible type profiles of his opponents, but he is completely indifferent. Or he might actually think Ψ_{-i} is the set of possible type profiles of his opponents and have preferences R.

Now say that two types θ_i and θ'_i are pairwise inseparable* if $\{\theta_i, \theta'_i\} \in \Xi^{**}_i$. Now theorem 1 can be generalized - allowing degeneracy - to show that types θ_i and θ'_i are strategically indistinguishable if and only if there pairwise inseparable*.

Allowing the possibility of degenerate preferences in our robust virtual implementation analysis is not very interesting. Suppose that degeneracy remains in the limit of the kth level inseparable sets described above in (47); in other words, suppose that there exists a type θ_i such that $\overline{R} \in \mathcal{R}_i \left(\theta_i, \Psi'_{-i} \right)$ for some $\Psi'_{-i} \in \Xi^*_{-i}$. Then θ_i will be contained in any set that survives the procedure for agent i, i.e., $\Psi_i \in \Xi^{**}_i \Rightarrow \theta_i \in \Psi_i$. Thus we will have θ_i is pairwise inseparable* from θ'_i for every $\theta'_i \in \Theta_i$. Thus robust virtual implementation will require that the social choice function be independent of agent i's type.

References

ABREU, D., AND H. MATSUSHIMA (1992a): "A Response to Glazer and Rosenthal," *Econometrica*, 60, 1439–1442.

——— (1992b): "Virtual Implementation in Iteratively Undominated Strategies: Complete Information," *Econometrica*, 60, 993–1008.

——— (1992c): "Virtual Implementation In Iteratively Undominated Strategies: Incomplete Information," Discussion paper, Princeton University and University of Tokyo.

——— (1994): "Exact Implementation," *Journal of Economic Theory*, 64, 1–19.

BATTIGALLI, P. (1998): "Rationalizability in Incomplete Information Games," Discussion paper, Princeton University.

BATTIGALLI, P., AND M. SINISCALCHI (2003): "Rationalization and Incomplete Information," *Advances in Theoretical Economics*, 3, Article 3.

BERGEMANN, D., AND S. MORRIS (2005a): "Robust Implementation: The Case of Direct Mechanisms," Discussion Paper Cowles Foundation Discussion Paper 1561, Yale University.

——— (2005b): "Robust Implementation: The Role of Large Type Spaces," Discussion Paper 1519, Cowles Foundation, Yale University.

——— (2005c): "Robust Mechanism Design," *Econometrica*, 73, 1771–1813.

BRANDENBURGER, A., AND E. DEKEL (1987): "Rationalizability and Correlated Equilibria," *Econometrica*, 55, 1391–1402.

CHUNG, K.-S., AND J. ELY (2001): "Efficient and Dominance Solvable Auctions with Interdependent Valuations," Discussion paper, Northwestern University.

CHUNG, K.-S., AND J. ELY (2007): "Foundations of Dominant Strategy Mechanisms," *Review of Economic Studies*, 74, 447–476.

CREMER, J., AND R. MCLEAN (1985): "Optimal Selling Strategies Under Uncertainty for a Discriminating Monopolist When Demands are Interdependent," *Econometrica*, 53, 345–361.

——— (1988): "Full Extraction of the Surplus in Bayesian and Dominant Strategy Auctions," *Econometrica*, 56, 1247–1258.

DASGUPTA, P., P. HAMMOND, AND E. MASKIN (1979): "The Implementation of Social Choice Rules. Some General Results on Incentive Compatibility.," *Review of Economic Studies*, 66, 185–216.

DASGUPTA, P., AND E. MASKIN (2000): "Efficient Auctions," *Quarterly Journal of Economics*, 115, 341–388.

DEKEL, E., D. FUDENBERG, AND S. MORRIS (2006): "Topologies on Types," *Theoretical Economics*, 1, 275–309.

GLAZER, J., AND R. ROSENTHAL (1992): "A Note on Abreu-Matsushima Mechanisms," *Econometrica*, 60, 1435–1438.

GUL, F., AND W. PESENDORFER (2005): "The Canonical Type Space for Interdependent Preferences," Discussion paper, Princeton University.

HEIFETZ, A., AND Z. NEEMAN (2006): "On the Generic (Im)Possibility of Full Surplus Extraction in Mechanism Design," *Econometrica*, 2006, 213–233.

JACKSON, M. (1991): "Bayesian Implementation," *Econometrica*, 59, 461–477.

JACKSON, M. (1992): "Implementation in Undominated Strategies: A Look at Bounded Mechanisms," *Review of Economic Studies*, 59, 757–775.

JEHIEL, P., B. MOLDOVANU, M. MEYER-TER-VEHN, AND B. ZAME (2006): "The Limits of Ex Post Implementation," *Econometrica*, 74, 585–610.

LEDYARD, J. (1979): "Dominant Strategy Mechanisms and Incomplete Information," in *Aggregation and Revelation of Preferences*, ed. by J.-J. Laffont, chap. 17, pp. 309–319. North-Holland, Amsterdam.

MORRIS, S. (1994): "Trade with Heterogeneous Prior Beliefs and Asymmetric Information," *Econometrica*, 62, 1327–1347.

MYERSON, R. (1981): "Optimal Auction Design," *Mathematics of Operations Research*, 6, 58–73.

NEEMAN, Z. (2004): "The Relevance of Private Information in Mechanism Design," *Journal of Economic Theory*, 117, 55–77.

POSTLEWAITE, A., AND D. SCHMEIDLER (1986): "Implementation in Differential Information Economies," *Journal of Economic Theory*, 39, 14–33.

SAMET, D. (1998): "Common Priors and Separation of Convex Sets," *Games and Economic Behavior*, 24, 172–174.

SEFTON, M., AND A. YAVAS (1996): "Abreu-Matsushima Mechanisms: Experimental Evidence," *Games and Economic Behavior*, 16, 280–302.

SERRANO, R., AND R. VOHRA (2001): "Some Limitations of Virtual Bayesian Implementation," *Econometrica*, 69, 785–792.

——— (2005): "A Characterization of Virtual Bayesian Implementation," *Games and Economic Behavior*, 50, 312–331.

WILSON, R. (1987): "Game-Theoretic Analyses of Trading Processes," in *Advances in Economic Theory: Fifth World Congress*, ed. by T. Bewley, pp. 33–70, Cambridge. Cambridge University Press.