Evaluating the Conditions for Robust Mechanism Design*

Takashi Kunimoto[†] and Roberto Serrano[‡]

First Version: March 2010
This Version: April 2010

Abstract

We assess the strength of the different conditions identified in the literature of robust mechanism design. We focus on three conditions: ex post incentive compatibility, robust monotonicity, and robust measurability. Ex post incentive compatibility has been shown to be necessary for any concept of robust implementation, while robust monotonicity and robust measurability have been shown to be necessary for robust (full) exact and virtual implementation, respectively. This paper shows that while violations of ex post incentive compatibility and robust monotonicity do not easily go away, we identify a mild condition on environments in which robust measurability is satisfied by all social choice functions over an open and dense subset of first-order types. We conclude that there is a precise sense in which robust virtual implementation can be significantly more permissive than robust exact implementation.
JEL Classification: C72, D78, D82.
Keywords: robust mechanism design, ex post incentive compatibility, robust monotonicity, robust measurability.

1 Introduction

Our attempt in this paper is to assess the strength of the different conditions identified in the literature of robust mechanism design. These include conditions relevant for partial implementation, as well as full implementation. Such assessment is important in the understanding of the possibilities and limitations in the design of decentralized institutions. By robustness, what is meant is that the assumption of common knowledge of the entire type space is not made, and hence the goal is that implementation results survive when applied to all type spaces whose higher-order beliefs are compatible with an original simpler common knowledge structure. Consistent with the robustness desideratum, the solution concept in which implementation is sought is the iterative elimination of strictly dominated strategies.

Three conditions are the crucial ones: ex post incentive compatibility, robust monotonicity and robust measurability. Ex post incentive compatibility has been shown to be necessary for robust partial implementation (Bergemann and Morris (2005)) and also for robust full implementation, both exactly and virtually (Bergemann and Morris (2009a,2010),

*Kunimoto gratefully acknowledges financial support from FQRSC and SSHRC of Canada, and Serrano from Spain's Ministry of Science and Innovation under grant Consolider 2010 CSD2006-0016.

[†]Department of Economics, McGill University and CIREQ, Montreal, Quebec, Canada; takashi.kunimoto@mcgill.ca

[‡]Department of Economics, Brown University, Providence, RI, U.S.A. and IMDEA-Social Sciences, Madrid, Spain; roberto_serrano@brown.edu

Artemov, Kunimoto and Serrano (2010)).[1] When one requires full implementation and this is sought to be exact, the condition of robust monotonicity, along with ex post incentive compatibility, crops up as necessary and almost sufficient (BM (2010)). And finally, if full implementation is still required, but exact implementation is relaxed to allow approximations of the social choice function (SCF), the so-called virtual implementation paradigm, robust measurability is the condition that emerges in the characterization (BM (2009a), AKS (2010)).

Ex post incentive compatibility is extremely demanding if one wishes to apply it over an unrestricted domain of environments (Jehiel, Meyer-ter-Vehn, Moldovanu and Zame (2006)). One way out from this negative result is the consideration of interesting subdomains in which the condition is still permissive (see BM (2009b) and the references therein). Another way out that one can conceivably consider is to study the case of robustness with respect to intermediate relaxations of the common knowledge assumption. For example, AKS (2010) consider finite sets of *first-order types*, each of which comprises a pair of payoff type and the first-order belief over the payoff type space. In that analysis, the relevant incentive compatibility condition applies to the first-order types that are present in the model. This notion is termed first-order incentive compatibility in AKS (2010). However, when one considers approximations of the unrestricted set of first-order beliefs, this notion does not make a difference. Indeed, we shall show in Theorem 1 that ex post incentive compatibility is equivalent to first-order incentive compatibility when imposed over any open and dense set of first-order beliefs. The equivalence is also extended to a locally uniform version of the condition.

Next, we take on robust monotonicity. Robust monotonicity is the requirement of Bayesian monotonicity in every type space. In Theorem 2 we show an equivalence between robust monotonicity when imposed over first-order beliefs in the interior of the probability simplex and a locally robust version of the condition. The result shows that a violation of robust monotonicity in one specific type space can be extended to an open ball of environments around it.

We learn from the first two results that violations of ex post incentive compatibility and robust monotonicity do not easily go away. When such violations are found, they will still remain in approximations of the environment. In contrast, Theorem 3 asserts that the same is not true about robust measurability in general environments.[2] That final result shows that, over weakly non-separable environments, robust measurability is satisfied by all SCFs over an open and dense subset of first-order type spaces. The proof relies on the set of first-order beliefs satisfying first-order type diversity, initially proposed in Serrano and Vohra (2005) and also used in AKS (2010).[3]

The rest of the paper proceeds as follows. Section 2 introduces preliminaries. Sections 3, 4 and 5 deal in turn with the incentive compatibility, monotonicity and measurability results. Section 6 closes the paper with two illustrative examples.

[1] In the sequel, we shall refer to these sets of authors as BM and AKS, respectively.

[2] BM (2009a,b, 2010) provide a set of results in which the gap between (robust) exact and virtual implementation vanishes. Our different conclusion stems from the fact that we shall work with weakly non-separable environments, arguably a mild condition when arbitrary utility functions are allowed.

[3] In those papers, the first-order type spaces are finite, something not assumed here.

2 Preliminaries

Let $N = \{1, \dots, n\}$ denote the set of agents and Θ_i be the set of *finite* payoff types of agent i. Denote $\Theta \equiv \Theta_1 \times \cdots \times \Theta_n$, and $\Theta_{-i} \equiv \Theta_1 \times \cdots \times \Theta_{i-1} \times \Theta_{i+1} \times \cdots \times \Theta_n$.[4] Let $q_i(\theta_{-i}|\theta_i)$ denote agent i's first-order belief that other agents receive the profile of types θ_{-i} when his payoff type is θ_i. Let Q_i be the set of such all probabilistic first-order beliefs of agent i. Note that Q_i is any subset of $\Delta(\Theta_{-i})$ for each agent i, where $\Delta(\Theta_{-i})$ denotes the set of probability distributions over Θ_{-i}. We call $T_i \equiv \Theta_i \times Q_i$ the set of *first-order types* of agent i. Agent i's first-order type t_i contains information about his payoff type θ_i and the first-order belief over Θ_{-i} conditional on θ_i.

Let A denote the set of pure outcomes, which are assumed to be independent of the information state. Suppose $A = \{a_1, \dots, a_K\}$ is finite. Let $\Delta(A)$ denote the set of probability distributions on A.

Agent i's state dependent von Neumann-Morgenstern utility function is denoted $u_i : \Delta(A) \times \Theta \to \mathbb{R}$.

We can now define an *environment* as $\mathcal{E} = (A, \{u_i, \Theta_i, Q_i\}_{i \in N})$, which is implicitly understood to be common knowledge among the agents. In particular, if Q_i is unrestricted for each i, that is, $Q_i = \Delta(\Theta_{-i})$, we call it a *payoff environment* denoted as $\mathcal{E}_\Delta = (A, \{u_i, \Theta_i\}_{i \in N})$.

We denote a type of agent i by τ_i and the agent i's set of types by $\mathcal{T}_i$. A *type* τ_i of agent i must include a description of his first-order type, which in turn includes a payoff type. Thus, there is a function $\hat{t}_i : \mathcal{T}_i \to T_i$, with $\hat{t}_i(\tau_i)$ being agent i's first-order type when his type is τ_i. We shall write $\hat{t}(\tau)$ to refer to the profile of first-order types when the type profile is τ. There is also a function $\hat{\theta}_i : \mathcal{T}_i \to \Theta_i$, with $\hat{\theta}_i(\tau_i)$ being agent i's payoff type when his type is τ_i. We shall write $\hat{\theta}(\tau)$ to denote the payoff type profile when the profile of types is τ. With some abuse of notation, let $\hat{\theta}_i(t_i)$ be agent i 's payoff type when his first-order type is t_i. A type τ_i of agent i must also include a description of his beliefs about the types of the other agents; thus, for any $\tau_{-i} \in \mathcal{T}_{-i}$, $\pi_i(\tau_{-i}|\tau_i)$ denotes the probability that agent i of type τ_i assigns to other agents having types τ_{-i}.

We require that types, first-order types and payoff types are coherent with each other. We express the coherence requirement in the following definition. A *type space* $\mathcal{T}$ is a collection:

$$\mathcal{T} = (\mathcal{T}_i, \hat{\theta}_i, \hat{t}_i, \pi_i)_{i \in N}.$$

Definition 1 *A type space $\mathcal{T} \equiv (\mathcal{T}_i, \hat{\theta}_i, \hat{t}_i, \pi_i)_{i \in N}$ is said to be* ***coherent*** *with an environment $\mathcal{E} = (A, \{u_i, \Theta_i, Q_i\}_{i \in N})$ if, for every $i \in N$ and every type $\tau_i \in \mathcal{T}_i$, the following two conditions must hold:*

1. *$\hat{\theta}_i(\tau_i) \in \Theta_i$ and $\hat{t}_i(\tau_i) \in \Theta_i \times Q_i$; and*
2. *For all $(\theta_i, q_i) \in \Theta_i \times Q_i$, $\hat{\theta}_i(\tau_i) = \theta_i$ whenever $\hat{t}_i(\tau_i) = (\theta_i, q_i)$.*

The first part of the definition is just the requirement that first-order type and payoff type be coherent with the agent's type. The second part requires similar coherence between

[4] Similar notation will be used for products of other sets.

first-order types and payoff types. These two requirements, in turn, imply that, for any $\tau_i \in \mathcal{T}_i$ with $\hat{t}_i(\tau_i) = (\theta_i, q_i)$ and $\theta_{-i} \in \Theta_{-i}$,

$$\int_{\tau_{-i}:\hat{\theta}_{-i}(\tau_{-i})=\theta_{-i}} \pi_i(\tau_{-i}|\tau_i)d\tau_{-i} = q_i(\theta_{-i}|\theta_i)$$

The reader is referred to AKS (2010) for the discussion of our coherence assumption.

A *social choice function* (SCF) is a function $f : \Theta \to \Delta(A)$. Note that the domain of the SCFs is not the true type space, but the payoff type space. Fix any coherent type space $\mathcal{T}$ throughout. The interim expected utility of agent i of type τ_i that pretends to be of type τ_i' corresponding to an SCF f is defined as:

$$U_i(f;\tau_i'|\tau_i) \equiv \int_{\mathcal{T}_{-i}} \pi_i(\tau_{-i}|\tau_i)u_i(f(\hat{\theta}(\tau_i', \tau_{-i})); \hat{\theta}(\tau_i, \tau_{-i}))d\tau_{-i}$$

Denote $U_i(f|\tau_i) = U_i(f;\tau_i|\tau_i)$.

Define $V_i(f;t_i'|t_i)$ to be the interim expected utility of agent i of first-order type t_i that pretends to be of first-order type t_i' corresponding to an SCF f as follows:

$$V_i(f;t_i'|t_i) = \sum_{\theta_{-i}\in\Theta_{-i}} q_i(\theta_{-i}|\theta_i)u_i(f(\hat{\theta}_i(t_i'), \theta_{-i}); \theta_i, \theta_{-i})$$

where $t_i \equiv (\theta_i, q_i) \in T_i = \Theta_i \times Q_i$ and $t_i' \equiv (\theta_i', q_i') \in T_i = \Theta_i \times Q_i$. Denote $V_i(f|t_i) = V_i(f;t_i|t_i)$.

We often use the following relationship between interim utility and first-order interim utility of agent i:

Lemma 1 (AKS (2010)) *For a given SCF $f : \Theta \to \Delta(A)$, $U_i(f;\tau_i'|\tau_i) = V_i(f;\hat{t}_i(\tau_i')|\hat{t}_i(\tau_i))$ for any coherent type space $\mathcal{T}$.*

A *mechanism* $\Gamma = ((M_i)_{i\in N}, g)$ describes a message space M_i for agent i and an outcome function $g : M \to \Delta(A)$, where $M = \times_{i\in N} M_i$. Let $\sigma_i : \mathcal{T}_i \to M_i$ denote a (pure) strategy for agent i and Σ_i his set of pure strategies.[5] Let

$$U_i(g \circ \sigma|\tau_i) \equiv \int_{\mathcal{T}_{-i}} \pi_i(\tau_{-i}|\tau_i)u_i(g(\sigma(\tau_{-i}, \tau_i)); \hat{\theta}(\tau_{-i}, \tau_i))d\tau_{-i}.$$

Given a mechanism $\Gamma = (M, g)$, let H_i be a subset of Σ_i. A strategy $\sigma_i \in H_i$ is *strictly dominated* for player i with respect to $H = \times_{j\in N} H_j$ if there exist $\tau_i \in \mathcal{T}_i$ and $\sigma_i' \in H_i$ such that for every $\sigma_{-i} \in \times_{j\neq i} H_j$,

$$U_i(g \circ (\sigma_i', \sigma_{-i})|\tau_i) > U_i(g \circ (\sigma_i, \sigma_{-i})|\tau_i).$$

Let $\mathcal{K}_i(H)$ denote the set of all undominated strategies for agent i with respect to $H = \times_{i\in N} H_i$. Let $\mathcal{K}(H) = \times_{i\in N}\mathcal{K}_i(H)$. Let $\mathcal{K}_i^0(\Sigma) = \Sigma_i$ and for each $k \geq 1$, $\mathcal{K}^k(\Sigma) = \times_{i\in N}\mathcal{K}_i^k(\Sigma)$, where $\Sigma = \times_{i\in N}\Sigma_i$ and $\mathcal{K}_i^k(\Sigma) = \mathcal{K}_i(\mathcal{K}^{k-1}(\Sigma))$. Let

$$\mathcal{K}^* \equiv \bigcap_{k=0}^{\infty} \mathcal{K}^k(\Sigma)$$

[5]To be exact, we must use the notation $\Sigma_i(\mathcal{T}_i)$ to make the underlying type space explicit. We, however, omit this dependence, since it is always clear from the context.

Definition 2 *A strategy profile $\sigma \in \Sigma$ is* ***iteratively undominated*** *if $\sigma \in \mathcal{K}^*$.*

An SCF f is said to be *exactly implementable* in iteratively undominated strategies for a type space $\mathcal{T}$ if there exists a mechanism $\Gamma = (M, g)$ such that for any $\sigma \in \mathcal{K}^*$, $g(\sigma(\tau)) = f(\hat{\theta}(\tau))$ for all $\tau \in \mathcal{T}$. We add the requirement that this definition should hold for every coherent type space $\mathcal{T}$ to obtain the definition of robust implementation:

Definition 3 *An SCF f is* ***robustly implementable*** *in iteratively undominated strategies if there exists a mechanism $\Gamma = (M, g)$ such that for any coherent type space $\mathcal{T}$ and any $\sigma \in \mathcal{K}^*$, $g(\sigma(\tau)) = f(\hat{\theta}(\tau))$ for every $\tau \in \mathcal{T}$.*

Consider the following uniform metric on SCFs:

$$d_f(f, h) = \max_{\theta \in \Theta} \max_{a \in A} |f(\theta|a) - h(\theta|a)|,$$

where the notation $f(\theta|a)$ refers to the probability with which f implements $a \in A$ in the payoff state θ.

An SCF f is said to be *ε-implementable* in iteratively undominated strategies for a coherent type space $\mathcal{T}$ if, there exists $\bar{\varepsilon} > 0$ such that for any $\varepsilon \in (0, \bar{\varepsilon}]$, there exists an SCF f^ε for which $d_f(f, f^\varepsilon) < \varepsilon$ and f^ε is exactly implementable in iteratively undominated strategies for the type space $\mathcal{T}$. The definition of robust virtual implementability now follows.

Definition 4 *An SCF f is* ***robustly ε-implementable*** *in iteratively undominated strategies if there exists $\bar{\varepsilon} > 0$ such that, for any $\varepsilon \in (0, \bar{\varepsilon}]$, there exists an SCF f^ε for which $d_f(f, f^\varepsilon) < \varepsilon$ and f^ε is robustly implementable in iteratively undominated strategies.*

3 Incentive Compatibility

In a setting that is robust to higher-order beliefs, the standard requirement of Bayesian incentive compatibility is given by the following definition:

Definition 5 *An SCF $f : \Theta \to \Delta(A)$ is said to satisfy* ***incentive compatibility*** *for a coherent type space $\mathcal{T}$ if for every $i \in N$, $\tau_i, \tau_i' \in \mathcal{T}_i$,*

$$U_i(f|\tau_i) \geq U_i(f; \tau_i'|\tau_i).$$

The notion of first-order type suggests the following definition, which turns out to be operationally useful:

Definition 6 *An SCF f satisfies* ***first-order incentive compatibility*** *if, for any $i \in N$, and any $t_i = (\theta_i, q_i), t_i' = (\theta_i', q_i') \in Q_i$,*

$$V_i(f|\theta_i, q_i) \geq V_i(f; \theta_i'|\theta_i, q_i).$$

The next lemma provides a useful link between these concepts and follows directly from Lemma 1:

Lemma 2 (AKS (2010)) *An SCF $f : \Theta \to \Delta(A)$ satisfies incentive compatibility for any coherent type space $\mathcal{T}$ if and only if it satisfies first-order incentive compatibility.*

The robust mechanism design literature has often justified the use of ex post incentive compatibility for attaining robust implementation (both for partial and full implementation). We provide this definition next:

Definition 7 *An SCF f satisfies **ex post incentive compatibility** if, for any $i \in N$, $\theta \in \Theta$, and $\theta_i^{'} \in \Theta_i$,*

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

It is easy to see that when $Q_i = \Delta(\Theta_{-i})$ for every agent $i \in N$, an SCF f is first-order incentive compatible if and only if it is ex post incentive compatible. The next result extends this observation slightly but in an important direction. The following result asserts that we cannot relax ex post incentive compatibility by restricting attention to an open dense subset of $\Delta(\Theta_{-i})$. Define $\Delta^0(\Theta_{-i}) = \{q_i \in \Delta(\Theta_{-i}) | q_i(\theta_{-i}) > 0 \ \forall \theta_{-i} \in \Theta_{-i}\}$ be the interior of $\Delta(\Theta_{-i})$.

Theorem 1 *Suppose that an environment $\mathcal{E} = (A, \{u_i, \Theta_i, Q_i\}_{i \in N})$ satisfies the property that $Q_i \equiv \Delta^*(\Theta_{-i})$ for each $i \in N$ and $\Delta^*(\Theta_{-i})$ is an open and dense subset of $\Delta^0(\Theta_{-i})$. Then, an SCF f satisfies first-order incentive compatibility if and only if it satisfies ex post incentive compatibility.*

Proof: It is straightforward to show that if an SCF is ex post incentive compatible, it is also first-order incentive compatible, for any first-order type space.

Hence, we focus on the other direction. Let f be a first-order incentive compatible SCF over an open and dense set $\Delta^*(\Theta_{-i})$. Suppose, by way of contradiction, that f is not ex post incentive compatible. This implies that there exist $i \in N$, $\theta \in \Theta$, and $\theta_i^{'} \neq \theta_i$ such that

$$u_i(f(\theta); \theta) < u_i(f(\theta_i^{'}, \theta_{-i}); \theta).$$

By the continuity of expected utility, we can construct $q_i \in \Delta^0(\Theta_{-i})$ such that $q_i(\theta_{-i}|\theta_i) = 1 - \varepsilon$ for $\varepsilon > 0$ small enough and

$$V_i(f|\theta_i, q_i) < V_i(f; \theta_i^{'}|\theta_i, q_i).$$

Once again, by the continuity of expected utility, there exist an open neighborhood $O_\delta(q_i) \subset \Delta^0(\Theta_{-i})$, i.e., a $\delta > 0$ small enough such that for any $dq_i \in \mathbb{R}^H$ with the property that $\|dq_i\| < \delta$,

$$V_i(f|\theta_i, q_i + dq_i) < V_i(f; \theta_i^{'}|\theta_i, q_i + dq_i)$$

where $H = |\Theta_{-i}|$. Note that the norm $\| \cdot \|$ is induced by the uniform metric d_q with the property that $d_q(q_i, q_i^{'}) = \max_{\theta_{-i} \in \Theta_{-i}} |q_i(\theta_{-i}) - q_i^{'}(\theta_{-i})|$ for any $q_i, q_i^{'} \in \Delta(\Theta_{-i})$. Thus, we have shown that any nearby first-order belief $q_i + dq_i \in O_\delta(q_i)$ satisfies the above inequality, and $O_\delta(q_i) \cap \Delta^*(\Theta_{-i}) \neq \emptyset$, which is a contradiction. ■

Jehiel et al (2006) and Hashimoto (2008) show that ex post incentive compatible SCFs are generically constant.[6] Therefore, ex post incentive compatibility is quite demanding if one allows an unrestricted domain of environments. While these results provide a limit for the success of robust implementation, there are some interesting subdomains of environments in which ex post incentive compatibility is still permissive. The reader is referred to BM (2009b) for such a class of environments where some positive results are obtained. Moreover, in auction environments, Dasgupta and Maskin (2000) and Bikhchandani (2006) also propose some subdomains of environments where ex post incentive compatibility is not restrictive.

If one imposes the first-order incentive compatibility condition when all agents' first-order beliefs are restricted to lie in $\Delta^0(\Theta_{-i})$, we shall refer to such a conditions as first-order incentive compatibility over Δ^0. By Theorem 1, an SCF satisfies ex post incentive compatibility if and only if it satisfies first-order incentive compatibility over Δ^0. Therefore, we can now consider the following local version of ex post incentive compatibility:

Definition 8 *An SCF f satisfies* ***locally uniform incentive compatibility*** *if for any agent $i \in N$ and any open set $Q_i^0 \subset \Delta^0(\Theta_{-i})$ such that for any $\theta_i, \theta_i' \in \Theta_i$, we have that:*

$$V_i(f|\theta_i, q_i) \geq V_i(f; \theta_i'|\theta_i, q_i) \;\; \forall \; q_i \in Q_i^0$$

And this leads easily to the next result:

Proposition 1 *An SCF f satisfies ex post incentive compatibility if and only if it satisfies locally uniform incentive compatibility.*

Proof: We can use Theorem 1 to know that ex post incentive compatibility is equivalent to first-order incentive compatibility over Δ^0. Now, clearly if f satisfies first-order incentive compatibility over Δ^0, it also satisfies locally uniform incentive compatibility.

To prove the other implication, assume that f violates first-order incentive compatibility over Δ^0. That is, there exist $i \in N, \theta_i, \theta_i' \in \Theta_i$, and $q_i \in \Delta^0(\Theta_{-i})$ such that

$$V_i(f|\theta_i, q_i) < V_i(f; \theta_i'|\theta_i, q_i).$$

Since $q_i \in \Delta^0$, by the continuity of expected utility, there exists $\delta > 0$ small enough such that for any $q_i + dq_i \in O_\delta(q_i)$,

$$V_i(f|\theta_i, q_i + dq_i) < V_i(f; \theta_i'|\theta_i, q_i + dq_i).$$

Setting $Q_i^0 = O_\delta(q_i)$, which is an open set, we can conclude that f also violates locally uniform incentive compatibility. ■

Remark: The above result shows that there is no difference between ex post incentive compatibility and locally uniform incentive compatibility. Therefore, if ex post incentive compatibility is very restrictive, it continues to be so for its local version. In other words, failures to satisfy ex post incentive compatibility will not easily go away.

[6]There is one difference with our setup because those papers focus on the case of a continuum of payoff types. However, this difference is immaterial because our Theorem 1 and Proposition 1 extend unchanged to a compact payoff type space. See also the remark after the proof of Lemma 5.

4 Monotonicity

A number of monotonicity conditions have been suggested in order to answer the question of (full) exact implementation. We begin this section with several standard definitions in the Bayesian implementation literature, suitably adapted to the robust setting.

For agent i, consider a mapping $\alpha_i = (\alpha_i(\theta_i))_{\theta_i \in \Theta_i} : \Theta_i \to \Theta_i$. A *deception* $\alpha = (\alpha_i)_{i \in N}$ is a collection of such mappings where at least one differs from the identity mapping.

Given an SCF f and a deception α, let $[f \circ \alpha]$ denote the following SCF: $[f \circ \alpha](\theta) = f(\alpha(\theta))$ for every $\theta \in \Theta$. That is, $[f \circ \alpha]$ is the SCF that would be implemented if the planner wanted to implement f but the agents were to use the deception α : then, in each payoff state θ, instead of realizing $f(\theta)$, the outcome $f(\alpha(\theta))$ would result.

For a payoff type $\theta_i \in \Theta_i$, an SCF f, and a deception α, let $f_{\alpha_i(\theta_i)}(\theta^{'}) = f(\theta^{'}_{-i}, \alpha_i(\theta_i))$ for all $\theta^{'} \in \Theta$. That is, the SCF $f_{\alpha_i(\theta_i)}$ is what would be implemented if the planner wished to implement f, all agents other than i were to be truthful, and agent i would report that his payoff type is $\alpha_i(\theta_i)$. We write $f \neq f \circ \alpha$ when there exists $\theta \in \Theta$ such that $f(\theta) \neq f(\alpha(\theta))$.

The following definition is borrowed from BM (2010):

Definition 9 *An SCF f satisfies* ***robust monotonicity*** *if for any deception α, whenever $f \neq f \circ \alpha$, there exist $i \in N$, $\theta_i \in \Theta_i$, and an SCF y such that:*

$$V_i(y \circ \alpha | \theta_i, q_i) > V_i(f \circ \alpha | \theta_i, q_i) \;\; \forall q_i \in \Delta(\Theta_{-i})$$

while

$$V_i(f | \theta^{'}_i, q^{'}_i) \geq V_i(y_{\alpha_i(\theta_i)} | \theta^{'}_i, q^{'}_i) \;\; \forall \theta^{'}_i \in \Theta_i, \;\; \forall q^{'}_i \in \Delta(\Theta_{-i}).$$

Note that the above definition for robust monotonicity, as written, does not exactly coincide with the one presented by BM (2010). Nevertheless, it can be shown that both are equivalent. Assume that $Q_i = \Delta(\Theta_{-i})$ for every $i \in N$. Then, robust monotonicity is equivalent to Bayesian monotonicity for every type space. By our Lemma 1, it is easy to see that the above definition is indeed the one for robust monotonicity.

Proposition 2 (BM (2010)) *Consider an environment $\mathcal{E}$ where $Q_i = \Delta(\Theta_{-i})$ for every $i \in N$. If an SCF f is robustly implementable in iteratively undominated strategies, it satisfies robust monotonicity.*

Remark: BM (2010) use the iterative deletion of never best responses as their solution concept. This solution concept is equivalent to iteratively undominated strategies in finite mechanisms. For the case of infinite mechanisms, iteratively undominated strategies is more stringent than iterative removal of never best responses. Thus, a fortiori, robust monotonicity is a necessary condition for robust implementation in iteratively undominated strategies.

We shall say that an SCF f satisfies robust monotonicity over Δ^0 if it satisfies robust monotonicity subject to all agents' first-order beliefs used in the condition being restricted to lie in $\Delta^0(\Theta_{-i})$. We note the following simple observation:

Remark: If an SCF f satisfies robust monotonicity, it satisfies robust monotonicity over Δ^0.

In particular, this implies that robust monotonicity over Δ^0 is also a necessary condition for robust exact implementation in iteratively undominated strategies.

Consider now the following local version of robust monotonicity:

Definition 10 *An SCF f satisfies **local robust monotonicity** if for any deception α, whenever $f \neq f \circ \alpha$, there exist $i \in N$, $\theta_i \in \Theta_i$, and an SCF y such that for every open set $Q_i^0 \subset \Delta(\Theta_{-i})$, we have that:*

$$V_i(y \circ \alpha | \theta_i, q_i) > V_i(f \circ \alpha | \theta_i, q_i) \ \ \forall q_i \in Q_i^0$$

while

$$V_i(f | \theta_i^{'}, q_i^{'}) \geq V_i(y_{\alpha_i(\theta_i)} | \theta_i^{'}, q_i^{'}) \ \ \forall \theta_i^{'} \in \Theta_i \ \ \forall q_i^{'} \in Q_i^0.$$

Remark: Maintaining the former ("reversal") clause for the definition of local robust monotonicity, we can strengthen the latter ("truth-telling") clause to $V_i(f | \theta_i^{'}, q_i^{'}) \geq V_i(y_{\alpha_i(\theta_i)} | \theta_i^{'}, q_i^{'}) \ \ \forall \theta_i^{'} \in \Theta_i \ \ \forall q_i^{'} \in \Delta^0(\Theta_{-i})$. In other words, we replace Q_i^0 with $\Delta^0(\Theta_{-i})$ for the range of possible $q_i^{'}$s. In particular, the proof of Theorem 2 below will not be affected by this change. Furthermore, we can also accommodate this change in the discussion of Example 2 in Section 6.

Using this definition, we state and prove our next result:

Theorem 2 *An SCF f satisfies robust monotonicity over Δ^0 if and only if it satisfies local robust monotonicity.*

Proof: Clearly, if f satisfies robust monotonicity over Δ^0, it also satisfies local robust monotonicity.

To prove the other implication, assume that f violates robust monotonicity over Δ^0. This means that there exists an environment with a specific first-order type space (with beliefs for each i in $\Delta^0(\Theta_{-i})$) over which f violates Bayesian monotonicity. That is, there exists a deception α with $f \neq f \circ \alpha$ such that for all $i \in N$ and for all $\theta_i \in \Theta_i$, there exists $q_i \in \Delta^0(\Theta_{-i})$ such that whenever one has that

$$V_i(f | \theta_i^{'}, q_i^{'}) \geq V_i(y_{\alpha_i(\theta_i)} | \theta_i^{'}, q_i^{'}) \ \ \forall \theta_i^{'} \in \Theta_i \ \ \forall q_i^{'} \in \Delta^0(\Theta_{-i}),$$

one also has that

$$V_i(y \circ \alpha | \theta_i, q_i) \leq V_i(f \circ \alpha | \theta_i, q_i).$$

Since expected utility preferences are continuous and q_i is in the interior of the probability simplex, the strictly upper contour sets are open and non-empty, and thus one can rewrite the last two inequalities as follows: whenever one has that

$$V_i(f | \theta_i^{'}, q_i^{'}) > V_i(y_{\alpha_i(\theta_i)} | \theta_i^{'}, q_i^{'}) \ \ \forall \theta_i^{'} \in \Theta_i \ \ \forall q_i^{'} \in \Delta^0(\Theta_{-i})$$

one also has that

$$V_i(y \circ \alpha | \theta_i, q_i) < V_i(f \circ \alpha | \theta_i, q_i).$$

Since these inequalities are strict, one can find an open neighborhood of q_i in which the same inequalities obtain. It follows that f violates local robust monotonicity. ■

Remark: The message of the above result is that, whenever one can find a violation of robust monotonicity, i.e., a violation of Bayesian monotonicity in some fixed first-order type space, such a violation can be extended to an open set of priors around the original one. Of course, if one found a violation of robust monotonicity on the boundary of $\Delta(\Theta_{-i})$, it may not be possible to extend it to an open set of priors around the original one; see however Example 2 in Section 6.

5 Measurability

This section deals with measurability, a condition that is key for virtual implementation in iteratively undominated strateigies. Roughly speaking, it requires that an SCF cannot vary in two payoff states whenever the types compatible with them have identical preferences. It was proposed by Abreu and Matsushima (1992), and hence, we shall refer to it as A-M measurability. Its robust version has been used in BM (2009a); see also AKS (2010).

Denote by Ψ_i a *partition* of the set of first-order types T_i, where ψ_i is a generic element of Ψ_i and $\Pi_i(t_i)$ is the element of Ψ_i that includes first-order type t_i.[7] Let $\Psi = \times_{i \in N} \Psi_i$ and $\psi = \times_{i \in N} \psi_i$. An SCF f is *measurable with respect to* Ψ if, for every $i \in N$ and every $t_i, t_i' \in T_i$, whenever $\Pi_i(t_i) = \Pi_i(t_i')$,

$$f(\hat{\theta}(t_i, t_{-i})) = f(\hat{\theta}(t_i', t_{-i})) \;\; \forall t_{-i} \in T_{-i}.$$

Measurability of f with respect to Ψ implies that for any player i, f does not distinguish between any pair of first-order types in the same cell of the partition Ψ_i.

For every $i \in N$, $t_i, t_i' \in T_i$, and $(n-1)$ tuple of partitions Ψ_{-i}, we say that t_i is *equivalent* to t_i' with respect to Ψ_{-i} if, for every f and every $\tilde{f}$ that are measurable with respect to $T_i \times \Psi_{-i}$,

$$V_i(f|t_i) \geq V_i(\tilde{f}|t_i) \Longleftrightarrow V_i(f|t_i') \geq V_i(\tilde{f}|t_i').$$

Let $\rho_i(t_i, \Psi_{-i})$ be the set of all elements of T_i that are equivalent to t_i with respect to Ψ_{-i}, and let

$$R_i(\Psi_{-i}) = \left\{ \rho_i(t_i, \Psi_{-i}) \subset T_i |\; t_i \in T_i \right\}.$$

Note that $R_i(\Psi_{-i})$ forms an equivalence class on T_i, that is, constitutes a partition of T_i. We define an infinite sequence of n-tuples of partitions, $\{\Psi^h\}_{h=0}^{\infty}$, where $\Psi^h = \times_{i \in N} \Psi_i^h$ in the following way. For every $i \in N$,

$$\Psi_i^0 = \{T_i\},$$

and recursively, for every $i \in N$ and every $h \geq 1$,

$$\Psi_i^h = R_i(\Psi_{-i}^{h-1}).$$

[7] With respect to Abreu and Matsushima (1992), recall that T_i is not necessarily finite in our current treatment.

Note that for every $h \geq 0$, Ψ_i^{h+1} is the same as, or finer than, Ψ_i^h. Define Ψ^* as follows:

$$\Psi^* \equiv \bigcup_{h=0}^{\infty} \Psi^h.$$

Definition 11 *An SCF f satisfies* ***A-M measurability*** *if it is measurable with respect to Ψ^*.*

Proposition 3 (AKS (2010)) *If an SCF f is robustly ε-implementable in iteratively undominated strategies, then it satisfies A-M measurability.*

When $Q_i = \Delta(\Theta_{-i})$ for each agent $i \in N$, adapting the above algorithm to the separation of types (instead of first-order types), BM (2009a) define the following property.

Definition 12 *An SCF f satisfies robust measurability whenever it satisfies A-M measurability for all type spaces coherent with the underlying payoff environment.*

Lemma 3 *Suppose $Q_i = \Delta(\Theta_{-i})$ for every $i \in N$ in an environment $\mathcal{E} = (A, \{u_i, \Theta_i, Q_i\}_{i\in N})$. Then, an SCF f satisfies A-M measurability if and only if it satisfies robust measurability.*

Proof: Since Q_i is unrestricted, robust measurability is equivalent to A-M measurability for all coherent type spaces. (Lemma 1 takes care of the details of the argument.) ■

We next formalize the idea that robust measurability is almost always satisfied by all SCFs.[8] We consider here unrestricted first-order type spaces.

Recall that the set of alternatives is $A = \{a_1, \dots, a_K\}$. Henceforth, we will find it convenient to identify a lottery $x \in \Delta(A)$ as a point in the $(K-1)$ dimensional simplex $\Delta^{K-1} = \{(x_1, \dots, x_K) \in \mathbb{R}_+^{K-1} | \sum_{k=1}^K x_k = 1\}$. Define $V_i^k(\theta_i, q_i)$ to be the interim expected utility of agent i of first-order type (θ_i, q_i) for the constant SCF that assigns a_k in each payoff state Θ, i.e.,

$$V_i^k(\theta_i, q_i) = \sum_{\theta_{-i}\in\Theta_{-i}} q_i(\theta_{-i}) u_i(a_k; \theta_i, \theta_{-i}).$$

Let $V_i(\theta_i, q_i) = (V_i^1(\theta_i, q_i), \dots, V_i^K(\theta_i, q_i))$. In the rest of the paper, we maintain the following regularity assumption imposed on the environments. An environment $\mathcal{E} = (A, \{u_i, \Theta_i, Q_i\}_{i\in N})$ is said to satisfy *first-order no-total-indifference (first-order NTI)* if for each $i \in N$ and each first-order type $t_i = (\theta_i, q_i)$, there exist two outcomes $a_k, a_{k'} \in A$ such that $V_i^k(\theta_i, q_i) \neq V_i^{k'}(\theta_i, q_i)$. Hence, in environments satisfying first-order NTI, without loss of generality, for each first-order type (θ_i, q_i), normalize expected utility by subtracting the constant $\min_k V_i^k(\theta_i, q_i)$ and dividing by the positive constant $\max_k V_i^k(\theta_i, q_i) - \min_k V_i^k(\theta_i, q_i)$.

Consider now the following definition:

[8] For finite environments, the argument can be found in AKS (2010).

Definition 13 *A payoff environment* $\mathcal{E}_\Delta = (A, \Theta_i, u_i)_{i \in N}$ *is* ***weakly non-separable*** *if, for any* $i \in N$*, any* θ_i *and* $\theta_i^{'} \in \Theta_i$ *with* $\theta_i \neq \theta_i^{'}$*, there exist* $a \in A$ *and* $\theta_{-i}, \theta_{-i}^{'} \in \Theta_{-i}$ *with* $\theta_{-i} \neq \theta_{-i}^{'}$ *such that:*

$$u_i(a; \theta_i, \theta_{-i}) - u_i(a; \theta_i, \theta_{-i}^{'}) \neq u_i(a; \theta_i^{'}, \theta_{-i}) - u_i(a; \theta_i^{'}, \theta_{-i}^{'}) \;\; (*).$$

It is easy to check that weak non-separability excludes private values environments. Outside of private values, when a payoff environment violates it, preferences are strongly separable, in that for at least two payoff types of an agent, the relative impact of interdependence on the change in ex-post utilities is the same and equals 1 for each alternative, and it is independent of –can be separated from– the payoff types of other agents. This justifies the term "weakly non-separable" environments.

The next definition is borrowed from AKS (2010):

Definition 14 *An environment* $\mathcal{E} = (A, \{u_i, \Theta_i, Q_i\}_{i \in N})$ *satisfies* ***first-order type diversity (FOTD)*** *if there do not exist* $i \in N$*,* $t_i = (\theta_i, q_i)$*,* $t_i^{'} = (\theta_i^{'}, q_i^{'}) \in T_i$ *with* $\theta_i \neq \theta_i^{'}$ *such that*

$$V_i(\theta_i, q_i) = V_i(\theta_i^{'}, q_i^{'}).$$

Without loss of generality, we focus only on agent i throughout. Since the payoff type space Θ is finite, we can denote $\Theta_{-i} = \{\theta_{-i}^h\}_{h=1}^H$.

Lemma 4 (The set of first-order beliefs under which FOTD holds is open) *Suppose that for any* $\theta_i, \theta_i^{'} \in \Theta_i$ *with* $\theta_i \neq \theta_i^{'}$*, and for any* $q_i, q_i^{'} \in \Delta^* \subseteq \Delta^0(\Theta_{-i})$*,*

$$V_i(\theta_i, q_i) \neq V_i(\theta_i^{'}, q_i^{'}).$$

Then, Δ^* *is open, i.e., for every* $q_i \in \Delta^*$ *there exists* $\delta > 0$ *such that for any* $dq_i \in \mathbb{R}^H$ *for which* $\sum_h dq_i(\theta_{-i}^h) = 0$ *and* $\|dq_i\| < \delta$*, we have that for any* $\theta_i, \theta_i^{'} \in \Theta_i$ *with* $\theta_i \neq \theta_i^{'}$*, and any* $q_i^{'} \in \Delta^*$*,*

$$V_i(\theta_i, q_i + dq_i) \neq V_i(\theta_i^{'}, q_i^{'})$$

where $q_i + dq_i \in \Delta^0(\Theta_{-i})$*.*

Proof: Pick $q_i \in \Delta^*$, a set of first-order beliefs over which FOTD holds. Recall that $\Delta(\Theta_{-i})$ is compact. Take an open cover of $\Delta(\Theta_{-i})$ as follows. The ε-open set O_ε in the open cover consists of all $q_i^{'}$'s such that

$$|V_i(q_i) - V_i(q_i^{'})| = \sum_{\theta_i} \sum_{\theta_i^{'} \neq \theta_i} \sum_k |V_i^k(\theta_i, q_i) - V_i^k(\theta_i^{'}, q_i^{'})| < \varepsilon.$$

Thus, $\Delta(\Theta_{-i}) \subseteq \bigcup_\varepsilon O_\varepsilon$. By compactness, take a finite subcover $\{O_1, O_2, \ldots, O_r\}$ such that $\Delta(\Theta_{-i}) \subseteq O_1 \cup \cdots \cup O_r$, which means that there exist a finite collection of increasing ε_i's with $\varepsilon_1 < \cdots < \varepsilon_r$ whose associated open sets also cover $\Delta(\Theta_{-i})$, and a fortiori, also cover Δ^*, a subset of $\Delta^0(\Theta_{-i})$, itself a subset of $\Delta(\Theta_{-i})$.

It follows that $\Delta^* = (Q^1 \cap \Delta^*) \cup \cdots \cup (Q^r \cap \Delta^*)$, where

$$\begin{aligned} Q^1 &= \{q_i^{'} : |V_i(q_i) - V_i(q_i^{'})| < \varepsilon_1\}; \\ Q^2 &= \{q_i^{'} : \varepsilon_1/2 < |V_i(q_i) - V_i(q_i^{'})| < \varepsilon_2\}; \\ Q^3 &= \{q_i^{'} : \varepsilon_2/2 < |V_i(q_i) - V_i(q_i^{'})| < \varepsilon_3\}; \\ \vdots &\quad \vdots \qquad\qquad \vdots \\ Q^r &= \{q_i^{'} : \varepsilon_{r-1}/2 < |V_i(q_i) - V_i(q_i^{'})| < \varepsilon_r\}. \end{aligned}$$

Without loss of generality, assume that $Q^2 \cap \Delta^* \neq \emptyset$. (If not, then we would have $\Delta^* = Q^1 \cap \Delta^*$. By choosing ε_1 small enough, thanks to FOTD, we can always make sure that $Q^2 \cap \Delta^* \neq \emptyset$.) For any $\delta > 0$, let $O_\delta(q_i) \equiv \{q_i + dq_i \in \Delta^0(\Theta_{-i}) : \|dq_i\| < \delta\}$ be a δ-neighborhood of q_i. Choose arbitrarily $q_i^{'}$ in the set $Q^2 \cap \Delta^*$ to satisfy that $\varepsilon_1/2 < |V_i(q_i) - V_i(q_i^{'})| < \varepsilon_2$, and also by FOTD, $|V_i(\theta_i, q_i) - V_i(\theta_i^{'}, q_i^{'})| > 0$ for any $\theta_i, \theta_i' \in \Theta_i$ with $\theta_i \neq \theta_i^{'}$. Due to the continuity of expected utility, we can choose $\delta(q_i^{'}) > 0$ sufficiently small so that for any $q_i + dq_i \in O_{\delta(q_i^{'})}(q_i)$, one has that $\varepsilon_1/2 < |V_i(q_i + dq_i) - V_i(q_i^{'})| < \varepsilon_2$ and $|V_i(\theta_i, q_i + dq_i) - V_i(\theta_i^{'}, q_i^{'})| > 0$ for any $\theta_i, \theta_i' \in \Theta_i$ with $\theta_i \neq \theta_i^{'}$. Define

$$\delta \equiv \inf_{q_i^{'} \in Q^2 \cap \Delta^*} \delta(q_i^{'}).$$

Note that $\delta > 0$ is well defined because $|V_i(q_i) - V_i(q_i^{'})| > \varepsilon_1/2$ for any $q_i^{'} \in Q^2 \cap \Delta^*$ and because of FOTD. This implies that for any $q_i + dq_i \in O_\delta(q_i)$, $\varepsilon_1/2 < |V_i(q_i + dq_i) - V_i(q_i^{'})| < \varepsilon_2$ and $|V_i(\theta_i, q_i + dq_i) - V_i(\theta_i^{'}, q_i^{'})| > 0$ for any $\theta_i, \theta_i' \in \Theta_i$ with $\theta_i \neq \theta_i^{'}$ and any $q_i^{'} \in Q^2 \cap \Delta^*$. Thus, we conclude that $O_\delta(q_i) \subseteq Q^2 \cap \Delta^* \subseteq \Delta^*$ and therefore, the set of first-order beliefs under which FOTD holds is open. ■

Lemma 5 (The set of first-order beliefs under which FOTD holds is dense) *Suppose that a payoff environment $\mathcal{E}_\Delta = (A, \Theta_i, u_i)_{i \in N}$ is weakly non-separable. Then, for any $\delta > 0$ small enough, there exists $dq_i \in \mathbb{R}^H$ with $\sum_h dq_i(\theta_{-i}^h) = 0$ and $\|dq_i\| < \delta$ such that for any $\theta_i, \theta_i^{'} \in \Theta_i$ with $\theta_i \neq \theta_i^{'}$,*

$$V_i(\theta_i, q_i + dq_i) \neq V_i(\theta_i^{'}, q_i^{'} + dq_i),$$

for any pair $q_i, q_i^{'} \in \Delta^0(\Theta_{-i})$.

Proof: Since $\Delta^0(\Theta_{-i})$ is separable, it contains a countable dense subset. Thus, it will suffice to base our arguments on a countable set of pairs $q_i, q_i^{'}$ for which there is a violation of FOTD. That is, consider payoff types $\theta_i, \theta_i^{'} \in \Theta_i$ with $\theta_i \neq \theta_i^{'}$, such that:

$$V_i(\theta_i, q_i) = V_i(\theta_i^{'}, q_i^{'})$$

for some $q_i, q_i^{'} \in \Delta^0(\Theta_{-i})$.

Fix arbitrarily an index set, $\Lambda = \{1, 2, \dots\}$. Assume that each $\lambda \in \Lambda$ corresponds to a pair of first-order types $\lambda = ((\theta_i^\ell, q_i^\ell), (\theta_i^m, q_i^m))$ that exhibits violations of FOTD. Since the payoff environment is weakly non-separable, for each such pair of relevant payoff types $\theta_i^\ell, \theta_i^m \in \Theta_i$ with $\theta_i^\ell \neq \theta_i^m$, there exist $\theta_{-i}, \theta'_{-i} \in \Theta_{-i}$ with $\theta_{-i} \neq \theta'_{-i}$ and $a_k \in A$ such that

$$u_i(a_k;\theta_i^\ell,\theta_{-i}) - u_i(a_k;\theta_i^\ell,\theta'_{-i}) \neq u_i(a_k;\theta_i^m,\theta_{-i}) - u_i(a_k;\theta_i^m,\theta'_{-i}).$$

We define $\theta_{-i}^{\lambda_1} \equiv \theta_{-i}$ and $\theta_{-i}^{\lambda_2} \equiv \theta'_{-i}$ and for each pair $(\theta_i^\ell, \theta_i^m)$ associated with each λ, fix such $\theta_{-i}^{\lambda_1}$ and $\theta_{-i}^{\lambda_2}$.

Define $dq_i \in \mathbb{R}^H$ as follows:

- $dq_i = \sum_{(\ell,m):(\theta_i^\ell,q_i^\ell),(\theta_i^m,q_i^m)\in\Lambda} dq_i[\ell,m]$;
- $dq_i[\ell,m](\theta_{-i}^{\lambda_1}) = \varepsilon^\lambda$ where $\varepsilon > 0$;
- $dq_i[\ell,m](\theta_{-i}^{\lambda_2}) = -\varepsilon^\lambda$;
- $dq_i[\ell,m](\theta_{-i}^{\tilde{h}}) = 0$ for any $\tilde{h} \neq \lambda_1, \lambda_2$

By construction, we note the following three facts: (1) $\sum_{\theta_{-i}} dq_i[\ell,m](\theta_{-i}) = 0$ for any $(\ell,m) \in \Lambda$; (2) $\sum_{\theta_{-i}} dq_i(\theta_{-i}) = 0$; and (3) $dq_i \neq 0$.

Fix $\delta > 0$ small enough. Since Θ is finite, we can choose $\varepsilon > 0$ small enough so that $\|dq_i\| < \delta$. For a sufficiently small $\delta > 0$, we guarantee that $q_i + dq_i \in \Delta^0(\Theta_{-i})$. By weak non-separability of the payoff environment and by construction of the specific dq_i, we get that for every $(\ell,m) \in \Lambda$,

$$V_i(\theta_i^\ell, q_i^\ell + dq_i) \neq V_i(\theta_i^m, q_i^m + dq_i).$$

So we conclude that for any $\delta > 0$ small enough there exists $dq_i \in \mathbb{R}^H$ with $\|dq_i\| < \delta$ such that for any $\theta_i, \theta'_i \in \Theta_i$ with $\theta_i \neq \theta'_i$, involved in a violation of FOTD, there exists a first-order belief in that δ-neighborhood for which such a violation ceases to exist.

Moreover, by the previous lemma, the set of first-order beliefs for which FOTD holds is open. Thus, for any $\delta > 0$ small enough and for any pair $(\theta_i, q_i), (\theta'_i, q'_i)$, by the continuity of expected utility, $V_i(\theta_i, q_i) \neq V_i(\theta'_i, q'_i)$ if and only if $V_i(\theta_i, q_i + dq_i) \neq V_i(\theta'_i, q'_i + dq_i)$.

It follows that the set of first-order beliefs for which FOTD holds is dense. ■

Remark: The finiteness of Θ seems to be essential for Lemma 5. As a consequence, the same comment applies to Theorem 3, Corollary 1, and Lemma 6 below. On the other hand, all the other results in the current paper, with minor modifications in their proofs, extend if one assumes a compact set of Θ_i. It follows that our results with bearing on partial or exact full implementation cover significantly more environments than those for virtual implementation. Having said that, we do not have a counterexample to Lemma 5 for the case of infinite compact payoff type spaces. Recall that Jehiel et al (2006) and Hashimoto (2008) define Θ_i to be a compact convex subset of a finite dimensional Euclidean space, and Bergemann and Morris (2009b) define it to be a compact interval in the real line.

The two lemmas together comprise the proof of the following result:

Theorem 3 *Suppose that the payoff environment $\mathcal{E}_\Delta = (A, \Theta_i, u_i)_{i\in N}$ is weakly non-separable. Then, robust measurability is generically a trivial condition. Specifically, for every $i \in N$, there exists an open and dense set $\Delta^*(\Theta_{-i}) \subset \Delta^0(\Theta_{-i})$ for which the property of first-order type diversity holds.*

Proof: This directly follows from the previous two lemmas, after observing that if an environment satisfies FOTD all first-order types can be separated in the first iteration of the measurability algorithm, implying that the final partition thereof, Ψ^*, is the finest partition, consisting of all singletons. ■

Let $V_i : \Theta_i \times \Delta(\Theta_{-i}) \to \mathbb{R}^K$ be an agent i's vector of first-order expected utilities over all constant SCFs. Recall our normalization of expected utility for each first-order type. Thus, for each (θ_i, q_i), $V_i(\theta_i, q_i) \in [0,1]^K$. Define $\mathcal{V}_i$ to be the set of agent i's normalized first-order expected utility functions. We endow $\mathcal{V}_i$ with the uniform metric.[9] Let $\mathcal{V} \equiv \mathcal{V}_1 \times \cdots \times \mathcal{V}_n$. Now, we can rephrase the above result in terms of payoffs as follows:

Corollary 1 *Suppose that the payoff environment $\mathcal{E}_\Delta = (A, \Theta_i, u_i)_{i\in N}$ is weakly non-separable. Then, there exists an open and dense set $\mathcal{V}^*$ of $\mathcal{V}$ such that for all $V \in \mathcal{V}^*$, the property of first-order type diversity holds.*

We close this section by extending our logic to higher-order beliefs. We make use of our coherence assumption:

Lemma 6 *Suppose that an environment $\mathcal{E} = (A, \{u_i, \Theta_i, Q_i\}_{i\in N})$ satisfies the property that $Q_i \equiv \Delta^*(\Theta_{-i})$ for each $i \in N$ where $\Delta^*(\Theta_{-i})$ is an open and dense subset of $\Delta^0(\Theta_{-i})$ in which the property of first-order type diversity holds. Then, for any coherent type space $\mathcal{T}$, there do not exist $i \in N$, $\tau_i, \tau_i^{'} \in \mathcal{T}_i$ with $\hat{\theta}_i(\tau_i) \neq \hat{\theta}_i(\tau_i^{'})$, such that*

$$(U_i^1(\tau_i), \ldots, U_i^K(\tau_i)) = (U_i^1(\tau_i^{'}), \ldots, U_i^K(\tau_i^{'})).$$

Proof: Fix an arbitrary coherent type space $\mathcal{T}$. As it will become clear, the argument does not depend on any particular type space coherent with the original environment $\mathcal{E}$. Consider agent i of type τ_i. Let $\hat{t}_i(\tau_i) \equiv t_i = (\theta_i, q_i)$. It follows from Lemma 1 that $U_i^k(\tau_i) = V_i^k(\theta_i, q_i)$ for each $k = 1, \ldots, K$.

Thus, we obtain $U_i^k(\tau_i) = V_i^k(\theta_i, q_i)$ whenever $\hat{t}_i(\tau_i) = (\theta_i, q_i)$. Similarly, consider agent i of type $\tau_i^{'}$. Let $\hat{t}_i(\tau_i^{'}) \equiv (\theta_i^{'}, q_i^{'})$. Then, we obtain $U_i^k(\tau_i^{'}) = V_i^k(\theta_i^{'}, q_i^{'})$ for each $k = 1, \ldots, K$ whenever $\hat{t}_i(\tau_i) = (\theta_i^{'}, q_i^{'})$. Having established this, first-order type diversity takes care of the rest of the argument because we define $Q_i \equiv \Delta^*(\Theta_{-i})$. ■

6 Examples

We shall close by revisiting briefly two examples, already contemplated in previous literature. The first one illustrates the assumption of weak non-separability, which we have used in the previous section:

Example 1 (How to generate weak non-separability) *Consider the example in BM (2009a, Section 3), also featured in AKS (2010, Section 8). We show next that although*

[9] In this case, the uniform metric d_V is defined by

$$d_V(V_i, \tilde{V}_i) = \max_{(\theta_i,q_i)\in\Theta_i\times\Delta(\Theta_{-i})} \max_{k=1,\ldots,K} |V_i^k(\theta_i, q_i) - \tilde{V}_i^k(\theta_i, q_i)|$$

it violates weak non-separability, a variant thereof will satisfy it. (To be faithful to the presentation of the example in the above papers, we do not normalize first-order expected utilities.)

For each agent $i \in N$, let Θ_i be a finite subset of $[0,1]$. If agent i receives the object, his ex post valuation for it is $h_i(\theta)$. Let $h_i : \Theta \to \mathbb{R}$ to be

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

Here $\gamma \geq 0$ is the interdependence parameter. Let a_i be the outcome that agent i obtains the object. Let a_0 denote the outcome that no agent obtains the object and the seller keeps it. Define $A^* \equiv \{a_0, a_1, \ldots, a_n\}$. Let

$$A \equiv A^* \times Y$$

where $Y \subset \mathbb{R}^n$ is a finite set such that $(y_1, \ldots, y_n)$ denotes the monetary transfers across agents. Then, we have

$$u_i((a, y_1, \ldots, y_n); \theta) = \begin{cases} h_i(\theta) + y_i & \text{if } a = a_i \\ y_i & \text{if } a \neq a_i \end{cases}$$

For any $i \in N$, $\theta_i \in \Theta_i$, $\theta_{-i}, \theta'_{-i} \in \Theta_{-i}$ with $\theta_{-i} \neq \theta'_{-i}$, $y = (y_1, \ldots, y_n) \in Y$, and any $a \in A^*$, we have $u_i((a, y); (\theta_i, \theta_{-i})) - u_i((a, y); (\theta_i, \theta'_{-i})) = \gamma \sum_{j \neq i} (\theta_j - \theta'_j)$ if $a = a_i$ or 0 if $a \neq a_i$. This does not depend on agent i's payoff type. Thus, weak-non-separability is not satisfied.

On the other hand, weak non-separability can be restored as follows. Note that the $h_i(\cdot)$ constructed is continuous and strictly increasing in θ_i. We slightly modify the previous specification by making the ex post utilities non-linear.

$$u_i((a, y_1, \ldots, y_n); \theta) \equiv v_i((a, y_1, \ldots, y_n), h_i(\theta)) = \begin{cases} [h_i(\theta) + y_i]^{\lambda_i(h_i(\theta))} & \text{if } a = a_i \\ y_i^{\lambda_i(h_i(\theta))} & \text{if } a \neq a_i \end{cases}$$

where $\lambda_i : \mathbb{R} \to (0,1)$ is an increasing function with typical term $\lambda_i(h_i(\theta)) \in (0,1)$.

$$\begin{aligned} & u_i((a, y); \theta_i, \theta_{-i}) - u_i((a, y); \theta_i, \theta'_{-i}) \\ = & \begin{cases} [h_i(\theta) + y_i]^{\lambda_i(h_i(\theta))} - \left[h_i(\theta_i, \theta'_{-i}) + y_i\right]^{\lambda_i(h_i(\theta_i, \theta'_{-i}))} & \text{if } a = a_i \\ y_i^{\lambda_i(h_i(\theta))} - y_i^{\lambda_i(h_i(\theta_i, \theta'_{-i}))} & \text{if } a \neq a_i \end{cases} \end{aligned}$$

This is indeed a class of environments proposed in BM (2009b) in which both robust monotonicity and robust measurability are equivalent to a condition called the contraction property.[10] *Here, we can restore the weak non-separability condition by making ex post utilities non-linear. We also observe that sufficiency results for robust virtual implementation –for example, Theorems 1 and 2 of AKS (2010)– will not be affected by this modification.*[11]

[10] See BM (2009b) for the definition of the contraction property. In the case of linear ex post utilities of this example, the contraction property is equivalent to the condition that $\gamma < 1/(n-1)$.

[11] Hashimoto (2008) succeeded in generalizing the genericity result of Jehiel et al (2006) to the environments that encompass similar non-linearities. Unlike these papers, note that our genericity argument does not need consumption externalities.

The next example shows that in some environments the difference between robust measurability and robust monotonicity is substantial, leading to a significant gap between the success of robust virtual implementation versus robust exact implementation.

Example 2 *[Only constant SCFs satisfy local robust monotonicity] We begin by slightly adapting the way the example is presented in Serrano (2004), an elaboration of the original one in Palfrey and Srivastava (1987), and we proceed to its robust analysis later. Let* $N = \{1, 2, 3, 4\}$*. There is a single commodity – money – and all consumers have one unit of the commodity as endowment in each state. The set of payoff types is* $\Theta_k = \{\theta_k, \theta_k^{'}, \theta_k^{''}\}$ *for* $k = 1, 2$*, while* $\Theta_j = \{\theta_j, \theta_j^{'}\}$ *for* $j = 3, 4$*. Let us define a subset of* Θ*:* $\Theta^* = \{\theta, \theta^{'}, \theta^{''}\}$*, where* $\theta = (\theta_1, \theta_2, \theta_3, \theta_4)$*,* $\theta^{'} = (\theta_1^{'}, \theta_2^{'}, \theta_3^{'}, \theta_4^{'})$*, and* $\theta^{''} = (\theta_1^{''}, \theta_2^{''}, \theta_3^{'}, \theta_4^{'})$*.*

Start by fixing a first-order belief for each agent. For each $k = 1, 2$*,*

$$q_k^*(\theta_{-k}|\theta_k) = q_k^*(\theta'_{-k}|\theta'_k) = q_k^*(\theta''_{-k}|\theta''_k) = 1$$

For $j = 3, 4$*,* $q_j^*(\theta_{-j}|\theta_j) = 1$*, but*

$$\begin{aligned} q_3^*(\theta'_{-3}|\theta'_3) &= 0.25 \text{ and } q_3^*(\theta''_{-3}|\theta'_3) = 0.75, \\ q_4^*(\theta'_{-4}|\theta'_4) &= 0.75 \text{ and } q_4^*(\theta''_{-4}|\theta'_4) = 0.25. \end{aligned}$$

Each agent i's state dependent ex post utility is as follows: for any $x \in \mathbb{R}_+$ *and any* $\theta \in \Theta$*,*

$$u_i(x; \theta) = x^{\lambda_i(\theta)}$$

where $\lambda_i(\theta) \in (0, 1)$*. For every* $i \in N$*, we assume that for every* $\theta, \theta' \in \Theta$ *with* $\theta \neq \theta'$*,* $\lambda_i(\theta) \neq \lambda_i(\theta')$*. This environment satisfies weak non-separability, which means that robust measurability is almost always a vacuous constraint.*

First, assume that the set of first-order beliefs is a singleton, i.e., $Q_i = \{q_i^*\}$ *for every agent* $i \in N$*. Note that incentive compatibility is not a constraint in this environment.*

Let f *be an SCF such that for some* $\theta, \theta' \in \Theta^*$ *with* $\theta' \neq \theta$*,* $f(\theta) \neq f(\theta')$*. We denote* $f(\theta)$ *as* $(f_1(\theta), f_2(\theta), f_3(\theta), f_4(\theta))$ *where* $f_i(\theta)$ *is the money that agent* i *is assigned by the SCF* f *in payoff state* θ*. Consider a deception* α *such that* $\alpha_i(\tilde{\theta}_i) = \theta_i$ *for every* $\tilde{\theta}_i \in \Theta_i$ *and every* $i \in N$*. For this deception,* $f \neq f \circ \alpha$ *since* $f \circ \alpha$ *is a constant SCF that assigns* $f(\theta)$ *in every payoff state. For any agent* $i \in N$*, any* $\tilde{\theta}_i \in \Theta_i$ *and any SCF* y*, it follows that*

$$V_i(f|\alpha_i(\tilde{\theta}_i), q_i^*) \geq V_i(y|\alpha_i(\tilde{\theta}_i), q_i^*) \Rightarrow f_i(\theta)^{\lambda_i(\theta)} \geq y_i(\theta)^{\lambda_i(\theta)} \Rightarrow f_i(\theta) \geq y_i(\theta).$$

Since $f \circ \alpha$ *and* $y \circ \alpha$ *specify* $f(\theta)$ *and* $y(\theta)$ *in every state, it follows that*

$$V_i(f \circ \alpha|\tilde{\theta}_i, q_i^*) \geq V_i(y \circ \alpha|\tilde{\theta}_i, q_i^*)$$

for any $\tilde{\theta}_i \in \Theta_i$*.*

Now, we perturb q^* *slightly. Agent* $k = 1, 2$*'s first-order beliefs over* Θ *are given by:*

$$q_k^{\varepsilon}(\tilde{\theta}_{-k}|\theta_k) = \begin{cases} 1 - \delta(\theta_k) & \text{if } \tilde{\theta}_{-k} = \theta_{-k} \\ \varepsilon & \text{otherwise} \end{cases}$$

$$q_k^{\varepsilon}(\tilde{\theta}_{-k}|\theta'_k) = \begin{cases} 1 - \delta(\theta'_k) & \text{if } \tilde{\theta}_{-k} = \theta'_{-k} \\ \varepsilon & \text{otherwise} \end{cases}$$

$$q_k^{\varepsilon}(\tilde{\theta}_{-k}|\theta'_k) = \begin{cases} 1 - \delta(\theta''_k) & \text{if } \tilde{\theta}_{-k} = \theta''_{-k} \\ \varepsilon & \text{otherwise} \end{cases}$$

where $\delta(\theta_k) = \delta(\theta'_k) = \delta(\theta''_k) = 11\varepsilon$. Agent $j = 3, 4$'s first-order beliefs over Θ are given by:

$$q_j^\varepsilon(\tilde{\theta}_{-j}|\theta_j) = \begin{cases} 1 - \delta(\theta_j) & \text{if } \tilde{\theta}_{-j} = \theta_{-j} \\ \varepsilon & \text{otherwise} \end{cases}$$

$$q_3^\varepsilon(\tilde{\theta}_{-3}|\theta'_3) = \begin{cases} 0.25 - \delta(\theta'_3) & \text{if } \tilde{\theta}_{-3} = \theta'_{-3} \\ 0.75 - \delta(\theta'_3) & \text{if } \tilde{\theta}_{-3} = \theta''_{-3} \\ \varepsilon & \text{otherwise} \end{cases}$$

$$q_4^\varepsilon(\tilde{\theta}_{-4}|\theta'_4) = \begin{cases} 0.75 - \delta(\theta'_4) & \text{if } \tilde{\theta}_{-4} = \theta'_{-4} \\ 0.25 - \delta(\theta'_4) & \text{if } \tilde{\theta}_{-4} = \theta''_{-4} \\ \varepsilon & \text{otherwise} \end{cases}$$

where $\delta(\theta_j) = 17\varepsilon$; and $\delta(\theta'_3) = \delta(\theta'_4) = 16\varepsilon$. For any agent $i \in N$, $\tilde{\theta}_i \in \Theta_i$ and any SCF y, assume $V_i(f|\alpha_i(\tilde{\theta}_i), q_i^\varepsilon) \geq V_i(y|\alpha_i(\tilde{\theta}_i), q_i^\varepsilon)$. By the continuity of expected utility, there exists $\bar{\varepsilon}_i > 0$ such that for any $\varepsilon \in (0, \bar{\varepsilon}_i]$, the above inequality implies $f_i(\theta) \geq y_i(\theta)$.

Let $\bar{\varepsilon} = \min\{\varepsilon_1, \varepsilon_2, \varepsilon_3, \varepsilon_4\}$. Define $Q_i^0 = \{q_i^\varepsilon\}_{\varepsilon \leq \bar{\varepsilon}}$ for each $i \in N$. Since $f \circ \alpha$ and $y \circ \alpha$ specify $f(\theta)$ and $y(\theta)$ in every state, we have that for any $i \in N$, any $\tilde{\theta}_i \in \Theta_i$, and any $q_i \in Q_i^0$,

$$V_i(f \circ \alpha|\tilde{\theta}_i, q_i) \geq V_i(y \circ \alpha|\tilde{\theta}_i, q_i).$$

In sum, for any $i \in N$ and any $\tilde{\theta}_i \in \Theta_i$, we conclude

$$V_i(f|\theta_i, q_i) \geq V_i(y_{\alpha_i(\tilde{\theta}_i)}|\theta_i, q_i) \;\; \forall q_i \in Q_i^0 \Rightarrow V_i(f \circ \alpha|\tilde{\theta}_i, q_i) \geq V_i(y \circ \alpha|\tilde{\theta}_i, q_i) \;\; \forall q_i \in Q_i^0.$$

Hence, the SCF f violates local robust monotonicity. In particular, only constant SCFs satisfy local robust monotonicity.

Finally, we discuss incentive compatibility. Assume that a free disposal technology is available. Let f be an arbitrary non-constant SCF over $\Theta^* = \{\theta, \theta', \theta''\}$. For every agent $i \in N$ and every $\tilde{\theta} \in \Theta$, define

$$\tilde{f}_i(\tilde{\theta}) = \begin{cases} f_i(\theta) & \text{if } \tilde{\theta}_{-i} = \theta_{-i} \\ f_i(\theta') & \text{if } \tilde{\theta}_{-i} = \theta'_{-i} \\ f_i(\theta'') & \text{if } \tilde{\theta}_{-i} = \theta''_{-i} \\ 0 & \text{otherwise.} \end{cases}$$

Define an SCF $\tilde{f}$ to be such that $\tilde{f}(\theta) = (\tilde{f}_1(\theta), \tilde{f}_2(\theta), \tilde{f}_3(\theta), \tilde{f}_4(\theta))$ for any $\theta \in \Theta$. By construction, $\tilde{f}$ is well defined (thanks to the free disposal technology) and satisfies ex post incentive compatibility. Besides, $\tilde{f}$ is equivalent to f over Θ^*.

References

Abreu, D. and H. Matsushima (1992), "Virtual Implementation in Iteratively Undominated Strategies: Incomplete Information," Mimeo, Princeton University.

Artemov, G., T. Kunimoto, and R. Serrano (2010), "Robust Virtual Implementation with Incomplete Information: Towards a Reinterpretation of the Wilson Doctrine," (Revised) Economics Working Paper 2007-06, Brown University.

Bergemann, D. and S. Morris (2005), "Robust Mechanism Design," *Econometrica* **73**, 1771-1813.

Bergemann, D. and S. Morris (2009a), "Robust Virtual Implementation," *Theoretical Economics* **4**, 45-88.

Bergemann, D. and S. Morris (2009b), "Robust Implementation in Direct Mechanisms," *Review of Economic Studies* **76**, 1175-1204.

Bergemann, D. and S. Morris (2010), "Robust Implementation in General Mechanisms," Mimeo.

Bikhchandani, S. (2006), "Ex Post Implementation in Environments with Private Goods," *Theoretical Economics* **1**, 369-393.

Dasgupta, P. and E. Maskin (2000), "Efficient Auctions," *Quarterly Journal of Economics* **115**, 341-388.

Hashimoto, T (2008), "Generic Impossibility of Partial Ex Post Implementation with General Utility Functions," Mimeo.

Jehiel, P., M. Meyer-ter-Vehn, B. Moldovanu and W.R. Zame (2006), "The Limits of Ex Post Implementation," *Econometrica* **74(3)**, 585-610.

Palfrey, T. and S. Srivastava (1987), "On Bayesian Implementable Allocations," *Review of Economic Studies*, **54**, 193-208.

Serrano, R. (2004), "The Theory of Implementation of Social Choice Rules," *SIAM Review* **46**, 377-414.

Serrano, R. and R. Vohra (2005), "A Characterization of Virtual Bayesian Implementation," *Games and Economic Behavior* **50**, 312-331.