11-2011

A new necessary condition for implementation in iteratively undominated strategies

Takashi KUNIMOTO
Singapore Management University, tkunimoto@smu.edu.sg

Roberto SERRANO

Citation

KUNIMOTO, Takashi and SERRANO, Roberto. A new necessary condition for implementation in iteratively undominated strategies. (2011). *Journal of Economic Theory*. 146, (6), 2583-2593.
Available at: https://ink.library.smu.edu.sg/soe_research/2004

Published in Journal of Economic Theory, Volume 146, Issue 6, November 2011, pp. 2583-2595.
http://doi.org/10.1016/j.jet.2011.02.005

A New Necessary Condition for Implementation in Iteratively Undominated Strategies *

Takashi Kunimoto† and Roberto Serrano‡

First Version: January 2010
This Version: February 2011

Abstract

We uncover a new necessary condition for implementation in iteratively undominated strategies by mechanisms that satisfy the "best element property" where for each agent, there exists a strategy profile that gives him the highest payoff in the mechanism. This class includes finite and regular mechanisms. We conclude that either the quasilinearity-like assumptions of available sufficiency results cannot be completely dispensed with or some mechanisms that do not satisfy the best element property must be employed. 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 by mechanisms satisfying the best element property 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.

*We are grateful to Xiao Luo for valuable comments and discussions. We are indebted to an associate editor and anonymous referees for detailed comments and suggestions that significantly improved this paper. 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 and Dept. of Economics, Hitotsubashi University, Kunitachi, Tokyo, Japan; takashikunimoto9@gmail.com

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

1 Introduction

Aside from incentive compatibility, the necessary conditions for implementation in iteratively undominated strategies are typically viewed as very permissive.[1] For example, in a standard Bayesian environment with incomplete information, in which type spaces are common knowledge, Abreu and Matsushima [1] [AM, henceforth] show that both incentive compatibility and their measurability condition (referred 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 [14], 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 [3] [BM from now on] and in Artemov, Kunimoto and Serrano [2] [AKS in the sequel].

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 ranking. BM make use of an economic assumption, which is essentially a robust analogue of AM's Assumption 2. Because of their robustness considerations, BM require 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 essentially postulate that the set of alternatives includes a numeraire, on which an arbitrarily small off-equilibrium penalty can be imposed. As a nice byproduct, these papers obtain sufficiency results using *finite* mechanisms, in which best responses always exist. In all three cases, the use of these assumptions in the sufficiency proofs is seemingly minor (only to employ 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 written without the aid of such assumptions.

In this paper, we show that such assumptions cannot be dropped if we restrict our attention to mechanisms satisfying the *best element property*: in a mechanism with this property, for each agent, there always exists a strategy profile that gives him the highest payoff in the mechanism. Note that finite mechanisms and *regular* mechanisms (See AM [1] for its formal

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

definition), which essentially require that best responses always exist, satisfy the best element property. Moreover, the best element property holds even in any mechanism that relies on a "standard" integer game where each agent has to announce an integer and can choose his best outcome if his integer is the highest. Standard integer games are often used in implementation theory (See Maskin [12] for example) and so, our results encompass such non-regular mechanisms as well.

Indeed, we uncover a new necessary condition, previously overlooked, that must be added. 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 (exactly or approximately) implementable SCFs in iteratively undominated strategies. Considered by itself, restricted deception-proofness can 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 [14, 15]. Bergemann, Morris, and Tercieux [5] and Serrano and Vohra [14, 15] use what we call a *stochastic integer game* as part of their implementing mechanisms. In a stochastic integer game, each agent has to announce an integer; the higher an integer he announces the higher the probability with which he can choose his best outcome; but, no matter how high an integer he announces, he cannot obtain his best outcome with probability 1. If an SCF f violates restricted deception-proofness, any implementing mechanism (e.g., the ones mentioned relying on stochastic integer games) cannot satisfy the best element property.[2] This explains our restriction on the class of mechanisms employed. We shall provide an example, which has appeared previously in the literature, to illustrate our points.

We study incomplete information environments. Two papers containing some related results for the complete information domain are Börgers [6] and Bergemann and Morris [4]. Bergemann and Morris [4] show a similar result for virtual implementation by finite mechanisms under complete information. Börgers [6] obtains an impossibility result under complete information when only deterministic finite mechanisms are allowed and all possible identical (strict) preferences are included in the domain of SCFs. We will discuss connections of our work to Börgers [6] at the end of the paper.

[2]In fact, Bergemann, Morris, and Tercieux [5] show in their Proposition 2 that no assumptions that amount to quasilinearity are needed for exact implementation in iteratively undominated strategies under complete information. We conjecture that a result similar to their Proposition 2 can be proven in incomplete information environments.

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, who 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 (nonempty) 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.[4]

Definition 1 (Strict Dominance) *A strategy $\sigma_i \in H_i$ is strictly dominated for agent i with respect to $H = \times_{j \in N} H_j$ if there exist $\theta_i \in \Theta_i$ and $\sigma_i' \in \Sigma_i$ such that for every $\sigma_{-i} \in \times_{j \neq i} H_j$,*

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

For any subsets $H, H' \subseteq \Sigma$, where $H' \subseteq H$, we use the notation $H \to H'$ (read: H is reduced to H') to mean that for any $\sigma \in H \backslash H'$, some σ_i is

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

[4] Our notation seems to assume that a message space M can be either finite or countable. However, we can also handle the case of uncountable M. In doing so, we must impose some suitable measurability condition on M so that the corresponding strategy space Σ_i and interim preferences $U_i(g \circ \sigma|\theta_i)$ are well-defined. See Duggan [9] and Serrano and Vohra [15] for this treatment.

strictly dominated with respect to H. Let λ^0 denote the *first* element in an *ordinal* Λ; let $\geq$ be the linear order on Λ; and let λ' denote a *successor* to λ in Λ.[5] Let $\{\mathcal{K}^\lambda\}_{\lambda\in\Lambda}$ be a finite, countably infinite, or uncountably infinite family of subsets of the strategy space Σ satisfying the following properties: (1) $\mathcal{K}^{\lambda_0} = \Sigma$; (2) $\mathcal{K}^\lambda \to \mathcal{K}^{\lambda'}$ where $\mathcal{K}^\lambda = \bigcap_{\lambda''<\lambda} \mathcal{K}^{\lambda''}$ for a limit ordinal λ and any successor λ'; and (3) $\mathcal{K}^* \equiv \bigcap_{\lambda\in\Lambda} \mathcal{K}^\lambda \to \mathcal{K}$ only for $\mathcal{K} = \mathcal{K}^*$.[6]

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

Remark: The above definition of iterative dominance allows for eliminating dominated strategies possibly by using strategies that have previously been eliminated.[7] This is equivalent to the standard definition of iterative dominance as long as we consider *finite* mechanisms. In fact, all the canonical mechanisms proposed by AM, BM and AKS are finite. Although the standard definition of iteratively undominated strategies uses only a countably infinite number of rounds of elimination, we could also allow for an *uncountable* number of rounds. This extension is sometimes necessary if we go beyond finite mechanisms. Fortunately, even allowing an uncountable number of rounds of elimination, Chen, Long, and Luo [7] show in their Theorem 1 that $\mathcal{K}^*$ always exists and is unique (i.e., order independent). This is true for *any* mechanism. Hence, $\mathcal{K}^*$ is well defined.

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 there exists a unique $\{\sigma\} = \mathcal{K}^*$ for which $g(\sigma(\theta)) = f(\theta)$ for all $\theta \in \Theta$.*

Remark: If σ is a unique iteratively undominated strategy profile, we gain nothing by allowing the agents to use "mixed strategies." Given the definition of implementability above, we automatically guarantee the existence of

[5] An ordinal Λ is a well-ordered set in the order-isomorphic sense. In particular, the well-ordered set of natural numbers is called the *first infinite ordinal*. By saying that λ' is a successor of λ, we mean that $\lambda' > \lambda$. A *limit ordinal* is an element in Λ which is not a successor.

[6] Although the concepts used here may look complex, the essential idea for iterative dominance is the same as the case of a countable number of eliminations. The reader is referred to AM [1] for the countable case of iterative dominance.

[7] Consider Example 5 of Dufwenberg and Stegeman [8] where there is only one player with $M_1 = (0, 1)$ and $u_1(x) = x$ for each $x \in M_1$. Following the standard definition of strict dominance, every $x \in (0, 1)$ can be iteratively undominated, i.e, the outcome induced by iterative dominance is order dependent. In finite games, it is well known that a strategy is a Nash equilibrium if it is the unique outcome of iterative dominance. However, this infinite game has no Nash equilibrium. If we instead use our definition of strict dominance, the set of iteratively undominated strategies is empty, which is consistent with the fact that there is no Nash equilibrium. This is the main reason why we allow for eliminating dominated strategies possibly by using strategies that have been previously eliminated.

a unique iteratively undominated strategy profile, which is a unique Bayesian Nash equilibrium as well. This equilibrium is furthermore in pure strategies (see Chen, Long and Luo [7]).

Consider the following metric on SCFs:

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

The notation $f(a|\theta)$ refers to the probability with which f implements $a \in A$ in the state θ.[8]

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 $i \in N$, $\theta_i, \theta_i' \in \Theta_i$,*

$$U_i(f|\theta_i) \geq U_i(f; \theta_i'|\theta_i)$$

Our definition of implementability implies implementability in (pure or mixed) Bayesian (Nash) equilibrium as well. It follows that incentive compatibility is necessary for "exact" implementation in iteratively undominated strategies. By a standard continuity argument, one can easily show that incentive compatibility is also necessary for "approximate" implementation in iteratively undominated strategies. Hence, we have the following:

Proposition 1 *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 (up to positive affine transformations) as under type θ_i' when evaluating any SCF that is measurable with respect to $\Theta_i \times \Psi_{-i}$.

[8] AM [1] and BM [3] use the ℓ_1-norm to measure the distance between two SCFs, while we use the ℓ_∞-norm. However, since we restrict our attention to finite environments (i.e., the set of pure alternatives and the type space are both finite), this change does not make any substantial difference.

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

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

Since Θ_i is finite for each agent $i \in N$, there exists a positive integer L such that $\psi^h = \psi^L$ for any $h \geq L$. Therefore, at most a finite number of iterations in the algorithm suffices for determining Ψ^*.

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

The following is an important result in AM [1]:

Proposition 2 *(AM [1]) If an SCF f is either exactly or approximately implementable in iteratively undominated strategies, then it satisfies AM-measurability.*

Remark: Although we assume that Θ_i is finite, we can extend the above result to both countably infinite and uncountably infinite Θ_i. To do so, we have to allow the use of transfinite induction in the AM-measurability algorithm as we did for the iterative removal procedure of strictly dominated strategies. See Kunimoto and Serrano [10] for details of this treatment.

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 $A = \{a_1, \ldots, a_K\}$. 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 $U_i(\theta_i) = (U_i^1(\theta_i), \ldots, U_i^K(\theta_i))$.

Here is the condition of *type diversity*, as stated in Serrano and Vohra [14]:

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 $\mathbb{R}^K$.

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

In this paper, we restrict attention to a class of mechanisms where for each agent, there always exists a strategy profile that gives him the highest payoff in the mechanism.

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

Definition 8 *Let $\Gamma = (M, g)$ be a mechanism. The profile $\sigma \in \Sigma_1(\Psi_1) \times \cdots \times \Sigma_n(\Psi_n)$ is agent i's **best strategy profile** with respect to Ψ if for all $\psi_i \in \Psi_i$, there exists some θ_i with $\theta_i \in \psi_i$ such that*

$$U_i(g \circ \sigma | \theta_i) \geq U_i(g \circ \tilde{\sigma} | \theta_i) \ \ \forall \tilde{\sigma} \in \Sigma(\Psi).$$

Definition 9 (Best Element Property) *A mechanism Γ is said to satisfy the **best element property** if, for each Ψ, for every agent $i \in N$, there exists a best strategy profile with respect to Ψ.*

Remark: See the introduction (specifically, its third paragraph) for the implications of the best element property.

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 (Identical Preferences) *An environment $\mathcal{E}$ satisfies **identical preferences** at a nonempty subset of the type space $\hat{\Theta} \subseteq \Theta$ if there exists a set Θ_0 with the following three properties:*

- *(1) for each agent $i \in N$, there exists a bijection $\phi_i : \Theta_0 \to \hat{\Theta}_i$ where $\hat{\Theta}_i \subseteq \Theta_i$;*

- *(2)* $\sum_{\theta_{-i} \in \hat{\Theta}_{-i}} q_i(\theta_{-i}|\theta_i) = 1$ *for each* $\theta_i \in \hat{\Theta}_i$*; and*
- *(3) there exists* $V : \mathcal{F} \times \Theta_0 \to \mathbb{R}$ *such that for each* $i \in N$*, there exist* $\beta_i > 0$ *and* $\gamma_i \in \mathbb{R}$ *for which* $U_i(\cdot|\theta_i) = \beta_i V(\cdot|\phi_i^{-1}(\theta_i)) + \gamma_i$ *for each* $\theta_i \in \hat{\Theta}_i$*.*

Property (1) says that, for each agent $i \in N$, there exists a set of types $\hat{\Theta}_i$ that can be embedded in the common type space Θ_0 across agents. Property (2) says that the event consisting of the n-fold Cartesian product of $\hat{\Theta} \equiv \hat{\Theta}_1 \times \cdots \times \hat{\Theta}_n$ is a belief-closed subspace of Θ. In words, $\hat{\Theta}$ can be treated separately from the rest of the type space. In particular, property (3) says that interim preferences are identical across agents within $\hat{\Theta}$. Note that when we consider *complete information* environments, each $\hat{\Theta}_i$ as well as Θ_0 become a singleton.

A *deception* is a profile of functions, $\alpha = (\alpha_i)_{i \in N}$, where $\alpha_i : \Theta_i \to \Theta_i$, $\alpha_i(\theta_i) \neq \theta_i$ for some $\theta_i \in \Theta_i$ for some $i \in N$. (Note that the identity function $I : \Theta \to \Theta$ 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. Note that $\mathcal{A}$ can be considered the entire strategy space Σ of the *direct* mechanism.

The following is the central 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 identical preferences at $\hat{\Theta}$, 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 $\theta_i \in \hat{\Theta}_i$.

Restricted deception-proofness means that, whenever the environment contains an informational event with 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. Importantly, Bergemann, Morris, and Tercieux [5] and Serrano and Vohra [14, 15] employ canonical mechanisms that do not satisfy the best element property for their sufficiency result if the SCF violates restricted deception-proofness. In those sufficiency results, the authors use what we call a *stochastic* integer game as part of the construction of the mechanisms where each agent has to announce an integer; the higher an integer he announces the higher the probability that he can choose his best outcome; however, no matter how high an integer he announces, each agent cannot choose his best outcome with probability 1.

We next present our main result:

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

Remark: The reader is referred to the introduction (specifically, its fourth paragraph) which explains why our result needs to qualify the class of mechanisms employed.

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

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

By our hypothesis of restricted deception-proofness, we consider an environment satisfying identical preferences at some nonempty belief-closed subspace $\hat{\Theta}$. Accordingly, we also have the associated common type space Θ_0 and each agent i's bijection $\phi_i : \Theta_0 \to \hat{\Theta}_i$. In what follows, we need the following notation:

$$H^{\Gamma,\hat{\Theta}} = \left\{ \tilde{f} \in \mathcal{F}^\Gamma \middle| \ \arg\max_{\tilde{f} \in \mathcal{F}^\Gamma} V(\tilde{f} | \phi_i^{-1}(\theta_i)) \ \ \forall\ \theta_i \in \hat{\Theta}_i, \ \forall i \in N \right\};$$

and

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

Note that we can guarantee that $H^{\Gamma,\hat{\Theta}} \neq \emptyset$ and $\hat{\Sigma}_i^{\Gamma,\hat{\Theta}} \neq \emptyset$ because the mechanism Γ satisfies the best element property. Define $[\mathcal{K}^k_{\hat{\Theta}}(\Sigma)]_i$ to be the set of agent i's strategies that are k-times iteratively undominated *when every agent's type space is restricted to the belief-closed subspace* $\hat{\Theta}$. Let $[\mathcal{K}^*_{\hat{\Theta}}(\Sigma)]_i$ be the corresponding strategies that are iteratively undominated. Let $\mathcal{K}^k_{\hat{\Theta}}(\Sigma) = \times_{i \in N} [\mathcal{K}^k_{\hat{\Theta}}(\Sigma)]_i$.[9]

We claim that $\hat{\Sigma}^{\Gamma,\hat{\Theta}} \subseteq \mathcal{K}^*_{\hat{\Theta}}(\Sigma)$. First, observe that $\hat{\Sigma}_i^{\Gamma,\hat{\Theta}} \subseteq [\mathcal{K}^0_{\hat{\Theta}}(\Sigma)]_i = \Sigma_i$ for each $i \in N$. We proceed by induction. According to the induction hypothesis, suppose that $\hat{\Sigma}^{\Gamma,\hat{\Theta}} \subseteq \mathcal{K}^k_{\hat{\Theta}}(\Sigma)$. Fix agent i arbitrarily. Our induction hypothesis guarantees that $\hat{\Sigma}_{-i}^{\Gamma,\hat{\Theta}} \subseteq [\mathcal{K}^k_{\hat{\Theta}}(\Sigma)]_{-i}$. Fix also $\hat{\sigma}_i \in \hat{\Sigma}_i^{\Gamma,\hat{\Theta}}$ arbitrarily.

[9]To save notation, here we only consider a countably infinite number of elimination of strategies. Whenever we need the transfinite induction to deal with infinite mechanisms, the induction argument below can be easily fixed accordingly.

By the induction hypothesis, there exists $\hat{\sigma}_{-i} \in \hat{\Sigma}_{-i}^{\Gamma,\hat{\Theta}} \subseteq [\mathcal{K}_{\hat{\Theta}}^{k}(\Sigma)]_{-i}$ such that for any $\theta_i \in \hat{\Theta}_i$,

$$V(g \circ (\hat{\sigma}_i, \hat{\sigma}_{-i})|\phi_i^{-1}(\theta_i)) \geq V(g \circ (\sigma_i^{'}, \hat{\sigma}_{-i})|\phi_i^{-1}(\theta_i)),$$

for any $\sigma_i^{'} \in \Sigma_i$. This implies $\hat{\Sigma}_i^{\Gamma,\hat{\Theta}} \subseteq [\mathcal{K}_{\hat{\Theta}}^{k+1}(\Sigma)]_i$. Since we have identical preferences at $\hat{\Theta}$, we can conclude that $\hat{\Sigma}^{\Gamma,\hat{\Theta}} \subseteq \mathcal{K}_{\hat{\Theta}}^{k+1}(\Sigma)$. This establishes that $\hat{\Sigma}^{\Gamma,\hat{\Theta}} \subseteq \mathcal{K}_{\hat{\Theta}}^{*}(\Sigma)$.

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

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

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

In particular, this implies that $f \in H^{\Gamma,\hat{\Theta}}$, and hence for each $\theta_i \in \hat{\Theta}_i$ and each $i \in N$, we have

$$V(f|\phi_i^{-1}(\theta_i)) = V(g \circ \hat{\sigma}|\phi_i^{-1}(\theta_i)) = \max_{\sigma \in \Sigma} V(g \circ \sigma|\phi_i^{-1}(\theta_i)) \geq \max_{\alpha \in \mathcal{A}} V(f \circ \alpha|\phi_i^{-1}(\theta_i)).$$

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 in iteratively undominated strategies by a mechanism satisfying the best element property, it satisfies restricted deception-proofness.*

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

Define

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

and

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

By our hypothesis of restricted deception-proofness, we consider an environment satisfying identical preferences at some nonempty subsets of the

type space $\hat{\Theta}$. Accordingly, we also have the associated common type space Θ_0 and each agent i's bijection $\phi_i : \Theta_0 \to \hat{\Theta}_i$.

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

$$H^{\Gamma_\varepsilon,\hat{\Theta}} = \left\{ \tilde{f} \in \mathcal{F}^{\Gamma_\varepsilon} \,\middle|\, \arg\max_{\tilde{f}} V(\tilde{f}|\phi_i^{-1}(\theta_i)) \;\; \forall\, \theta_i \in \hat{\Theta}_i, \;\; \forall i \in N \right\};$$

and

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

Once again, we can guarantee that $H^{\Gamma_\varepsilon,\hat{\Theta}} \neq \emptyset$ and $\hat{\Sigma}_i^{\Gamma_\varepsilon,\hat{\Theta}} \neq \emptyset$ because the mechanism Γ_ε satisfies the best element property. Define $H^{\Gamma,\hat{\Theta}}$ and $\hat{\Sigma}_i^{\Gamma,\hat{\Theta}}$ as the limits of $H^{\Gamma_\varepsilon,\hat{\Theta}}$ and $\hat{\Sigma}_i^{\Gamma_\varepsilon,\hat{\Theta}}$, respectively.[10] With the definitions so adapted, the rest of the proof proceeds as the proof of the previous proposition. ■

4 Discussion

4.1 An Example

At this point it will be useful to consider an example that first appeared in Palfrey and Srivastava ([13], Example 3) and that was extensively analyzed in Serrano and Vohra ([14], 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:

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

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

[10] We take a sequence along which such limits exist.

Using this, it can be checked that in this environment, only constant SCFs satisfy AM-measurability.

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.[11] As argued in Serrano and Vohra [14], the modified example satisfies type diversity, and hence, all SCFs now satisfy AM-measurability. However, AM's sufficiency result cannot be applied to any non-constant SCF even then.[12]

We have identified a new necessary condition for exact or approximate implementation in iteratively undominated strategies, and we show next that in the three-alternative example there are SCFs that satisfy incentive compatibility and AM-measurability, 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 by a mechanism satisfying the best element property. Extra conditions (either on the environment, like the AM, BM and AKS papers used; or on the SCF itself) must be imposed.

For instance, let $q = 99/100$ and consider the following SCF f:

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

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

Noting the fact that all agents have identical preferences, we first check that f satisfies incentive compatibility:

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

which is strictly greater than

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

And

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

[11] 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.

[12] See again the introduction (specifically, its second paragraph).

which is strictly greater than

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

As can be checked, the environment satisfies identical preferences at Θ (the entire type space) 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:

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

and

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

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

Suppose that the SCF f is exactly implementable in iteratively undominated strategies by a mechanism $\Gamma = (M, g)$ satisfying the best element property. Since all agents have identical preferences at Θ and the mechanism Γ satisfies the best element property, there must exist a strategy profile $\sigma \in \Sigma$ such that $U_i(g \circ \sigma|\theta_i) \geq \max_{\tilde{\alpha} \in \mathcal{A}} U_i(f \circ \tilde{\alpha}|\theta_i)$ for any $\theta_i \in \Theta_i$ and any $i \in N$. By our Proposition 3, we have shown that such σ is indeed an iteratively undominated strategy profile, i.e., $\sigma \in \mathcal{K}^*$. By our hypothesis of implementability, we must have $U_i(g \circ \sigma|\theta_i) = U_i(f|\theta_i)$ for each $\theta_i \in \Theta_i$ and each $i \in N$. However, this contradicts the fact that, for the deception α of the previous paragraph, $U_i(f \circ \alpha|\theta_i) > U_i(f|\theta_i)$ for each $\theta_i \in \Theta_i$ and each $i \in N$.

4.2 The Connection with a Result of Börgers's (1995)

Börgers [6] shows the following remarkable result under complete information:

Theorem [Börgers [6]] Let $f : \Theta \to A$ be a deterministic full-range SCF where $|A| \geq 2$. Suppose that every agent has strict preferences over A in each state $\theta \in \Theta$. Assume further that the domain of SCFs Θ contains all possible unanimous strict preferences over A. Then, the SCF f is exactly implementable in iteratively undominated strategies by a finite mechanism if and only if it is dictatorial.

Note that the "if"-part is easy to establish. The striking part of the result is its the "only if"-part. Assume that an SCF f is exactly implementable by a finite mechanism. We can decompose the (only if)-part of the proof

into the following three steps: (Step 1): $f(\theta) = a$ if all agents have identical preferences at θ with a being the best outcome; (Step 2): For each outcome $a \in A$, there exists exactly one agent $i(a) \in N$ who can either enforce or exclude outcome a; and (Step 3): $i(a)$ must be the same agent across all outcomes. Hence, he is a dictator.

Under the full range assumption, our Proposition 3 can be considered an extension of Step 1 of Börgers [6] from complete information to incomplete information. In doing so, we expand the applicability of Börgers's argument in Step 1 to encompass a much richer class of environments: (1) we can take care of von Neumann-Morgenstern utility functions over the lottery space; (2) we can deal with stochastic SCFs; (3) we can include infinite mechanisms, as long as they satisfy the best element property; and (4) we can show that the result holds for approximate implementation as well as exact implementation.

However, the good news stop there. As already pointed out by Börgers [6], the use of stochastic mechanisms seems to create real difficulties to the rest of the argument. We have not been able to prove the extension of the result, but we have not been able to find a counterexample either. In this regard, the connections with Majumdar and Sen [11] and their impossibility result for ordinally Bayesian incentive compatibility seem quite relevant. We leave this intriguing issue as an important open question.

References

[1] D. Abreu and H. Matsushima, Virtual implementation in iteratively undominated strategies: incomplete Information, Mimeo, (1992), Princeton University.

[2] G. Artemov, T. Kunimoto, and R. Serrano, Robust virtual implementation with incomplete information: towards a reinterpretation of the Wilson doctrine, Mimeo, (2010), University of Melbourne, McGill University and Brown University.

[3] D. Bergemann and S. Morris, Robust virtual implementation, *Theoretical Economics* 4 (2009), 45-88.

[4] D. Bergemann and S. Morris, Rationalizable implementation, Mimeo, (2009), Yale University and Princeton University

[5] D. Bergemann, S. Morris and O. Tercieux, Rationalizable implementation, forthcoming in *J. Econ. Theory* (2010).

[6] T. Börgers, A note on implementation and strong dominance, W. Barnett et. al (Eds.), Social Choice, Welfare, and Ethics, Cambridge University Press, 1995, pp. 277-287.

[7] Y-C. Chen, N-V. Long, and X. Luo, Iterated strict dominance in general games, *Games Econ. Behav.* 61 (2007), 299-315.

[8] M. Dufwenberg and M. Stegeman, Existence and uniqueness of maximal reductions under iterated strict dominance, *Econometrica* 70 (2002), 2007-2023.

[9] J. Duggan, Virtual Bayesian implementation, *Econometrica* 65 (1997), 1175-1199.

[10] T. Kunimoto and R. Serrano, Evaluating the conditions for robust mechanism design, Mimeo, (2010), McGill University and Brown University.

[11] D. Majumdar and A. Sen, Ordinally Bayesian incentive compatible voting rules, *Econometrica* 72 (2004), 523-540.

[12] E. Maskin, Nash equilibrium and welfare optimality, *Rev. Econ. Stud.*, 66 (1999), 23-38.

[13] T. Palfrey and S. Srivastava, Mechanism design with incomplete information: a solution to the implementation problem, *J. Polit. Economy* 97 (1989), 668-691.

[14] R. Serrano and R. Vohra, A characterization of virtual Bayesian implementation, *Games Econ. Behav.* 50 (2005), 312-331.

[15] R. Serrano and R. Vohra, Multiplicity of mixed equilibria in mechanisms: a unified approach to exact and approximate implementation, *J. Math. Econ.* 46 (2010), 775-785.