10-2013

Robust virtual implementation with almost complete information

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

Robust Virtual Implementation with Almost Complete Information*

Takashi Kunimoto†

First Version: February 2008
This Version: October 2013

Abstract

Artemov, Kunimoto, and Serrano (2013a,b, henceforth, AKS) study a mechanism design problem where arbitrary restrictions are placed on the set of first-order beliefs of agents. Calling these restrictions Δ, they adopt Δ-rationalizability (Battigalli and Siniscalchi (2003)) and show that Δ-incentive compatibility and Δ-measurability are necessary and sufficient conditions for robust virtual implementation. By appropriately defining Δ in order to restrict attention to *complete information* environments, I exploit the implications of AKS and show that the permissive implementation result of Abreu and Matsushima (1992a) is robust to how the underlying type space is specified. However, AKS need to fix a complete information environment throughout their analysis and therefore does not enable us to ask if robust virtual implementation results are "robust" to the relaxation of the complete information environment. The main result of this paper shows that permissive robust virtual implementation results can be extended to nearby incomplete information environments.

JEL Classification: C72, D78, D82.

Keywords: complete information, first-order belief, incentive compatibility, measurability, robust virtual implementation, rationalizable strategies, .

---

*I am grateful to an associate editor and two anonymous referees for their comments and suggestions that substantially improve the paper. I am thankful to Yi-Chun Chen, Jihong Lee, Daisuke Oyama, Stephen Morris, Roberto Serrano, and Satoru Takahashi for their comments. This paper also benefits from the comments of the audience at Concordia University, McGill University, National University of Singapore, Osaka University, Seoul National University, and the SAET conference in Brisbane. I gratefully acknowledge financial support from Japan Society for Promotion of Science, Japan Center for Economic Research, and Yamada Academic Research Fund.

†Department of Economics, Hitotsubashi University, Kunitachi, Tokyo, JAPAN; takashi_kunimoto@econ.hit-u.ac.jp

1 Introduction

The theory of implementation or mechanism design attempts to identify the conditions under which a social choice rule may be decentralized through some institution (or mechanism); that is, when agents, acting on their own self-interest, can arrive at the outcomes prescribed by the social choice rule.[1] This paper addresses the question of *full implementation*, which requires that the set of outcomes prescribed by a given solution concept coincide with the social choice rule. *Virtual implementation* means that the planner contents himself with implementing the social choice rule with arbitrarily high probability. This is an approximate version of *exact* implementation, which insists on implementing the social choice rule with probability 1.

Typically, achieving a correct design of institutions depends on the knowledge of key parameters in the environment and each agent privately possesses some information about these parameters. An agent's private information is summarized by the notion of *type*. For an agent, a type specifies (i) his private information about his own preferences and/or the preferences of others (*payoff type*), (ii) his belief about the payoff types of others (*first-order belief*), (iii) his belief about others' first-order beliefs (*second-order belief*), and so on, leading to a hierarchy of beliefs *ad infinitum*. A basic assumption of the classic approach to mechanism design is that the underlying spaces of types are common knowledge among the planner and the agents. In making this assumption, one effectively assumes that each first-order belief corresponds to a unique infinite hierarchy of beliefs.

This common-knowledge assumption is often seen as unrealistic. To make their analysis *robust* to the specification of higher-order beliefs, Artemov, Kunimoto, and Serrano (2012a,b) (henceforth, AKS) use a type-free solution concept of rationalizability – Δ-*rationalizability* (Battigalli and Siniscalchi (2003)) – that guarantees that the predictions are the same for any higher-order beliefs, as long as those predictions are consistent with their Δ restriction on the first-order beliefs. *Robust implementation* is the requirement that implementation survive any specification of higher-order beliefs consistent with the common knowledge structure of the environment (i.e., consistent with Δ-restriction). AKS show that in quasi-transferable environments (soon to be defined), a social choice function (henceforth, SCF) is robustly virtually implementable if and only if it satisfies Δ-*incentive compatibility* and Δ-*measurability*.

The current paper aims at exploiting the implications of AKS by restricting attention to *complete information* environments, which have received a lot of special attention in the literature. More specifically, I characterize a complete information environment as Δ-restrictions on the set of first-order beliefs. Thus, this paper treats complete information environments in the same way as it does incomplete information environments, while the implementation literature commonly treat these two environments in separate papers.[2] This paper

[1]For surveys on implementation, see, for example, Jackson (2001) and Serrano (2004).

[2]Notable exceptions are A&M (1992b) and Mookherjee and Reichelstein (1992).

offers two main results. First, in complete information environments where there are at least three agents, *any* SCF is robustly virtually implementable. In fact, Abreu and Matsushima (A&M, henceforth, 1992a,b) already established the same result but restricted their attention to a fixed finite type space. So, I show that A&M's (1992a) result is robust to how one fixes the underlying type space. Second, I show that permissive robust virtual implementation results can be extended to nearby incomplete information environments. Since the set of complete information environments is "non-generic" relative to the set of all incomplete information environments, it is desirable to check whether or not the implementation results are robust to a small amount of incomplete information. This robustness question is answered in the affirmative: when there are at least three agents, *any* SCF is robustly virtually implementable under *almost* complete information.

This exhibits a stark contrast with the implementation literature using refinements of Nash equilibrium. It is well known that almost any SCF is *exactly* implementable using refinements of Nash equilibrium. Despite these permissive implementation results, Chung and Ely (2003) show that if a mechanism implements a non-Maskin monotonic SCF in undominated Nash equilibrium, there are a nearby incomplete information environment and an undominated Bayesian Nash equilibrium that is "not" close to any of undominated Nash equilibria.[3] *Maskin monotonicity* is known to be a necessary condition for Nash implementation (See Maskin (1999)) and it is quite demanding in some contexts.[4] Muller and Satterthwaite (1977) states that any onto, ex post efficient SCF defined on the domain of all strict preferences over a finite set of alternatives is dictatorial if it satisfies Maskin monotonicity. Maskin (1999) shows that with only two agents, this result extends to social choice correspondences. Moreover, Aghion, Fudenberg, Holden, Kunimoto, and Tercieux (2012, henceforth, AFHKT) show that (1) if an SCF is implementable in subgame perfect equilibrium by a *Moore-Repullo mechanism*, its unique subgame perfect equilibrium cannot be approximated by "any" equilibrium in some nearby incomplete information games [5] and (2) if a mechanism implements a non-Maskin monotonic SCF in subgame perfect equilibrium, there are a nearby incomplete information environment and a sequential equilibrium that is "not" close to any subgame perfect equilibria.[6]

Both Chung and Ely (2003) and AFHKT (2012) essentially show that if an SCF is Nash implementable, it is robustly implementable under almost complete information. However, their robustness requirement is much weaker than that adopted in this paper: these authors (i) fix a finite type space all the time; (ii) only perturb the common prior distribution over that fixed

[3] Undominated (Bayesian) Nash equilibrium is a (Bayesian) Nash equilibrium in which no player uses weakly dominated actions.

[4] See Section 6 for the definition of Maskin monotonicity.

[5] See AFHKT (2012) for the definition of Moore-Repullo mechanisms.

[6] Kunimoto (2010) proposes an even smaller class of perturbations than that of Chung and Ely (2003) and AFHKT (2012) such that the undominated Nash equilibrium correspondence is always guaranteed to have a closed graph in the limit of complete information.

type space; and (iii) require that the equilibrium correspondence be nonempty for any nearby environment and have a closed graph in the limit of complete information. On the other hand, this paper adopts a type-free solution concept – Δ-rationlizability – under complete information so that all the results of this paper do not depend upon how the underlying type space is specified. Moreover, this paper's results do not depend upon whether agents possess prior beliefs over the type space or those prior beliefs are common. Thus, virtual implementation can be considered an appropriate approach if one insists on both permissive implementation results and the robustness to almost complete information.

The rest of this paper is organized as follows: in Section 2, I introduce the preliminary notation and definitions. In Section 3, I define a complete information environment as an incomplete information environment with Δ restrictions on the set of first-order beliefs. I exploit the implications of necessary conditions for robust virtual implementation in complete information environments (Propositions 2 and 3). In Section 4, I provide a characterization of robust virtual implementation in quasi-transferable environments with complete information (Theorem 3 and Corollary 1). In Section 5, I show that there is a precise sense in which all the permissive robust virtual implementation results are "robust" to the relaxation of the complete information environments (Theorem 4 and Corollary 2). Section 6 concludes.

2 Preliminaries

Throughout the paper, I extensively follow the notation and setup of AKS (2013a). I refer the reader to their paper for most concepts discussed here. Let $N = \{1, \ldots, n\}$ denote the set of agents and Θ_i be the set of *finite* payoff-relevant (or, simply, *payoff*) types of agent i. Denote $\Theta \equiv \Theta_1 \times \cdots \times \Theta_n$, and $\Theta_{-i} \equiv \Theta_1 \times \cdots \times \Theta_{i-1} \times \Theta_{i+1} \times \cdots \times \Theta_n$.[7] Let $q_i(\theta_{-i})$ denote agent i's first-order belief that other agents receive the profile of payoff types θ_{-i}. For each $\theta_i \in \Theta_i$, let $Q_i[\theta_i] \subseteq \Delta(\Theta_{-i})$ be the (nonempty) set of admissible first-order beliefs of agent i of payoff type θ_i. A pair (θ_i, q_i) is called agent i's *first-order type*.

Let A denote the set of pure outcomes, which are assumed to be independent of the information state. Suppose $A = \{a_1, \ldots, a_K\}$ 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 $u_i : \Delta(A) \times \Theta \to \mathbb{R}$.

I can now define an *environment* as $\mathcal{E} = (A, \{u_i, \Theta_i, (Q_i[\theta_i])_{\theta_i \in \Theta_i}\}_{i \in N})$, which is implicitly understood to be common knowledge among the agents. AKS assume $Q_i[\theta_i]$ is defined uniformly over all payoff types θ_i and denote it by Q_i. Bergemann and Morris (2009, henceforth, B&M) not only consider the case that the set of first-order beliefs is uniform across payoff types but also assume the unrestricted set of first-order belief, $Q_i = \Delta(\Theta_{-i})$ for every agent

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

$i \in N$. The reason why AKS and B&M assume this uniformity is that this assumption makes it very difficult for the planner to elicit the agents' payoff types from their first-order beliefs and by this assumption, one can obtain robust implementation as a result. On the contrary, as I argue in the next section, this paper needs to assume that the set of first-order beliefs must depend upon its' payoff types to take care of the case of complete information.

Throughout the paper, I impose the following assumption on environments:

Definition 1 (Quasi-Transferability) *An environment $\mathcal{E}$ satisfies **quasi-transferability** if there exists a collection of lotteries $\{\bar{a}_i\}_{i\in N}$ and $\{\underline{a}_i\}_{i\in N}$ in $\Delta(A)$ such that for any $\theta \in \Theta$,*

1. $u_i(\bar{a}_i; \theta) > u_i(\underline{a}_i; \theta)$ *for any* $i \in N$;
2. $u_i(\underline{a}_j; \theta) \geq u_i(\bar{a}_j; \theta)$ *for any* $i, j \in N$ *with* $i \neq j$.

Remark: This assumption allows the agents to (partially) transfer their utilities among them. By making this assumption, I essentially postulate that A includes a numeraire, which can be transferred across agents. Quasi-transferability is by no means innocuous and I will discuss some implications of this assumption in Section 6.3. In what follows, I will make it clear wherever I don't need this assumption.

A *social choice function* (SCF) is a function $f : \Theta \to \Delta(A)$. Note that the domain of the SCFs is the "payoff" type space. Define $V_i(f; \theta_i^{'}|\theta_i, q_i)$ to be the interim expected utility of agent i of first-order type (θ_i, q_i) that pretends to be of payoff type $\theta_i^{'}$ corresponding to an SCF f as follows:[8]

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

where $\theta_i, \theta_i^{'} \in \Theta_i$ and $q_i \in Q_i[\theta_i]$. Denote $V_i(f|\theta_i, q_i) = V_i(f; \theta_i|\theta_i, q_i)$.

A *mechanism* $\Gamma = ((M_i)_{i\in N}, g)$ describes a (nonempty) *finite* message space M_i for each agent i and an outcome function $g : M \to \Delta(A)$, where $M = \times_{i\in N} M_i$.[9]

Next I define a message correspondence profile $S = (S_1, \ldots, S_n)$ where for each $i \in N$,

$$S_i : \Theta_i \to 2^{M_i},$$

and I write $\mathcal{S}$ for the collection of message correspondence profiles. The collection $\mathcal{S}$ is a lattice with the natural ordering of set inclusion: $S \subseteq S'$ if $S_i(\theta_i) \subseteq S_i^{'}(\theta_i)$ for all $i \in N$ and $\theta_i \in \Theta_i$. The largest element is $\bar{S} =$

[8]Note how, since the SCF does not depend on first-order beliefs, the misrepresentation of q_i into $q_i^{'}$ is of no consequence.

[9]The implementation literature often uses infinite message spaces. However, as long as quasi-transferability is assumed, the restriction to finite mechanisms can be made without loss of generality in this paper.

$(\bar{S}_1, \ldots, \bar{S}_n)$, where $\bar{S}_i(\theta_i) = M_i$ for all $i \in N$ and $\theta_i \in \Theta_i$. The smallest element is $\underline{S} = (\underline{S}_1, \ldots, \underline{S}_n)$, where $\underline{S}_i(\theta_i) = \emptyset$ for all $i \in N$ and $\theta_i \in \Theta_i$.

I define an operator $b = (b_1, \ldots, b_n)$ to iteratively eliminate never best responses. To this end, I denote the belief of agent i over message and payoff type profiles of the remaining agents by $\mu_i \in \Delta(\Theta_{-i} \times M_{-i})$. Most importantly, I introduce some restrictions on agents' first-order beliefs. For any $q_i \in \Delta(\Theta_{-i})$, define

$$\Delta^{q_i}(\Theta_{-i} \times M_{-i}) \equiv \{\mu_i \in \Delta(\Theta_{-i} \times M_{-i}) |\ \mathrm{marg}_{\Theta_{-i}} \mu_i = q_i\},$$

where $\mathrm{marg}_{\Theta_{-i}} \mu_i(\theta_{-i}) \equiv \sum_{m_{-i}} \mu_i(\theta_{-i}, m_{-i})$ for each $\theta_{-i} \in \Theta_{-i}$. The operator $b : \mathcal{S} \to \mathcal{S}$ is now defined as follows: for each $i \in N$ and $\theta_i \in \Theta_i$,

$$b_i(S)[\theta_i] \equiv \left\{ m_i \left| \begin{array}{c} \exists q_i \in Q_i[\theta_i]\ \exists \mu_i \in \Delta^{q_i}(\Theta_{-i} \times M_{-i}) \text{ s.t.} \\ \mu_i(\theta_{-i}, m_{-i}) > 0 \Rightarrow m_j \in S_j(\theta_j)\ \forall j \neq i; \text{ and} \\ m_i \in \arg\max_{m_i' \in M_i} \sum_{\theta_{-i}, m_{-i}} \mu_i(\theta_{-i}, m_{-i}) u_i(g(m_i', m_{-i}); \theta_i, \theta_{-i}) \end{array} \right. \right\}$$

This is an incomplete information version of rationalizability, proposed by Battigalli and Siniscalchi (2003). They call it Δ-*rationalizability* and denote by Δ restrictions on the set of first-order beliefs $Q_i[\theta_i]$. I observe that b is increasing by definition: $S \leq S' \Rightarrow b(S) \leq b(S')$. By Tarski's fixed point theorem, there is a largest fixed point of b, which I label S^Γ. Thus, I have that (i) $b(S^\Gamma) = S^\Gamma$ and (ii) $b(S) = S \Rightarrow S \leq S^\Gamma$. Since the mechanism is finite, I can define

$$S_i^\Gamma(\theta_i) \equiv \bigcap_{k \geq 1} b_i(b^k(\bar{S}))[\theta_i] \neq \emptyset,$$

Thus $S_i^\Gamma(\theta_i)$ are the set of messages surviving iterated deletion of never best responses; equivalently, $S_i^\Gamma(\theta_i)$ is the set of messages that player i with payoff type θ_i might send consistent with common certainty of rationality, but with some restrictions on the first-order beliefs. Note that, since the message space M is finite, $S_i^\Gamma(\theta_i) \neq \emptyset$; it is also unique. I refer to $S_i^\Gamma(\theta_i)$ as the Δ-*rationalizable* messages of payoff type θ_i of agent i in mechanism Γ.

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

$$\|y - y'\| \equiv \sum_{a \in A} |y(a) - y'(a)|,$$

where $y(a)$ and $y'(a)$ each denotes the probability that outcome a is realized. The following is the definition of robust virtual implementation.

Definition 2 (Robust Virtual Implementation) *An SCF f is **robustly virtually implementable** if, for any $\varepsilon > 0$, there exists a mechanism $\Gamma^\varepsilon = (M^\varepsilon, g^\varepsilon)$ for which for any $\theta \in \Theta$ and $m \in M^\varepsilon$,*

$$S^{\Gamma^\varepsilon}(\theta) \neq \emptyset \ \textit{and}\ m \in S^{\Gamma^\varepsilon}(\theta) \Rightarrow \|g^\varepsilon(m) - f(\theta)\| \leq \varepsilon.$$

3 Complete Information

This section and the next section focus on *complete information* environments, which are a special class of incomplete information environments.

Definition 3 (Complete Information) $\mathcal{E}^* = (A, (\Theta_i, (Q_i^*[\theta_i])_{\theta_i \in \Theta_i}, u_i)_{i \in N})$ *is said to be a **complete information** environment if there exists a set* Θ_0 *with the following two conditions:*

1. *for each agent* $i \in N$*, there exists a bijective map* $\phi_i : \Theta_i \to \Theta_0$*; and*
2. *for each* $i \in N$*,* $\theta_i \in \Theta_i$*,* $q_i \in Q_i^*[\theta_i]$*, and* $\theta_{-i} \in \Theta_{-i}$*,*

$$q_i(\theta_{-i}) = \begin{cases} 1 & \text{if } \phi_j(\theta_j) = \phi_i(\theta_i) \text{ for each } j \neq i \\ 0 & \text{otherwise} \end{cases}$$

Therefore, I can now easily characterize complete information as restrictions on the set of first-order beliefs. This also justifies including the set of first-order beliefs as part of the environment. In what follows, I denote by $\mathcal{E}^*$ a complete information environment satisfying quasi-transferability.

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

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

Since the payoff profile is common knowledge in a complete information environment $\mathcal{E}^*$, θ_i's first-order belief q_i is not needed for the expression for $V_i^k(\theta_i, q_i)$. This is very different from the incomplete information counterpart where AKS need to elicit each agent's first-order belief as part of message as well as his payoff type. In what follows, I write $V_i^k(\theta_i)$ for this:

$$V_i^k(\theta_i) = u_i(a_k; \theta_i, \phi_{-i}^{-1}(\theta_0)),$$

where $\phi_{-i}^{-1}(\theta_0) = (\phi_1^{-1}(\theta_0), \ldots, \phi_{i-1}^{-1}(\theta_0), \phi_{i+1}^{-1}(\theta_0), \ldots, \phi_n^{-1}(\theta_0))$ and $\theta_0 = \phi_i(\theta_i) \in \Theta_0$. Let $V_i(\theta_i) = (V_i^1(\theta_i), \ldots, V_i^K(\theta_i))$.

Throughout the paper, I take for granted that in complete information environments, any pair of distinct payoff profiles induces distinct (cardinal) preferences over the set of lotteries, at least for *some* agent. Formally, I assume the following.

Definition 4 *A complete information environment* $\mathcal{E}^*$ *is **non-redundant** if there do not exist* $\theta_0, \theta_0' \in \Theta_0$ *with* $\theta_0 \neq \theta_0'$ *such that for all* $i \in N$*, there exist* $\beta_i > 0$ *and* $\gamma_i \in \mathbb{R}$ *for which:*

$$V_i(\phi_i^{-1}(\theta_0)) = \beta_i V_i(\phi_i^{-1}(\theta_0')) + \gamma_i e,$$

where e *is the unit vector in* $\mathbb{R}^K$*.*

Remark: Given the very nature of the definition of the payoff type space, this assumption is innocuous. One needs an ordinal version of this domain restriction for virtual Nash implementation of Abreu and Sen (1991) as well.[10]

Define $\Theta^* = \{\theta \in \Theta | \ \phi_1(\theta_1) = \cdots = \phi_n(\theta_n) \in \Theta_0\}$ as the subset of the payoff type space in which each agent receives the "same" signal. The concern for implementation should be only on the set Θ^* because any state outside of Θ^* is supposed not to be realized under complete information. This leads to the use of the following concept: two SCFs f and h are said to be *equivalent* if $f(\theta) = h(\theta)$ for every $\theta \in \Theta^*$.[11] Let $f \approx h$ denote that f and h are equivalent.

3.1 Incentive Compatibility

The following is the standard interim incentive compatibility condition applied to the set of first-order beliefs $Q_i[\theta_i]$:

Definition 5 (Δ-Incentive Compatibility) *An SCF $f : \Theta \to \Delta(A)$ satisfies **Δ-incentive compatibility** if for every $i \in N$, $\theta_i \in \Theta_i$, $\theta_i^{'} \in \Theta_i$, and $q_i \in Q_i[\theta_i]$,*

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

AKS identify Δ-incentive compatibility as a necessary condition for robust virtual implementability:

Proposition 1 (AKS) *If an SCF is robustly virtually implementable, then it satisfies Δ-incentive compatibility.*

The next proposition formalizes the usual argument in the implementation literature under complete information: when there are at least three agents, incentive compatibility becomes a vacuous constraint.

Proposition 2 (A&M (1992b)) *Let $\mathcal{E}^*$ be a complete information environment where there are at least three agents. Then, for any SCF f, there exists an SCF $\hat{f} \approx f$ such that $\hat{f}$ satisfies Δ-incentive compatibility.*

Remark: This result does not depend upon quasi-transferability.

Proof: The reader is referred to Section 6 of A&M (1992b).■

3.2 Measurability

In an important paper, A&M (1992b) uncover a condition that they call *measurability* (I shall refer to it as A&M measurability) that is necessary for virtual

[10]See Section 3.1 of Abreu and Sen (1991) for the detail.
[11]See Jackson (1991) for the argument on equivalent SCFs.

implementation in iteratively undominated strategies over a standard environment that fixes a Bayesian type space.[12] AKS adapt A&M measurability to their robust virtual implementation and show that a version of measurability, called Δ-*measurability*, is also a necessary condition for "robust" virtual implementation. This section exploits the implication of Δ-measurability condition imposed on a complete information environment $\mathcal{E}^*$.

Denote by Ψ_i a *partition* of the set of payoff types Θ_i, where ψ_i is a generic element of Ψ_i and $\Psi_i(\theta_i)$ denotes the element of Ψ_i that includes payoff type θ_i. Let $\Psi = \times_{i\in N}\Psi_i$ and $\psi = \times_{i\in N}\psi_i$.

Definition 6 *An SCF f is* ***measurable with respect to*** Ψ *if, for every $i \in N$ and every $\theta_i, \theta_i^{'} \in \Theta_i$ with $\theta_i \neq \theta_i'$, whenever $\Psi_i(\theta_i) = \Psi_i(\theta_i')$,*

$$f(\theta_i, \theta_{-i}) = f(\theta_i', \theta_{-i}) \ \ \forall \theta_{-i} \in \Theta_{-i}.$$

Measurability of f with respect to Ψ implies that for any agent i, f does not distinguish between any pair of payoff types that lie in the same cell of the partition Ψ_i.

I can now provide the definition of *equivalent* payoff types. Note that, since agent $i \in N$ distinguishes all his payoff types, I consider a partition $\Theta_i \times \Psi_{-i} \equiv \{\{\theta_i\}_{\theta_i\in\Theta_i}\} \times \Psi_{-i}$ in that definition.

Definition 7 *For every $i \in N$, $\theta_i, \theta_i^{'} \in \Theta_i$ with $\theta_i \neq \theta_i'$, and $(n-1)$ tuple of partitions Ψ_{-i}, we say that θ_i is* ***equivalent*** *to θ_i' (denoted by $\theta_i \sim \theta_i^{'}$) with respect to Ψ_{-i} if, for any pair of SCFs f and $\tilde{f}$ which are measurable with respect to $\Theta_i \times \Psi_{-i}$,*

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

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

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

Note that $R_i(\Psi_{-i})$ forms an equivalence class on Θ_i, that is, it constitutes a partition of Θ_i. I 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 $h \geq 1$,

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

[12] Iteratively undominated strategies is the set of strategies that survive the iterated deletion of strictly dominated strategies. As long as finite mechanisms are considered, iteratively undominated strategies are the same as rationalizable strategies.

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

$$\Psi^* \equiv \bigvee_{h=0}^{\infty} \Psi^h,$$

where $\bigvee$ denotes the join on $\{\Psi^h\}_{h=0}^{\infty}$ associated with $\geq$. 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 $h \geq L$. Now, I am ready to define Δ-measurability.

Definition 8 *An SCF f satisfies **Δ-measurability** if it is measurable with respect to Ψ^*.*[13]

AKS show that Δ-measurability is a necessary condition for robust virtual implementation:

Theorem 1 (AKS) *If an SCF f is robustly virtually implementable, then it satisfies Δ-measurability.*

The next result shows that in complete information environments, Δ-measurability is a vacuous constraint.

Proposition 3 (A&M (1992b)) *Consider a complete information environment $\mathcal{E}^*$. For any SCF f, there exists an SCF $\hat{f} \approx f$ such that $\hat{f}$ satisfies Δ-measurability.*

Remark: This result does not depend upon quasi-transferability.

Proof: We can easily adapt here the proof of A&M (1992b, Section 6) and therefore omit the proof.■

4 A Characterization under Complete Information

AKS provide a characterization of robust virtual implementation by means of Δ-incentive compatibility and Δ-measurability.

Theorem 2 (AKS) *An SCF f is **robustly virtually implementable** if and only if it satisfies Δ-incentive compatibility and Δ-measurability.*

Recall that I denote by $\mathcal{E}^*$ a complete information environment satisfying quasi-transferability. In what follows, I only focus on the first and second iteration of Δ-measurability algorithm. Recall that Ψ_i^h denotes the partition over Θ_i that is derived from the h-th iteration of Δ-measurability algorithm and that $\Psi_i^h(\theta_i)$ denotes an element of Ψ_i^h that includes payoff type θ_i.

[13]See AKS for an intuitive exposition of Δ-measurability.

I already show by Proposition 3 that Δ-measurability is a vacuous constraint in complete information environments. In addition, the next lemma shows that one can construct an SCF x, which is measurable with respect to Ψ^2 and fully separates all payoff types.

Lemma 1 *Consider a complete information environment $\mathcal{E}^*$. Then, there exists an SCF x that is measurable with respect to Ψ^2 with the following two properties: (1) for each $i \in N$, $\theta_i \in \Theta_i$ and $\psi_i^2 \in \Psi_i^2 \backslash \Psi_i^2(\theta_i)$,*

$$V_i(x; \Psi_i^2(\theta_i)|\theta_i) > V_i(x; \psi_i^2|\theta_i);$$

and (2) for every $\theta, \theta' \in \Theta$ with $\theta \neq \theta'$, there exists $i \in N$,

$$V_i\left(x; \Psi_i^2(\theta_i)|\theta_i\right) > V_i(x; \Psi_i^2(\theta_i')|\theta_i).$$

Proof: For each $i \in N$ and $h = 1, 2$, let

$$F(\Psi_i^h \times \Psi_{-i}^{h-1}) = \left\{ f : \Theta \to A \middle| \ f \text{ is measurable with respect to } \Psi_i^h \times \Psi_{-i}^{h-1} \right\}$$

be the set of "deterministic" SCFs that are measurable with respect to $\Psi_i^h \times \Psi_{-i}^{h-1}$. Recall that A and Θ are finite. So, the functional space $F(\Psi_i^h \times \Psi_{-i}^{h-1})$ is finite as well. Let $\bar{x}_i^h : \Theta \to \Delta(A)$ be a social choice function that assigns equal probability to each element of $F(\Psi_i^h \times \Psi_{-i}^{h-1})$. That is, for each $\theta \in \Theta$,

$$\bar{x}_i^h(\theta) = \frac{1}{K^h} f^1(\theta) + \cdots + \frac{1}{K_h} f^{K_h}(\theta),$$

where K_h denotes the cardinality of $F(\Psi_i^h \times \Psi_{-i}^{h-1})$. By construction, $\bar{x}_i^h$ is measurable with respect to $\Psi_i^h \times \Psi_{-i}^{h-1}$ and, abusing notation, we can write $\bar{x}_i^h(\theta) = \bar{x}_i^h(\Psi^h(\theta))$. Following Lemma D.1 of AKS (2013b), we obtain the following: for every $i \in N$, there exists a collection of SCFs $\{x_i^h[\psi_i^h]\}_{\psi_i^h \in \Psi_i^h}$ that are measurable with respect to $\Psi_i^h \times \Psi_{-i}^{h-1}$, close to $\bar{x}_i^h$, such that for every $\theta_i, \in \Theta_i$ and $\psi_i^h \in \Psi_i^h \backslash \Psi_i^h(\theta_i)$,

$$V_i(x_i^h[\Psi_i^h(\theta_i)]|\theta_i) > V_i(x_i^h[\psi_i^h]|\theta_i).$$

Define an SCF x as follows: for each $\psi^2 \in \Psi^2$ with $\psi^2 \subseteq \psi^1 \in \Psi^1$,

$$x(\psi^2) = \frac{1-\delta}{n} \sum_{i \in N} x_i^1[\psi_i^1] + \frac{\delta}{n} \sum_{i \in N} x_i^2[\psi_i^2](\psi_i^2, \psi_{-i}^1),$$

where $\delta > 0$ is determined later. By construction, the SCF x is measurable with respect to Ψ^2. It only remains to show that the SCF x satisfies the two properties in the statement of the lemma. For every SCF $y : \Theta \to \Delta(A)$, define

$$G(y) = \max_{i \in N} \max_{\theta, \theta' \in \Theta} \left| u_i(y(\theta'); \theta) - u_i(y(\theta_i, \theta'_{-i}); \theta) \right|.$$

Choose $\delta > 0$ small enough such that

$$(1-\delta) \min_{i \in N, \theta_i \in \Theta_i, \psi_i^h \neq \Psi_i^h(\theta_i)} V_i(x_i^1[\Psi_i^h(\theta_i)]|\theta_i) - V_i(x_i^1[\psi_i^1]|\theta_i) > \delta \sum_{j \in N} G(x_j^2).$$

Fix $i \in N$, $\theta_i \in \Theta_i$, and $\psi_i^2 \in \Psi_i^2 \backslash \Psi_i^2(\theta_i)$. There are two cases we have to consider. First, assume $\psi_i^1 \notin \Psi_i^1(\theta_i)$ where $\psi_i^2 \subseteq \psi_i^1$. Then, due to the choice of δ, we must have

$$V_i(x_i^1[\Psi_i^1(\theta_i)]|\theta_i) > V_i(x_i^1[\psi_i^1]|\theta_i) \Rightarrow V_i(x; \Psi_i^1(\theta_i)|\theta_i) > V_i(x; \psi_i^1|\theta_i).$$

Assume that $\psi_i^1 \in \Psi_i^1(\theta_i)$ but $\psi_i^2 \notin \Psi_i^2(\theta_i)$. Then, due to the choice of δ, we have

$$\begin{aligned} & V_i(x_i^1[\Psi_i^1(\theta_i)]|\theta_i) = V_i(x_i^1[\psi_i^1]|\theta_i) \text{ and } V_i(x_i^2[\Psi_i^2(\theta_i)]|\theta_i) > V_i(x_i^2[\psi_i^2]|\theta_i) \\ \Rightarrow \quad & V_i(x; \Psi_i^2(\theta_i)|\theta_i) > V_i(x; \psi_i^2|\theta_i). \end{aligned}$$

Hence, inequality (1) holds. It is easy to see that the non-redundancy condition guarantees that inequality (2) holds.■

I am now ready to state and prove the main result of this section:

Theorem 3 (Characterization under Complete Information) *Consider a complete information environment $\mathcal{E}^*$. An SCF f satisfies Δ-incentive compatibility if and only if there exists an SCF $\hat{f} \approx f$ such that $\hat{f}$ is robustly virtually implementable.*

Remark: The proof is essentially an adaptation of the proof of Theorem 4 of AKS to the complete information environments. The most important fact of which this paper makes use is that the implementing mechanism is finite.

Proof: By Theorem 1, we know that Δ-incentive compatibility is necessary for robust virtual implementation. So, we focus on the sufficiency part. Suppose f satisfies Δ-incentive compatibility. By the same argument of Proposition 3, we can construct an SCF $\hat{f} \approx f$ such that $\hat{f}$ both satisfies Δ-incentive compatibility and Δ-measurability. The rest of the proof is essentially a straightforward modification of the proof of Theorem 4 of AKS (2013a,b). The only modification we have to make there is that (1) each agent i's message space only consists of J repetition of i's "payoff" type space, i.e., $M_i = (\Theta_i)^J$, while the message space of AKS also contain the announcement of first-order beliefs; and (2) we need to replace Lemma D.1 (AKS (2013b)) with Lemma 1 of the current paper to use it as the separation term of the outcome function.■

Finally, I conclude this section with a permissive implementation result.

Corollary 1 *Let $\mathcal{E}^*$ be a complete information environment where there are at least three agents. Then, for **any** SCF f, there exists an SCF $\hat{f} \approx f$ such that $\hat{f}$ is robustly virtually implementable.*

Remark: When there are only two agents, Δ-incentive compatibility can be shown to be equivalent to the *intersection property* (Abreu and Sen (1991)). I refer the reader to A&M (1992b, Section 6) for this equivalence. This condition says that whenever two agents disagree about the state, there is a lottery that is (weakly) worse than the social choice for both agents, whichever the true state is.[14] Unfortunately, quasi-transferability does not help us weaken the requirement of Δ-incentive compatibility. Hence, robust virtual implementation is much more restrictive for the case of two agents.

Proof: This directly follows from Theorem 3 and Proposition 2.■

This corollary significantly extends the main result of A&M (1992a). As I explain in the fourth paragraph of the introduction, A&M (1992a) fix a finite type space. Indeed, this corollary shows that A&M's main result holds irrespective of how one specifies a type space. Moreover, it is a refinement of the main result of Abreu and Sen (1991), provided that one assumes the expected utility hypothesis and quasi-transferability. See Section 6.2 for more details.

5 A Characterization under Almost Complete Information

Until the previous section, I have fixed a complete information environment $\mathcal{E}^*$ throughout. Since the set of complete information environments is "non-generic" relative to the set of all incomplete information environments, the permissive nature of robust virtual implementation might crucially rely on the complete information assumption. Thus, I ask the following robustness question: are the results robust to the relaxation of the complete information assumption? This question cannot be answered by AKS because AKS have to fix the set of first-order beliefs $Q_i[\theta_i]$ throughout, while the current paper wants to perturb this very set. In answering this question, I am motivated by (but do not exactly follow) the robust equilibrium analysis, which is first proposed by Kajii and Morris (1997) and later generalized by Oyama and Tercieux (2010) who extend its analysis to the non-common priors environments.

5.1 Type Space

In the rest of the paper, if X is a topological space, I treat it as a measurable space with its Borel σ-algebra and I denote by $\Delta(X)$ the space of Borel probability measures on X. If X is countable, it is treated as a topological space endowed with the discrete topology. Moreover, if X and Y are Borel

[14]Formally, an SCF f satisfies the intersection property if, for any $\theta, \theta' \in \Theta^*$, there exists a lottery $z(\theta,\theta') \in \Delta(A)$ such that $u_1(f(\theta');\theta') \geq u_1(z(\theta,\theta');\theta')$ and $u_2(f(\theta);\theta) \geq u_2(z(\theta,\theta');\theta)$.

measurable spaces, a product space $X \times Y$ is equipped with the associated product topology and product Borel σ-algebra.

When the robustness with respect to nearby incomplete information is the issue, I need a framework that allows for the perturbations of the complete information environments. I propose a *type space* as such a framework that accommodates nearby incomplete information environments:

Definition 9 *A **type space** is a tuple $\mathcal{T} \equiv (\mathcal{T}_i, \hat{\theta}_i, \hat{q}_i, \pi_i)_{i \in N}$ where for each player $i \in N$,*

1. *$\mathcal{T}_i$ is a countable space;*
2. *$\hat{\theta}_i : \mathcal{T}_i \to \Theta_i$ is an onto (or surjective) mapping[15];*
3. *$\hat{q}_i : \mathcal{T}_i \to \Delta(\Theta_{-i})$; and*
4. *$\pi_i : \mathcal{T}_i \to \Delta(\mathcal{T}_{-i})$.*

I denote a type of agent i by τ_i and the agent i's countable set of types by $\mathcal{T}_i$.[16] A *type* τ_i of agent i must include a description of his payoff type and first-order belief. Thus, $\hat{\theta}_i(\tau_i)$ is agent i's payoff type when his type is τ_i and $\hat{q}_i[\tau_i]$ is the first-order belief of agent i of type τ_i. $\pi_i[\tau_i]$ denotes the belief of agent i of type τ_i about the types of all other agents.

The next concept of coherence allows the complete information environment to be embedded into a type space.

Definition 10 *A type space $\mathcal{T} = (\mathcal{T}_i, \hat{\theta}_i, \hat{q}_i, \pi_i)_{i \in N}$ is **coherent** with an environment $\mathcal{E} = (A, (\Theta_i, (Q_i[\theta_i])_{\theta_i \in \Theta_i}, u_i)_{i \in N})$ if, for each $i \in N$ and $\tau_i \in \mathcal{T}_i$, the marginal of $\pi_i[\tau_i]$ on Θ_{-i} through $\hat{\theta}_{-i}$ is $\hat{q}_i[\tau_i]$ which is an element in $Q_i[\hat{\theta}_i(\tau_i)]$: for any $i \in N$, $\tau_i \in \mathcal{T}_i$, and $\theta_{-i} \in \Theta_{-i}$,*

$$\sum_{\tau_{-i} : \hat{\theta}_{-i}(\tau_{-i}) = \theta_{-i}} \pi_i[\tau_i](\tau_{-i}) = \hat{q}_i[\tau_i](\theta_{-i}).$$

Recall that $\mathcal{E}^*$ denotes a complete information environment satisfying quasi-transferability. A type space $\mathcal{T}^* = (\mathcal{T}_i^*, \hat{\theta}_i^*, \hat{q}_i^*, \pi_i^*)_{i \in N}$ is said to be a *complete information structure* if it is coherent with $\mathcal{E}^*$. Fix a mechanism $\Gamma = (M, g)$ throughout this section unless mentioned otherwise. Moreover, let $\Gamma(\mathcal{T}^*)$ denote a *complete information game* associated with $\mathcal{T}^*$.

5.2 Rationalizability under a Type Space $\mathcal{T}$

In a type space $\mathcal{T} = (\mathcal{T}_i, \hat{\theta}_i, \hat{q}_i, \pi_i)_{i \in N}$ and a mechanism $\Gamma = (M, g)$, I define agent i's message correspondence $S^{\mathcal{T}} = (S_1^{\mathcal{T}}, \ldots, S_n^{\mathcal{T}})$ where for each $i \in N$,

$$S_i^{\mathcal{T}} : \mathcal{T}_i \to 2^{M_i}.$$

[15] This guarantees that the type space is at least as large as the payoff type space.

[16] I could extend the analysis to more general spaces, but at the cost of additional technical complexity.

and I write $\mathcal{S}^{\mathcal{T}}$ for the collection of message correspondence profiles of the game $\Gamma(\mathcal{T})$ with the natural set-inclusion order.

I define an operator $b = (b_1, \ldots, b_n)$ to iteratively eliminate never best responses. The operator $b : \mathcal{S}^{\mathcal{T}} \to \mathcal{S}^{\mathcal{T}}$ is now defined as follows: for any $i \in N$, $\tau_i \in \mathcal{T}_i$,

$$b_i(S^{\mathcal{T}})[\tau_i] \equiv \left\{ m_i \left| \begin{array}{c} \exists \mu_i \in \Delta(\mathcal{T}_{-i} \times M_{-i}) \text{ s.t.} \\ \mu_i(\theta_{-i}, m_{-i}) > 0 \Rightarrow m_j \in S_j^{\mathcal{T}}(\tau_j) \ \forall j \neq i; \text{ and} \\ m_i \in \arg\max_{m_i' \in M_i} \sum_{\tau_{-i}, m_{-i}} \mu_i(\tau_{-i}, m_{-i}) u_i(g(m_i', m_{-i}); \hat{\theta}(\tau_i, \tau_{-i})) \end{array} \right. \right\}$$

Since the message space is finite, I have that for each $i \in N$ and $\tau_i \in \mathcal{T}_i$,

$$S_i^{\Gamma(\mathcal{T})}(\tau_i) \equiv \bigcap_{k \geq 1} b_i(b^k(\bar{S}))[\tau_i] \neq \emptyset.$$

Thus $S_i^{\Gamma(\mathcal{T})}(\tau_i)$ are the set of messages surviving iterated deletion of never best responses; equivalently, $S_i^{\Gamma(\mathcal{T})}(\tau_i)$ is the set of messages that player i with type τ_i might send consistent with common certainty of rationality.

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:

$$U_i(f; \tau_i' | \tau_i) \equiv \sum_{\theta_{-i}} \hat{q}_i[\tau_i](\theta_{-i}) u_i(f(\hat{\theta}_i(\tau_i'), \theta_{-i}); \hat{\theta}_i(\tau_i), \theta_{-i})$$

Denote $U_i(f|\tau_i) = U_i(f; \tau_i|\tau_i)$. For $\theta_i = \hat{\theta}_i(\tau_i)$, $\theta_i' = \hat{\theta}_i(\tau_i')$, and $q_i = \hat{q}_i[\tau_i]$, I have

$$V_i(f; \theta_i' | \theta_i, q_i) = U_i(f; \tau_i' | \tau_i).$$

Therefore, I obtain the following equivalence between Δ-rationalizability under complete information and rationlaizability under a type space $\mathcal{T}^* = (\mathcal{T}_i^*, \hat{\theta}_i^*, \hat{q}_i^*, \pi_i)_{i \in N}$: for any $\tau \in \mathcal{T}^*$,

$$S^{\Gamma}(\hat{\theta}^*(\tau)) = S^{\Gamma(\mathcal{T}^*)}(\tau).$$

Hence, as far as one is concerned with rationalizability under complete information, each agent's behavior does not depend upon how the underlying type space $\mathcal{T}^*$ is described. This is exactly the point exemplified by Corollary 1.

5.3 The Robustness to Almost Complete Information

Fix a type space $\mathcal{T} = (\mathcal{T}_i, \hat{\theta}_i, \hat{q}_i, \pi_i)_{i \in N}$. For each measurable $E_{-i} \subseteq \mathcal{T}_{-i}$, define the *p-belief operator* for agent i, $B_i^p : 2^{\mathcal{T}_{-i}} \to 2^{\mathcal{T}_i}$, is defined by

$$B_i^p(E_{-i}) = \left\{ \tau_i \in \mathcal{T}_i \left| \ \pi_i[\tau_i](E_{-i}) \geq p \right. \right\},$$

where $\pi_i[\tau_i](E_{-i}) \equiv \sum_{\tau_{-i} \in E_{-i}} \pi_i[\tau_i](\tau_{-i})$. That is, $B_i^p(E_{-i})$ is the set of types of agent i where he believes with probability at least p that event E_{-i} is true. Let $B_*^p(E) = E \cap \times_{i \in N} B_i^p(E_{-i}) = E \cap \{\tau \in \mathcal{T} | \ \pi_i[\tau_i](E_{-i}) \geq p \ \forall i \in N\}$ be the set of states where event E is indeed true and every agent believes with probability at least p that event E is true. At a state τ, an event E is *mutual p-believed* at order L if $\tau \in \bigcap_{\ell=1}^{L} [B_*^p]^{\ell}(E)$, where $[B_*^p]^{\ell}(\cdot)$ is defined recursively by $[B_*^p]^{\ell}(E) = B_*^p([B_*^p]^{\ell-1}(E))$ for every $\ell \geq 1$. Finally, at state τ, an event E is *common p-believed* if $\tau \in \bigcap_{\ell=0}^{\infty} [B_*^p]^{\ell}(E)$.

In what follows, for two type spaces $\mathcal{T} = (\mathcal{T}_i, \hat{\theta}_i, \hat{q}_i, \pi_i)_{i \in N}$ and $\mathcal{T}' = (\mathcal{T}_i', \hat{\theta}_i', \hat{q}_i', \pi_i')_{i \in N}$, I say $\mathcal{T} \supseteq \mathcal{T}'$ if, σ-algebra on each $\mathcal{T}_i'$ is the relative σ-algebra obtained from each $\mathcal{T}_i$, and for all $i \in N$ and $\tau_i \in \mathcal{T}_i'$:

$$\pi_i[\tau_i](E_{-i}) = \pi_i'[\tau_i](\mathcal{T}_{-i}' \cap E_{-i}) \text{ for any measurable } E_{-i} \subseteq \mathcal{T}_{-i}.$$

Now, I introduce a class of perturbations of the complete information structure used for this paper:

Definition 11 *A type space $\mathcal{T}^*[\eta, L]$ is an (η, L)-**perturbation** of $\mathcal{T}^*$ if $\mathcal{T}^*[\eta, L] \supseteq \mathcal{T}^*$ and for all $i \in N$ and all $\tau_i \in \mathcal{T}_i^*$,*

$$\pi_i[\tau_i] (E_{-i}) \geq 1 - \eta,$$

*where $E \equiv \bigcap_{\ell=1}^{L} [B_*^{1-\eta}]^{\ell} (\mathcal{T}^*) \subseteq \mathcal{T}^*[\eta, L]$, which denotes the set of all states in which $\mathcal{T}^*$ is mutually $(1-\eta)$-believed at order L. Such a perturbation is denoted $\mathcal{T}^*[\eta, L] \supseteq \mathcal{T}^*$ and $\Gamma(\mathcal{T}^*[\eta, L])$ is called a (nearby) **incomplete information game**.*

Remark: This perturbation is different from those considered by Kajii and Morris (1997) and Oyama and Tercieux (2010) because they take the approximation of outcome distributions from the *ex ante* point of view and include "crazy types," each of which exhibits distinct preferences in the perturbations, hence much richer perturbations. In this paper, I perturb only the set of first-order beliefs $Q_i^*[\theta_i]$ within the richer embedding space and take the approximation of outcome distributions from the *ex post* point of view. Indeed, this follows and generalizes the robustness approach by Chung and Ely (2003) and Kunimoto (2010) who discuss the robustness of undominated Nash implementation and by AFHKT (2012) who discuss the robustness of subgame perfect implementation. However, this paper's robustness analysis is much more demanding than these papers because these authors have to fix a *finite* type space all the time, while this paper allows for all countable (both finite and countably infinite) type spaces. This perturbation is also different from that of Oury and Tercieux (2012). They instead consider the perturbation in terms of the product topology of weak convergence of infinite hierarchies of probabilistic beliefs (as I describe in the second paragraph of the Introduction) in the universal type space.[17] This is a much richer perturbation than that

[17] See Oury and Tercieux (2012) for the construction of the universal type space.

of this paper. However, if one is only concerned with partial implementation, this richer perturbation suffices. The reason why I need a smaller class of perturbations is that I am concerned with *full* implementation that is robust to almost complete information.

I introduce the definition of robust virtual implementation under almost complete information:

Definition 12 *An SCF f is robustly virtually implementable under* ***almost complete information*** *if, for any $\varepsilon > 0$, there exist $\bar{\eta} > 0$, $\underline{L} \geq 1$, and a mechanism $\Gamma = (M, g)$ such that for each $\eta \in (0, \bar{\eta})$, $L \geq \underline{L}$, $\mathcal{T}^*[\eta, L] \supseteq \mathcal{T}^*$, $\tau \in \mathcal{T}^*$, and $m \in M$,*

$$S^{\Gamma(\mathcal{T}^*[\eta,L])}(\tau) \neq \emptyset \text{ and } m \in S^{\Gamma(\mathcal{T}^*[\eta,L])}(\tau) \Rightarrow \|g(m) - f(\hat{\theta}^*(\tau))\| \leq \varepsilon.$$

Remark: I can also define robust implementation under almost complete information for other solution concepts. I will elaborate on this in Section 6.1.

Theorem 4 (Characterization under "Almost" Complete Information) *An SCF f satisfies Δ-incentive compatibility if and only if there exists an SCF $\hat{f} \approx f$ such that $\hat{f}$ is robustly virtually implementable under almost complete information.*

Proof: We have already shown by Theorem 1 that Δ-incentive compatibility is a necessary condition for robust virtual implementation under complete information. This must continue to be the case under almost complete information. Thus, we focus on the sufficiency part. Since f satisfies Δ-incentive compatibility, by Theorem 3, there exists an SCF $\hat{f} \approx f$ such that $\hat{f}$ is robustly virtually implementable under complete information. Moreover, this is so by a "finite" mechanism.

Fix an arbitrarily small $\varepsilon > 0$. By the previous argument, we can fix a finite implementing mechanism $\Gamma = (M, g)$ with the property that for any $\theta \in \Theta^*$ and any $m \in M$,

$$S^{\Gamma}(\theta) \neq \emptyset \text{ and } m \in S^{\Gamma}(\theta) \Rightarrow \|g(m) - \hat{f}(\theta)\| \leq \varepsilon/2.$$

This also implies that for any complete information structure $\mathcal{T}^* = (\mathcal{T}_i^*, \hat{\theta}_i^*, \hat{q}_i^*, \pi_i^*)_{i \in N}$, $\tau \in \mathcal{T}^*$, and $m \in M$,

$$S^{\Gamma(\mathcal{T}^*)}(\tau) \neq \emptyset \text{ and } m \in S^{\Gamma(\mathcal{T}^*)}(\tau) \Rightarrow \|g(m) - \hat{f}(\hat{\theta}^*(\tau))\| \leq \varepsilon/2.$$

Recall that we write $\mathcal{S}$ for the collection of message correspondence profiles and it is a lattice with the natural ordering of set inclusion. For each $\eta > 0$, we define on operator $b^\eta : \mathcal{S} \to \mathcal{S}$ as follows: for each $i \in N$ and each $\theta_i \in \Theta_i$,

$$b_i^\eta(S)[\theta_i] \equiv \left\{ m_i \left| \begin{array}{c} \exists q_i \in Q_i^*[\theta_i] \; \exists \mu_i \in \Delta^{q_i}(\Theta_{-i} \times M_{-i}) \text{ s.t.} \\ \mu_i(\theta_{-i}, m_{-i}) > 0 \Rightarrow m_j \in S_j(\theta_j) \; \forall j \neq i; \text{ and} \\ m_i \in \arg\max_{m_i'} \sum_{\theta_{-i}, m_{-i}} \mu_i(\theta_{-i}, m_{-i}) u_i(g(m_i', m_{-i}); \theta_i, \theta_{-i}) - \eta \bar{u} \end{array} \right. \right\},$$

where $\bar{u} \equiv \max_{i \in N} \max_{\theta \in \Theta} \max_{m \in M} u_i(g(m); \theta)$. Since the mechanism Γ is finite, the iterated deletion of never best responses stops at a finite number. Call such a number $\underline{L}$. Moreover, by the finiteness of Γ and Θ and the continuity of expected utility, for each $\ell \in \{1, \ldots, \underline{L}\}$, one can choose $\eta_\ell > 0$ small enough so that for each $i \in N$ and $\theta_i \in \Theta_i$,

$$\bigcap_{k=1}^{\ell} b_i(b^k(\bar{S}))[\theta_i] = \bigcap_{k=1}^{\ell} b_i^{\eta_{\ell+1}}(b^{k,\eta_k})(\bar{S})[\theta_i].$$

This means that, due to the finiteness of Γ and Θ, never best responses under complete information continue to be so even under a small payoff perturbation.

Define $\eta' = \min\{\eta_1, \ldots, \eta_{\underline{L}}\}$. Fix any type space $\mathcal{T}^*[\eta', \underline{L}]$, which is an $(\eta', \underline{L})$-perturbation of $\mathcal{T}^*$. Then, for any $\tau \in \mathcal{T}^*$, we have

$$\tau \in \bigcap_{\ell=1}^{\underline{L}} [B_*^{1-\eta'}]^\ell(\mathcal{T}^*) \Rightarrow S^{\Gamma(\mathcal{T}^*[\eta', \underline{L}])}(\tau) = S^{\Gamma(\mathcal{T}^*)}(\tau) = S^{\Gamma}(\hat{\theta}^*(\tau)).$$

where $\bigcap_{\ell=1}^{\underline{L}} [B_*^{1-\eta'}]^\ell(\mathcal{T}^*) \subseteq \mathcal{T}^*[\eta', \underline{L}]$. Since $\mathcal{T}^*[\eta', \underline{L}]$ is an $(\eta', \underline{L})$-perturbation of $\mathcal{T}^*$, we have that for each $i \in N$ and $\tau_i \in \mathcal{T}_i^*$,

$$\pi_i^{\eta', \underline{L}}[\tau_i](E_{-i}) \geq 1 - \eta',$$

where $E \equiv \bigcap_{\ell=1}^{\underline{L}} [B_*^{1-\eta'}]^\ell(\mathcal{T}^*) \subseteq \mathcal{T}^*[\eta', \underline{L}] = (\mathcal{T}_i^{\eta', \underline{L}}, \hat{\theta}_i^{\eta', \underline{L}}, \hat{q}_i^{\eta', \underline{L}}, \pi_i^{\eta', \underline{L}})_{i \in N}$. By the finiteness of Γ and the continuity of expected utility, we can choose $\eta'' \in (0, \eta']$ small enough so that for each $\tau \in \mathcal{T}^*$ and $m \in M$,

$$m \in S^{\Gamma(\mathcal{T}^*)}(\tau) \Rightarrow m \in S^{\Gamma(\mathcal{T}^*[\eta'', \underline{L}])}(\tau).$$

This implies that for each $\tau \in \mathcal{T}^*$ and $\mathcal{T}^*[\eta'', \underline{L}]$,

$$S^{\Gamma(\mathcal{T}^*)}(\tau) \neq \emptyset \Rightarrow S^{\Gamma(\mathcal{T}^*[\eta'', \underline{L}])}(\tau) \neq \emptyset.$$

Moreover, by the continuity of expected utility, one can easily see that for any $\eta \in (0, \eta'']$ and $L \geq \underline{L}$,

$$S^{\Gamma(\mathcal{T}^*[\eta'', \underline{L}])}(\tau) \neq \emptyset \Rightarrow S^{\Gamma(\mathcal{T}^*[\eta, L])}(\tau) \neq \emptyset.$$

In sum, we have that for any $\eta \in (0, \eta'']$, $L \geq \underline{L}$, $\mathcal{T}^*[\eta, L]$, and $\tau \in \mathcal{T}^*$,

$$S^{\Gamma(\mathcal{T}^*[\eta, L])}(\tau) \neq \emptyset.$$

So, the message correspondence is guaranteed to be nonempty for any nearby type space.

At any state $\tau \in \mathcal{T}^*$, all agents believe with probability at least $(1 - \eta'')$ that event E is true, i.e., $S^{\Gamma(\mathcal{T}^*[\eta'', \underline{L}])}(\tau) = S^{\Gamma(\mathcal{T}^*)}(\tau)$. This implies that the probability of event E being true is at least $(1 - \eta'')$. Therefore, we can choose

$\bar{\eta} = \min\{\varepsilon/2, \eta''\}$ such that for any $\tau \in \mathcal{T}^*$, $\mathcal{T}^*[\bar{\eta}, \underline{L}] \supseteq \mathcal{T}^*$, $m \in S^{\Gamma(\mathcal{T}^*[\bar{\eta},\underline{L}])}(\tau)$, and $m^* \in S^{\Gamma(\mathcal{T}^*)}(\tau)$,

$$\|g(m) - g(m^*)\| \leq \bar{\eta}\bar{a} = \bar{\eta} \leq \varepsilon/2,$$

where $\bar{a} \equiv \sup_{\alpha,\alpha' \in \Delta(A)} \|\alpha - \alpha'\| = 1$.[18]

Furthermore, for any $\tau \in \mathcal{T}^*$, $\eta \in (0, \bar{\eta}]$, $L \geq \underline{L}$, $m \in S^{\Gamma(\mathcal{T}^*[\eta,L])}(\tau)$, and $m^* \in S^{\Gamma(\mathcal{T}^*)}(\tau)$, we have

$$\|g(m) - g(m^*)\| \leq \bar{\eta} \leq \varepsilon/2.$$

Then, for any $\eta \in (0, \bar{\eta})$, $L \geq \underline{L}$, $\tau \in \mathcal{T}^*$, $\mathcal{T}^*[\eta, L] \supseteq \mathcal{T}^*$, $m \in S^{\Gamma(\mathcal{T}^*[\eta,L])}(\tau)$, and $m^* \in S^{\Gamma(\mathcal{T}^*)}(\tau)$, we have

$$\begin{aligned}
\|g(m) - \hat{f}(\hat{\theta}^*(\tau))\| &= \|g(m) - g(m^*) + g(m^*) - \hat{f}(\hat{\theta}^*(\tau)\| \\
&\leq \|g(m) - g(m^*)\| + \|g(m^*) - \hat{f}(\hat{\theta}^*(\tau)\| \;\; (\because \text{ triangle inequality}) \\
&\leq \varepsilon/2 + \varepsilon/2 = \varepsilon.
\end{aligned}$$

Summarizing all the arguments thus far, we have that there exist $\bar{\eta} > 0$ and $\underline{L} \geq 1$ such that for each $\eta \in (0, \bar{\eta}]$, $L \geq \underline{L}$, $\mathcal{T}^*[\eta, L] \supseteq \mathcal{T}^*$, $\tau \in \mathcal{T}^*$, and $m \in M$,

$$S^{\Gamma(\mathcal{T}^*[\eta,L])}(\tau) \neq \emptyset \text{ and } m \in S^{\Gamma(\mathcal{T}^*[\eta,L])}(\tau) \Rightarrow \|g(m) - \hat{f}(\hat{\theta}^*(\tau))\| \leq \varepsilon.$$

This completes the proof.■

The following corollary is the main result of this paper:

Corollary 2 *Suppose that there are at least three agents. Then, for* ***any*** *SCF f, there exists an SCF $\hat{f} \approx f$ such that $\hat{f}$ is robustly virtually implementable under almost complete information.*

Proof: This directly follows from Theorem 4 and Proposition 2.■

6 Concluding Remarks

In conclusion, I discuss some related papers and connect these papers to the current paper.

[18] Recall that $\|\alpha - \alpha'\| \equiv \sum_{a \in A} |\alpha(a) - \alpha'(a)|$ and $\alpha(a)$ and $\alpha'(a)$ each denotes the probability that pure outcome a is realized. Note also that the finiteness of A is not essential here and moreover, if A is a complete separable space, all the arguments go through by focusing on a countable dense subset of it.

6.1 Chung and Ely (2003) and Aghion, Fudenberg, Holden, Kunimoto, and Tercieux (2012)

By requiring solution concepts to have a closed graph in the limit of complete information in a fixed finite type space, Chung and Ely (2003) and AFHKT (2012) show that only Maskin monotonic social choice rules can be robustly implementable under almost complete information. Here I provide the definition of Maskin monotonicity:

Definition 13 *An SCF f satisfies* ***Maskin monotonicity*** *if, for any of states $\theta, \theta' \in \Theta$, whenever*

$$\forall i \in N, \ \forall b \in \Delta(A): \ u_i(f(\theta);\theta) \geq u_i(b;\theta) \Rightarrow u_i(f(\theta);\theta') \geq u_i(b;\theta'),$$

then $f(\theta) = f(\theta')$.

These authors' robustness requirement is much weaker than that adopted in this paper because the following three restrictions are imposed on the class of perturbations: (i) $\underline{L} = \infty$; (ii) $\mathcal{T}^*[\eta,\infty] = \Theta$ for each $\eta > 0$; and (iii) $\mathcal{T}^* = \Theta^*$.[19] So, their robustness requires that one fix a finite type space and only perturb the prior beliefs over that fixed type space. On the other hand, this paper's result remain intact regardless of what (potentially infinite) type spaces are considered.

AFHKT (2012) show that when there are at least three agents, if an SCF f satisfies Maskin monotonicity and no-veto-power, there exists a mechanism $\Gamma = (M, g)$ such that (1) there exist a sequence of strategy profiles $\{\sigma^\eta\}_{\eta>0}$ and a sequence of beliefs $\{\mu^\eta\}_{\eta>0}$ with $(\sigma^\eta, \mu^\eta) \in S^{\Gamma(\mathcal{T}^*[\eta,\infty])}$ such that $\lim_{\eta\to 0} g(\sigma^\eta(\tau)) = f(\hat{\theta}^*(\tau))$ for each $\tau \in \mathcal{T}^*$ and (2) for any $(\sigma^\eta, \mu^\eta) \in S^{\Gamma(\mathcal{T}^*[\eta,\infty])}$, we have $\lim_{\eta\to 0} g(\sigma^\eta(\tau)) = f(\hat{\theta}^*(\tau))$ for each $\tau \in \mathcal{T}^*$ where $S^{\Gamma(\mathcal{T}^*[\eta,\infty])}$ stands for either the undominated (Bayesian) Nash equilibrium correspondence used by Chung and Ely (2003) or the subgame perfect (or sequential) equilibrium correspondence used by AFHKT (2012).[20] Here, μ^η denotes the profile of all agents' beliefs over the type space and the set of histories of the game. When undominated Nash equilibrium is used, there is no need to include the agent's belief profile μ^η as part of the equilibrium. This is needed for sequential equilibrium.

At the same time, these authors also show that if a mechanism $\Gamma = (M, g)$ implements a non-Maskin monotonic SCF in either undominated Nash or subgame perfect equilibrium, there is a sequence of strategy profiles $\{\sigma^\eta\}_{\eta>0}$ and a sequence of belief profiles $\{\mu^\eta\}_{\eta>0}$ with $(\sigma^\eta, \mu^\eta) \in S^{\Gamma(\mathcal{T}^*[\eta,\infty])}$ such that $\lim_{\eta\to 0} g(\sigma^\eta(\tau)) \neq f(\hat{\theta}^*(\tau))$ for some $\tau \in \mathcal{T}^*$. This implies that Maskin monotonicity is a necessary condition for robust implementation under almost complete information, regardless of whether one uses either undominated Nash or subgame perfect equilibrium as a solution concept.

[19] Recall that $\Theta^* = \{\theta \in \Theta | \phi_1(\theta_1) = \cdots = \phi_n(\theta_n) \in \Theta_0\}$.

[20] To be exact, this result needs one more condition, the *no-worst-alternative* condition (NWA). See Online Appendix of AFHKT (2012) for the precise definitions of no-veto-power and NWA.

6.2 Abreu and Sen (1991)

Abreu and Sen (1991) show that when there are at least three agents, *any* social choice function is virtually implementable in Nash equilibrium. This paper shows that A&M's permissive virtual implementation results are robust to how the underlying type space is fixed, while Abreu and Sen's is certainly not. Nevertheless, this improvement comes at some cost. This paper's sufficiency results need quasi-transferability, while Abreu and Sen do not need. While I explicitly need the expected utility hypothesis, Abreu and Sen do not. They instead require that individual preferences over lotteries be *monotone* in the sense that any shift of probability weight from a less preferred to a more preferred pure alternative yields a lottery which is preferred. The monotone preferences are weaker than and implied by the expected utility representation.

6.3 Börgers (1995) and Kunimoto and Serrano (2011)

Quasi-transferability is the the single assumption on which this paper's results crucially rely. It is very natural to ask the extent to which the permissive results can be extended to more general environments. Börgers (1995) shows that only dictatorial SCFs are "exactly" implementable in rationalizable strategies, provided that all players could have all possible identical preferences over a finite set of outcomes and only finite, deterministic mechanisms are considered. This possibility, however, is explicitly avoided by quasi-transferability, which excludes any possibility that all agents have identical preferences. Therefore, one might attribute the permissive result of virtual implementation (at least partly) to this domain restriction. Kunimoto and Serrano (2011) indeed confirm this point. They uncover a new necessary condition that implies that quasi-transferability cannot be completely dispensed with for implementation in rationalizable (iteratively undominated) strategies. They term the condition "restricted deception-proofness." It requires that, in environments with identical preferences, the SCF be immune to all deceptions, making it then stronger than incentive compatibility. As I argue above, quasi-transferability excludes the case that all agents have identical preferences. However, even if one can exclude any possibility that all agents have identical preferences, there might be some other conditions needed for general environments in characterizing virtual implementation in rationalizable strategies. This is an interesting open question and I leave it for the future work.

6.4 Oury and Tercieux (2012)

Oury and Tercieux (2012) is related to the current paper. They propose *continuous implementation* where the planner would like that, in any perturbation of his initial model, there is some equilibrium that yields the desired outcome, not only at types in the initial model, but also types that are close to it. What they mean by "two types are close to each other" is that these two types are close in the product topology of weak convergence of infinite hierarchies of beliefs in the universal type space. They consider *partial* (as opposed to full) Bayesian (resp., Nash) implementation under incomplete (resp., complete) information.[21] Proposition 2 of their paper is one of their main results and is the most relevant here: An SCF is continuously "partially" virtually implementable by a finite mechanism if and only if it is fully virtually implementable in rationalizable strategies by a finite mechanism. Hence, in quasi-transferable environments where there are at least three agents, *any* SCF is robustly virtually implementable under almost complete information and at the same time, *any* SCF is virtually continuously implementable under complete information by a finite mechanism.

References

[1] Abreu, D. and H. Matsushima, "Virtual Implementation in Iteratively Undominated Strategies: Complete Information," *Econometrica* vol. 60, (1992a), 993-1008.

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

[3] Abreu, D. and A. Sen, "Virtual Implementation in Nash Equilibrium," *Econometrica* vol. 59, (1991), 997-1021.

[4] Aghion, P., D. Fudenberg, R. Holden, T. Kunimoto, and O. Tercieux, "Subgame-Perfect Implementation under Information Perturbations," *Quarterly Journal of Economics*, vol. 127, (2012), 1843-1881.

[5] Artemov, G., T. Kunimoto, and R. Serrano, "Robust Virtual Implementation: Toward a Reinterpretation of the Wilson Doctrine," *Journal of Economic Theory*, vol. 148, (2013a), 424-447.

[6] Artemov, G., T. Kunimoto, and R. Serrano, "Online Appendix to "Robust Virtual Implementation: Toward a Reinterpretation of the Wilson Doctrine"," (2013b), available at `https://sites.google.com/site/tkunimoto73/home/research`.

[7] Battigalli, P. and M. Siniscalchi, "Rationalization and Incomplete Information," *Advances in Theoretical Economics* 3(1), (2003), Article 3.

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

[9] Chung, K-S. and J. Ely, "Implementation with Near-Complete Information," *Econometrica* vol. 71, (2003), 857-87.

[10] Börgers, T, "A Note on Implementation and Strong Dominance," W. Barnett et. al (eds.) "*Social Choice, Welfare, and Ethics*," (1995), Cambridge University Press.

[21]An SCF is partially implementable in (Bayesian) Nash equilibrium if there exists a mechanism such that there is a (Bayesian) Nash equilibrium whose outcome coincides with that of the SCF.

[11] Jackson, M. O., "Bayesian Implementation," *Econometrica* vol. 59, (1991), 461-477.

[12] Jackson, M. O., "A Crash Course in Implementation Theory," *Social Choice and Welfare*, vol. 18, (2001), 655-708.

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

[14] Kunimoto, T., "How Robust is Undominated Nash Implementation?," (2010), *mimeo*.

[15] Kunimoto, T. and R. Serrano, "A New Necessary Condition for Implementation in Iteratively Undominated Strategies," *Journal of Economic Theory* vol. 146, (2011), 2583-2595

[16] Maskin, E., "Nash Equilibrium and Welfare Optimality," *Review of Economic Studies*, vol. 66, (1999), 23-38.

[17] Mookherjee D. and S. Reichelstein, "The Revelation Approach to Nash Implementation," *Economics Letters*, 39, (1992), 309-313.

[18] Muller, E., and M. Satterthwaite, "The Equivalence of Strong Positive Association and Strategy-Proofness," *Journal of Economic Theory*, vol. 14, (1977), 412-418.

[19] Oury, M. and O. Tercieux, "Continuous Implementation," *Econometrica*, vol. 80, (2012), 1605-1637.

[20] Oyama, D. and O. Tercieux, "Robust Equilibria under Non-Common Priors," *Journal of Economic Theory*, vol. 145, (2010), 752-784.

[21] Serrano, R., "The Theory of Implementaion of Social Choice Rules," *SIAM Review*, vol. 46, (2004), 377-414.