A New Necessary Condition for Implementation in Iteratively Undominated Strategies *

Takashi Kunimoto†and Roberto Serrano‡

This Version: January 2010

Abstract

Implementation in iteratively undominated strategies relies on permissive conditions. However, for the sufficiency results available, authors have relied on assumptions that amount to quasilinear preferences on a numeraire. We uncover a new necessary condition that implies that such assumptions cannot be dispensed with. We term the condition "restricted deception-proofness." It requires that, in environments with identical preferences, the social choice function be immune to all deceptions, making it then stronger than incentive compatibility. In some environments the conditions for (exact or approximate) implementation are more restrictive than previously thought.

JEL Classification: C72, D78, D82.

Keywords: mechanism design, exact and approximate implementation, iteratively undominated strategies, restricted deception-proofness, incentive compatibility, measurability.

1 Introduction

The conditions for implementation in iteratively undominated strategies are typically viewed as very permissive.[1] For example, in a standard Bayesian

*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.

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

‡Dept. of Economics, Brown University, Providence, RI, USA and IMDEA-Social Sciences, Madrid, Spain; roberto_serrano@brown.edu

[1]Here, "iteratively undominated strategies" refers to the iterative removal of *strictly* dominated strategies.

environment with incomplete information in which type spaces are common knowledge, Abreu and Matsushima (1992) [AM, henceforth] show that both incentive compatibility and their measurability condition (which we shall refer to as AM measurability from now on) are necessary for (exact or approximate) implementation in iteratively undominated strategies. Incentive compatibility is the central restriction in the economic theory of information, and it can sometimes be quite demanding. However, as argued for instance in AM or in Serrano and Vohra (2005), AM measurability is usually very weak: interim preferences of the different types are almost always distinct from each other, and then, AM measurability amounts to no restriction at all. These necessity results are generalized to robust environments, in which weaker common knowledge requirements are made, in Bergemann and Morris (2009a) [BM from now on] and in Artemov, Kunimoto and Serrano (2009) [AKS in the sequel].[2]

In the three papers afore mentioned (AM, BM and AKS), additional conditions are used to prove the corresponding sufficiency results. AM's Assumption 2 states that, for each agent i and each state, there exist two ex-post lotteries that i ranks strictly, and for which all other agents have the (weakly) opposite preferences. BM make use of an economic assumption, which is essentially a robust analogue of AM's Assumption 2. Due to their robustness considerations, BM need the assumption that for each agent i, there exists a constant lottery z_i that i strictly prefers to the uniform lottery $\bar{y}$, and for which all other agents have the (weakly) opposite preferences, "regardless of the underlying payoff types." Finally, AKS assume directly the existence of quasilinear preferences over a numeraire. In all three cases, the use of these assumptions in the sufficiency proofs is seemingly minor, in order to allow infinitesimal punishments out of equilibrium. Thus, one might have thought that such conditions could be dispensed with and that new proofs of the authors' sufficiency results could be engineered without the aid of such assumptions. In this paper, we show that such a hope is misplaced. Indeed, such assumptions cannot be dropped because a new necessary condition that the literature had overlooked must be added.

We identify such a condition, and we term it *restricted deception-proofness*. It says that in environments in which preferences are identical across agents, the social choice function (SCF) must be immune to all manipulations via deceptions. As such, the condition is then stronger than incentive compatibility and sometimes strictly so, leading to a new restriction on the SCFs that can be (exactly or approximately) implementable in iteratively undominated strategies. Considered by itself, restricted deception-proofness can

[2] As noted, the BM and AKS papers study the robust implementation problem, but their conclusions can be applied to a fixed Bayesian type space setting as a particular case. Going the other way, the new necessary condition in the current paper can be readily extended to make it applicable in the robust setting.

be substantially more restrictive than AM measurability or the conditions of virtual monotonicity and its mixed counterpart (the latter two found in Serrano and Vohra (2005, 2009)). We shall provide an example, which has appeared previously in the literature, to illustrate our points. We close by noting that we study incomplete information environments; two papers containing some related results for the complete information domain are Börgers (1995) and Bergemann and Morris (2009b). Bergemann and Morris (2009b) show a similar result for virtual implementation under complete information. Börgers (1995) obtains some impossibility result under complete information when only deterministic mechanisms are allowed and all possible identical preferences are included as part of the domain.

2 Preliminaries

Let $N = \{1, \dots, n\}$ denote the set of agents and Θ_i be the set of finite types of agent i. Denote $\Theta \equiv \Theta_1 \times \cdots \times \Theta_n$, and $\Theta_{-i} \equiv \Theta_1 \times \cdots \times \Theta_{i-1} \times \Theta_{i+1} \times \cdots \times \Theta_n$.[3] Let $q_i(\theta_{-i}|\theta_i)$ denote agent i's belief that other agents receive the profile of types θ_{-i} when his type is θ_i.

Let A denote the set of pure outcomes, which are assumed to be independent of the information state. For simplicity, 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.

A *social choice function* (SCF) is a function $f : \Theta \to \Delta(A)$. The interim expected utility of agent i of type θ_i that pretends to be of type θ_i' corresponding to an SCF f is defined as:

$$U_i(f; \theta_i'|\theta_i) \equiv \sum_{\theta_{-i} \in \Theta_{-i}} q_i(\theta_{-i}|\theta_i) u_i(f(\theta_i', \theta_{-i})); (\theta_i, \theta_{-i})).$$

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

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 $\sigma_i : \Theta_i \to M_i$ denote a (pure) strategy for agent i and Σ_i his set of pure strategies. Let

$$U_i(g \circ \sigma|\theta_i) \equiv \sum_{\theta_{-i} \in \Theta_{-i}} q_i(\theta_{-i}|\theta_i) u_i(g(\sigma(\theta_{-i}, \theta_i)); (\theta_{-i}, \theta_i)).$$

Given a mechanism $\Gamma = (M, g)$, let H_i be a subset of Σ_i.

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

Definition 1 (Strict Dominance) *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 $\sigma_{-i} \in \times_{j \neq i} H_j$,*

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

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

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

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

Definition 3 (Exact Implementability) *An SCF f is said to be exactly implementable in iteratively undominated strategies if there exists a mechanism $\Gamma = (M, g)$ such that for any $\sigma \in \mathcal{K}^*$, $g(\sigma(\theta)) = f(\theta)$ for all $\theta \in \Theta$.*

Consider the following metric on SCFs:

$$d(f, h) = \sup \left\{ |f(\theta|a) - h(\theta|a)| \;\middle|\; \theta \in \Theta,\ a \in A \right\}$$

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

Definition 4 (Approximate Implementability) *An SCF f is said to be virtually or approximately 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 $d(f, f^{\varepsilon}) < \varepsilon$ and f^{ε} is exactly implementable in iteratively undominated strategies.*

The next standard definition is very important in the entire economic theory of information:

Definition 5 (Incentive Compatibility) *An SCF $f : \Theta \to \Delta(A)$ is said to satisfy incentive compatibility if for every $i \in N$, $\theta_i, \theta_i^{'} \in \Theta_i$,*

$$U_i(f|\theta_i) \geq U_i(f; \theta_i^{\prime}|\theta_i)$$

As is well-known (e.g., see AM (1992)), the next proposition identifies incentive compatibility as a necessary condition for implementability:

Proposition 1 *(AM (1992)) If an SCF f is either exactly or approximately implementable in iteratively undominated strategies, then it satisfies incentive compatibility.*

For the next definition we require some more notation. Let Ψ_{-i} be a partition of Θ_{-i}. Say that θ_i is equivalent to θ_i' with respect to Ψ_{-i} when agent i's interim expected utility under type θ_i is exactly the same as under type θ_i' when evaluating any SCF that is measurable with respect to $\Theta_i \times \Psi_{-i}$.

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

$$R_i(\Psi_{-i}) = \{\rho_i(\theta_i, \Psi_{-i}) \subset \Theta_i |\ \theta_i \in \Theta_i\}.$$

Note that $R_i(\Psi_{-i})$ forms an equivalence class on Θ_i, that is, it constitutes a partition of Θ_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 $i \in N$,

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

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

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

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 6 (AM Measurability) *An SCF f is said to satisfy AM-measurability if it is measurable with respect to Ψ^*.*

The following result is also shown in AM (1992):

Proposition 2 *(AM (1992)) If an SCF f is either exactly or approximately implementable in iteratively undominated strategies, then it satisfies AM-measurability.*

To easily check AM-measurability, it is often possible to finish the algorithm in the first iteration. When this happens, we say that the environment satisfies type diversity. To define this condition, recall that $A = \{a_1, \ldots, a_K\}$. Henceforth, we will find it convenient to identify a lottery $x \in \Delta(A)$ as a point in the $(K-1)$ dimensional unit simplex $\Delta^{K-1} = \{(x_1, \ldots, x_K) \in \mathbb{R}_+^{K-1} \mid \sum_{k=1}^{K} x_k = 1\}$. Define $U_i^k(\theta_i)$ to be the interim expected utility of agent i of type θ_i for the constant SCF that assigns a_k in each state in Θ, i.e.,

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

Let $U_i(\theta_i) = (U_i^1(\theta_i), \dots, U_i^K(\theta_i))$.

Here is the condition of *type diversity* of Serrano and Vohra (2005):

Definition 7 (Type Diversity) *An environment $\mathcal{E}$ satisfies type diversity (TD) if there do not exist $i \in N$, $\theta_i, \theta_i^{'} \in \Theta_i$ with $\theta_i \neq \theta_i^{'}$, $\beta \in \mathbb{R}_{++}$ and $\gamma \in \mathbb{R}$ such that*

$$U_i(\theta_i) = \beta U_i(\theta_i^{'}) + \gamma e,$$

where e is the unit vector in Δ^{K-1}.

Clearly, under type diversity, the measurability algorithm stops after the first iteration, leading to the finest partion possible – all types are separated. As a result, all SCFs satisfy AM-measurability.

In this paper, we restrict our attention to well behaved mechanisms where best responses are always well defined. The next definitions are borrowed from AM (1992):

For every $i \in N$ and every partition Ψ_i, let $\Sigma_i(\Psi_i)$ denote the set of mixed strategies of player i that are measurable with respect to Ψ_i.

Definition 8 *The profile $\sigma \in \Sigma_1(\Psi_1) \times \cdots \times \Sigma_n(\Psi_n)$ is a* ***pseudo-Bayesian equilibrium*** *with respect to Ψ if for all $i \in N$ and all $\psi_i \in \Psi_i$, there exists some θ_i with $\theta_i \in \psi_i$ such that*

$$U_i(g \circ \sigma | \theta_i) \geq U_i(g \circ (\sigma_i^{'}, \sigma_{-i}) | \theta_i) \ \ \forall \sigma_i^{'} \in \Sigma_i$$

Definition 9 (Regular Mechanisms) *A mechanism Γ is said to be* ***regular*** *if, for each Ψ, there exists a pseudo-Bayesian equilibrium with respect to Ψ.*

In particular, finite mechanisms - like the ones constructed in AKS, AM, and BM - are regular. Mechanisms that rely on the use of integer games are not regular. More importantly, Bergemann, Morris, and Tercieux (2010) do employ such non-regular mechanisms for their sufficiency result.

3 Restricted Deception-Proofness

This section introduces a new property of SCFs and contains our main result.

Let $\mathcal{F}$ be the set of all SCFs.

Definition 10 (Strategically Identical Preferences) *An environment $\mathcal{E}$ satisfies* ***strategically identical preferences*** *at the set of types Θ_0 and the admissible class of mechanisms $\tilde{\Gamma}$ if the following four properties are satisfied:*

- *(1) $\Theta_0 \subseteq \Theta_i$ for each $i \in N$;*
- *(2) $q_i(\theta_{-i}|\theta_i) = 0$ for each $\theta_i \in \Theta_0$ whenever $(\theta_i, \theta_{-i}) \notin \Theta_0^n$, where $\Theta_0^n \equiv \Theta_0 \times \cdots \times \Theta_0$;*
- *(3) there exists $V : \mathcal{F} \times \Theta_0 \to \mathbb{R}$ such that for each $i \in N$ and for each $\theta_0 \in \Theta_0$, there exist $\beta_i > 0$ and $\gamma_i \in \mathbb{R}$ such that $U_i(\cdot|\theta_0) = \beta_i V(\cdot|\theta_0) + \gamma_i$; and*
- *(4) for each mechanism $\Gamma = ((M_i)_{i \in N}, g) \in \tilde{\Gamma}$, there exists a strategy profile $\hat{\sigma}$ such that $V(g \circ \hat{\sigma}|\theta_0) \geq V(g \circ \sigma|\theta_0)$ for every σ and every $\theta_0 \in \Theta_0$.*

This definition says that, for each agent there exists a set of types Θ_0 that is exactly the same across agents. Moreover, the event consisting of the n-fold Cartesian product of Θ_0 is common knowledge among all agents. In particular, for each type $\theta_0 \in \Theta_0$, interim preferences are identical across agents. Finally, interim preferences may differ across different types in Θ_0, but, as in a pure coordination game, for any mechanism these agents could play within a certain class, there always exists a strategy profile that yields an outcome that is placed at the top of every type's interim preferences, a "common top property" for all types θ_0 within Θ_0. This last property can be automatically satisfied if we consider regular mechanisms.

A *deception* is a profile of functions, $\alpha = (\alpha_i)_{i \in N}$, where $\alpha_i : \Theta_i \mapsto \Theta_i$, $\alpha_i(\theta_i) \neq \theta_i$ for some $\theta_i \in \Theta_i$ for some $i \in N$. (Note that the identity function I on Θ is not a deception.) For an SCF f and a deception α, $f \circ \alpha$ denotes the SCF such that for each $\theta \in \Theta$, $[f \circ \alpha](\theta) = f(\alpha(\theta))$. Let $\mathcal{A}$ be the set of all deceptions union with the identity function on Θ.

The following is the main definition of this paper:

Definition 11 (Restricted Deception-Proofness) *An SCF f satisfies the restricted deception-proofness property if, whenever an environment $\mathcal{E}$ satisfies strategically identical preferences at Θ_0 and the direct mechanism for f, it follows that*

$$U_i(f|\theta_i) = \max_{\alpha \in \mathcal{A}} U_i(f \circ \alpha|\theta_i)$$

for each $i \in N$ and $\theta_i \in \Theta_0$.

Restricted deception-proofness means that, whenever the environment contains an informational event with strategically identical preferences over the strategic situation described by the SCF's direct mechanism, the SCF has a "common top" property for all types of all agents. Indeed, among all possible manipulations of the SCF, embodied by all deceptions, no type of any agent would like to use that coordinated effort to depart from truth-telling. We shall illustrate the definition in the next section.

We next present our main result:

Proposition 3 *If an SCF f is exactly implementable by a regular mechanism in iteratively undominated strategies, it satisfies restricted deception-proofness.*

Proof: Let $\Gamma = (M, g)$ be an implementing mechanism that is regular. Let $\mathcal{F}^{\Gamma}$ be the set of SCFs associated with Γ. That is,

$$\mathcal{F}^{\Gamma} = \left\{ \tilde{f} \in \mathcal{F} \middle| \; \tilde{f} = g \circ \sigma \text{ for some } \sigma \in \Sigma \right\}.$$

Since the implementing mechanism Γ is regular, property (4) of the definition of strategically identical preferences is satisfied for Γ. By our hypothesis of restricted deception-proofness, we consider an environment satisfying strategically identical preferences at Θ_0 and the mechanism Γ. In what follows, we need the following notation:

$$H^{\Gamma,\Theta_0} = \left\{ \tilde{f} \in \mathcal{F}^{\Gamma} \middle| \; \arg\max_{\tilde{f} \in \mathcal{F}^{\Gamma}} V(\tilde{f}|\theta_0) \;\; \forall \, \theta_0 \in \Theta_0 \right\} \neq \emptyset;$$

and

$$\hat{\Sigma}_i^{\Gamma,\Theta_0} = \left\{ \sigma_i \in \Sigma_i | \; g \circ \sigma \in H^{\Gamma,\Theta_0} \text{ for some } \sigma_{-i} \in \Sigma_{-i} \right\} \neq \emptyset.$$

Note that the non-emptyness of H^{Γ,Θ_0} and $\hat{\Sigma}_i^{\Gamma,\Theta_0}$ are guaranteed because the mechanism Γ is regular. Define $[\mathcal{K}^k_{\Theta_0}(\Sigma)]_i$ to be the set of agent i's strategies that are k-times iteratively undominated *when every agent's type space is restricted to* Θ_0. Let $[\mathcal{K}^*_{\Theta_0}(\Sigma)]_i$ be the corresponding strategies that are iteratively undominated. Let $\mathcal{K}^k_{\Theta_0}(\Sigma) = \times_{i \in N} [\mathcal{K}^k_{\Theta_0}(\Sigma)]_i$ and $\mathcal{K}^*_{\Theta_0}(\Sigma) = \times_{i \in N} [\mathcal{K}^*_{\Theta_0}(\Sigma)]_i$.

We claim that $\hat{\Sigma}^{\Gamma,\Theta_0} \subset \mathcal{K}^*_{\Theta_0}(\Sigma)$. First, observe that $\hat{\sigma}_i^{\Gamma,\Theta_0} \subset [\mathcal{K}^0_{\Theta_0}(\Sigma)]_i = \Sigma_i$ for each $i \in N$. We proceed by induction. According to the induction hypothesis, suppose that $\hat{\Sigma}^{\Gamma,\Theta_0} \subset \mathcal{K}^k_{\Theta_0}(\Sigma)$. Fix agent i arbitrarily. Our induction hypothesis guarantees that $\hat{\Sigma}_{-i}^{\Gamma,\Theta_0} \subseteq [\mathcal{K}^k_{\Theta_0}(\Sigma)]_{-i}$. Fix also $\hat{\sigma}_i \in \hat{\Sigma}_i^{\Gamma,\Theta_0}$ arbitrarily. By the induction hypothesis, $\hat{\sigma}_i$ is undominated with respect to $[\mathcal{K}^k_{\Theta_0}(\Sigma)]_{-i}$. And combining the strategically identical preferences assumption and the induction hypothesis, for any $\theta_0 \in \Theta_0$, there exists $\hat{\sigma}_{-i} \in \hat{\Sigma}_{-i}^{\Gamma,\Theta_0} \subset [\mathcal{K}^k_{\Theta_0}(\Sigma)]_{-i}$ such that

$$V(g \circ (\hat{\sigma}_i, \hat{\sigma}_{-i})|\theta_0) \geq V(g \circ (\sigma_i', \hat{\sigma}_{-i})|\theta_0),$$

for any $\sigma_i' \in [\mathcal{K}^k_{\Theta_0}(\Sigma)]_i$. This implies that $\hat{\Sigma}_i^{\Gamma,\Theta_0} \subset [\mathcal{K}^{k+1}_{\Theta_0}(\Sigma)]_i$. Since i was chosen arbitrarily, this shows that $\hat{\Sigma}^{\Gamma,\Theta_0} \subset \mathcal{K}^{k+1}_{\Theta_0}(\Sigma)$. This establishes that $\hat{\Sigma}^{\Gamma,\Theta_0} \subseteq \mathcal{K}^*_{\Theta_0}(\Sigma)$.

Since f is implementable in iteratively undominated strategies, we have that

$$g \circ \hat{\Sigma}^{\Gamma,\Theta_0} \subseteq g \circ \mathcal{K}^*_{\Theta_0}(\Sigma) = (f(\theta))_{\theta \in \Theta_0^n}.$$

Therefore, we can choose $\hat{\sigma} \in \mathcal{K}^*$ such that $g \circ \hat{\sigma} = f$ and $\hat{\sigma}_i \in \hat{\Sigma}_i^{\Gamma,\Theta_0}$ for all $i \in N$.

In particular, this implies that $f \in H^{\Gamma,\Theta_0}$, and hence

$$V(f|\theta_0) = V(g \circ \hat{\sigma}|\theta_0) = \max_{\sigma \in \Sigma} V(g \circ \sigma|\theta_0) \geq \max_{\alpha \in \mathcal{A}} V(f \circ \alpha|\theta_0) \text{ for each } \theta_0 \in \Theta_0.$$

Here, the last inequality follows because the set $\mathcal{F}^\Gamma$ contains the set of SCFs associated with the direct mechanism for f (i.e., f itself union with the set of $f \circ \alpha$ for all deceptions α). Thus, f satisfies restricted deception-proofness. This completes the proof. ■.

The next result is a simple, but important extension of the previous one:

Proposition 4 *If an SCF f is approximately implementable by a regular mechanism in iteratively undominated strategies, it satisfies restricted deception-proofness.*

Proof: Let $\Gamma_\varepsilon = ((M_i)_{i \in N}, g_\varepsilon)$ denote the implementing regular mechanism when the approximation is $\varepsilon > 0$. Fix $\bar{\varepsilon}$ to be small enough and consider the class of mechanisms $\tilde{\Gamma} = \bigcup_{0 \leq \varepsilon \leq \bar{\varepsilon}} \Gamma_\varepsilon$.

Define

$$\mathcal{F}^\Gamma_\varepsilon = \left\{ \tilde{f} \in \mathcal{F} \middle| \ \tilde{f} = g_\varepsilon \circ \sigma \text{ for some } \sigma \in \Sigma \right\}$$

and

$$\mathcal{F}^\Gamma = \limsup_{\varepsilon \to 0} \mathcal{F}^\Gamma_\varepsilon.$$

By our hypothesis of restricted deception-proofness, we consider an environment satisfying strategically identical preferences at Θ_0 and at the class of mechanisms $\tilde{\Gamma}$.

For each $\varepsilon \leq \bar{\varepsilon}$, let

$$H^{\Gamma_\varepsilon,\Theta_0} = \left\{ \tilde{f} \in \mathcal{F}_\varepsilon{}^\Gamma \middle| \ \arg\max_{\tilde{f}} V(\tilde{f}|\theta_0) \ \forall\, \theta_0 \in \Theta_0 \right\} \neq \emptyset;$$

and

$$\hat{\Sigma}_i^{\Gamma_\varepsilon,\Theta_0} = \left\{\sigma_i \in \Sigma_i | \; g_\varepsilon \circ \sigma \in H^{\Gamma_\varepsilon,\Theta_0} \text{ for some } \sigma_{-i} \in \Sigma_{-i}\right\} \neq \emptyset.$$

Once again, the non-emptyness of $H^{\Gamma_\varepsilon,\Theta_0}$ and $\hat{\Sigma}_i^{\Gamma_\varepsilon,\Theta_0}$ are guaranteed because the mechanism Γ_ε is regular. Define H^{Γ,Θ_0} and $\hat{\Sigma}_i^{\Gamma,\Theta_0}$ as the limits of $H^{\Gamma_\varepsilon,\Theta_0}$ and $\hat{\Sigma}_i^{\Gamma_\varepsilon,\Theta_0}$, respectively.

With the definitions so adapted, the rest of the proof proceeds as the proof of the previous proposition. ■

4 Discussion

At this point it will be useful to consider an example that first appeared in Palfrey and Srivastava (1989, Example 3) and that was extensively analyzed in Serrano and Vohra (2005, Section 5).

There are two alternatives, $A = \{a, b\}$ and three agents. Each agent has two possible types, $\Theta_i = \{\theta_a, \theta_b\}$ and each type is drawn independently with $q_i(\theta_b) = q$ for all i and $q^2 > 0.5$. Agents have identical preferences, given by

$$u_i(a,\theta) = \begin{cases} 1 & \text{if at least two agents are of type } \theta_a \\ 0 & \text{otherwise} \end{cases}$$

$$u_i(b,\theta) = \begin{cases} 1 & \text{if at least two agents are of type } \theta_b \\ 0 & \text{otherwise} \end{cases}$$

For each agent, the corresponding interim utilities for the constant SCFs assigning alternatives a and b are:

$$\begin{aligned} U_i^a(\theta_a) &= 1 - q^2, \qquad & U_i^b(\theta_a) &= q^2, \\ U_i^a(\theta_b) &= (1-q)^2, \qquad & U_i^b(\theta_b) &= 1 - (1-q)^2. \end{aligned}$$

Since $q^2 > 0.5$, this implies that $U_i^b(\theta_i) > U_i^a(\theta_i)$ for all i and $\theta_i \in \Theta_i$.

Using this, it can be checked that in this environment, only constant SCFs satisfy AM-measurability. On the other hand, as argued in Serrano and Vohra (2005), all SCFs satisfy virtual monotonicity, a necessary condition for approximate implementation in Bayesian equilibrium. Furthermore, appealing to the results in Serrano and Vohra (2009), all SCFs in this environment also satisfy mixed virtual monotonicity. It follows that all SCFs that are incentive compatible are approximately implementable in (mixed) Bayesian equilibrium. However, since AM-measurability is necessary for implementation in regular mechanisms, we know that the implementing mechanism in Bayesian equilibrium must involve the use of integer games or devices alike.

For us, what is more interesting now is the modification of the example by adding a third alternative c, which for instance gives a zero payoff to all agents in all states.[4] As argued in Serrano and Vohra (2005), the modified example satisfies type diversity, and hence, all SCFs now satisfy AM-measurability (AM (1992)). However, AM's sufficiency result cannot be applied to any non-constant SCF even then. AM use an assumption (their assumption 2) which requires that in each state the *ex-post* preferences (over lotteries) of the agents are different, which is clearly not the case in the present example. Similarly, the sufficiency results in BM –based on their economic assumption– or AKS –based on quasilinear preferences– cannot be applied either, as this example violates both of them. AM's Assumption 2, BM's economic assumption and AKS's quasilinearity feature as sufficient conditions, and until now, it was not known whether such assumptions were necessary.

We have identified a new necessary condition for exact or approximate implementation in iteratively undominated strategies, and we show next that this extra condition has some bite. Indeed, in the three-alternative example there are SCFs that are incentive compatible and AM-measurable, but that violate the restricted deception-proofness property. Thus, it is not possible to show a sufficiency result for approximate implementation in iteratively undominated strategies that relies only on incentive compatibility and AM-measurability. Extra conditions (either on the environment, like the AM, BM and AKS papers used; or on the SCF itself) must be imposed.

For many allowable values of parameter q, restricted deception-proofness boils down to incentive compatibility, and hence it does not represent a reduction in the set of implementable SCFs – although as a necessary condition by itself, it is substantially more restrictive than AM-measurability, which is trivially satisfied by all SCFs in this case.

However, there are values of q for which restricted deception-proofness has additional bite. For instance, let $q = 99/100$ and consider the following SCF f:

$$f(\theta_a, \theta_a, \theta_a) = b,$$

$$f(\theta_a, \theta_b, \theta_a) = 0.9a + 0.1b,$$

$$f(\theta_a, \theta_a, \theta_b) = 0.9a + 0.1b,$$

$$f(\theta_b, \theta_a, \theta_a) = 0.9a + 0.1b,$$

$$f(\theta_a, \theta_b, \theta_b) = 0.1a + 0.9b,$$

$$f(\theta_b, \theta_a, \theta_b) = 0.1a + 0.9b,$$

$$f(\theta_b, \theta_b, \theta_a) = 0.1a + 0.9b,$$

$$f(\theta_b, \theta_b, \theta_b) = 0.1a + 0.9b.$$

[4] All that is needed is a third alternative to ensure type diversity. No assumption regarding a universally bad outcome or anything of that sort is needed here.

We first check that f satisfies incentive compatibility:

$$U(f|\theta_a) = (99/100)^2 0.9 + 2(99/10000)0.9 = 0.89991,$$

which is strictly greater than

$$U(f, \theta_b|\theta_a) = (1/10000)0.9 + 2(99/10000)0.1 + (99/100)^2 0.9 = 0.88416.$$

And

$$U(f|\theta_b) = 0.9,$$

which is strictly greater than

$$U(f, \theta_a|\theta_b) = 2(99/10000)0.1 + (99/100)^2 0.9 = 0.88407.$$

As it can be checked, the environment satisfies strategically identical preferences at Θ (the entire payoff type space) and at the direct mechanism for f, but f violates restricted deception-proofness. Indeed, consider the deception α such that $\alpha_i(\theta_a) = \alpha_i(\theta_b) = \theta_a$ for $i = 1, 2, 3$. Note that $f \circ \alpha(\theta) = b$ for every $\theta \in \Theta$. We next compute the interim expected utilities of each of the two types for this manipulated version of the SCF:

$$U(f \circ \alpha|\theta_a) = (99/100)^2 = 0.9801 > 0.89991 = U(f|\theta_a),$$

and

$$U(f \circ \alpha|\theta_b) = 1 - (1/100)^2 = 0.9999 > 0.9 = U(f|\theta_b).$$

So, both types of each agent have an incentive to manipulate the SCF by using the proposed deception, instead of truth-telling.

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 (2009), "Robust Virtual Implementation with Incomplete Information: Towards a Reinterpretation of the Wilson Doctrine," Mimeo, University of Melbourne, McGill University and Brown University.

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

Bergemann, D. and S. Morris (2009b), "Rationalizable Implementation," Mimeo, Yale University and Princeton University

Bergemann, D., S. Morris and O. Tercieux (2010), "Rationalizable Implementation," Mimeo, Yale University, Princeton University and Paris School of Economics.

Börgers, T. (1995), "A Note on Implementation and Strong Dominance," in edited by W. Barnett et. al (eds.) "*Social Choice, Welfare, and Ethics*," Cambridge University Press.
Palfrey, T. and S. Srivastava (1989), "Mechanism Design with Incomplete Information: a Solution to the Implementation Problem," *Journal of Political Economy* **97**, 668-691.
Serrano, R. and R. Vohra (2005), "A Characterization of Virtual Bayesian Implementation," *Games and Economic Behavior* **50**, 312-331.
Serrano, R. and R. Vohra, (2009), "Multiplicity of Mixed Equilibria in Mechanisms: A Unified Approach to Exact and Approximate Implementation," Mimeo, Brown University.