Robust Virtual Implementation with Incomplete Information: Towards a Reinterpretation of the Wilson Doctrine*

Georgy Artemov†, Takashi Kunimoto‡ and Roberto Serrano§

First Version: May 2007
This Version: August 2011

Abstract

Placing any arbitrary restrictions on sets of first-order beliefs, we study robust virtual implementation (RVI) compatible with such restrictions. Using Δ-rationalizability as our solution concept, we show that first-order incentive compatibility (FOIC) and Abreu-Matsushima measurability (AMM) are necessary for RVI. In quasi-transferable environments: (i) under the generic condition of first-order type diversity, FOIC is sufficient for RVI; and (ii) FOIC and AMM characterize RVI. The connections with the case of unrestricted RVI are discussed at some length.

JEL Classification: C72, D78, D82.

Keywords: Wilson doctrine, mechanism design, robust virtual implementation, Δ-rationalizability, incentive compatibility, measurability, type diversity.

---

*This paper has benefited greatly from the comments of an Associate Editor and three anonymous referees and from conversations with Murali Agastya, Dirk Bergemann, Antonio Cabrales, Jeff Ely, Philippe Jehiel, Vijay Krishna, Hitoshi Matsushima, Stephen Morris, Tomas Sjöström, Rani Spiegler, Ronald Stauber, Takeshi Suzuki and Takehiko Yamato. Artemov gratefully acknowledges financial support from the University of Melbourne through ERC grant, Kunimoto from FQRSC and SSHRC of Canada, and Serrano from Spain’s Ministry of Science and Innovation under grant Consolider 2010 CSD2006-0016. Serrano also thanks CEMFI in Madrid for its hospitality.

†Department of Economics, the University of Melbourne, Australia; gartemov@unimelb.edu.au

‡Department of Economics, Hitotsubashi University, Kunitachi, Tokyo, Japan; takashikunimoto9@gmail.com

§Department of Economics, Brown University, Providence, RI, U.S.A. and IMDEA-Social Sciences Institute, Madrid, Spain; roberto_serrano@brown.edu
“Game Theory has a great advantage in explicitly analyzing the consequences of trading rules that presumably are really common knowledge; it is deficient to the extent it assumes other features to be common knowledge, such as one player’s probability assessment about another’s preferences or information. I foresee the progress of game theory as depending on successive reductions in the base of common knowledge required to conduct useful analysis of practical problems. Only by repeated weakening of common knowledge assumptions will the theory approximate reality.”


1 Introduction

The usual approach to mechanism design under incomplete information assumes that the underlying type spaces are common knowledge among the planner and the agents. When designing mechanisms the planner can then make use of such knowledge, something that is often criticized as unrealistic, as the opening quote from Wilson makes clear. Previous papers in the literature known as robust mechanism design counter this shortcoming by going to the polar opposite assumption, i.e., that the planner has no information about any belief-component of agents’ types – only the payoff type spaces are assumed to be common knowledge. In this paper we take an intermediate approach, by allowing the planner to have some additional information concerning agents’ beliefs, beyond the payoff type spaces. This extra information is added to the common knowledge structure and constitutes our reinterpretation of the Wilson doctrine. This provides a robustness check of some of those previous results. Although we do not model where this partial information comes from, we find it reasonable to imagine that the planner may have already acquired some partial knowledge that allows her to rule out some first-order beliefs, perhaps through past experience and/or analogies to similar problems. Of course, it would be an interesting question to explicitly model the planner’s endogenous information acquisition, which is not our aim here.

If one fixes the set of payoff-relevant types or payoff types – ex post preferences –, the usual interpretation of the Wilson doctrine asks that the conclusions of the theory be robust to arbitrary beliefs over the payoff types. Because agents perform the analysis at the interim stage, agents’ interim preferences are the key. Thus, we propose a suitable reinterpretation of the doctrine in which, along with payoff types, first-order beliefs over payoff types (but not complex higher-order beliefs over beliefs) are central in obtaining a clear understanding of the bounds of implementation. In this sense, arbitrary first-order beliefs are elevated to the level of a fundamental – like payoff types are. Details will follow soon below.

The robust implementation analysis has strengthened the limitations of the theory, by turning conditions such as incentive compatibility or Bayesian monotonicity into much stronger ones, like ex post incentive compatibility or robust monotonicity, respectively. Our reason to focus on virtual implementation (Matsushima (1988), Abreu and Sen (1991)) in this paper is that the measurability condition uncovered by Abreu and Matsushima
AM measurability from now on—central to virtual implementation under incomplete information, is usually perceived as being extremely permissive in a standard Bayesian environment. The crux of the current paper is to suggest a way to evaluate the exact strength of its extension to the robust setting. In doing so, we complement the analysis performed in Bergemann and Morris (henceforth, BM (2009)) in assessing the restrictiveness of their robust measurability condition. In environments in which the appropriate incentive compatibility condition is somewhat permissive, settling this question is crucial to evaluate the possibilities of (approximate) implementation.

Indeed, in a series of papers, BM (2005, 2008, 2009) seek for robust implementation results. Their work relevant to the current study is contained in their latter two papers, which deal with full implementation. Faithful to the Wilson doctrine, they construct mechanisms that rely exclusively on the use of payoff types, and require that implementation must obtain for any type space coherent with the original payoff type space. When insisting on robust exact implementation, BM (2008) identify ex post incentive compatibility and robust monotonicity as necessary and almost sufficient conditions. These can be very demanding. For instance, in some settings ex post incentive compatibility would generically require a social choice function (SCF) to be constant (Jehiel et al. (2006)); in others, it still leaves room for non-trivial SCFs (see our Subsection 4.1, or BM (2009, Section 3)). In these environments, robust monotonicity needs to be satisfied as well and that amounts to requiring Bayesian monotonicity in every type space.\(^\text{1}\) In a paper directly related to ours, BM (2009) also consider robust virtual implementation and identify ex post incentive compatibility and robust measurability as the corresponding key conditions.

We shall assume that for each agent \(i\) there is a prespecified set \(Q_i\) of allowed first-order beliefs (when \(Q_i\) is the entire unrestricted simplex, we would be studying the case considered by BM (2009)). The solution concept we shall employ is \(\Delta\)-rationalizability (Battigalli and Siniscalchi (2003)), where \(\Delta\) expresses that the first-order beliefs are restricted to lie in the sets \(Q_i\).\(^\text{2}\) The combination of payoff type and first-order belief for a player will comprise our notion of a first-order type.\(^\text{3}\) Therefore, we shall fix a (typically large) space of first-order types, which we will assume to be common knowledge among the agents, and we shall require that implementation obtain for all higher-order beliefs coherent with our original first-order type space.

To begin with, robust virtual implementation will be limited by (interim) incentive compatibility imposed on the first-order types present in the model. We shall refer to such

---

\(^\text{1}\)See also the related paper by Saijo, Sjöström and Yamato (2007). They propose secure implementation as a double implementation in dominant strategies and in Nash equilibrium, and characterize it in terms of strategy-proofness and the rectangular property. In particular, they also discuss robust secure implementation, which requires a securely implementing mechanism to work for all possible common priors (Section 5 of their paper).

\(^\text{2}\)Following BM (2008), Brandenburger and Dekel (1987) and Dekel, Fudenberg and Morris (2006), we can also characterize the solution concept—iteratively elimination of never best responses—in terms of interim rationalizability which, in turn, is equivalent to the Bayesian equilibria in all coherent type spaces. There are, however, two reasons why our definition of interim rationalizability is more demanding than that of BM (2008, 2009) and Dekel, Fudenberg, and Morris (2006). First, we include the sets \(Q_i\) of first-order beliefs over the payoff type space as part of the environment which is assumed to be common knowledge. Second, at each round of elimination of never best responses, we explicitly require agents’ first-order beliefs to lie in the sets \(Q_i\).

\(^\text{3}\)We thank Stephen Morris for suggesting this name, which replaces our previous, less transparent “pseudo-type.”
a condition as first-order incentive compatibility, and Theorem 1 shows it to be necessary for robust virtual implementation in Δ-rationalizable strategies. Since BM (2009) require robustness with respect to all type spaces, they must require ex post incentive compatibility, on whose strength we have already commented above. In our approach we may rule out some first-order types by assumption, implying that our incentive compatibility will be correspondingly weaker. But this is not the major difference between the two papers: if we allow for a rich enough set of first-order types, the additional incentive compatibility constraints lead to a negligible gap with the ex post version of the condition.

Our Theorem 2 shows that the original AM measurability is also necessary for robust virtual implementation in our sense. The major difference with BM (2009) concerns the difference between the AM measurability and robust measurability conditions. We shall argue that AM measurability is trivially satisfied by every SCF if one imposes it over almost every type space, instead of every type space, which is what robust measurability does. To show the difficulties of robust virtual implementation, BM (2008, 2009) construct a very specific type space in which the interim preferences of all types are aligned. Serrano and Vohra (2001) earlier observed that virtual Bayesian implementation may fail for exactly that reason in a standard Bayesian environment with a fixed type space, but such failures are arguably “rare.”

To make our point, and generalizing the approach in Serrano and Vohra (2005), we begin by proposing a condition that we term first-order type diversity (FOTD). We then consider quasi-transferable environments satisfying this condition and show (Theorem 3) that an SCF is robustly virtually implementable as long as it satisfies first-order incentive compatibility. The FOTD condition is generic in our settings when there are at least three alternatives (see Subsection 4.3 for the case of finite sets of first-order beliefs, and Appendix A for a proof in the continuum case); thus, in quasi-transferable environments, one almost never needs to rely on any additional condition beyond the appropriate version of incentive compatibility.

Next, we seek to obtain a characterization. We extend the work of AM (1992c) to our settings. Theorem 4 provides the sufficiency argument to show that first-order incentive compatibility and AM measurability are necessary and sufficient conditions for robust virtual implementation in Δ-rationalizable strategies. Moreover, we elaborate on the connection between FOTD and AM measurability: as hinted in the original paper by AM (1992c), FOTD is associated with the first iteration of the measurability algorithm, which, in general, may have multiple steps. The algorithm determines the maximum possible separation of types – or first-order types – on the basis of their interim preferences.

We learn from Theorems 3 and 4 and the genericity discussion in Subsection 4.3 that

---

4While to prove our sufficiency results (Theorems 3 and 4) we employ the assumption that Qi’s are finite sets, this is not essential. Indeed, one can adapt the canonical “maximally revealing” mechanism in BM (2009) to the specific Qi’s assumed. The finiteness assumption makes the argument in Theorem 3 especially transparent, while Theorem 4 extends its argument to the case when the measurability algorithm does not end in the first round.

5While we owe much to this important work (AM (1992c)), our analysis exhibits several differences. Our Theorem 3, based on FOTD, is based on a heuristic argument in AM (1992c), but dispenses with the private values assumption. Our solution concept is different, because of our restrictions on first-order beliefs. Our mechanism does not necessarily isolate a unique Δ-rationalizable strategy profile yet implementation is successful. Having said all this, one way to understand our work is that the original result of AM (1992c) is the “robust” characterization of virtual implementation in rationalizable strategies in the robust setting.
robust measurability, which is AM measurability in every type space, is a trivial condition if one weakens it slightly and imposes it over almost every type space. When FOTD is imposed on a finite model, such a model can be viewed as an approximation of the atomless continuum of first-order types. In continuum settings, while FOTD continues to be generic (see Appendix A), it is fair to say that, in conjunction with other usual assumptions, such as convexity of the set of first-order beliefs, it becomes less appealing, as it implies a strong separation of beliefs, determined by the payoff types (see again Subsections 4.1 and 4.2).

A final word is called for regarding the nature of our mechanisms and the connection with virtual implementation in Bayesian equilibrium. First, the distinction between implementation in pure- or mixed-strategy equilibria is of no significance, once we ask for robustness with respect to type spaces. Our sufficiency results apply a fortiori to virtual implementation in mixed-strategy Bayesian equilibrium. Virtual implementation in Bayesian equilibrium is typically more permissive than virtual implementation in iteratively undominated strategies. However, the difference is “small” in that it concerns environments violating FOTD. Furthermore, the additional SCFs so implemented must rely on the use of non-regular mechanisms (e.g., using integer games and devices alike): following a result of AM (1992c) for a fixed type space, AM measurability is necessary for robust virtual implementation in Bayesian equilibrium if one uses regular mechanisms.

Our mechanisms are finite, and best responses always exist. The mechanisms we use adapt the one in AM (1992c) to our robust setting. The latter was itself an adaptation of the mechanism in AM (1992a) to incomplete information. One could raise objections to the realism of all these mechanisms. In particular, Glazer and Rosenthal (1992) criticize the mechanism in AM (1992a), a criticism that also applies to ours. These criticisms mostly involved the complexities of high-order beliefs on endogenous or strategic uncertainty, and they were addressed in AM (1992b). On the other hand, to the extent that we show that our result is independent of – robust to – higher-order belief considerations on the exogenous uncertainty over type spaces, the iterative elimination of never best responses does not involve high levels of complexity in this different sense. We view this as a strong defence of Abreu-Matsushima type of mechanisms.

The paper is organized as follows: In Section 2 we introduce the preliminary notation and definitions. In Section 3 we present Theorems 1 and 2, showing that first-order incentive compatibility and AM measurability are necessary conditions for robust virtual implementation in $\Delta$-rationalizable strategies. Section 4 discusses the relationship of our approach with that in which the planner has no information about the set of first-order types; in it, we go over an example that first appeared in BM (2009), and discuss the generality issue. Our first sufficiency result (Theorem 3), based on FOTD, appears in Section 5. Section 6 provides the characterization result (Theorem 4). We conclude in Section 7. Appendix A contains the proof of genericity of FOTD for continuum settings and Appendix B discusses the connection with virtual implementation in Bayesian equilibrium.

---

6For each fixed type space, this follows since virtual monotonicity (Serrano and Vohra (2005)) is strictly weaker than AM measurability. Also, the sufficiency results for virtual Bayesian implementation have been obtained without the use of the quasi-transferability assumption. As shown in Kunimoto and Serrano (2011), this cannot be dispensed with in our setting.
2 Preliminaries

Let \( N = \{1, \ldots, n\} \) denote the set of agents and \( \Theta_i \) be the set of finite payoff types of agent \( i \). Denote \( \Theta = \Theta_1 \times \cdots \times \Theta_n \), and \( \Theta_{-i} = \Theta_1 \times \cdots \times \Theta_{i-1} \times \Theta_{i+1} \times \cdots \times \Theta_n \). Let \( q_i(\theta_{-i}) \) denote agent \( i \)'s first-order belief that other agents receive the profile of payoff types \( \theta_{-i} \). For an abstract finite set \( S \), we will denote the set of probability distributions over \( S \) by \( \Delta(S) \). Let \( Q_i \subseteq \Delta(\Theta_{-i}) \) be the set of allowed first order beliefs of agent \( i \). We call \( T_i = \Theta_i \times Q_i \) the set of first-order types of agent \( i \).

Let \( A \) denote the set of pure outcomes, which is assumed to be independent of the information state. Suppose \( A = \{a_1, \ldots, a_K\} \) is finite.

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

We can now define an environment as \( \mathcal{E} = (A, \{u_i, \Theta_i, Q_i\}_{i \in N}) \), which is implicitly understood to be common knowledge among the agents. In particular, if \( Q_i \) is unrestricted for each \( i \), that is, \( Q_i = \Delta(\Theta_{-i}) \), we call it a payoff environment denoted as \( \mathcal{E}_\Delta = (A, \{u_i, \Theta_i\}_{i \in N}) \). This is the environment that BM (2008, 2009) consider when they explore the notion of robustness. Our approach adopts an intermediate robustness criterion, as it allows \( Q_i \) to be an arbitrary set of first-order beliefs. In particular, our model escapes the criticism in Neeman (2004) of "beliefs-determine-preferences," as one can allow a rich set of payoff and first-order belief types. While \( Q_i \) consists of all possible beliefs that agents could potentially have in our model, no prior on that set, common or not, needs to be assumed.

A social choice function (SCF) is a function \( f : \Theta \rightarrow \Delta(A) \). Note that, as is standard in the literature on robust implementation, 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 \( (\theta_i, q_i) \) that pretends to be of first-order type \( (\theta'_i, q'_i) \) corresponding to an SCF \( f \) as follows:

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

where \( (\theta_i, q_i), (\theta'_i, q'_i) \in T_i = \Theta_i \times Q_i \). 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 agent \( i \) and an outcome function \( g : M \rightarrow \Delta(A) \), where \( M = \times_{i \in N} M_i \).

Next we define the solution concept of \( \Delta \)-rationalizability that we use in the paper.

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

\[
S_i : \Theta_i \rightarrow 2^{M_i}
\]

---

\(^7\)Similar notation will be used for products of other sets.

\(^8\)Note that we do not introduce the notion of types for the agents and use in the sequel a type-free solution concept of rationalizability.

\(^9\)If \( A \) were an arbitrary separable metric space, we would work with its countable dense subset. The reader is referred to Section 6 of Abreu and Sen (1991) or to Duggan (1997) for more details. See footnotes 14 and 22 in the sequel, when this assumption is invoked.

\(^10\)Note how, since the SCF does not depend on first-order beliefs, the misrepresentation of \( q_i \) into \( q'_i \) is of no consequence. However, some separation arguments below will justify our current definition.
and we write $S$ for the collection of message correspondence profiles. The collection $S$ 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 $\overline{S} = (\overline{S}_1, \ldots, \overline{S}_n)$, where $\overline{S}_i(\theta_i) = M_i$ for all $i \in N$ and $\theta_i \in \Theta_i$. The smallest element is $S = (S_1, \ldots, S_n)$, where $S_i(\theta_i) = \emptyset$ for all $i \in N$ and $\theta_i \in \Theta_i$.

We define an operator $b = (b_1, \ldots, b_n)$ to iteratively eliminate never best responses. To this end, we 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, we introduce some restrictions on agents’ first-order beliefs. For any $q_i \in Q_i$, define

$$\Delta_q(\Theta_{-i} \times M_{-i}) \equiv \{ \mu_i \in \Delta(\Theta_{-i} \times M_{-i}) \mid \text{marg}_{\Theta_{-i}} \mu_i = q_i \},$$

where 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 $b : S \rightarrow S$ is now defined as follows:

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

This is an incomplete information version of rationalizability, proposed by Battigalli and Siniscalchi (2003). They call it $\Delta$-rationalizability and denote by $\Delta$ restrictions on the set of first-order beliefs. When $Q_i = \Delta(\Theta_{-i})$, this rationalizability is equivalent to the one used by BM (2009). We 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 we label $S^\Gamma$. Thus, we have that (i) $b(S^\Gamma) = S^\Gamma$ and (ii) $b(S) = S \Rightarrow S \leq S^\Gamma$. Since the message space is finite, we have

$$S^\Gamma_i(\theta_i) \equiv \bigcap_{n \geq 1} b_i(b^n(S))[\theta_i].$$

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

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

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

**Definition 1 (Robust Virtual Implementation)** An SCF $f$ is robustly virtually implementable if there exists $\varepsilon > 0$ such that, for any $\varepsilon \in (0, \varepsilon]$, there exists a mechanism $\Gamma^\varepsilon = (M^\varepsilon, g^\varepsilon)$ for which for any $\theta \in \Theta$ and $m \in M^\varepsilon$,

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

Note that when $\varepsilon$ is taken to be 0 in the above definition, the corresponding concept would be robust exact implementation.

---

11To avoid heavy notation, we ignore the fact that the message correspondence depends on the underlying mechanism $\Gamma$. 

7
3 Necessity for Robust Virtual Implementation

In this section we discuss the necessary conditions for robust virtual implementation when the environment specifies an arbitrary set $Q_i$ of first-order beliefs for each agent $i$. These conditions are necessary independently of the more specific assumptions made on the environment.\footnote{In a standard Bayesian environment with a fixed type space, Kunimoto and Serrano (2011) identify another necessary condition if one uses finite or regular mechanisms, which a fortiori also applies to our robust settings. This condition is vacuously satisfied in the presence of quasi-transferability (to be defined in Section 5), and hence, given our results later in the current paper, there is no need to state it here.}

3.1 Incentive Compatibility

The notion of first-order type suggests the following definition, which is the standard interim incentive compatibility condition applied to the first-order types present in the model, as specified by the sets $Q_i$:

**Definition 2 (First-Order Incentive Compatibility)** An SCF $f : \Theta \rightarrow \Delta(A)$ satisfies first-order incentive compatibility if, for any $i \in N$ any $(\theta_i, q_i) \in T_i = \Theta_i \times Q_i$ and any $\theta_i \in \Theta_i$,

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

We shall say that an SCF $f$ satisfies strict first-order incentive compatibility if all the inequalities in the preceding definition are strict whenever $\theta_i \neq \theta_i'$.

For a fixed mechanism $\Gamma = (M, g)$, we define agent $i$’s (pure) strategy $\sigma_i : \Theta_i \rightarrow M_i$. The next theorem identifies first-order incentive compatibility as a necessary condition for implementability:

**Theorem 1** If an SCF $f$ is robustly virtually implementable, then it satisfies first-order incentive compatibility.

**Proof**: By our hypothesis, for each $\varepsilon > 0$ sufficiently small, there exists a corresponding mechanism $\Gamma$ such that for all $\theta \in \Theta$, $m \in S^f(\theta) \Rightarrow \|g(m) - f(\theta)\| \leq \varepsilon$.

Fix $\sigma_{-i} : \Theta_{-i} \rightarrow M_{-i}$ such that $\sigma_{-i}(\theta_{-i}) \in S^f_{-i}(\theta_{-i})$ for each $\theta_{-i} \in \Theta_{-i}$. For any $m'_i \in S^f_i(\theta'_i)$, robust virtual implementation requires that for any $\theta_{-i} \in \Theta_{-i}$,

$$\|g(m'_i, \sigma_{-i}(\theta_{-i})) - f(\theta'_i, \theta_{-i})\| \leq \varepsilon. \tag{1}$$

Suppose that agent $i$ is of first-order type $(\theta_i, q_i)$ and he holds the belief $\mu_i \in \Delta_q(\Theta_{-i} \times M_{-i})$ such that for each $\theta_{-i}$ with $q_i(\theta_{-i}) > 0$ and each $m_{-i} \in M_{-i}$, $\mu_i(\theta_{-i}, m_{-i}) > 0$ if and only if $m_{-i} = \sigma_{-i}(\theta_{-i})$. Let $m_i$ be any message that is a best response to the belief $\mu_i$. Then if $m_i \in S^f_i(\theta_i)$, virtual implementation implies that for any $\theta_{-i} \in \Theta_{-i}$,

$$\|g(m_i, \sigma_{-i}(\theta_{-i})) - f(\theta_i, \theta_{-i})\| \leq \varepsilon. \tag{2}$$

By the best response property of $m_i$ and the construction of $\mu_i$,

$$\sum_{\theta_{-i}, m_{-i}} \mu_i(\theta_{-i}, m_{-i}) \left[u_i(g(m_i, m_{-i}); \theta_i, \theta_{-i}) - u_i(g(m'_i, m_{-i}); \theta_i, \theta_{-i})\right] \geq 0.$$
Once again, by the construction of $\mu_i$, we can rewrite the above inequality as follows:

$$
\sum_{\theta_{-i}} q_i(\theta_{-i}) \left[ u_i(g(m_i, \sigma_{-i}(\theta_{-i})); \theta_i, \theta_{-i}) - u_i(g(m'_i, \sigma_{-i}(\theta_{-i})); \theta_i, \theta_{-i}) \right] \geq 0.
$$

(3)

Due to the fact that $\Theta$ and $A$ are finite, (1), (2), and (3) together imply the following: there exists $C > 0$ such that

$$
\sum_{\theta_{-i}} q_i(\theta_{-i}) \left[ u_i(f(\theta_i, \theta_{-i}); \theta_i, \theta_{-i}) - u_i(f(\theta'_i, \theta_{-i}); \theta_i, \theta_{-i}) \right] \geq -\varepsilon C.
$$

Since $\varepsilon$ can be chosen arbitrarily small due to the requirement of robust virtual implementation, we obtain

$$
\sum_{\theta_{-i}} q_i(\theta_{-i}) \left[ u_i(f(\theta_i, \theta_{-i}); \theta_i, \theta_{-i}) - u_i(f(\theta'_i, \theta_{-i}); \theta_i, \theta_{-i}) \right] \geq 0.
$$

This can be written as:

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

This establishes that $f$ satisfies first-order incentive compatibility. ■

When $Q_i = \Delta(\Theta_{-i})$ for every $i \in N$, it is easy to see that first-order incentive compatibility is equivalent to ex post incentive compatibility:

**Definition 3 (Ex Post Incentive Compatibility)** An SCF $f : \Theta \to \Delta(A)$ satisfies ex post incentive compatibility if for any $i \in N$, $\theta_{-i} \in \Theta_{-i}$, and $\theta_i, \theta'_i \in \Theta_i$,

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

### 3.2 Measurability

In an important paper, AM (1992c) uncovered a condition that they termed measurability (we shall refer to it as AM measurability) that was necessary for virtual implementation in iteratively undominated strategies over a standard environment that fixes a Bayesian type space. In this section we revisit the AM measurability condition by applying it to our robust implementation analysis.

Denote by $\Psi_i$ a partition of the set of first-order types $T_i$, where $\psi_i$ is a generic element of $\Psi_i$ and $\Psi_i(t_i)$ denotes the element of $\Psi_i$ that includes first-order type $t_i = (\theta_i, q_i)$. Let $\Psi = \times_{i \in N} \Psi_i$ and $\psi = \times_{i \in N} \psi_i$.

**Definition 4** An SCF $f$ is measurable with respect to $\Psi$ if, for every $i \in N$ and every $t_i = (\theta_i, q_i), t'_i = (\theta'_i, q'_i) \in T_i$ with $\theta_i \neq \theta'_i$, whenever $\Psi_i(t_i) = \Psi_i(t'_i)$,

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

Measurability of $f$ with respect to $\Psi$ 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 $\Psi_i$.  


Definition 5 A strategy $\sigma_i$ for player $i$ is measurable with respect to $\Psi_i$ if for every $t_i = (\theta_i, q_i), t_i' = (\theta_i', q_i') \in T_i$ with $\theta_i \neq \theta_i'$,

$$\Psi_i(t_i) = \Psi_i(t_i') \implies \sigma_i(\theta_i) = \sigma_i(\theta_i').$$

A strategy profile $\sigma$ is measurable with respect to $\Psi$ if, for every $i \in N$, $\sigma_i$ is measurable with respect to $\Psi_i$.

We can now provide the definition of equivalent (first-order) types. Note that, since agent $i \in N$ distinguishes all its first order types, we consider a partition $T_i \times \Psi_{-i}$ in that definition, unlike the definition of measurable strategies.

Definition 6 For every $i \in N$, $t_i = (\theta_i, q_i), t_i' = (\theta_i', q_i') \in T_i$ with $\theta_i \neq \theta_i'$, and $(n-1)$ tuple of partitions $\Psi_{-i}$, we say that $t_i$ is equivalent to $t_i'$ (denoted by $t_i \sim t_i'$) with respect to $\Psi_{-i}$ if, for any pair of SCFs $f$ and $\tilde{f}$ which are measurable with respect to $T_i \times \Psi_{-i}$,

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

Remark: What we aim to distinguish here are “payoff types.” In particular, we consider any two first-order types $(\theta_i, q_i)$ and $(\theta_i', q_i)$ as (essentially) equivalent if $\theta_i = \theta_i'$.

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

$$R_i(\Psi_{-i}) = \{ \rho_i(t_i, \Psi_{-i}) \subseteq T_i | t_i \in T_i \}.\$$

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

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

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

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

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

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

where $\bigvee$ denotes the join on $\{\Psi_i^h\}_{h=0}^\infty$ associated with $\geq$.

Since $\Theta$, is finite for each agent $i \in N$, there exists a positive integer $L$ such that $\Psi^h = \Psi^L$ for any $h \geq L$. We can write $\Psi^* = \Psi^L$.

Definition 7 An SCF $f$ satisfies AM measurability if it is measurable with respect to $\Psi^*$.  

10
Note how the partitions $\Psi_0, \Psi_1, ..., \Psi_*$ used in AM measurability are simply derived as a property of the environment. The aim is to “treat equally” those first-order types that are “indistinguishable” according to their interim preferences. Thus, we start considering constant SCFs, i.e., SCFs that are measurable with respect to the coarsest possible partition, and we separate first-order types who have different payoff types and different interim preferences over this class of SCFs. This gives us a new partition of the set of first-order types for each agent (iteration 1). Next, we consider SCFs measurable with respect to these new partitions, and ask the same question: are there first-order types that, having the same preferences over constant SCFs, now can be separated because, having different payoff types, they exhibit different interim preferences over the enlarged class of SCFs considered? If the answer is No, the process ends and we have found $\Psi_*$. If it is Yes, we proceed to make the induced finer partition of each set of first-order types (iteration 2), and so on. The process ends with the identification of $\Psi_*$, which provides the maximum possible degree of first-order type separation or distinguishability in terms of interim preferences. AM measurability simply asks that the SCF not distinguish between different first-order types that are “indistinguishable” according to $\Psi_*$. 

To compare robust and AM measurabilities, we follow the notation of BM (2009). Consider the set of interim preferences that an agent might have over constant lotteries if his payoff type is $\theta_i$ and he knows that the type $\theta_j$ of each opponent $j$ belongs to some subset $\xi_j$ of his possible types $\Theta_j$. We define

$$R_i(\theta_i, \xi_{-i}) = \{ R \in \mathcal{R} \mid R = R_{\theta_i, \lambda_i} \text{ for some } \lambda_i \in \Delta(\xi_{-i}) \},$$

where $\mathcal{R}$ denotes the set of preference relations on $\Delta(A)$ and for any $y, y' \in \Delta(A)$,

$$y \mathcal{R}_{\theta_i, \lambda_i} y' \iff \sum_{\theta_{-i} \in \Theta_{-i}} \lambda_i(\theta_{-i}) u_i(y; \theta_i, \theta_{-i}) \geq \sum_{\theta_{-i} \in \Theta_{-i}} \lambda_i(\theta_{-i}) u_i(y'; \theta_i, \theta_{-i}).$$

We say that $\xi_{-i}$ separates $\xi_i$ if there exist $\theta_i, \theta'_i \in \xi_i$ such that $R_i(\theta_i, \xi_{-i}) \cap R_i(\theta'_i, \xi_{-i}) = \emptyset$, i.e., whatever those realized preferences are, we can rule out at least one possible type of $i$. Set

$$\Xi^0 = 2^{\Theta_i};$$

and

$$\Xi^{k+1} = \{ \xi_i \in \Xi^k \mid \xi_{-i} \text{ does not separate } \xi_i \text{ for some } \xi_{-i} \in \Xi^k \}$$

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

$$\Xi^* = \bigcap_{k \geq 0} \Xi^k.$$ 

An SCF $f$ is measurable with respect to $\Xi$ if, for any $i \in N$ and $\theta_i, \theta'_i \in \Theta_i$, whenever there exists $\xi_i \subseteq \Theta_i$ such that $\{ \theta_i, \theta'_i \} \subseteq \xi_i \in \Xi_i$, $f(\theta_i, \theta_{-i}) = f(\theta'_i, \theta_{-i})$ for all $\theta_{-i} \in \Theta_{-i}$. An $f$ is robustly measurable if it is measurable with respect to $\Xi^*$.

$^{13}$BM (2009) use $\Psi_{-i}$ for $\xi_{-i}$. We change the notation here to avoid confusion with our use of $\Psi$. 

11
Lemma 1  If an SCF $f$ satisfies robust measurability, it also satisfies AM measurability.

Remark: The converse statement is not necessarily true, as examples of Sections 4.1 and 4.2 show. However, we will show later that the converse holds when the set of first-order beliefs is unrestricted.

Proof: It suffices to show that for each $k \geq 1$, whenever two payoff types $\theta_i$ and $\theta'_i$ are separated at step $k$ of the robust measurability algorithm, for any $q_i, q'_i \in Q_i$, two first-order types $(\theta_i, q_i)$ and $(\theta'_i, q'_i)$ are also separated by step $k$ of the AM measurability algorithm.

We prove this by induction. Fix $k = 1$ and suppose that $\theta_i$ and $\theta'_i$ are separated at step 1 of the robust measurability algorithm. Since $\xi_{-i}^0 = \times_{j \neq i} 2^{\Theta_j}$, this implies

$$R_i(\theta_i, \xi_{-i}^0) \cap R_i(\theta'_i, \xi_{-i}^0) = \emptyset.$$ 

Since $\Psi_{-i}^0 = \times_{j \neq i} \{\Theta_j\}$, this implies that, for any $\lambda_i, \lambda'_i \in \Delta(\Theta_{-i})$, we have

$$\rho_i(\theta_i, \lambda_i, \Psi_{-i}^0) \cap \rho_i(\theta'_i, \lambda'_i, \Psi_{-i}^0) = \emptyset.$$ 

Therefore, for any $q_i, q'_i \in Q_i$, two first-order types $(\theta_i, q_i)$ and $(\theta'_i, q'_i)$ are separated in the first round of the AM measurability algorithm. Hence, if $f$ is measurable with respect to $\Xi^1$, it is also measurable with respect to $\Psi^1$.

Take an arbitrary $k \geq 1$ and suppose, by the induction hypothesis, that $f$ is measurable with respect to both $\Xi^k$ and $\Psi^k$. Assume further that $\theta_i$ and $\theta'_i$ are “not” yet separated until step $k$ of the robust measurability algorithm but they are separated at step $k + 1$. Assume further that for any $q_i, q'_i \in Q_i$, two first-order types $(\theta_i, q_i)$ and $(\theta'_i, q'_i)$ are “not” yet separated at step $k$ of the AM measurability algorithm. This last assumption is made without loss of generality because otherwise, the proof is all the easier. By our induction hypothesis, for any $\xi_{-i}^k \in \Xi_{-i}^k$, we have

$$R_i(\theta_i, \xi_{-i}^k) \cap R_i(\theta'_i, \xi_{-i}^k) = \emptyset.$$ 

Since $f$ is measurable with respect to both $\Xi^k$ and $\Psi^k$ by our induction hypothesis, we have that for any $q_i, q'_i \in Q_i$, $x_{-i}^k \in \Xi_{-i}^k$, we have

$$\rho_i(\theta_i, q_i, \Psi_{-i}^0) \cap \rho_i(\theta'_i, q'_i, \Psi_{-i}^0) = \emptyset.$$ 

Thus, for any $q_i, q'_i \in Q_i$, two first-order types $(\theta_i, q_i)$ and $(\theta'_i, q'_i)$ are separated in step $k + 1$ of the AM measurability algorithm. Therefore, if $f$ is measurable with respect to $\Xi^{k+1}$, it is also measurable with respect to $\Psi^{k+1}$. This completes the proof. $\blacksquare$

In the next lemma, we show that the difference between AM measurability and robust measurability disappears if we consider unrestricted first-order beliefs.

Lemma 2 Let $Q_i = \Delta(\Theta_{-i})$ for each $i \in N$. Then, an SCF $f$ satisfies robust measurability if and only if it satisfies AM measurability.
Proof: Given the previous lemma, it suffices to show that for each \( k \geq 1 \), whenever for any \( q_i, q'_i \in Q_i = \Delta(\Theta_{-i}) \), two first-order types \((\theta_i, q_i)\) and \((\theta'_i, q'_i)\) are separated in step \( k \) of the AM measurability algorithm, two payoff types \( \theta_i \) and \( \theta'_i \) are separated in step \( k \) of the robust measurability algorithm.

We prove this by induction. Fix \( k = 1 \) and suppose that for any \( q_i, q'_i \in Q_i = \Delta(\Theta_{-i}) \), two first-order types \((\theta_i, q_i)\) and \((\theta'_i, q'_i)\) are separated in step 1 of the AM measurability algorithm. That is, for any \( q_i, q'_i \in Q_i = \Delta(\Theta_{-i}) \), we have

\[
\rho_i(\theta_i, q_i, \Psi^0_{-i}) \cap \rho_i(\theta'_i, q'_i, \Psi^0_{-i}) = \emptyset.
\]

Since \( Q_i \) is unrestricted, \( \Psi^0_{-i} = \times_{j \neq i} \{\Theta_j\} \), and \( \xi^0_{-i} = \times_{j \neq i} 2^{\Theta_j} = \Xi^0_{-i} \), this hypothesis implies that

\[
R_i(\theta_i, \xi^0_{-i}) \cap R_i(\theta'_i, \xi^0_{-i}) = \emptyset.
\]

Then, two payoff types \( \theta_i \) and \( \theta'_i \) are separated in step 1 of the robust measurability algorithm. Hence, if \( f \) is measurable with respect to \( \Psi^1 \), it is measurable with respect to \( \Xi^1 \).

Take an arbitrary \( k \geq 1 \) and suppose, by the induction hypothesis, that \( f \) is measurable with respect to both \( \Xi^k \) and \( \Psi^k \). Assume further that for any \( q_i, q'_i \in Q_i = \Delta(\Theta_{-i}) \), two first-order types \((\theta_i, q_i)\) and \((\theta'_i, q'_i)\) are “not” yet separated until step \( k \) of the AM measurability algorithm and two payoff types \( \theta_i \) and \( \theta'_i \) are not yet separated until step \( k \) of the robust measurability algorithm but these two first-order types \((\theta_i, q_i)\) and \((\theta'_i, q'_i)\) are separated in step \( k+1 \) of the AM measurability algorithm. Our assumption implies that for any \( q_i, q'_i \in Q_i = \Delta(\Theta_{-i}) \),

\[
\rho_i(\theta_i, q_i, \Psi^k_{-i}) \cap \rho_i(\theta'_i, q'_i, \Psi^k_{-i}) = \emptyset.
\]

Since \( f \) is measurable with respect to both \( \Psi^k \) and \( \Xi^k \) and \( Q_i \) is unrestricted, for any \( \xi^k_{-i} \in \Xi^k_{-i} \), we have

\[
R_i(\theta_i, \xi^k_{-i}) \cap R_i(\theta'_i, \xi^k_{-i}) = \emptyset.
\]

This implies that \( \theta_i \) and \( \theta'_i \) are separated in step \( k+1 \) of the robust measurability algorithm. We conclude that \( f \) is measurable with respect to \( \Xi^{k+1} \), which completes the proof. \( \blacksquare \)

AM (1992c) show that in a Bayesian environment with a fixed type space, AM measurability is a necessary condition for virtual implementation in iteratively undominated strategies. We establish a robust analogue of this result:

Theorem 2 If an SCF \( f \) is robustly virtually implementable, then it satisfies AM measurability.

Proof: Since \( f \) is robustly virtually implementable, there exists a mechanism \( \Gamma = (M, g) \) such that whenever \( m \in S^f(\theta) \), \( \|g(m) - f(\theta)\| \leq \varepsilon \) for \( \varepsilon > 0 \). For each \( h \geq 1 \), let \( K^h = \times_{i \in N} K^h_i \) be the sets of strategies that survive \( h \) rounds of iterative elimination of never best responses.
Consider an arbitrary constant strategy profile \( \sigma[0] \in \mathcal{K}^0 \) which is measurable with respect to \( \times_{i \in \mathbb{N}} \{T_i\} \). Then, either (1) \( \|g(\sigma[0](\theta)) - f(\theta)\| \leq \varepsilon \) for every \( \theta \) or (2) \( g(\sigma[0](\theta)) - f(\theta)\| > \varepsilon \) for some \( \theta \in \Theta \).

In case (1), because \( \varepsilon \) can be chosen arbitrarily, \( f \) is a constant SCF. It is then measurable with respect to \( \times_{i \in \mathbb{N}} \{T_i\} \), hence with respect to \( \Psi^* \) as well. Thus, \( f \) satisfies AM measurability and we complete the proof.

In case (2), by the definition of \( \Psi^i \) and our hypothesis that \( f \) is robustly virtually implementable, it follows that for every \( i \in \mathbb{N} \), there exists \( \sigma_i[1] \in \mathcal{K}_i \) that is a best response to \( \sigma_{-i}[0] \) and is measurable with respect to \( \Psi^i_i \). Hence, \( \sigma_i[1] \in \mathcal{K}^1_i \).

There are again two possibilities: suppose \( \|g(\sigma[1](\theta)) - f(\theta)\| \leq \varepsilon \) for every \( \theta \in \Theta \). Then \( g \circ \sigma[1] \) is measurable with respect to \( \Psi^1 \). Consider \( t_i = (\theta_i, q_i), t'_i = (\theta'_i, q'_i) \in T_i \) such that \( \theta_i \neq \theta'_i \) and \( \Psi^i_i(t_i) = \Psi^i_i(t'_i) \). Note that \( \Psi^i_i(t_i) \) is the element of \( \Psi^i_i \) that includes \( t_i \). By the previous hypothesis, we have that for any \( \theta_{-i} \), \( \|g(\sigma[1](\theta_i, \theta_{-i}) - f(\theta_i, \theta_{-i}))\| \leq \varepsilon \) and \( \|g(\sigma[1](\theta'_i, \theta_{-i}) - f(\theta'_i, \theta_{-i}))\| \leq \varepsilon \). Since \( \sigma[1](\theta_i, \theta_{-i}) = \sigma[1](\theta'_i, \theta_{-i}) \) for \( \theta_{-i} \) by measurability with respect to \( \Psi^1 \), we have \( \|f(\theta_i, \theta_{-i}) - f(\theta'_i, \theta_{-i})\| \leq 2\varepsilon \). Since this must be true for any \( \varepsilon > 0 \), we obtain \( f(\theta_i, \theta_{-i}) = f(\theta'_i, \theta_{-i}) \) for any \( \theta_{-i} \). Hence, \( f \) satisfies AM measurability.

Suppose, on the other hand, that \( \|g(\sigma[1](\theta)) - f(\theta)\| > \varepsilon \) for some \( \theta \in \Theta \), in which case at least one type finds his strategy \( \sigma_i[1] \) as never a best response given \( \mathcal{K}^1 \). We then repeat the argument to arrive at either AM measurability of \( f \) or at a conclusion that some strategy is never a best response given \( \mathcal{K}^2 \).

Take an arbitrary \( h = 2, 3, \ldots \), and suppose that there exists a strategy profile \( \sigma[h-1] \in \mathcal{K}^{h-1} \) that is measurable with respect to \( \Psi^{h-1} \). Again, there are two possibilities: if \( \|g(\sigma[h-1](\theta)) - f(\theta)\| \leq \varepsilon \) for every \( \theta \in \Theta \), by the argument in the previous paragraph, we can show that \( f \) satisfies AM measurability. Otherwise, since \( f \) is robustly virtually implementable by our hypothesis, for every \( i \in \mathbb{N} \), there exists \( \sigma_i[h] \in \mathcal{K}_i \) that is a best response to \( \sigma_{-i}[h-1] \) and is measurable with respect to \( \Psi^{h}_i \).

Let \( \sigma^* \) be a strategy profile that survives the iterative elimination of never best responses in the implementing mechanism \( \Gamma \). Then, the preceding argument implies that \( \sigma^* \) is measurable with respect to \( \Psi^* \). It follows that \( g \circ \sigma^* \) is measurable with respect to \( \Psi^* \). By our hypothesis that \( f \) is robustly virtually implementable, we have \( \|g(\sigma^*(\theta)) - f(\theta)\| \leq \varepsilon \) for any \( \theta \in \Theta \). Consider \( t_i = (\theta_i, q_i), t'_i = (\theta'_i, q'_i) \in T_i \) such that \( \theta_i \neq \theta'_i \) and \( \Psi^i_i(t_i) = \Psi^i_i(t'_i) \). Once again, by our hypothesis that \( f \) is robustly virtually implementable, we can show that \( \|f(\theta_i, \theta_{-i}) - f(\theta'_i, \theta_{-i})\| \leq 2\varepsilon \) for any \( \theta_{-i} \in \Theta_{-i} \). Since this must be true for any \( \varepsilon > 0 \), it follows that \( f(\theta_i, \theta_{-i}) = f(\theta'_i, \theta_{-i}) \) for any \( \theta_{-i} \). Thus, \( f \) satisfies AM measurability.

To illustrate the implications of AM measurability, we shall introduce a weak regularity assumption on environments. To do so, we need some notation. Recall that \( A = \{a_1, \ldots, a_K\} \). Define \( V^k_i(t_i) \) to be the interim expected utility of agent \( i \) of first-order type \( t_i = (\theta_i, q_i) \) for the constant SCF that assigns \( a_k \) in each state in \( \Theta \), i.e.,

\[
V^k_i(t_i) = \sum_{\theta_{-i} \in \Theta_{-i}} q_i(\theta_{-i})u_i(a_k; \theta_i, \theta_{-i}).
\]

Let \( V_i(t_i) = (V^1_i(t_i), \ldots, V^K_i(t_i)) \).

Next, we define the condition of first-order type diversity in an environment, which will play an important role in our analysis:
Definition 8 (FOTD) An environment $\mathcal{E}$ satisfies first-order type diversity (FOTD) if there do not exist $i \in N$, $t_i = (\theta_i, q_i), t'_i = (\theta'_i, q'_i) \in T_i$ with $\theta_i \neq \theta'_i$, $\beta \in \mathbb{R}_{++}$ and $\gamma \in \mathbb{R}$ such that

$$V_i(t_i) = \beta V_i(t'_i) + e,$$

where $e$ is the unit vector in $\mathbb{R}^K$.\(^{14}\)

First-order type diversity is an extension of the type diversity condition for a standard Bayesian environment, used in Serrano and Vohra (2005). The reader is referred to that paper to find an appraisal of the connections of type diversity with the conditions of interim value distinguished types (Palfrey and Srivastava (1993, definition 6.3)), incentive consistency (Duggan (1997)), and with the algorithm behind measurability due to AM (1992c). As discussed below, the condition is especially compelling in finite environments, although its definition does not rely on finite sets of first-order types.

There is a tight connection between FOTD and the measurability algorithm. When we have considered FOTD, we have defined a vector $V_i(t_i)$ of agent $i$’s valuations of each alternative $a_k$. When the algorithm that determines $\Psi^*$ does not stop in the first step, we need to consider a more complicated “version” of $a_k$, that we define below. Define

$$F = \{h \mid h(\theta) \text{ is a degenerate lottery for all } \theta \in \Theta\}.$$  

Recall that $\Theta$ and $A$ are finite. Then, $F$ becomes a finite functional space. Define also

$$F(\Psi) = \{h \in F \mid h \text{ is measurable with respect to } \Psi\}.$$  

Let $|F(T_i \times \Psi_{-i})| = K$.\(^{15}\) Define $V^k_i(t_i; \Psi_{-i})$ to be the interim expected utility of agent $i$ of first-order type $t_i = (\theta_i, q_i)$ for each SCF $f^k \in F(T_i \times \Psi_{-i})$, i.e.,

$$V^k_i(t_i; \Psi_{-i}) = \sum_{\theta_{-i} \in \Theta_{-i}} q_i(\theta_{-i}) u_i(f^k(\theta_i, \theta_{-i}); \theta_i, \theta_{-i}).$$

Let $V_i(t_i; \Psi_{-i}) = (V^1_i(t_i; \Psi_{-i}), \ldots, V^K_i(t_i; \Psi_{-i})).$

The next lemma follows simply from the definitions of $F(\Psi)$ and of equivalent first-order types. Its proof is omitted:

**Lemma 3** Let $t_i = (\theta_i, q_i), t'_i = (\theta'_i, q'_i) \in T_i$ with $\theta_i \neq \theta'_i$. Then, $t_i$ is equivalent to $t'_i$ with respect to $\Psi_{-i}$ if and only if there exist $\beta > 0$ and $\gamma \in \mathbb{R}$ such that

$$V_i(t_i; \Psi_{-i}) = \beta V_i(t'_i; \Psi_{-i}) + e,$$

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

---

\(^{14}\) If $A$ is a separable metric space, let $A^* = \{a_1, a_2, \ldots\}$ be a countable dense subset of $A$. Now, we can define

$$V_i(t_i) = (V^k_i(t_i))_{k=1}^{\infty} \in \mathbb{R}^\infty.$$

We also define $e$ as the countable unit base in $A$ with $\|e\| = 1$. With these qualifications, FOTD is also well defined for separable metric spaces.

\(^{15}\) This is a slight abuse of notation, since $K$ was defined in previous sections as the finite number of alternatives in the set $A$. This should not cause any confusion.
The following is a characterization of FOTD in terms of the measurability construction:

**Corollary 1** An environment \( \mathcal{E} \) satisfies FOTD if and only if there do not exist \( i \in N \) and \( t_i = (\theta_i, q_i), t_i' = (\theta_i', q_i') \in T_i \) with \( \theta_i \neq \theta_i' \) such that \( t_i \) is equivalent to \( t_i' \) with respect to \( \Psi_{0-i}^0 \). It follows that \( \Psi_i^1 = T_i \) for each agent \( i \in N \), and \( \Psi^* = T \).

In light of Corollary 1, one can make the following useful observation (see Serrano and Vohra (2005) for a similar assertion concerning their type diversity):

**Lemma 4 (FOTD \( \Rightarrow \) AM measurability)** Suppose an environment \( \mathcal{E} \) satisfies FOTD. Then, every SCF satisfies AM measurability.

That is, if the environment satisfies FOTD, the algorithm that separates payoff types in the definition of measurability arrives at the finest partition in the first round.

### 4 The Relationship with the Case of Unrestricted First-Order Types

BM (2009) study robust virtual implementation without specifying first-order type spaces as part of the common knowledge structure in the environment. Under an economic assumption on the domain, they characterize robust virtual implementation by means of ex post incentive compatibility and robust measurability. Robust measurability amounts to AM measurability in every type space, and BM (2009) assess it as a very demanding condition, which leads them to conclude that the restrictions to robust virtual implementation are severe. This section elaborates on this assessment, and in doing so, compares their results to ours. We shall organize it in three subsections: the first two are based on an example and the third discusses genericity issues more generally.

#### 4.1 An Example under First-Order Type Diversity

We find it useful to adopt the example from Section 3 in BM (2009). It describes the classic problem of allocating one unit of an indivisible good. Most importantly, it will help underscore the differences between the two papers.

Let the set of payoff types be a finite subset of \([0, 1]\). For simplicity, let us consider the case in which there are only two payoff types for each agent, \( \theta_i = 0 \) and \( \theta_i = 1 \). If agent \( i \) receives the object, his ex post valuation for it is \( \theta_i + \gamma \sum_{j \neq i} \theta_j \). Here, \( \gamma \geq 0 \) is the interdependence parameter.

Our focus is on SCFs that allocate the object efficiently, that is, to the agent with the highest ex post valuation. It can be shown that when \( \gamma > 1 \), even the standard interim incentive compatibility condition cannot be met by any such SCF. Thus, exact and virtual implementation of this important class of SCFs are impossible in this case.

Suppose then that \( \gamma \leq 1 \). BM (2009) show that robust virtual implementation is possible in this example if there is not too much interdependence in preferences across agents (specifically, when \( \gamma < 1/(n - 1) \)). For this case, BM (2005, 2009) construct a direct mechanism where truth-telling is the unique rationalizable action, and hence the desired
outcome is robustly virtually implementable (the mechanism implements the desired allocation with arbitrarily high probability and the winner pays the “pivotal” price, whereas a random allocation is implemented with the rest of probability).\footnote{Also when $\gamma < 1/(n - 1)$, Chung and Ely (2001) had earlier shown that truth-telling is the unique strategy surviving iterative deletion of weakly dominated strategies in the direct mechanism that uses only the pivotal price.}

On the other hand, BM (2009) show that robust virtual implementation is impossible, also in the intermediate range of $\gamma$’s ($1/(n - 1) \leq \gamma \leq 1$), this time because of a failure of robust measurability. In trying to understand the “size” of this failure, we shall show that, under some standard assumptions, for almost every specification of the set of first-order beliefs $Q_i$, the necessary conditions for robust virtual implementation (i.e., first-order incentive compatibility and AM measurability) are very permissive in the example, thanks to the FOTD condition. We proceed to details.

For simplicity in the writing of expressions below, let $n = 3$. Suppose that the first-order types for each agent are independent.\footnote{This independence assumption is made also for the sake of simplicity. Essentially the same argument will go through even if there is correlation.} Recall that there are two payoff types for each agent (0 and 1) and that we are interested in SCFs that allocate the good efficiently. The specific SCF we consider allocates the good to that agent who announces the highest payoff type (in the event of a tie, the object is allocated at random among the highest announcements, using equal probabilities). To calculate the prices at which the good will be sold, denote by $p_k$ the price that corresponds to $k = 0, 1, 2, 3$ announcements of the high type $\theta_i = 1$. Denote by $q$ (resp., $q'$) the probability that agent $i$ of payoff type $\theta_i = 0$ (resp., $\theta_i = 1$) believes that agent $j$ is of the low payoff type.

Then, the incentive compatibility constraint for payoff type $\theta_i = 0$ is
\[
q^2(1/3)(-p_0) \geq q^2(-p_1) + q(1 - q)(\gamma - p_2) + (1 - q)^2(2\gamma - p_3),
\]
and the one for $\theta_i = 1$ is
\[
q'^2(1 - p_1) + q'(1 - q')(1 + \gamma - p_2) + (1 - q')^2(1/3)(1 + 2\gamma - p_3) \geq q'^2(1/3)(1 - p_0).
\]

So, for example, if one adopts a pricing rule so that $p_0 = p_1 = 0$, $p_2 = \gamma$ and $p_3 = 2\gamma$, these constraints are met for all values of $q$ and $q'$. Thus, the ex post efficient allocation of the object, together with these prices, satisfies ex post incentive compatibility, and therefore, it also satisfies first-order incentive compatibility for any specification of the $Q_i$’s.

Next, we turn our attention to FOTD. First, we claim that for $1/2 > \gamma > 0$, the environment satisfies FOTD. Given our pricing rule, there are nine constant alternatives of relevance:

• $a_1$: the object is allocated to agent 1 for a price of 0;
• $a_2$: the object is allocated to agent 1 for a price of $\gamma$;
• $a_3$: the object is allocated to agent 1 for a price of $2\gamma$;
• $a_k$, $k = 4, \ldots, 9$: the object is allocated to either agent 2 or 3 for each of the three prices.
Therefore, the last six entries in each nine-dimensional vector for agent 1’s interim expected utility are all zeros. We write these vectors of interim expected utility for the first-order types of agent 1 (the ones for agents 2 and 3 are similar, but alter the location of the zero components):

\[ V_1(0, q) = (2\gamma(1 - q), 2\gamma(1 - q) - \gamma, 2\gamma(1 - q) - 2\gamma, 0, \ldots, 0) \]
\[ V_1(1, q') = (1 + 2\gamma(1 - q'), 1 + 2\gamma(1 - q') - \gamma, 1 + 2\gamma(1 - q') - 2\gamma, 0, \ldots, 0) \]

When \( \gamma \in (0, 1/2) \), it can be easily checked that none of these vectors are positive affine transformations of one another. Thus, FOTD always holds in this case, no matter what sets \( Q_i \) of first-order beliefs are picked. This strong separation of first-order types helps to explain the permissive result in BM (2009).

In contrast, suppose now that \( 1 \geq \gamma \geq 1/2 \). For this case, the claim in BM (2009) is that robust virtual implementation in their sense is impossible. Let us explain why. Of course, our SCF of interest still satisfies ex post incentive compatibility. The failure identified in BM (2009) concerns their robust measurability condition. For us, note that the vectors of interim expected utility written above still apply. In particular, two first-order types with a different payoff type could have positive affine collinear vectors only when

\[ q' - q = \frac{1}{2\gamma} \]

Therefore, if the set of first-order beliefs \( Q_i \) excludes these first-order belief pairs, the environment satisfies FOTD, and every SCF satisfies AM measurability. It follows that the failure of robust measurability is due only to the presence of such “non-generic” pairs of first-order types. That is, even in a model with a continuum of first-order types, violations of FOTD are restricted to a set of measure 0, and thus a robust version of measurability is a trivial condition, satisfied by all SCFs, if one imposes it over a full measure set of first-order types. Subsection 4.3 and the Appendix A elaborate on this.

Having said that, in conjunction with other standard conditions, such as convexity of the set of first-order beliefs, FOTD imposes strong separation requirements, as it implies that the lowest value \( q' \) for the interval of first-order beliefs accompanying payoff type 1 must be at a distance from the highest value \( q \) of the allowed first-order belief for payoff type 0 of at least \( 1/(2\gamma) \), implying something akin to the “belief-determined preferences” assumption that Neeman (2004) rightly criticizes. In this sense, while FOTD is still generic in continuum settings, it is much less appealing in some of them.

4.2 The Example beyond First-Order Type Diversity

In this subsection we address what happens in the example in environments that do not satisfy FOTD. Again, to simplify our expressions, suppose that \( n = 3 \). Then, the relevant range for \( \gamma \) is \([1/2, 1]\). For an environment to violate FOTD, recall that, for at least one pair of first-order types present in the model, \( q' - q = 1/(2\gamma) \), where \( q \) (\( q' \)) represents the

---

18Recall that for ease of presentation, we are writing our expressions for \( n = 3 \). The general condition here is \( 1 \geq \gamma \geq 1/(n - 1) \). A similar comment applies to the previous paragraph, for which the general condition is \( \gamma \in (0, 1/(n - 1)) \).
Thus, these two first-order types would self-reveal themselves if offered the choice between
preferences over that class. However, if we allow SCFs to depend on reports of agent
these types cannot be separated by using constant SCFs, as they have the same interim
payoff types of an agent \( i \neq j \) can be separated in the second round of the algorithm.
Consider a pair of first-order types of agent \( i \), \((0, q), (1, q')\), such that \( q' = 1/(2\gamma) \).
These types cannot be separated by using constant SCFs, as they have the same interim
assumptions, be separated if one goes only one step further in the AM measurability
algorithm. First, suppose that there is an agent \( j \) whose payoff types are fully separated in
of agent \( i \) does not contain \( q, q' \) such that \( q' = 1/(2\gamma) \)).
We will show that all payoff types of an agent \( i \neq j \) can be separated in the second round of the algorithm.
We claim that even these “non-generic” pairs of first-order types may, under some
We have shown above how to construct SCFs that separate first-order types \((0, q)\) and \((1, q')\) prefers
\( y_i \) to \( x_i \), we compute interim utilities of these two types:

\[
\begin{align*}
V_i(x_i | (0, q)) &= 1/2(1 - q)(1 - q) \times 2\gamma + q \times \gamma \\
V_i(y_i | (0, q)) &= q(1 - q)\gamma.
\end{align*}
\]

\[
\begin{align*}
V_i(y_i | (1, q')) &= q'[1 - q')(1 + \gamma) + q'] \\
V_i(x_i | (0, q)) &\equiv 1/2(1 - q')(1 - q') \times 2\gamma + q'(1 + \gamma).
\end{align*}
\]

Note that, as \( q' = 1/(2\gamma) \) and \( \gamma \in [1/2, 1) \), it follows that \( q' > 1/2 > 1 - q' \) and
Note that \( x_i \) and \( y_i \) separate any pair of first-order types such that \( q' = 1/(2\gamma) \).
Therefore, as long as there exists an agent (such as \( j \)) whose finest partition is reached in the
first round of the algorithm, all payoff types of every other agent can be separated in the
second round of the algorithm. In that case, the final partition of the measurability
algorithm is the finest partition of all singletons for every agent, even for \( \gamma \in [1/2, 1] \).
Therefore, robust virtual implementation is not restricted at all by robust measurability,
which becomes a trivial condition in our model: every SCF satisfies it.
We have shown above how to construct SCFs that separate first-order types \((0, q)\) and
\((1, q')\); let us now demonstrate how we construct SCFs that separate all first-order types of agent \( i \). For agent \( i \), one can construct a collection of constant SCFs \( x^1_i \in \{\ell_i(t_i)\}_{t_i} \)
to separate the different classes of equivalent first-order types in the first iteration of the
measurability algorithm. Further, one can find \( x^2_i \in \{x_i, y_i\} \) to separate the first-order
types that form non-singleton atoms of the partition. Then, an SCF that is measurable
with respect to \( T_i \times \Psi^1_i \), essentially \( (1 - \delta)x^1_i + \delta x^2_i \) that, for \( \delta > 0 \) small enough, will
separate all first-order types: because of the strict inequalities on the \( x^1_i \), the first-order
probability that an agent of payoff type \( \theta_i = 0 \) (\( \theta_i = 1 \)) believes another agent to be of the low payoff type.
We claim that even these “non-generic” pairs of first-order types may, under some
assumptions, be separated if one goes only one step further in the AM measurability
algorithm. First, suppose that there is an agent \( j \) whose payoff types are fully separated in
the first round of the algorithm (that is, \( Q_j \) does not contain \( q, q' \) such that \( q' = 1/(2\gamma) \)).
We will show that all payoff types of an agent \( i \neq j \) can be separated in the second round of the algorithm.
Consider a pair of first-order types of agent \( i \), \((0, q), (1, q')\), such that \( q' = 1/(2\gamma) \).
These types cannot be separated by using constant SCFs, as they have the same interim
preferences over that class. However, if we allow SCFs to depend on reports of agent \( j \),
these two types can be separated. Let \( x_i \) be an SCF that gives the object to agent \( i \) for
free with probability 1/2 if \( \theta_i = 0, \theta_j = 1 \); with the rest of probability and in all other
cases it gives the object to \( k \neq i, j \). Similarly, \( y_i \) gives the object for free to agent \( i \) if
\( \theta_i = 1, \theta_j = 0 \); in all other cases it gives the object to \( k \neq i, j \). Note that these SCFs
satisfy AM measurability, because \( T_j \) is partitioned in all singletons after the first round of
the measurability algorithm. To show that first-order type \((0, q)\) prefers \( x_i \) to \( y_i \) and \((1, q')\)
prefer \( y_i \) to \( x_i \), we compute interim utilities of these two types:

\[
\begin{align*}
V_i(x_i | (0, q)) &= 1/2(1 - q)(1 - q) \times 2\gamma + q \times \gamma \\
V_i(y_i | (0, q)) &= q(1 - q)\gamma.
\end{align*}
\]

\[
\begin{align*}
V_i(y_i | (1, q')) &= q'[1 - q'](1 + \gamma) + q'] \\
V_i(x_i | (0, q)) &\equiv 1/2(1 - q')(1 - q') \times 2\gamma + q'(1 + \gamma).
\end{align*}
\]
types that are separated in the first iteration of the algorithm stick to truth-telling for small enough \( \delta \). For the rest, each pair of first-order types that form an atom in the partition have identical preferences over constant SCFs (\( x_1^t \) are constant – each such first-order type will choose their most preferred SCF from this set of functions). These types are separated by the \( x_2^t \), as shown in the above argument.

Let us now turn to the case where \( Q_i \) of every agent \( i \) has first-order beliefs \( q' - q = 1/(2\gamma) \). In that case, the measurability algorithm stops in the first round and separation is impossible. The AM measurability would then require that SCFs to be implemented must be constant across \((0, q)\) and \((1, q')\). As the SCF depends only on payoff types, this implies that only constant SCFs are robustly virtually implementable.

The reason for this lack of separation is easy to see. We do not impose any restrictions on second-order beliefs in the paper. In particular, these “non-generic” first-order type pairs \((0, q), (1, q')\) of agent \( i \) may believe that first-order types of agents \( j, k \) are always either \((0, q_l)\) or \((1, q'_l)\) with \( q'_l - q_l = 1/(2\gamma) \), for \( l = j, k \). Such a belief does not violate any assumptions on the environment, as long as agent \( i \) of type \((0, q)\) \((1, q')\) believes agent \( j \) or \( k \) is \((0, q)\) with probability \( q \) \((q')\). The pairs \((0, q), (1, q')\) are not separable in the first round of the algorithm and form elements \( \psi_j \in \Psi^1_j \) and \( \psi_k \in \Psi^1_k \) of partitions of \( T_j, T_k \).

SCFs that separate \((0, q), (1, q')\) need to be measurable with respect to the partitions \( \Psi^1_j \times \Psi^1_k \). It then implies that separating SCFs in the second round of the algorithm are constant on \( \{ \psi_j, \psi_k \} \). As agent \( i \) assigns probability 1 on first-order types of \( j, k \neq i \) being in \( \psi_j, \psi_k \), a variation of SCFs outside of \( \{ \psi_j, \psi_k \} \) is irrelevant. Thus, \((0, q), (1, q')\) would need to be separated by constant SCFs, but this is impossible as these types were not separated in the first round. Hence, if second-order beliefs are unrestricted, robust virtual implementation is very limited. It can be shown that, by imposing some restrictions on the second-order beliefs, such limitations can be removed. We shall skip the details, as such restrictions on second-order beliefs are foreign to the paper.

### 4.3 Genercity of First-Order Type Diversity

In this subsection we abandon the example and make a more general point. The result in BM (2009) can be understood as uncovering robust measurability as an additional restriction for robust virtual implementation beyond ex post incentive compatibility. Recall that, given a class of environments indexed by different type spaces, robust measurability amounts to AM measurability on every type space in the class. We shall provide here an argument of genericity of the FOTD assumption when the sets \( Q_i \) are finite.\(^{20}\)

Recall that \( A \) is a finite set consisting of \( K \) alternatives, and recall our definition of the first-order interim utility \( V_i(t_i) = (V^K_i(t_{ij}))_{j=1,\ldots,K} \). Let \( K \geq 3 \) for this subsection (if \( K = 2 \), a violation of first-order TD happens when ordinal preferences are the same across types, a property that is certainly preserved for small perturbations).

Let \( V_i : \Theta_i \times \Delta(\Theta_{-i}) \to \mathbb{R}^K \) be an agent \( i \)’s first-order expected utilities over all constant SCFs. For each first-order type \( t_i = (\theta_i, q_i) \), assume there exist two alternatives \( a_k, a_k' \in A \) such that \( V^K_i(t_i) < V^K_i(t_i) \), and choosing one such pair of alternatives with extreme values, normalize expected utilities so that \( V^K_i(\theta, q_i) = 0 \), \( V^K_i(\theta, q_i) = 1 \), and \( V_i(\theta_i, q_i) \in [0,1]^K \).

\(^{19}\)For notational simplicity, we shall use below the same values of \( q, q' \) for agents \( i, j, k \).

\(^{20}\)A similar genericity argument can be provided for the infinite case; see the Appendix A and also Kunimoto and Serrano (2010).
Let \(|T_i|\) denote the cardinality of the set of first-order types for agent \(i\). Call \(S = \sum_{i \in N} |T_i|\). With this notation, one can associate a normalized environment \(\mathcal{E}\) with a point on \(\Omega\), the unit cube in \(\mathbb{R}^{(K-2)S}\) with vertices at the points \((0, \ldots, 0)\) and \((1, \ldots, 1)\). Endow \(\Omega\) with the uniform metric, and define open balls using this metric relative to \(\Omega\). Since the property of first-order TD is defined by a finite number of inequalities, one can easily see that the set of points in \(\Omega\) satisfying it is an open and dense subset of \(\Omega\). That is,

- for each environment in \(\Omega\) that satisfies first-order TD, there exists an open ball around it containing only the environments in which the property is maintained, and
- for each environment in \(\Omega\) that violates first-order TD and for each open ball around it, there always exists an environment satisfying first-order TD in that ball.

Suppose therefore that the planner does not know which payoff types or first-order types will be chosen by nature, i.e., which point in \(\Omega\) will be chosen, and suppose she can specify an ex-ante probability measure over such nature choices. The assumption that she can confine herself to \(\Omega\) uses the innocuous normalization of expected utilities and assumes further that she knows that she will be dealing only with “finite worlds,” a finite number of payoff types for each agent and a finite set of possible first-order beliefs (perhaps due to complexity issues, in specifying payoffs and probabilities, agents stop after a finite number of decimals). It then follows from our Theorem 1 below and from the afore discussion that she will be able to robustly virtually implement any ex post incentive compatible SCF with ex-ante probability 1. In this sense, the robust measurability restriction is generically trivial in our settings.

We remark again that, while the genericity of FOTD continues to hold in the continuum, as the key is to rule out “rare” pairs of first-order types, in conjunction with other standard assumptions in the continuum, such as convexity of the set of first-order beliefs, FOTD may imply a strong association of payoff types and first-order beliefs. Nonetheless, if in the unrestricted continuum model, the planner is forced to sample at most a finite number of first-order beliefs, our finite model analysis applies, in which FOTD is much more compelling.

### 5 Sufficiency for Robust Virtual Implementation

So far, we have focused on necessary conditions for robust virtual implementation. In the process, we have identified first-order incentive compatibility and AM measurability as relevant conditions. In the previous section, we have argued that AM measurability is generically a trivial condition. We shall now establish sufficiency results for robust virtual implementation. To do so, we introduce the assumption that the set of first-order beliefs for every agent is finite. This finiteness will be maintained in the rest of the paper, and we use it to obtain more transparent arguments; we shall comment on how to relax it after Theorem 4.\(^{21}\)

While not required for necessity arguments obtained so far, for the sufficiency results of the paper we maintain the following assumption on environments:

\(^{21}\)We may motivate the finiteness of the set \(Q\), simply because there is a limit on how fine the beliefs could be. Alternatively, we simply take a discrete approximation of the unrestricted set of first-order beliefs.
Lemma 6 Suppose 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 $\theta \in \Theta$, 
\begin{enumerate}
    \item $u_i(\bar{a}_i; \theta) > u_i(\underline{a}_i; \theta)$ for any $i \in N$;
    \item $u_i(\bar{a}_i; \theta) \geq u_i(\bar{a}_i; \theta)$ for any $i, i' \in N$ with $i \neq i'$.
\end{enumerate}

Remark: This is an exact analogue of AM's (1992c) Assumption 2. This assumption allows the agents to (partially) transfer their utilities among them. By making this assumption, we essentially postulate that $A$ includes a numeraire, which can be transferred across agents. Moreover, this assumption cannot be completely dispensed with as long as we seek for implementation by finite mechanisms. See Kunimoto and Serrano (2011) for the detail.

From the linearity of expected utility and since $\Theta$ is finite, we obtain the following lemma:

Lemma 5 Suppose that an environment $\mathcal{E}$ satisfies quasi-transferability. Then, there exists $\bar{\eta} > 0$ such that for any $\eta \in (0, \bar{\eta}]$, there exists a collection of lotteries $\{\bar{a}_i(\eta)\}_{i \in N}$ and $\{\underline{a}_i(\eta)\}_{i \in N}$ in $\Delta(A)$ such that for any $\theta \in \Theta$, 
\begin{enumerate}
    \item $0 < u_i(\bar{a}_i; \theta) - u_i(\underline{a}_i; \theta) \leq \eta$ for any $i \in N$;
    \item $u_i(\bar{a}_i; \theta) \geq u_i(\bar{a}_i; \theta)$ for any $i, i' \in N$ with $i \neq i'$.
\end{enumerate}

Proof: Since $\Theta$ is finite, we can define
\[ \bar{\eta} \equiv \max_{i \in N} \max_{\theta \in \Theta} u_i(\bar{a}_i; \theta) - u_i(\underline{a}_i; \theta) > 0. \]

For any $\eta \in (0, \bar{\eta}]$ and any $i \in N$, define the following:
\[ \bar{a}_i(\eta) = \frac{\eta}{\bar{\eta}} \bar{a}_i + \left(1 - \frac{\eta}{\bar{\eta}}\right) \bar{x} \]
\[ \underline{a}_i(\eta) = \frac{\eta}{\bar{\eta}} \underline{a}_i + \left(1 - \frac{\eta}{\bar{\eta}}\right) \bar{x}, \]
where $\bar{x}$ is the uniform lottery over $A$. It is immediate to check that these lotteries satisfy the inequalities in the statement. ■

5.1 A Preliminary Separation Lemma

In environments satisfying quasi-transferability, FOTD and the finiteness of $Q_i$, we next show the following critical lemma, arguably a generalization of Lemma 1 in Serrano and Vohra (2005).

Lemma 6 Suppose an environment $\mathcal{E}$ satisfies quasi-transferability, FOTD and that for all $i \in N$ the set $Q_i$ is finite. Then there exist constant SCFs $\left(\ell_i(\theta_i, q_i)\right)_{\theta_i \in \Theta, q_i \in Q_i}$ such that for every $i \in N$ and $t_i = (\theta_i, q_i), t_i' = (\theta_i', q_i') \in T_i$ with $\theta_i \neq \theta_i'$,
\[ V_i(\ell_i(\theta_i, q_i)|\theta_i, q_i) > V_i(\ell_i(\theta_i', q_i)|\theta_i, q_i). \]
we have been able to find which satisfy (4). In the latter case, it is clear that we can choose two constant SCFs assigning equal probability to each alternative in \( A \), i.e., \( \bar{x}(t) = (1/K, \ldots, 1/K) \) for all \( t \in T \). We will use induction on the number of first-order types of agent \( i \).

First, we show that for \( i \in N \), and for two first-order types \( t_i = (\theta_i, q_i) \), \( t_i' = (\theta_i', q_i') \in T_i \) with \( \theta_i \neq \theta_i' \), there exist constant SCFs \( x \) and \( x' \), close to \( \bar{x} \), such that
\[
V_i(x|t_i) > V_i(x'|t_i) \quad \text{and} \quad V_i(x'|t_i') > V_i(x|t_i').
\] (4)

The interim indifference curve of agent \( i \) of first-order type \( t_i = (\theta_i, q_i) \) through \( \bar{x} \) is described by a hyperplane, \( H \), in \( \mathbb{R}^{K-1}_+ \):
\[
H = \left\{ (x_1, \ldots, x_{K-1}) \in \mathbb{R}^{K-1}_+ \left| \sum_{k=1}^{K-1} p_k(t_i)x_k = \bar{u} \right. \right\},
\]
where \( p_k(t_i) = (V^K_i(t_i) - V^K_i(t_i')) \) for \( k = 1, \ldots, K - 1 \).

Let \( p(t_i) = (p_1(t_i), \ldots, p_{K-1}(t_i)) \in \mathbb{R}^{K-1}_+ \). Consider the interim indifference hyperplane through \( \bar{x} \) of agent \( i \) of first-order type \( t_i' = (\theta_i', q_i') \) where \( \theta_i \neq \theta_i' \):
\[
H' = \left\{ (x_1, \ldots, x_{K-1}) \in \mathbb{R}^{K-1}_+ \left| \sum_{k=1}^{K-1} p_k(t_i')x_k = \bar{u}' \right. \right\}
\]

Given quasi-transferability, we must have \( p(t_i) \neq 0 \) and \( p(t_i') \neq 0 \). We claim that \( p(t_i) \neq cp(t_i') \) for any \( c > 0 \). Suppose not; that is, there is \( c > 0 \) such that \( p(t_i) = cp(t_i') \). This implies that \( V_i(t_i) = cV_i(t_i') + \gamma \), which contradicts FOTD. Thus, either \( p(t_i) = cp(t_i') \) where \( c < 0 \) or there does not exist \( c \neq 0 \) such that \( p(t_i) = cp(t_i') \). In the former case, it is easy to see (using quasi-transferability) that any point which lies above \( H \) must be below \( H' \) and, choosing two points (one above \( H \) and one below it) close to \( \bar{x} \), one finds constant SCFs which satisfy (4). In the latter case, it is clear that we can choose two constant SCFs which satisfy (4).

Recall that \( Q_i \) is assumed to be finite. So, let us state the following induction hypothesis: suppose that for the first \( |T_i| - 1 \) first-order types of agent \( i \), i.e., for all \( t_i \in T_i \setminus \{ t_i^0 \} \), we have been able to find \( |T_i| - 1 \) constant SCFs near \( \bar{x} \), say \( x(t_i) \), such that for every \( t_i = (\theta_i, q_i) \in T_i \setminus \{ t_i^0 \} \), \( V_i(x(t_i)|t_i) > V_i(x(t_i')|t_i) \) for every \( t_i' = (\theta_i', q_i') \in T_i \setminus \{ t_i^0, t_i \} \) with \( \theta_i \neq \theta_i' \). Consider first-order type \( t_i^0 \). Choose the constant SCF among the collection \( (x(t_i))_{t_i \in T_i \setminus \{ t_i^0 \}} \) that is ranked highest by first-order type \( t_i^0 \) (without loss of generality, there is only one). Call it \( x(t_i) \). By arguments similar to the ones in the previous paragraph, because of quasi-transferability and FOTD, one can find a constant SCF near \( x(t_i) \), call it \( x(t_i^0) \), such that first-order types \( t_i = (\theta_i, q_i) \) and \( t_i^0 = (\theta_i^0, q_i^0) \) with \( \theta_i \neq \theta_i^0 \) satisfy (4). To
do this, we construct a set of new constant SCFs \( \ell_i[t_i] \) satisfying (4) as follows. Consider compounding lotteries \( (x[t_i])_{t_i \in T_i \setminus \{t_i\}} \) and \( x[t_i] \) (to obtain \( (\ell_i[t_i])_{t_i \in T_i \setminus \{t_i\}} \)) and \( x[t_i^0] \) (to obtain \( \ell_i[t_i^0] \)). Since all inequalities concerning \( (x[t_i])_{t_i \in T_i \setminus \{t_i\}} \) are strict and \( Q_i \) is finite, the weights can be chosen so that the collection of constant SCFs \( (\ell_i(t_i))_{t_i \in T_i} \) satisfy all the inequalities in the statement of the lemma, so the proof is complete. \(^{22}\)

5.2 A Robust Canonical Mechanism

This subsection introduces a mechanism that will be used to obtain a robust virtual implementation result. The mechanism is finite, and hence best replies are always well defined.

The mechanism \( \Gamma = (M, g) \) uses the collection of constant SCFs \( \ell_i \) of Lemma 6. \(^{23}\) The construction is as follows: Every agent \( i \) makes \( (J + 1) \) simultaneous announcements (where \( J \) is to be determined in the proof), each of which is of his own first-order type

\[
M_i = M_i^0 \times M_i^1 \times \cdots \times M_i^J = T_i^0 \times T_i \times \cdots \times T_i^{J+1}.
\]

Denote

\[
m_i = (m_i^0, \ldots, m_i^J) \in M_i, \quad m_i^s \in M_i^s \quad \forall s = 0, \ldots, J
\]

\[
m = (m^0, \ldots, m^J) \in M, \quad m^s = (m_i^s)_{i \in N} \in M^s = \times_{i \in N} M_i^s
\]

\[
m^s = (m_0^s, m_Q^s), \quad m_0^s \in \Theta, \quad m_Q^s \in Q, \quad \forall s = 0, \ldots, J
\]

Using Lemma 5, we introduce the following punishment lottery to reward a consistent announcement from each agent:

\[
\xi(i, m) = \begin{cases} 
    a_i(\eta) & \text{if } \exists j \in \{1, \ldots, J\} \text{ s.t. } m_i^j \neq m_i^0 \text{ for } i \in N \\
    \text{and } m^s = m^0 \forall s \in \{0, \ldots, j - 1\} \\
    \bar{a}_i(\eta) & \text{otherwise}
\end{cases}
\]

Thus, the punishment lottery punishes any player who, in the announcement rounds 1, 2, \ldots, \( J \), first deviates from the announcement that he makes in round 0. We shall choose the parameters of the mechanism so that it works for any arbitrarily small fine \( \eta > 0 \).

Define \( \ell : T \to \Delta(A) \) as follows:

\[
\ell(t) = \frac{1}{n} \sum_{i \in N} \ell_i(t_i),
\]

\(^{22}\)If \( A \) is a separable metric space, the modification we must make to the previous argument is the way we define the lottery \( \bar{x}(t) \):

\[
\bar{x}(t) = (\bar{x}_k(t))_{k=1}^{\infty}
\]

where \( \bar{x}_k(t) = (1 - \delta) \delta^{k-1} \), and \( 0 < \delta < 1 \).

\(^{23}\)It is inspired by the heuristic section of AM (1992c) that precedes their formal analysis. We dispense with one important assumption made there: private values. The reader is referred to footnote 5 in the introduction for the differences from AM (1992c).
where \( \{\ell_i(t_i)\}_{t_i \in T, i \in N} \) are constructed as in Lemma 6. Given an SCF \( f \), for any profile of agents’ messages \( m \), the outcome function of the mechanism is

\[
g(m) = \varepsilon \ell(m^0) + \frac{\varepsilon^2}{n} \sum_{i \in N} \xi(i, m) + \frac{1 - \varepsilon - \varepsilon^2}{J} \sum_{s=1}^{J} \{\varepsilon^2 \ell(m^s) + (1 - \varepsilon^2) f(m^s)\}
\]

where \( \varepsilon > 0 \) will be chosen small enough and \( J \) large enough.

This outcome function has three terms: the first, weighted by a probability of \( \varepsilon \), depends only on \( m^0 \) and consists of the SCFs from Lemma 6 that induce the appropriate separation of first-order types; the second, weighted by \( \varepsilon^2 \), is the punishment lottery we have just constructed; the third term, having the remaining weight, depends on the rest of the announcements \( m^1, \ldots, m^J \) and consists of the (slightly modified) SCF \( f \) being implemented.

Note how the mechanism performs two main roles. The first concerns information revelation, and it is accomplished through the first term of the outcome function thanks to FOTD. If the planner just wanted to implement the SCF’s \( \ell_i \), she could rely on FOTD to succeed: this would be similar to a private values environment.\(^{24}\) However, the SCF of interest is \( f \). Therefore, the second role of the outcome function is the approximation of \( f \), which in principle may allow for complex interdependences, making it manipulable beyond private values considerations. As will be shown, the mechanism suggested will accomplish these different goals. In particular, the introduction of the small penalty \( \eta \), through an induction argument, will have large incentive effects and the desired \( f \) will approximately obtain, and all this will be done regardless of complexities concerning higher-order beliefs over types.

Equivalently, the outcome function can be expressed as

\[
g(m) = \varepsilon \ell(m^0) + \frac{\varepsilon^2}{n} \sum_{i \in N} \xi(i, m) + \frac{1 - \varepsilon - \varepsilon^2}{J} \sum_{s=1}^{J} \tilde{f}(m^s),
\]

where

\[
\tilde{f}(m^s) = \varepsilon^2 \ell(m^s) + (1 - \varepsilon^2) f(m^s).
\]

Note that if \( f : \Theta \rightarrow \Delta(A) \) satisfies first-order incentive compatibility, \( \tilde{f} : T \rightarrow \Delta(A) \) satisfies strict first-order incentive compatibility. This is because of the addition of the \( \ell_i \) terms constructed in Lemma 6. Also, \( \tilde{f} \) is close to \( f \) for small \( \varepsilon > 0 \).

In closing our comments about the mechanism, we note an important difference with the analysis of AM (1992c). Because of the way the FOTD property is stated, it is not required that each first-order type reports truthfully its identity; it suffices for each first-order type to report truthfully its payoff type.

---

\(^{24}\)The reader should be warned that it would be incorrect to use a first-order interim utility function \( V_i \) as a private values type. There is certainly loss of information if we reduce our setup to this “private values” environment of the first-order interim utility functions, as agents would not be able to evaluate non-constant SCFs.
5.3 Sufficiency under first order type diversity

**Theorem 3** Suppose an environment $E$ satisfies FOTD, quasi-transferability, and that $Q_i$ is finite for every $i \in N$. If an SCF $f$ satisfies first-order incentive compatibility, it is robustly virtually implementable.

**Proof:** The proof consists of two claims using the mechanism of the previous subsection.

**Claim 3.1:** For every agent $i \in N$ and every $\theta_i \in \Theta_i$, if $m_i \in S_i^\Omega(\theta_i)$, then there exists some $q_i \in Q_i$ such that $m_i^0 = (\theta_i, q_i) \in T_i$.

**Proof of Claim 3.1:** We begin by noting Fact 1:

**Fact 1:** For any $\eta > 0$, we can choose $\varepsilon > 0$ small enough so that

$$\min_{i \in N, t_i = (\theta_i, q_i), t_i' = (\theta'_i, q'_i) \in T_i, \theta'_i \neq \theta_i} \{V_i(\ell_i(t_i)|t_i) - V_i(\ell_i(t'_i)|t_i)\} > \varepsilon\eta$$

This fact follows because $N$ is finite and $T_i$ is finite for every $i \in N$.

Then, by our choice of $\varepsilon$ in Fact 1, given an arbitrary $\eta$, for any $i \in N$, we have

$$\frac{\varepsilon}{n} \{V_i(\ell_i(t_i)|t_i) - V_i(\ell_i(t'_i)|t_i)\} > \frac{\varepsilon^2}{n} \eta \quad \forall \ t_i = (\theta_i, q_i), t'_i = (\theta'_i, q'_i) \in T_i, \theta'_i \neq \theta_i.$$

Recall the outcome function of the mechanism, and notice that announcement $m_i^0$ affects only the first term and possibly the second through the punishment lottery. According to the last inequality, the payoff loss from misreporting one’s first-order type in $m_i^0$ exceeds the maximum possible gain from the second term, whatever strategies are used by the other agents. Thus, player $i$ will be strictly better off by reporting the true payoff type in the 0-th announcement, even if he were to misrepresent the rest of his announcements.

Formally, we argue by contradiction. Fix agent $i$ of first-order type $t_i = (\theta_i, q_i)$. Let $m_i$ be a message sent by agent $i$ such that $m_i^0 = t'_i = (\theta'_i, q'_i)$ where $\theta'_i \neq \theta_i$.

Define $\hat{m}_i$ as follows:

$$\hat{m}_i^s = m_i^s \ \forall s \geq 1,$$

$$\hat{m}_i^0 = (\theta_i, q_i) \in T_i.$$

We compare below the interim utilities of agent $i$ of first-order type $t_i = (\theta_i, q_i)$ when he employs $m_i$ and $\hat{m}_i$ against any $\mu_i \in \Delta^N(\Theta_{-i} \times M_{-i})$:

$$\sum_{\theta_{-i}, m_{-i}} \mu_i(\theta_{-i}, m_{-i}) \left[u_i(g(\hat{m}_i, m_{-i}); \theta_i, \theta_{-i}) - u_i(g(m_i, m_{-i}); \theta_i, \theta_{-i})\right]$$

$$= \frac{\varepsilon}{n} \{V_i(\ell_i(\theta_i, q_i)|\theta_i, q_i) - V_i(\ell_i(\theta'_i, q'_i)|\theta_i, q_i)\}$$

$$+ \frac{\varepsilon^2}{n} \sum_{\theta_{-i}, m_{-i}} \mu_i(\theta_{-i}, m_{-i}) \sum_{i' \in N} \left\{u_i(\xi(i', \hat{m}_i, m_{-i}); \theta_i, \theta_{-i}) - u_i(\xi(i', m_i, m_{-i}); \theta_i, \theta_{-i})\right\}$$

$$> 0 \quad \text{(: Fact 1 and quasi-transferability)}$$

Thus, $m_i \notin S_i^\Omega(\theta_i)$, which is a contradiction. ■
Claim 3.2: Let \( m \in S^T(\theta) \). Suppose that for all \( i \in N \) and all \( \theta_i \in \Theta_i \) there exist \( q_i \in Q_i \) such that \( m_i^s = (\theta_i, q_i) \) for all \( s \in \{0, \ldots, j\} \), where \( 0 \leq j \leq J - 1 \). Then

\[
m_i^{j+1} = m_i^0 = (\theta_i, q_i) \quad \forall i \in N.
\]

**Proof of Claim 3.2:** We need some additional pieces of notation for the proof. Consider any profile of functions, \( \alpha = (\alpha_i)_{i \in N} \), where \( \alpha_i : T_i \rightarrow T_1 \). Consider the extended SCF \( \tilde{f} : T \rightarrow \Delta(A) \), which is constructed in the mechanism \( \Gamma \) and a first-order type \( t_i \in T_i \). Let \( \tilde{f} \circ \alpha(t) = \tilde{f}(\alpha(t)) \) for all \( t \in T \). Define the following:

\[
\gamma_i(t_i) \equiv \max_{\alpha} V_i(\tilde{f} \circ \alpha | t_i) - \min_{\alpha} V_i(\tilde{f} \circ \alpha | t_i)
\]

\[
\gamma_i(t_i) \equiv \max_{t_i \in T_i} \gamma_i(t_i)
\]

\[
\gamma \equiv \max_{i \in N} \gamma_i > 0
\]

The number \( \gamma \) is well defined because \( T_i \) is finite for every \( i \in N \) and because \( \tilde{f} \) satisfies strict first-order incentive compatibility.

Consider agent \( i \) of first-order type \( (\theta_i, q_i) \). Suppose, by way of contradiction, that \( m_i^{j+1} = (\theta'_i, q'_i) \) with \( \theta'_i \neq \theta_i \).

Define \( \bar{m}_i \) such that

\[
\bar{m}_i^s = m_i^s \quad \forall s \neq j + 1,
\]

and

\[
\bar{m}_i^{j+1} = (\theta_i, q_i).
\]

We then claim that for any \( \mu_i \in \Delta^q(\Theta_i \times M_{-i}) \),

\[
\sum_{\theta_{-i}, m_{-i}} \mu_i(\theta_{-i}, m_{-i}) [u_i(g(\bar{m}_i, m_{-i}); \theta_i, \theta_{-i}) - u_i(g(m_i, m_{-i}); \theta_i, \theta_{-i})] > 0.
\]

Fix \( \mu_i \in \Delta^q(\Theta_i \times M_{-i}) \). Suppose that for any \( \theta_{-i} \) and \( m_{-i} \), if \( \mu_i(\theta_{-i}, m_{-i}) > 0 \), then \( m_i^{j+1} = (\theta_{-i}, q_{-i}) \in \Theta_i \times Q_{-i} \) for some \( q_{-i} \). Then, by strict incentive compatibility of \( \tilde{f} \) and the induction hypothesis, \( \bar{m}_i \) yields higher payoff than \( m_i \) in the \( j + 1 \)-st term of the third part of the outcome function. In addition, the punishment second term cannot get worse by using \( \bar{m}_i \) instead of \( m_i \). This is because quasi-transferability guarantees that any other agent’s punishment in a subsequent round does not hurt agent \( i \) at all. Indeed, here agent \( i \) does not have an incentive to misreport his payoff type in round \( j + 1 \) in order to preempt other agents’ punishment in a subsequent round.\(^{25}\) Thus, in this case, \( \bar{m}_i \) has a higher expected payoff than \( m_i \). Therefore, \( m_i \notin S_i^T(\theta_i) \), which is a contradiction.

On the other hand, suppose that there exists \( (\theta_{-i}, m_{-i}) \) with \( \mu_i(\theta_{-i}, m_{-i}) > 0 \) such that there exist agent \( i' \neq i \) and \( (\theta'_{i'}, q'_{i'}) \) for whom \( m_i^{j+1} = (\theta_{-i}, q_{-i}) \in \Theta_i \times Q_{-i} \) for some \( q_{-i} \). Then, by construction of \( \gamma \), for any \( \mu_i \in \Delta^q(\Theta_i \times M_{-i}) \) under the induction hypothesis, we have

\[
\gamma \geq \sum_{\theta_{-i}, m_{-i}} \mu_i(\theta_{-i}, m_{-i}) \left[ u_i(\tilde{f}(m_i^{j+1}, m_{-i}^{j+1}); \theta_i, \theta_{-i}) - u_i(\tilde{f}(\bar{m}_i^{j+1}, m_{-i}^{j+1}); \theta_i, \theta_{-i}) \right].
\]

Next, we note Fact 2:

\(^{25}\)A similar comment applies to the parallel step in the proof of Claim 4.2.
Fact 2: We can choose $J$ large enough so that

$$\frac{\varepsilon^2}{n\eta} > \frac{1 - \varepsilon - \varepsilon^2}{J}\gamma.$$ 

Then, using Fact 2, we have

$$\frac{\varepsilon^2}{n\eta} > \frac{1 - \varepsilon - \varepsilon^2}{J}\gamma$$

$$\geq \frac{1 - \varepsilon - \varepsilon^2}{J} \left\{ \sum_{\theta_{-i},m_{-i}} \mu_i(\theta_{-i}, m_{-i}) \left[ u_i(\tilde{f}(m_{-i}^{j,i+1}, m_{-i}^{j,i+1}); \theta_i, \theta_{-i}) - u_i(\tilde{f}(\tilde{m}_{-i}^{j,i+1}, m_{-i}^{j,i+1}); \theta_i, \theta_{-i}) \right] \right\}.$$ 

Then, by improving his payoff in the punishment term, $\tilde{m}_i$ yields higher payoff than $m_i$. That is, for any $\mu_i \in \Delta^q(\Theta_{-i} \times M_{-i})$ under the induction hypothesis, we have

$$\sum_{\theta_{-i},m_{-i}} \mu_i(\theta_{-i}, m_{-i}) \left[ u_i(g(\tilde{m}_i, m_{-i}); \theta_i, \theta_{-i}) - u_i(g(m_i, m_{-i}); \theta_i, m_{-i}) \right] > 0.$$ 

In other words, under the induction hypothesis, it is always better for agent $i$ of payoff type $\theta_i$ to wait for one more round to misrepresented his payoff type so that other players misrepresented their payoff types first, thereby avoiding the punishment involved in the second term of the outcome function. This, however, contradicts our hypothesis that $m_i \in S^r_i(\theta_i)$.

Claims 3.1 and 3.2 together show that for any $\theta \in \Theta$, any message profile $m \in S^r(\theta)$ surviving the iterative elimination of never best responses has the property that there exists $q \in Q$ such that $m^s = (\theta, q) \in \Theta \times Q$ for every $s \in \{0, 1, \ldots, J\}$. The resulting outcome is

$$(1 - \varepsilon^2)(1 - \varepsilon - \varepsilon^2) f(\theta) + \frac{\varepsilon + (1 - \varepsilon - \varepsilon^2)\varepsilon^2}{n} \sum_{i \in N} \ell_i(\theta_i, q_i) + \varepsilon^2 \bar{x}.$$ 

In addition, no monetary penalties $\eta$ are levied. This outcome is arbitrarily close to $f(\theta)$ for every $\theta \in \Theta$ when $\varepsilon > 0$ is chosen to be small enough. This completes the proof of Theorem 3.

6 A Characterization of Robust Virtual Implementation

When we introduced quasi-transferability, we have noted that it is, effectively, an indispensable restriction on the environment. In its absence, an additional necessary condition crops up (Kunimoto and Serrano (2011)). The characterization we obtain in this section is for quasi-transferable environments.

For a fixed type space, AM (1992c) establish that AM measurability and incentive compatibility are necessary and sufficient for virtual implementation in iteratively undominated strategies. Following their steps, we establish the robust analog of that result in our environments satisfying that the sets $Q_i$ are finite. We shall remark on the general unrestricted case of $Q_i$’s after the proof of Theorem 4.
Given our results so far, we know that AM measurability is generically a trivial condition, since it can be completely dispensed with in quasi-transferable environments satisfying FOTD. For the rest of environments, AM measurability imposes additional restrictions, and sometimes those restrictions are so severe that only constant SCFs can be virtually implemented (see Serrano and Vohra (2001), BM (2009)). We turn to details now.

Recall the recursive construction behind AM measurability, and, in particular, the partitions $\Psi_i^h$ for $i \in N$ and $h = 0, 1, \ldots$. For each $i \in N$, $t_i \in T_i$, and $h \geq 0$, let $\Psi_i^h(t_i)$ be the element of $\Psi_i^h$ that includes $t_i$.

As we will be using a mechanism similar to the one in Subsection 5.2, our initial task is to construct the first -- separating -- term of the outcome function. The next lemma provides SCFs that will help us separate first-order types, as allowed by the $h$-th iteration in the measurability construction. It is a generalization of Lemma 6.

Lemma 7 Suppose an environment $E$ satisfies quasi-transferability and that the sets $Q_i$ are finite. Then, for every $i \in N$ and every $h = 1, 2, \ldots, L$, there exist SCFs $x_i^h[\psi_i^h]: \Theta \rightarrow \Delta(A)$, which are measurable with respect to $\Psi_i^h \times \Psi_i^{h-1}$, and such that for every $t_i = (\theta_i, q_i) \in T_i$ and $\psi_i^h \in \Psi_i^h \setminus \Psi_i^1(t_i)$ for which there exists $t'_i = (\theta'_i, q'_i) \in \psi_i^h$ with $\theta_i \neq \theta'_i$,

$$V_i\left(x_i^h[\psi_i^h](t_i)|\theta_i, q_i\right) > V_i\left(x_i^h[\psi_i^h]; \psi_i^h|\theta_i, q_i\right),$$

where

$$V_i(x_i^h[\cdot]; |\theta_i, q_i) \equiv \sum_{\theta_{-i} \in \Theta_{-i}} q_i(\theta_{-i})u_i(x_i^h[\cdot](\theta_i, \theta_{-i}); \theta_i, \theta_{-i}).$$

Proof: Again we recall that $A$ is finite. Fix iteration $h$ in the AM measurability algorithm. Consider the SCF $\bar{x}^h$, which prescribes in each state the lottery $\bar{x}^h$, assigning equal probability to each SCF in $F(\Psi_i^h \times \Psi_{-i}^{h-1})$, the space of degenerate SCFs measurable with respect to $\Psi_i^h \times \Psi_{-i}^{h-1}$. That is,

$$\bar{x}^h(t) = \frac{1}{K^h}f^1(t) + \ldots + \frac{1}{K^h}f^{K^h}(t)$$

for all $t \in T$. Here, $|F(\Psi_i^h \times \Psi_{-i}^{h-1})| = K^h$. By construction, $\bar{x}^h$ is measurable with respect to $\Psi_i^h \times \Psi_{-i}^{h-1}$, and, abusing notation, we can write $\bar{x}^h(t) = \bar{x}^h(\Psi_i^h(t))$.  

We claim that for every $i \in N$, every $t_i, t'_i \in T_i$, with $\Psi_i^h(t_i) \neq \Psi_i^h(t'_i)$, there exist SCFs $x_i^h[\Psi_i^h(t_i)]$ and $x_i^h[\Psi_i^h(t'_i)]$ that are measurable with respect to $\Psi_i^h \times \Psi_{-i}^{h-1}$, close to $\bar{x}^h$, such that

$$V_i(x_i^h[\Psi_i^h(t_i)]|t_i) > V_i(x_i^h[\Psi_i^h(t'_i)]; t'_i|t_i) \text{ and } V_i(x_i^h[\Psi_i^h(t'_i)]|t'_i) > V_i(x_i^h[\Psi_i^h(t_i)]; t_i|t'_i).$$

We can prove this claim by using the same argument as in Lemma 6. That is, consider the $(K^h-1)$-dimensional unit simplex, whose extreme points are the elements of the functional space $F(\Psi_i^h \times \Psi_{-i}^{h-1})$. Note how the first-order interim expected utility of each extreme

\[\text{In fact, given the mechanism we construct below, in which agents report atoms of the partition } \psi_i^* \text{ and not first-order types, this will be a convenient way to write the argument of an extended SCF. Therefore, we shall use this repeatedly in the rest of this section.}\]
point is well defined for each first-order type, and thus, one can consider the corresponding hyperplanes as the level curves of such interim utility. By construction of the $h$-th iteration of measurability, first-order types $t_i$ and $t'_i$ can be separated in their interim preferences over SCFs in $F(\Psi^h_i \times \Psi^{h-1}_i)$ whenever $\Psi^h_i(t_i) \neq \Psi^h_i(t'_i)$. Then, one can find two SCFs to separate the two first-order types as written in (5). The rest of the argument is based on an induction step on the number of elements of $\Psi^h_i$, exactly as in the proof of Lemma 6. ■

We are now ready to state and prove the main result of this section:

**Theorem 4 (A Characterization of Robust Virtual Implementation)** Suppose an environment $E$ satisfies quasi-transferability and that the sets $Q_i$ are finite. An SCF $f$ is robustly virtually implementable if and only if it satisfies first-order incentive compatibility and AM measurability.

**Proof:** By Theorems 1 and 2, first-order incentive compatibility and AM measurability are necessary conditions (even for general environments). Under the finiteness of $Q_i$ and quasi-transferability, we shall now establish that they are also sufficient, by constructing a canonical implementing mechanism. We note that the construction of the canonical mechanism of this section is a generalization of that in Theorem 3 once we take into account that the measurability algorithm may not stop at the first step.

In the mechanism $\tilde{\Gamma}$, every agent $i$ makes $(J+1)$ simultaneous announcements; in each the agent announces an atom in the partition $\psi^*_i \in \Psi^*_i$:

$$M_i = M^0_i \times M^1_i \times \cdots \times M^J_i = \Psi^*_i \times \cdots \times \Psi^*_i$$

for an integer $J$ to be defined below. Correspondingly, the truthful $s$-th announcement for agent $i$ of first-order type $t_i$ is $m^s_i = \Psi^*_i(t_i)$.

Define $x : T \to \Delta(A)$ by

$$x(t) = \frac{\alpha}{n} \sum_{i \in N} \sum_{h=0}^L \delta^h x^h_i[\Psi^h_i(t_i)][\hat{\theta}(t)] \quad \forall t \in T$$

where $\alpha$ is defined as

$$\alpha \equiv \frac{1}{1 + \delta + \delta^2 + \cdots + \delta^L}.$$ 

and $x^h_i[\Psi^h_i(t_i)]$ are arbitrary constant SCFs for $h = 0$, and are as constructed in Lemma 7 for each $h > 0$; $0 < \delta < 1$.

We shall use the same punishment lotteries $\xi(i, m)$ as in the mechanism of Theorem 3.

For any $i \in N$, define also

$$\ell_i(t) = \alpha \sum_{h=0}^L \delta^h x^h_i[\Psi^h_i(t_i)][\hat{\theta}(t)]$$

for any $t \in T$.

Let the outcome function of the mechanism $\tilde{\Gamma}$ be $\tilde{g}$, defined as follows:

$$\tilde{g}(m) = \varepsilon x(m^0) + \frac{\varepsilon^2}{n} \sum_{i \in N} \xi(i, m) + \frac{(1 - \varepsilon - \varepsilon^2)}{J} \sum_{s=1}^J \tilde{f}(m^s)$$

30
For any Fact 3: the statement of Fact 3, construct an induction step on the number of announcements. Theorem 3, although it is somewhat more complicated. Specifically, the proof will require reporting $\Psi$ using a message that survives the iterative elimination of never best responses, he must be the canonical mechanism for each agent. This serves to establish that if each agent (unlike the $\ell$ induction with respect to $m$) exists.

Claim 4.1, a second induction argument is required, this time on $\epsilon, \delta$. Then, we note the following fact:

**Fact 3:** For any $\eta > 0$, we can choose $\varepsilon > 0$ small enough so that for any $h = 1, \ldots, L$,

$$
\frac{\varepsilon}{n} \alpha \min_{i \in N, t_i \in T_i, \psi_i^h \neq \Psi_i^h(t_i)} \left\{ \delta^h V_i(x_i^h[\Psi_i^h(t_i)]; \psi_i^h|t_i) - \delta^h V_i(x_i^h[\psi_i^h]; \psi_i^h|t_i) \right\} > \frac{\varepsilon^2}{n} \eta.
$$

Fix all of these variables, $\varepsilon, \delta$, and the monetary penalty $\eta$ at the specified levels.

The rest of the argument in the proof relies on two steps, as Claims 3.1 and 3.2 in Theorem 3, although it is somewhat more complicated. Specifically, the proof will require double use of mathematical induction. Claims 4.1 and 4.2 below, similar to Claims 3.1 and 3.2 of Theorem 3, construct an induction step on the number of announcements $j$ in the canonical mechanism for each agent. This serves to establish that if each agent $i$ is using a message that survives the iterative elimination of never best responses, he must be reporting $\Psi_i^h(t_i)$ ($J + 1$) times when his first-order type is $t_i$. However, to establish Claim 4.1, a second induction argument is required, this time on $h$, the rounds of iteration in the AM measurability algorithm. This is needed because the functions $x_i^h[\cdot]$ that are used to separate first-order types are not independent of the announcements made by others (unlike the $\ell_j$’s functions of Theorem 3). Now we proceed to complete the argument.

**Claim 4.1:** Suppose that $m \in S^T(\theta)$. Then, for any $i \in N$ and $h = 0, 1, \ldots, L$, we have $m_i^0 \subseteq \Psi_i^h(\theta_i, q_i)$ for some $q_i \in Q_i$. In other words, $m_i^0 = \Psi_i^h(\theta_i, q_i)$ with some $q_i \in Q_i$.

**Proof of Claim 4.1:** Consider agent $i$ of first-order type $(\theta_i, q_i)$. We prove this step by induction with respect to $h$. Suppose $h = 0$. Then, $\Psi_i^0(t_i) = T_i$ for any $t_i \in T_i$. Therefore, the statement $m_i^0(\theta_i) \subseteq \Psi_i^0(\theta_i, q_i)$ in Claim 4.1 is trivially satisfied.

Suppose that $m_i^0 \subseteq \Psi_i^h(\theta_i, q_i)$ for any $h \leq L - 1$. What we want to show is that $m_i^0 \subseteq \Psi_i^{L}(\theta_i, q_i)$, which equals $\Psi_i^{L}(\theta_i, q_i)$. Suppose, by way of contradiction, that $m_i^0 \subseteq \Psi_i^{L-1}(\theta_i, q_i)\setminus\Psi_i^{L}(\theta_i, q_i)$. Consider agent $i$’s alternative message $\tilde{m}_i$ with the following properties:

$$
m_i^j = \tilde{m}_i^j \forall j \geq 1,
m_i^0 = \Psi_i^0(\theta_i, q_i).
$$
With Lemma 7 in mind, we compare below the interim utilities of agent \( i \) of first-order type \((\theta_i,q_i)\) when he employs \( m_i \) and \( \tilde{m}_i \) against any \( \mu_i \in \Delta^q(\Theta_{-i} \times M_{-i}) \):

\[
\sum_{\theta_{-i},m_{-i}} \mu_i(\theta_{-i},m_{-i}) \{ u_i(g(\tilde{m}_i,m_{-i});\theta_i,\theta_{-i}) - u_i(g(m_i,m_{-i});\theta_i,\theta_{-i}) \} \\
= \alpha \frac{\delta L}{n} \left\{ V_i(x^L_i|\tilde{m}^0_i)|\theta_i,q_i| - V_i(x^L_i|m^0_i)|\theta_i,q_i| \right\} \\
+ \frac{\varepsilon^2}{n} \sum_{\theta_{-i},m_{-i}} \mu_i(\theta_{-i},m_{-i}) \sum_{i' \in N} \{ u_i(\xi(i',\tilde{m}_i,m_{-i});\theta_i,\theta_{-i}) - u_i(\xi(i',m_i,m_{-i});\theta_i,\theta_{-i}) \} \\
> 0 \quad (\because \text{Fact 3 and quasi-transferability}).
\]

This is because no \( x^i, h < L \), is affected by this message change and because for each \( i' \neq i \), \( x^L_{i'} \) is measurable with respect to \( \Psi_{i'}^L \times \Psi_{i'}^{L-1} \) — recall that \( m^0_i \subseteq \Psi_{i'}^{L-1}(\theta_i,q_i) \setminus \Psi_{i'}^L(\theta_i,q_i) \).

Thus, what agent \( i \) of first-order type \((\theta_i,q_i)\) loses from the first term of the outcome function by misreporting in the 0-th announcement is greater than the punishment he would get from the second term, regardless of the other agents’ announcements.

The above inequality implies that player \( i \) will be strictly better off by telling the truth in the 0-th announcement, even if he misrepresents the rest of his announcements and pays the penalty when he is the first deviator from a coherent announcement. Therefore, \( m_i \) is never a best response, which contradicts the hypothesis that \( m \in S^F(\theta) \). This completes the proof of Claim 4.1. \( \blacksquare \)

**Claim 4.2:** Let \( m \in S^F(\theta) \). Suppose that there exists \( q \in Q \) such that \( m^s = \Psi^*(\theta,q) \) for all \( s \in \{0, \ldots, j\} \), where \( 0 \leq j \leq J - 1 \). Then

\[
m^{j+1} = \Psi^*(\theta,q).
\]

**Proof of Claim 4.2:** By Claim 4.1, we have proved that each agent reports an atom of the final partition that is consistent with his payoff type at the 0-th announcement. Thus, \( f : T \rightarrow \Delta(A) \) satisfies strict first-order incentive compatibility if \( f : \Theta \rightarrow \Delta(A) \) satisfies first-order incentive compatibility. Also, we can focus on agent \( i \) of first-order type \((\theta_i,q_i)\) with \( m^0_i = \Psi^*_i(\theta_i,q_i) \).

Suppose, by way of contradiction, that there exists \( (\theta'_i,q'_i) \in T_i \) with \( \theta_i \neq \theta'_i \) such that \( (\theta'_i,q'_i) \in m^{j+1}_i \neq \Psi^*_i(\theta_i,q_i) \) for some agent \( i \) of some payoff type \( \theta_i \in \Theta_i \). So, by the construction of the punishment lottery, he has to pay the penalty \( \eta \). Define \( \tilde{m}_i \) such that

\[
\tilde{m}^s_i = m^s_i \forall s \neq j + 1, \\
\tilde{m}^{j+1}_i = \Psi^*_i(\theta_i,q_i).
\]

What we want to show is that for any \( \mu_i \in \Delta^q(\Theta_{-i} \times M_{-i}) \),

\[
\sum_{\theta_{-i},m_{-i}} \mu_i(\theta_{-i},m_{-i}) \{ u_i(g(\tilde{m}_i,m_{-i});\theta_i,\theta_{-i}) - u_i(g(m_i,m_{-i});\theta_i,\theta_{-i}) \} > 0.
\]

Fix \( \mu_i \in \Delta^q(\Theta_{-i} \times M_{-i}) \). Suppose that for any \( \theta_{-i} \) and \( m_{-i} \), if \( \mu_i(\theta_{-i},m_{-i}) > 0 \), \( m^{j+1}_i = \Psi^*_i(\theta_{-i},q_{-i}) \) for some \( q_{-i} \). Under the induction hypothesis, by strict incentive
compatibility of \( \tilde{f} \) and by the definition of the punishment lottery \( \xi(i, m) \), \( \tilde{m}_i \) yields higher payoff than \( m_i \).

On the other hand, suppose that there exists \((\theta_{-i}, m_{-i})\) with \( \mu_i(\theta_{-i}, m_{-i}) > 0 \) such that there exists agent \( i' \neq i \) for whom \( m_{i+1} \notin \Psi^*(\theta_{i'}, q_{r'}) \) for any \( q_{r'} \in Q_{r'} \). Then, define \( \gamma' \) exactly as we defined \( \gamma \) in the proof of Claim 3.2, but with respect to the extended SCF \( \tilde{f} \) of the mechanism \( \tilde{\Gamma} \) in the current proof. Thus, we note the following:

**Fact 4:** We can choose \( J \) large enough so that

\[
\frac{\varepsilon^2}{n} \eta > \frac{1 - \varepsilon - \varepsilon^2}{J} \gamma'
\]

By Fact 4 and the definition of \( \gamma' \),

\[
\frac{\varepsilon^2}{n} \eta > \frac{1 - \varepsilon - \varepsilon^2}{J} \gamma' \geq \frac{1 - \varepsilon - \varepsilon^2}{J} \sum_{\theta_{-i}, m_{-i}} \mu_i(\theta_{-i}, m_{-i}) \left\{ u_i(\tilde{f}(m_{i+1}, m_{-i}); \theta_i, \theta_{-i}) - u_i(\tilde{f}(\bar{m}_{i+1}, m_{-i}); \theta_i, \theta_{-i}) \right\}.
\]

Then, \( \tilde{m}_i \) yields higher payoff than \( m_i \) for any \( \mu_i \in \Delta(\Theta_{-i} \times M_{-i}) \), which contradicts the hypothesis that \( m_i \in S^f(\theta_i) \). This completes the proof of Claim 4.2. ■

Claims 4.1 and 4.2 together show that any message profile \( m \) surviving the iterative elimination of never best responses has the property that: for any \( \theta \in \Theta \),

\[
m \in S^f(\theta) \Rightarrow \exists q \in Q \text{ s.t. } m^s = \Psi^*(\theta, q) \text{ for any } s \in \{0, 1, \ldots, J\}.
\]

and no monetary penalties \( \eta \) are levied.

Since the SCF \( f \) satisfies AM measurability, the resulting outcome is the same as

\[
(1 - \varepsilon^2)(1 - \varepsilon - \varepsilon^2)f(\theta) + \varepsilon(1 - \varepsilon)(1 + \varepsilon)^2x(\theta, q) + \varepsilon^2\bar{x}.
\]

This is arbitrarily close to \( f(\theta) \) for any \( \theta \in \Theta \) whenever \( \varepsilon > 0 \) is chosen small enough. This completes the proof of Theorem 4. ■

**Remark:** If the sets \( Q_i \) are not finite, but arbitrary sets, an adaptation of the maximally revealing mechanism of BM (2009) to the appropriately restricted notion of measurability would provide the proof of Theorem 4 for this case. We choose to present the proof based on finite \( Q_i \)'s as a way to illustrate how the argument must be built up from our Theorem 3, our result based on FOTD. Of course, the general characterization of robust virtual implementation so obtained, by means of first-order incentive compatibility and AM measurability for any arbitrary collection of sets \( Q_i \), boils down to the characterization theorem in BM (2009), in terms of ex post incentive compatibility and robust measurability, when \( Q_i = \Delta(\Theta_{-i}) \) for all \( i \in N \).

### 7 Conclusion

By proposing a reinterpretation of the Wilson doctrine – the planner can rely on restrictions on first-order beliefs, which can be made common knowledge in addition to payoff types –
we have shown that robust virtual implementation is often as powerful as it can possibly be. Indeed, with first-order type diversity, the limits of implementation are given by incentive compatibility, but every incentive compatible SCF can be robustly virtually implemented. Thus, even if one insists on robustness of implementation results, there is a gap between the results offered by exact implementation and those offered by the virtual approach. For both, the main restriction is the appropriate kind of incentive compatibility. In this respect, both are equivalent, so when many types are present in the model, interim incentive compatibility may become quite restrictive, although one can find environments (see the example in Sections 4.1 and 4.2) in which even ex post incentive compatibility is still permissive. The real difference, though, stems from the extra conditions that tackle the “multiplicity of equilibrium” problem, key to full implementation. While robust monotonicity (BM (2008)) is often quite limiting, we have argued that robust measurability is not. Indeed, in our settings, robust measurability – AM measurability over the allowed type spaces – is a trivial condition if it is imposed over almost every type space.

Appendix A: Genericity of FOTD in Continuum Settings

We extend here our argument on the genericity of FOTD in Section 4.3 to some settings in the continuum. Without loss of generality, we focus only on agent i throughout. We denote by $\Delta^0(\Theta_{-i})$ the interior of $\Delta(\Theta_{-i})$.

Our quasi-transferability assumption guarantees the following no total indifference condition: for each $i \in N$ and each first-order type $t_i = (\theta_i, q_i)$, there exist two outcomes $a_k, a_{k'} \in A$ such that $V_i^k(\theta_i, q_i) \neq V_i^{k'}(\theta_i, q_i)$. We also make the following assumption:

**Definition 10 (Ex Post Type Diversity)** A payoff environment $E_\Delta = (A, \Theta_i, u_i)_{i \in N}$ satisfies **ex post type diversity** if, for every $i \in N$, every $\theta_i, \theta'_i \in \Theta_i$ with $\theta_i \neq \theta'_i$, there exist $a \in A$ and $\theta_{-i} \in \Theta_{-i}$ such that $u_i(a; \theta_i, \theta_{-i}) \neq u_i(a; \theta'_i, \theta_{-i})$.

**Remark**: Note that this assumption is significantly weak in the sense that when $|A| \geq 3$, it is generically a vacuous condition due to the finiteness of $\Theta$. See also Section 4.3 for the argument. Private values environments are essentially the only class of environments excluded.

We are now ready to state the main result of this section that shows FOTD generically holds in the continuum setups.

**Theorem 5** Suppose that the payoff environment $E_\Delta = (A, \Theta_i, u_i)_{i \in N}$ satisfies ex post type diversity and no total indifference. Then, for every $i \in N$, there exists a residual subset in $\Delta^0(\Theta_{-i})$ for which FOTD holds.\(^{27}\)

**Proof**: Fix $(\theta^\ell, \theta^m)$ with $\theta^\ell \neq \theta^m$. Taking into account the no-total-indifference condition, we define

$$B_{(\ell,m)}(q_i) = \{q_i' \in \Delta^0(\Theta_{-i}) | V_i(\theta^\ell_i, q_i) = V_i(\theta^m_i, q_i')\}.$$  

\(^{27}\)A set is meager if it contains a countable union of nowhere dense sets. The complement of a meager set is called a residual set. A residual set is a usual topological notion of a generic set.
Then, we claim the following:

**Claim A**: For any $\theta^i_\ell, \theta^m_i \in \Theta_i$ with $\theta^i_\ell \neq \theta^m_i$ and any $q_i \in \Delta^0(\Theta_{-i})$, $B_{(\ell,m)}(q_i)$ is either empty or a countable union of isolated points in $\Delta^0(\Theta_{-i})$.

**Proof of Claim A**: Suppose, on the contrary, that there exists $q'_i \in B_{(\ell,m)}(q_i)$ such that for any neighborhood $V$ of $q'_i$ in $\Delta^0(\Theta_{-i})$ such that $V \backslash \{q'_i\} \cap B_{(\ell,m)}(q_i) \neq \emptyset$. Ex post type diversity allows us to choose a neighborhood $V$ of $q'_i$ such that for any $q^i_\ell \in V$, $V_i(\theta^i_\ell, q_i) \neq V_i(\theta^m_i, q_i)$. This implies that $V \backslash \{q'_i\} \cap B_{(\ell,m)}(q_i) = \emptyset$, which is a contradiction. The reason why $B_{(\ell,m)}(q_i)$ is at most a countable union of isolated points is that $\Delta^0(\Theta_{-i})$ is a separable metric space and contains a countable dense subset in it. ■

Assume that $B_{(\ell,m)}(q_i)$ is nonempty. Then, it is a closed set so that the closure of $B_{(\ell,m)}(q_i)$ is the same as $B_{(\ell,m)}(q_i)$. However, $B_{(\ell,m)}(q_i)$ has no interior points in it. Therefore, $B_{(\ell,m)}(q_i)$ is nowhere dense.

Since $\Delta^0(\Theta_{-i})$ is a separable metric space, it contains a countable dense subset in it. We define $\{q^\lambda_i\}_\lambda^{\infty}$ as such a countable dense subset in $\Delta^0(\Theta_{-i})$. Define

$$B \equiv \bigcup_{\ell,m: \ell \neq m} \bigcup_{\lambda=1}^{\infty} B_{(\ell,m)}(q^\lambda_i).$$

Since countable unions of countable sets are countable, the above set $B$ is a countable union of nowhere dense sets, i.e., a meager set. Once again, since $\Delta^0(\Theta_{-i})$ is a separable metric space and it contains a countable dense subset in it, this set $B$ characterizes the set of all first-order beliefs such that FOTD is violated. Thus, the complement of this set is

$$\Delta^* \equiv \Delta^0(\Theta_{-i}) \backslash B,$$

which characterizes the set of all first-order beliefs in $\Delta^0(\Theta_{-i})$ such that FOTD holds. Thus, this set $\Delta^*$ is a residual set in $\Delta^0(\Theta_{-i})$. This completes the proof. ■

**Appendix B: The Relationship with Virtual Bayesian Implementation**

All our results have been obtained using the very weak solution concept of $\Delta$-rationalizability. When robustness with respect to type spaces is a concern, it follows that there must be a connection with the approach that uses Bayesian equilibrium in every type space. This appendix explores this connection.

**Definition 11** A type space is a tuple $T \equiv (T_i, \hat{\theta}_i, \pi_i)_{i \in N}$ where

1. for each $i \in N$, $T_i$ is a countable space;
2. for each $i \in N$, $\hat{\theta}_i : T_i \rightarrow \Theta_i$ is an onto mapping;
3. for each $i \in N$, $\pi_i : T_i \rightarrow \Delta(T_{-i})$; and
for each $i \in N$ and $\tau_i \in T_i$, the marginal of $\pi_i[\tau_i]$ on $\Theta_{-i}$ through $\hat{\theta}_{-i}$ is $\text{marg}_{\Theta_{-i}} \pi_i[\tau_i]$ which is an element in $Q_i$.

The fourth requirement above says that each agent’s beliefs are reduced to some element of the set of his first-order beliefs if it is marginalized on other agents’ payoff type space. Note also that when $Q_i$ is unrestricted, i.e., $Q_i = \Delta(\Theta_{-i})$, this coherence requirement is a vacuous condition.\(^{28}\)

For an incomplete information game $\Gamma(T)$, let $\sigma_i : T_i \to \Delta(M_i)$ denote agent $i$’s (mixed) strategy and $\sigma_i(m|\tau_i)$ denote the probability with which agent $i$ of type $\tau_i$ use $m_i$. Besides, let us denote $\sigma(m|\tau) = \prod_{j \in N} \sigma_j(m_j|\tau_j)$. Let $B(\Gamma(T))$ be the set of mixed-strategy Bayesian equilibria of the mechanism $\Gamma(T)$.

**Definition 12 (Robust Virtual Implementation in Bayesian Equilibrium)** An SCF $f$ is robustly virtually implementable in mixed-strategy Bayesian equilibrium if there exists $\varepsilon > 0$ such that, for any $\varepsilon \in (0, \bar{\varepsilon}]$, there exists a mechanism $\Gamma^\varepsilon = (M^\varepsilon, g^\varepsilon)$ such that for any type space $T$, $B(\Gamma^\varepsilon(T)) \neq \emptyset$, any $\sigma^* \in B(\Gamma^\varepsilon(T))$, and any $\tau \in T$,

$$\|g^\varepsilon(\sigma^*(\tau)) - f(\hat{\theta}(\tau))\| \leq \varepsilon,$$

where $g(\sigma^*(\tau)) = \sum_{m \in M} \sigma^*(m|\tau) g(m) \in \Delta(A)$.

We now state several simple corollaries that link our earlier results to implementation in Bayesian equilibrium. Recall Theorem 3. As an important by-product, we obtain the following result for the environments satisfying FOTD.

**Corollary 2 (Robust Virtual Bayesian Implementation)** Suppose an environment $E$ satisfies FOTD, quasi-transferability and that the sets $Q_i$ are finite. If an SCF satisfies first-order incentive compatibility, then it is robustly virtually implementable in mixed strategy Bayesian equilibrium.

Next, with the same argument, one can provide the following simple corollary to Theorem 4 if one does not assume FOTD:

**Corollary 3 (A Sufficient Condition for Robust Virtual Bayesian Implementation)** Suppose an environment $E$ satisfies quasi-transferability and that the sets $Q_i$ are finite. An SCF $f$ is robustly virtually implementable in mixed strategy Bayesian equilibrium if it satisfies first-order incentive compatibility and AM measurability.

It is important to note that AM measurability is not necessary for robust virtual implementation in mixed strategy Bayesian equilibrium. To make this point, an elaboration of the example in Section 5 of Serrano and Vohra (2005) would suffice.\(^{29}\) However, when the implementing mechanism is required to be regular, to be defined next, AM measurability becomes necessary for robust virtual implementation in mixed strategy Bayesian equilibrium.

\(^{28}\)We could extend the analysis to more general spaces $T$, but at the cost of additional technical complexity.

\(^{29}\)Although Serrano and Vohra (2005) restricts attention to implementation in pure strategies, the argument can be extended to also cover mixed strategies.
The next definitions are borrowed from AM (1992c): For every \( i \in N \) and every partition \( \Psi_i \), let \( \Sigma_i(\Psi_i) \) denote the set of mixed strategies of player \( i \) that are measurable with respect to \( \Psi_i \).

**Definition 13 (First-Order Bayesian Equilibrium)** The profile \( \sigma \in \Sigma_1(\Psi_1) \times \cdots \times \Sigma_n(\Psi_n) \) is a first-order Bayesian equilibrium with respect to \( \Psi \) in \( \Gamma(T) \) for a coherent type space \( T \) if for all \( i \in N \) and all \( \psi_i \in \Psi_i \), there exists some \( \tau_i \in T_i \) with \( \hat{t}_i(\tau_i) \in \psi_i \), such that for any \( m_i \in M_i \) with \( \sigma_i(m_i|\tau_i) > 0 \),

\[
\sum_{\tau_{-i}} \sum_{m_{-i}} \left( \prod_{j \neq i} \sigma_j(m_j|\tau_j) \right) \left[ u_i(g(m_i, m_{-i}); \hat{\theta}(\tau)) - u_i(g(m'_i, m_{-i}); \hat{\theta}) \right] \geq 0 \quad \forall m'_i \in M_i
\]

**Definition 14 (Regular Mechanisms)** A mechanism \( \Gamma(T) \) is said to be regular if for each \( \Psi \) there exists a first-order Bayesian equilibrium with respect to \( \Psi \) in \( \Gamma(T) \) for any type space \( T \).

In particular, finite mechanisms – like the ones constructed in the proofs of Theorems 3 and 4 – are regular. Mechanisms that rely on the use of integer games – e.g., like the one constructed in Serrano and Vohra (2005) – are not regular.

The next result extends a result in AM (1992c) to our settings:

**Proposition 1** If an SCF is robustly virtually implementable in mixed strategy Bayesian equilibrium by a regular mechanism, then it satisfies AM measurability.

**Proof:** Since \( f \) is robustly virtually implementable in Bayesian equilibrium by a regular mechanism, there exists a regular mechanism \( \Gamma = (M, g) \) such that for any \( T \) and any \( \sigma \in B(\Gamma(T)) \),

\[
\|g(\sigma(\tau)) - f(\hat{\theta}(\tau))\| \leq \varepsilon \quad \forall \tau \in T.
\]

Fix an arbitrary coherent type space \( T \). Let \( \sigma \in \times_{i \in N} \Sigma_i(\Psi^*_i) \) be a first-order Bayesian equilibrium with respect to \( \Psi^* \). The existence of such \( \sigma^* \) is guaranteed by the regularity of the mechanism \( \Gamma \). Note that \( \sigma \) is measurable with respect to \( \Psi^* \). What we want to show here is that \( \sigma \) is a Bayesian equilibrium as well.

If \( m_i = \sigma_i(\tau_i) \) is a best response for player \( i \) of type \( \tau_i \), then \( m_i \) is also a best response for player \( i \) of any \( \tau_i' \) such that \( \hat{t}_i(\tau_i') \in \rho_i(t_i, \Psi^*_i) \). That is, this implies that for any \( \psi_i \in \Psi^*_i \), for any \( \tau_i, \tau_i' \in T_i \) such that \( \hat{t}_i(\tau_i), \hat{t}_i(\tau_i') \in \psi_i \) and \( \hat{\theta}_i(\tau_i) = \hat{\theta}_i(\tau_i') \), the best responses of player \( i \) of type \( \tau_i \) and \( \tau_i' \) to any \( \sigma_{-i} \) that is measurable with respect to \( \Psi^*_{-i} \) are the same. Then, it follows that any first-order Bayesian equilibrium \( \sigma \) that is measurable with respect to \( \Psi^* \) is in fact a Bayesian equilibrium. By our hypothesis of robust virtually implementation, we have that for any \( \tau_{-i} \in T_{-i} \),

\[
\|g(\sigma_i(\tau_i), \sigma_{-i}(\tau_{-i})) - f(\hat{\theta}_i(\tau_i), \tau_{-i})\| \leq \varepsilon \quad \text{and} \quad \|g(\sigma_i(\tau_i'), \sigma_{-i}(\tau_{-i})) - f(\hat{\theta}_i(\tau_i'), \tau_{-i})\| \leq \varepsilon.
\]

Since \( \sigma_i(\tau_i) = \sigma_i(\tau_i') \), we can conclude that \( \|f(\theta_i, \theta_{-i}) - f(\theta_i', \theta_{-i})\| \leq 2\varepsilon \) for any \( \theta_{-i} \in \Theta_{-i} \). Since this must be true for any \( \varepsilon > 0 \), we obtain \( f(\theta_i, \theta_{-i}) = f(\theta_i', \theta_{-i}) \) for any \( \theta_{-i} \in \Theta_{-i} \). Hence, \( f \) satisfies AM measurability. 

37
For a type space $\mathcal{T}$, Define $U_i(f; \tau'_i|\tau_i)$ to be the interim utility of agent $i$ of type $\tau_i$ who pretends to be of type $\tau'_i$ corresponding to an SCF $f$ as follows:

$$U_i(f; \tau'_i|\tau_i) \equiv \sum_{\theta_{-i} \in \Theta_{-i}} \text{marg}_{\theta_{-i}} \pi_i[\tau_i](\theta_{-i}) u_i(f(\hat{\theta}_i(\tau'_i), \theta_{-i}); \hat{\theta}_i(\tau_i), \theta_{-i}).$$

Denote $U_i(f|\tau_i) = U_i(f; \tau_i|\tau_i)$. Then, we say that an SCF $f$ satisfies incentive compatibility for a type space if for any $i \in \mathcal{N}$, $\tau_i, \tau'_i \in \mathcal{T}_i$, $U_i(f|\tau_i) \geq U_i(f; \tau'_i|\tau_i)$. Therefore, it is easy to observe that an SCF satisfies first-order incentive compatibility if and only if it satisfies incentive compatibility for any type space.

Putting together this observation, Proposition 1 and Theorem 4, we arrive at the following:

**Corollary 4 (A Characterization of Robust Virtual Bayesian Implementation)**

Suppose an environment $\mathcal{E}$ satisfies quasi-transferability and that the sets $Q_i$ are finite. An SCF $f$ is robustly virtually implementable in mixed strategy Bayesian equilibrium by a regular mechanism if and only if it satisfies first-order incentive compatibility and AM measurability.

On the other hand, the usual approach for a fixed type space to (exact and virtual) Bayesian implementation has ruled out the consideration of mixed strategies. We show next that if one includes robustness considerations with respect to type spaces, the distinction between pure and mixed strategy equilibrium implementation is of no significance:

**Proposition 2** An SCF is robustly virtually implementable in pure-strategy Bayesian equilibrium if and only if it is robustly virtually implementable in mixed-strategy Bayesian equilibrium.

**Proof:** That full implementation in mixed strategy equilibrium implies full implementation in pure equilibrium is easy to see. We argue the opposite direction.

Suppose not. There exists an SCF $f$ that is robustly virtually implementable in pure Bayesian equilibrium that is not robustly virtually implementable in mixed equilibrium. This means that any mechanism that virtually implements $f$ in pure equilibrium over every coherent type space has an equilibrium in properly mixed strategies whose outcome does not approximate $f$. But then, one can construct a sufficiently large coherent type space and perform a purification of that equilibrium. The result is a pure-strategy Bayesian equilibrium of the mechanism whose outcome is far from $f$. This contradicts that $f$ is robustly virtually implementable in pure-strategy equilibrium.

Thus, while implementation in pure or mixed equilibrium may give different answers for a fixed type space, that difference goes away when one requires robustness in implementation with respect to type spaces.

---

30 Duggan (1997) and Serrano and Vohra (2010) are exceptions.

31 This purification is possible because our robustness analysis does allow for infinite type spaces.
References


