3-2013

Robust virtual implementation: Toward a reinterpretation of the Wilson doctrine

Georgy ARTEMOV

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

Roberto SERRANO

DOI: https://doi.org/10.1016/j.jet.2012.12.015

Published in Journal of Economic Theory, Volume 148, Issue 2, March 2013, pp. 424-447.
http://doi.org/10.1016/j.jet.2012.12.015

Robust Virtual Implementation: Toward a Reinterpretation of the Wilson Doctrine*

Georgy Artemov†, Takashi Kunimoto‡ and Roberto Serrano§

First Version: May 2007
This Version: July 2012

Abstract

We study a mechanism design problem where arbitrary restrictions are placed on the sets of first-order beliefs of agents. Calling these restrictions Δ, we use Δ-rationalizability (Battigalli and Siniscalchi (2003)) as our solution concept, and require that a mechanism virtually implement a socially desirable outcome. We obtain two necessary conditions, Δ-incentive compatibility and Δ-measurability and show that the latter is satisfied as long as a particular zero-measure set of first order beliefs is ruled out. In environments allowing small transfers of utility among agents, these two conditions are also sufficient.
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 input from Murali Agastya, Dirk Bergemann, Antonio Cabrales, Pedro Dal Bo, Geoffroy de Clippel, Kfir Eliaz, Jeff Ely, Sven Feldmann, Philippe Jehiel, Vijay Krishna, Judith Levi, Simon Loertscher, Hitoshi Matsushima, Claudio Mezzetti, Stephen Morris, Tomas Sjöström, Rani Spiegler, Ronald Stauber, Takeshi Suzuki, Takashi Ui and Takehiko Yamato. Artemov gratefully acknowledges financial support from the University of Melbourne through ECR grant, Kunimoto from FQRSC, SSHRC of Canada, and Japan Society for the Promotion of Science (23830024), 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.; 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."

Robert Wilson (1987)

1 Introduction

The correct design of institutions can be decisive for achieving desirable economic goals. Typically, achieving such a correct design depends on the knowledge of key parameters in the environment, most of which are often unknown by the economic authority. The theory of implementation or mechanism design looks in a systematic way at the design of rules for social interaction that do not assume detailed knowledge of the fundamentals by those with power to impose social outcomes. In attempting to bring further realism to the theory, one can assume that not only the economic authority (or mechanism designer, or planner) does not know those parameters, but the same holds true for the agents in the system, who may, for example, not know the true preferences of others. These are called *incomplete information environments*.

In such environments, an agent's private information is summarized by the notion of a *type*. For an agent, a type specifies (i) his private information about his own preferences and/or the preferences of others (*payoff type*), (ii) his belief about the payoff types of others (*first-order belief*), (iii) his belief about others' first-order beliefs (*second-order belief*), and so on, leading to a hierarchy of beliefs *ad infinitum*.

A basic assumption of the classic approach to mechanism design under incomplete information is that the underlying spaces of types are common knowledge among the planner and the agents. In making this assumption, as is often the case in the literature, one effectively assumes that each first-order belief corresponds to a unique infinite hierarchy of beliefs. Hence, the designer does not need to consider higher-order beliefs in the analysis.[1] This common-knowledge assumption is often seen as unrealistic, and, as the opening quote from Wilson suggests, the theory should be revised with more attention given to this issue. Carrying out such a research program

[1]For instance, to be able to extract full surplus from the agents, Cremer and McLean (1988) exploit the property that in a fixed finite type space with a generic common prior, the planner can infer agents' preferences by eliciting their first-order beliefs. See Neeman (2004) and Heifetz and Neeman (2006) for further details.

is said to be following the *Wilson doctrine*.

This paper proposes a reinterpretation of the Wilson doctrine. It provides a different framework, rich enough to allow flexibility in higher-order beliefs, and within that framework, it studies *virtual* (i.e, approximate) implementation. We start with the assumption that the designer may have some information that allows her to rule out certain first-order beliefs of the agents; this information may come through previous experience or from some information acquisition process. We denote by Δ such a restriction on first-order beliefs, and we assume it to be common knowledge.[2]

To make our analysis *robust* to the specification of higher-order beliefs, we use a type-free solution concept of rationalizability – Δ-*rationalizability* (Battigalli and Siniscalchi (2003))– that guarantees that the predictions are the same for any higher-order beliefs, as long as those predictions are consistent with our Δ restriction on the first-order belief set.

Suppose the planner wishes to (approximately) implement a particular set of socially optimal outcomes, summarized by a *social choice function (SCF)*. We derive both necessary and sufficient conditions for robust virtual implementation (RVI) of SCFs in Δ-rationalizable strategies. The conditions consist of a version of incentive compatibility, which we call Δ-*incentive compatibility*, and an additional condition, which we call Δ-*measurability*.

Furthermore, we show that *every* SCF satisfies Δ-measurability as long as the Δ-restriction rules out a particular zero-measure set of first-order beliefs; the set includes all the beliefs where two *different* payoff types have *identical* interim preferences. Thus, generically, the appropriate version of incentive compatibility is the only condition restricting RVI of SCFs.

This result ought to be contrasted with another result obtained for RVI for the case of no restrictions on first-order beliefs ($\Delta = \emptyset$).[3] Then, our characterization reduces to the one obtained by Bergemann and Morris (2009) (we use the initials B&M to refer to these authors from now on), who conclude that virtual implementation is severely restricted beyond their version of incentive compatibility. B&M (2009) argue that, for an important class of preferences, the conditions for RVI are equivalent to the conditions for robust *exact* implementation, which are commonly assessed as restrictive. Our analysis shows that this conclusion hinges on a zero-measure set, and hence, that the usual distinction between the exact and virtual implementation approaches remains, even in the robust setting, when such a set is removed.

We provide further detail in the rest of the introduction. We begin by reviewing different notions of implementation in order to frame our contribution more precisely. We then explain the extra restrictions on beliefs brought about by the

[2]We use Δ to designate our restriction following a convention in the literature, but this is not to be confused with a probability simplex, for which we also use the symbol Δ. No confusion should arise in context. Note also that the case $\Delta = \emptyset$ (i.e., no restriction at all) is covered.

[3]See the recent book on robust mechanism design by Bergemann and Morris (2012) which surveys this literature, and also collects all their papers on the topic. See also the papers cited in its introduction and a related paper by Saijo, Sjöström, and Yamato (2007, Section 5).

planner's additional information and describe necessary and sufficient conditions for implementation in our sense.

1.1 Notions of Implementation

The current paper studies *full, virtual, and robust implementation* of SCFs.[4] *Full implementation* is the requirement that the set of outcomes prescribed by a given solution concept coincide with the SCF. One considers full implementation after ensuring that partial implementation – which allows other outcomes that are solutions to a game but not socially optimal – is possible. *Virtual implementation* means that the planner contents herself with implementing the SCF with arbitrarily high probability. This is an approximate version of exact implementation, which insists on implementing the SCF with probability 1. Finally, *robust implementation* is the requirement that implementation survive any specification of higher-order beliefs consistent with the common knowledge structure of the environment (i.e., consistent with the Δ-restriction). In contrast, what we shall call the *classic* approach to implementation assumes that the entire type space is common knowledge; this approach thus pins down the higher-order beliefs uniquely, a significantly stronger assumption.

When one requires robust implementation, one should expect the conditions on SCFs to become stronger than in the classic approach. Indeed, for exact implementation, the robust analysis is known to turn conditions that are already somewhat restrictive into much stronger ones.[5]

In view of the strength of the conditions required for robust exact implementation, one might prefer to consider its approximate version: virtual implementation (Matsushima (1988), Abreu and Sen (1991)). For classic environments where the type spaces are common knowledge, Abreu and Matsushima (1992a, b) (henceforth, we use A&M to refer to these authors) characterize virtual implementation in iterative elimination of strictly dominated strategies. This solution concept places weak requirements on agents' rationality and is closely related to rationalizability, which we use. A&M (1992b) find that incentive compatibility and a measurability condition, which we shall refer to as *A&M measurability*, are necessary conditions. In quasi-transferable environments, which allow small utility transfers among agents, both are also shown to be sufficient.

Incentive compatibility is clearly unavoidable in full implementation, since it is

[4]For surveys on implementation, see, for example, Palfrey and Srivastava (1993), Jackson (2001), and Serrano (2004).

[5]In the environments where $\Delta = \emptyset$, the robust analogues of conditions for exact implementation are ex-post incentive compatibility and robust monotonicity. For appraisals of these conditions, see, e.g., Jehiel et al. (2006) and B&M(2011), respectively. Jehiel et al. (2006) show that in some settings ex post incentive compatibility would generically require a social choice function (SCF) to be constant. In others, it still leaves room for nontrivial SCFs (see our Subsection 4.1, or B&M (2009, Section 3)). Robust monotonicity amounts to requiring Bayesian monotonicity – a necessary condition in the classic approach – in every type space, which is a very restrictive condition (B&M, 2011).

necessary even for partial implementation. Hence, if one considers full implementation, as we do, one should focus on the measurability condition. A&M measurability stipulates that the SCF must be constant on each element of the suitably defined, finest possible partition of the payoff-type space, which corresponds to the maximum possible separation of payoff types according to their interim preferences.[6] In a classic Bayesian environment, A&M measurability is usually perceived as very permissive, and even close to trivial. Hence, in such classic settings, the limitations of virtual implementation essentially amount to the incentive compatibility constraint.

Yet the role that fixed finite type spaces play in the constructions presented in A&M (1992b) has not been fully understood; one conjecture (B&M (2009, p. 49)) is that this role is crucial, as it is for a number of other results in the literature. Indeed, in the robust analysis when $\Delta = \emptyset$, B&M (2009) characterize RVI by means of *ex post incentive compatibility* and *robust measurability*, and they assess the latter to be an extremely restrictive condition. To show the restrictiveness of their version of measurability, B&M (2009) construct a very specific type space in which the interim preferences of all types *must be* aligned. In such a type space, payoff types cannot be separated.[7] In contrast, the arbitrary Δ-restriction, which is the crux of the current paper, suggests a way to evaluate the limitations of RVI. By using this Δ-restriction, our analysis both complements the one in B&M (2009) by assessing the restrictiveness of measurability-like conditions, and challenges some of its conclusions.

1.2 The Δ-restriction on First-order Beliefs

To be specific, the Δ-restriction assumes 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, our analysis reduces to the case considered by B&M (2009).) Consistent with this restriction, the solution concept of Δ-rationalizability simply imposes the requirement that first-order beliefs lie in the sets Q_i on top of standard rationalizability.simply imposes the requirement that first-order beliefs lie in the sets Q_i.[8]

We proceed as follows. We fix a (typically rich) space of *first-order types*, which are the combinations of payoff types and first-order beliefs. We assume this space to be common knowledge among the agents. Then, we require that implementation in Δ-rationalizable strategies obtain for all higher-order beliefs coherent with our original first-order type space.

[6] To find such a partition, one constructs an iterative algorithm that gradually refines the payoff-type space; see A&M (1992b) and our Subsection 3.2 for details.

[7] Serrano and Vohra (2001) earlier observed that virtual Bayesian implementation may fail in a classic Bayesian environment for exactly that reason, but they argued later (Serrano and Vohra (2005)) that such failures are arguably "rare."

[8] Battigalli and Siniscalchi (2003) show that Δ-rationalizability coincides with Bayesian equilibria in all coherent type spaces and Battigalli et al. (2011) show that (a suitably defined) Δ-rationalizability is equivalent to the interim correlated rationalizability of Dekel, Fudenberg, and Morris (2007).

1.3 Necessary Conditions

To begin with, RVI in Δ-rationalizable strategies is limited by the (interim) incentive compatibility imposed on the first-order types present in the model. We refer to such a condition as Δ-*incentive compatibility*, and Theorem 1 shows it to be necessary for RVI in Δ-rationalizable strategies. In Theorem 2 we then show that a version of A&M's original measurability is also necessary for RVI in our sense. We propose Δ-*measurability*, which we define as A&M measurability subject to the Δ-restriction.

Generalizing the approach in Serrano and Vohra (2005), we propose next a condition that we term Δ-*type diversity*. In environments satisfying Δ-type diversity, all payoff types can be separated from one another in the first round of the measurability algorithm. This renders Δ-measurability a trivial condition. Importantly, almost every environment in our settings satisfies the Δ-type diversity condition when there are at least three alternatives.

1.4 Sufficient Conditions and Implementing Mechanisms

To construct implementing mechanisms, we first consider quasi-transferable environments satisfying Δ-type diversity, and we show (Theorem 3) that an SCF is robustly virtually implementable as long as it satisfies Δ-*incentive compatibility*.[9] Since Δ-type diversity generically holds, it follows that in quasi-transferable environments, one almost never needs to rely on any additional condition beyond the appropriate version of incentive compatibility.

Next, we obtain a characterization without using the Δ-type diversity assumption, and thereby extend the work of A&M (1992b) to our setting and solution concept. The proof of Theorem 4 provides the sufficiency argument to show that Δ-*incentive compatibility* and Δ-*measurability* are necessary and sufficient for RVI in Δ-rationalizable strategies. To further underscore the tight connection with A&M (1992b), we follow A&M's arguments closely in the construction of our mechanisms.

In summary, our results demonstrate that, even in the robust analysis, the limitations of virtual implementation are almost always confined to the relevant notion of incentive compatibility. The measurability condition drops out because it is generically satisfied by all SCFs. Even if a zero-measure set of first-order beliefs is ruled out, Δ-incentive compatibility is as restrictive as ex-post incentive compatibility. For measurability conditions, in contrast, ruling out certain zero-measure sets turns a restrictive condition into a trivial one. In an area plagued with very negative results, and with the understanding that the incentive-compatibility requirements may indeed be quite limiting, our contribution clarifies the possibilities of RVI, and thus offers a piece of good news to increase the permissiveness of the theory.

[9]To prove our sufficiency results (Theorems 3 and 4) we employ the assumption that Q_i's are finite sets, but this is not essential. Indeed, as detailed after the statement of Theorem 4, one can adapt the canonical "maximally revealing" mechanism in B&M (2009) to the specific Q_i's assumed. The finiteness assumption makes the argument in Theorem 3 and its connections to mechanisms in A&M (1992b) and in Theorem 4 especially transparent.

1.5 Plan of the Paper

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 Δ-incentive compatibility and Δ-measurability are necessary conditions for RVI in Δ-rationalizable strategies. Section 4 discusses the relationship between our approach and the case in which the planner has no information about the set of first-order beliefs, and presents our genericity arguments. Sufficiency results are presented in Section 5. We conclude in Section 6. Appendix A contains the proof of genericity of Δ-type diversity for continuum settings. A discussion on the connection with virtual implementation in Bayesian equilibrium, as well as all the omitted proofs and formal details in the paper, can be found in the online appendix (Artemov, Kunimoto, and Serrano (2012), not for publication in print).

2 Preliminaries

Let $N = \{1, \ldots, n\}$ denote the set of agents and Θ_i be the set of *finite* payoff types of agent i. Denote $\Theta \equiv \Theta_1 \times \cdots \times \Theta_n$, and $\Theta_{-i} \equiv \Theta_1 \times \cdots \times \Theta_{i-1} \times \Theta_{i+1} \times \cdots \times \Theta_n$.[10] Let $q_i(\theta_{-i})$ denote agent i's first-order belief that other agents receive the profile of payoff types θ_{-i}. For an abstract finite set X, we will denote the set of probability distributions over X by $\Delta(X)$. Let $Q_i \subseteq \Delta(\Theta_{-i})$ be the set of allowed first-order beliefs of agent i. We call $T_i \equiv \Theta_i \times Q_i$ 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.[11]

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

We can now define an *environment* as $\mathcal{E} = (A, \{u_i, \Theta_i, Q_i\}_{i \in N})$, which is implicitly understood to be common knowledge among the agents. In particular, if Q_i is unrestricted for each i, that is, $Q_i = \Delta(\Theta_{-i})$, we call it a *payoff environment* denoted as $\mathcal{E}_\Delta = (A, \{u_i, \Theta_i\}_{i \in N})$. This is the environment that B&M (2009, 2011) 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, as one can allow a rich set of payoff and first-order belief types, our model escapes the criticism in Neeman (2004) of "beliefs-determine-preferences", the problem we have alluded to in the opening paragraph of our paper. 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 $f : \Theta \to \Delta(A)$. Note that, as is standard in the literature on robust implementation, the domain of the SCFs is the

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

[11] 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 footnote 15 below, when this assumption is invoked.

payoff type space.

Define $V_i(f;\theta_i'|\theta_i,q_i)$ to be the interim expected utility of agent i of first-order type (θ_i,q_i) that pretends to be of first-order type (θ_i',q_i') corresponding to an SCF f as follows:[12]

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

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 $g:M\to\Delta(A)$, where $M=\times_{i\in N}M_i$.

Next we define the solution concept of Δ-rationalizability that we use in the paper.

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

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

and we write $\mathcal{S}$ for the collection of message correspondence profiles.[13] The collection $\mathcal{S}$ is a lattice with the natural ordering of set inclusion: $S\subseteq S'$ if $S_i(\theta_i)\subseteq S_i'(\theta_i)$ for all $i\in N$ and $\theta_i\in\Theta_i$. The largest element is $\bar{S}=(\bar{S}_1,\ldots,\bar{S}_n)$, where $\bar{S}_i(\theta_i)=M_i$ for all $i\in N$ and $\theta_i\in\Theta_i$. The smallest element is $\underline{S}=(\underline{S}_1,\ldots,\underline{S}_n)$, where $\underline{S}_i(\theta_i)=\emptyset$ for all $i\in N$ and $\theta_i\in\Theta_i$.

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

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

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

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

This is an incomplete information version of rationalizability, proposed by Battigalli and Siniscalchi (2003). They call it Δ-*rationalizability* and denote by Δ restrictions on the set of first-order beliefs. When $Q_i=\Delta(\Theta_{-i})$, this rationalizability is equivalent to the one used by B&M (2009). We observe that b is increasing by

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

[13]To avoid heavy notation, we ignore the fact that the message correspondence depends on the underlying mechanism Γ.

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^Γ. 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

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

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

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

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

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

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

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

3 Necessity for Robust Virtual Implementation

In this section we discuss the necessary conditions for RVI 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.[14]

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:

[14] 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.

Definition 2 (Δ-incentive compatibility) *An SCF $f : \Theta \to \Delta(A)$ satisfies* ***Δ-incentive compatibility*** *if, for any $i \in N$ any $(\theta_i, q_i) \in T_i = \Theta_i \times Q_i$ and any $\theta_i' \in \Theta_i$,*

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

We shall say that an SCF f satisfies *strict* Δ-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 $\sigma_i : \Theta_i \to M_i$. The next theorem identifies Δ-incentive compatibility as a necessary condition for implementability:

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

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

Fix $\sigma_{-i} : \Theta_{-i} \to M_{-i}$ such that $\sigma_{-i}(\theta_{-i}) \in S^\Gamma_{-i}(\theta_{-i})$ for each $\theta_{-i} \in \Theta_{-i}$. For any $m_i' \in S_i^\Gamma(\theta_i')$, RVI 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 (θ_i, q_i) and he holds the belief $\mu_i \in \Delta^{q_i}(\Theta_{-i} \times M_{-i})$ such that for each θ_{-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 μ_i. Then if $m_i \in S_i^\Gamma(\theta_i)$, RVI 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 μ_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 μ_i, we can rewrite the above inequality as follows:

$$\sum_{\theta_{-i}} q_i(\theta_{-i}) \left[u_i(g(m_i, \sigma_{-i}(\theta_{-i})); \theta_i, \theta_{-i}) - u_i(g(m_i', \sigma_{-i}(\theta_{-i})); \theta_i, \theta_{-i})\right] \geq 0. \tag{3}$$

Due to the fact that Θ 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 ε can be chosen arbitrarily small due to the requirement of RVI, 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 Δ-incentive compatibility. ■

When $Q_i = \Delta(\Theta_{-i})$ for every $i \in N$, it is easy to see that Δ-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, A&M (1992b) have uncovered a condition that they have termed *measurability* (we shall refer to it as A&M measurability) that is necessary for virtual implementation in iteratively undominated strategies over a standard environment that fixes a Bayesian type space. In this section we revisit the A&M measurability condition by adapting it to our RVI analysis.

Denote by Ψ_i a *partition* of the set of first-order types T_i, where ψ_i is a generic element of Ψ_i and $\Psi_i(t_i)$ denotes the element of Ψ_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** Ψ 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 Ψ implies that for any agent i, f does not distinguish between any pair of payoff types that lie in the same cell of the partition Ψ_i.

Definition 5 *Fix a mechanism $\Gamma = (M, g)$. A strategy $\sigma_i : \Theta_i \to M_i$ for player i is **measurable with respect to** Ψ_i if for every $t_i = (\theta_i, q_i), t_i' = (\theta_i', q_i') \in T_i$ with $\theta_i \neq \theta_i'$,*

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

*A strategy profile σ is **measurable with respect to** Ψ if, for every $i \in N$, σ_i is measurable with respect to Ψ_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 $T_i \times \Psi_{-i} \equiv \{\{t_i\}_{t_i \in T_i}\} \times \Psi_{-i}$ 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 Ψ_{-i}, we say that t_i is **equivalent** to t_i' (denoted by $t_i \sim t_i'$) with respect to Ψ_{-i} if, for any pair of SCFs f and $\tilde{f}$ which are measurable with respect to $T_i \times \Psi_{-i}$,*

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

Remark: What we aim to distinguish are "payoff types." In particular, we consider two first-order types (θ_i, q_i) and (θ_i', q_i') as 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 Ψ_{-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 $h \geq 0$, Ψ_i^{h+1} is the same as, or finer than, Ψ_i^h. Thus, we have a partial order $\geq$ as $\Psi_i^{h+1} \geq \Psi_i^h$. Define Ψ^* as follows:

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

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

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

Definition 7 *An SCF f satisfies **Δ-measurability** if it is measurable with respect to Ψ^*.*

Note how the partitions Ψ^0, Ψ^1, ..., and hence, the final partition Ψ^* used in Δ-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 Ψ^*. 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 Ψ^*, which provides the maximum possible degree of first-order type separation or distinguishability in terms of interim preferences. Δ-measurability simply asks that the SCF not distinguish between different first-order types that are "indistinguishable" according to Ψ^*.

Δ-measurability, when applied over the unrestricted set of first-order beliefs, yields robust measurability, a condition introduced by B&M (2009), who also note that relation. Robust measurability implies Δ-measurability, but the examples in the next section show that they are distinct conditions. The connections between the two conditions are detailed in the online appendix.

It should be apparent that Δ-measurability is akin to A&M measurability subject to the Δ restriction on first-order types, but in the end is concerned with the separation of payoff types. A&M (1992b) show that in a Bayesian environment with a fixed type space, A&M 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 Δ-measurability.*

Proof: Since f is robustly virtually implementable, there exists a mechanism $\Gamma = (M, g)$ such that whenever $m \in S^\Gamma(\theta)$, $\|g(m) - f(\theta)\| \leq \varepsilon$ for $\varepsilon > 0$. Recall that $\sigma_i : \Theta_i \to M_i$ is defined as agent i's pure strategy. For each $h \geq 1$, let $\mathcal{K}^h = \times_{i \in N} \mathcal{K}_i^h$ 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 $\times_{i \in N}\{T_i\}$. Then, either (1) $\|g(\sigma[0](\theta)) - f(\theta)\| \leq \varepsilon$ for every θ or (2) $\|g(\sigma[0](\theta)) - f(\theta)\| > \varepsilon$ for some $\theta \in \Theta$.

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

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

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 Ψ^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 $\Psi_i^1(t_i) = \Psi_i^1(t_i')$. Note that $\Psi_i^1(t_i)$ is the element of Ψ_i^1 that includes t_i. By the previous hypothesis, we have that for any θ_{-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 θ_{-i} by measurability with respect to Ψ^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 θ_{-i}. Hence, f satisfies Δ-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 Δ-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 Ψ^{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 Δ-measurability. Otherwise, since f is robustly virtually implementable by our hypothesis, for every $i \in N$, 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 Ψ_i^h.

Let σ^* be a strategy profile that survives the iterative elimination of never best responses in the implementing mechanism Γ. Then, the preceding argument implies that σ^* is measurable with respect to Ψ^*. It follows that $g \circ \sigma^*$ is measurable with respect to Ψ^*. 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 $\Psi_i^*(t_i) = \Psi_i^*(t_i')$. 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 θ_{-i}. Thus, f satisfies Δ-measurability.■

To illustrate the implications of Δ-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 $V_i^k(t_i)$ 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 Θ, i.e.,

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

Let $V_i(t_i) = (V_i^1(t_i), \ldots, V_i^K(t_i))$.

Next, we define the condition of Δ-type diversity in an environment, which will play an important role in our analysis:

Definition 8 (Δ-TD) *An environment $\mathcal{E}$ satisfies Δ-**type diversity (Δ-TD)** 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*

$$V_i(t_i) = \beta V_i(t_i') + \gamma e,$$

where e is the unit vector in $\mathbb{R}^K$.[15]

[15] 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

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

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

Δ-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 A&M (1992b). 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 Δ-TD and the measurability algorithm. In Δ-TD we have defined a vector $V_i(t_i)$ of agent i's valuations of each alternative a_k. When the algorithm that determines Ψ^* does not stop in the first step, we need to consider a more complicated "version" of a_k, that we define below. Let $\mathcal{F}$ denote the set of all SCFs. Define

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

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

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

Let $|F(T_i \times \Psi_{-i})| = K$.[16] Define $V_i^k(t_i; \Psi_{-i})$ 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.,

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

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

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

Lemma 1 *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' ($t_i \sim t_i'$) with respect to Ψ_{-i} if and only if there exist $\beta > 0$ and $\gamma \in \mathbb{R}$ such that*

$$V_i(t_i; \Psi_{-i}) = \beta V_i(t_i'; \Psi_{-i}) + \gamma e,$$

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

The following is a characterization of Δ-TD in terms of the measurability construction:

Corollary 1 *An environment $\mathcal{E}$ satisfies Δ-TD 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' ($t_i \sim t_i'$) with respect to Ψ_{-i}^0. It follows that $\Psi_i^1 = T_i \backslash \sim$ for each agent $i \in N$, and $\Psi^* = T \backslash \sim$, where each $T_i \backslash \sim$ and $T \backslash \sim$ denotes a quotient space generated by the equivalence relation $\sim$.*

[16] 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.

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 2 (Δ-TD $\Rightarrow$ Δ-measurability) *Suppose an environment $\mathcal{E}$ satisfies Δ-TD. Then,* ***every*** *SCF satisfies Δ-measurability.*

That is, if the environment satisfies Δ-TD, 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

B&M (2009) study RVI 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 RVI by means of ex post incentive compatibility and robust measurability. Robust measurability amounts to A&M measurability in every type space, and B&M (2009) assess it as a very demanding condition, which leads them to conclude that the limitations to RVI 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 Δ-type diversity

We find it useful to adopt the example from Section 3 in B&M (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 $\gamma \leq 1$. B&M (2009) show that RVI 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, B&M (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).[17]

On the other hand, B&M (2009) show that RVI is impossible, also in the intermediate range of γ'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 RVI (i.e., Δ-incentive compatibility and Δ-measurability) are very permissive in the example, thanks to the Δ-TD 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.[18] 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 Δ-incentive compatibility for any specification of the Q_i's.

Next, we turn our attention to Δ-TD. First, we claim that for $1/2 > \gamma > 0$, the environment satisfies Δ-TD. 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 γ;

[17] 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.

[18] This independence assumption is made also for the sake of simplicity. Essentially the same argument will go through even if there is correlation. The result for a fully general case of correlated first-order beliefs is available upon request.

- a_3: the object is allocated to agent 1 for a price of 2γ;
- 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):

$$\begin{array}{rcrrrl} V_1(0,q) & = & (2\gamma(1-q), & 2\gamma(1-q)-\gamma, & 2\gamma(1-q)-2\gamma, & 0,\ldots,0) \\ V_1(1,q') & = & (1+2\gamma(1-q'), & 1+2\gamma(1-q')-\gamma, & 1+2\gamma(1-q')-2\gamma, & 0,\ldots,0) \end{array}$$

When $\gamma \in (0, 1/2)$, it can be easily checked that none of these vectors are positive affine transformations of one another. Thus, Δ-TD 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 B&M (2009).

In contrast, suppose now that $1 \geq \gamma \geq 1/2$.[19] For this case, the claim in B&M (2009) is that RVI 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 B&M (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 Δ-TD, and every SCF satisfies Δ-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 Δ-TD 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 beliefs. Subsection 4.3 and Appendix A elaborate on this.

Having said that, in conjunction with other standard conditions, such as convexity of the set of first-order beliefs, Δ-TD 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

[19] Recall 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 $\gamma \in (0, 1/(n-1))$.

sense, while Δ-TD is still generic in continuum settings, it is much less appealing in some of them.

4.2 The Example beyond Δ-type diversity

In this subsection we address what happens in the example in environments that do not satisfy Δ-TD. Again, to simplify our expressions, suppose that $n = 3$. Then, the relevant range for γ is $[1/2, 1]$. For an environment to violate Δ-TD, 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 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 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 $q' - q = 1/(2\gamma)$). 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, $(0, q)$ and $(1, q')$, such that $q' - q = 1/(2\gamma)$. 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 $j \neq i$, 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 Δ-measurability, because Θ_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 $(1, q')$ prefers y_i to x_i, we compute interim utilities of these two types:

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

$$\begin{aligned} V_i(y_i|(1,q')) &= q'[(1-q')(1+\gamma)+q'] \\ V_i(x_i;(0,q)|(1,q')) &= 1/2(1-q')[(1-q')(1+2\gamma)+q'(1+\gamma)]. \end{aligned}$$

Note that, as $q' - q = 1/(2\gamma)$ and $\gamma \in [1/2, 1]$, it follows that $q' \geq 1/2 \geq 1 - q'$ and $q \leq 1/2 \leq 1 - q$. Then $(0, q)$ prefers x_i to y_i because $(1-q)\gamma + 1/2 \times q\gamma > (1-q)\gamma$, while $(1, q')$ prefers y_i to x_i because $(1-q')(1+\gamma)+q' > (1-q')(1/2+\gamma)+q'\times(1+\gamma)/2$. Thus, these two first-order types would self-reveal themselves if offered the choice between x_i and y_i.

Note that x_i and y_i separate any pair of first-order types $(0, q)$ and $(1, q')$ such that $q' - q = 1/(2\gamma)$. 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 in Θ_i for every agent i, even for $\gamma \in [1/2, 1]$. Therefore, RVI is not restricted at all by Δ-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 $x_i^1 \in \{\ell_i(t_i)\}_{t_i}$ to separate the different classes of equivalent first-order types in the first iteration of the measurability algorithm. Further, one can find $x_i^2 \in \{x_i, y_i\}$ to separate the first-order types that form non-singleton atoms of the partition in Θ. Then, an SCF that is measurable with respect to $T_i \times \Psi_{-i}^1$, essentially $(1-\delta)x_i^1 + \delta x_i^2$ that, for $\delta > 0$ small enough, will separate all first-order types: because of the strict inequalities on the x_i^1, the first-order types that are separated in the first iteration of the algorithm stick to truth-telling for small enough δ. For the rest, each pair of first-order types that form an atom in the partition have identical preferences over constant SCFs (x_i^1 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_i^2, 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. Δ-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 $l = j, k$.[20] 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 $\psi_j \in \Psi_j^1$ and $\psi_k \in \Psi_k^1$ of partitions of T_j, T_k.

SCFs that separate $(0, q), (1, q')$ need to be measurable with respect to the partitions $\Psi_j^1 \times \Psi_k^1$. 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 ψ_j, ψ_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, RVI is very limited. It can be shown that, by imposing some

[20] For notational simplicity, we shall use below the same values of q, q' for agents i, j, k.

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 Genericity of Δ-type diversity

In this subsection we abandon the example and make a more general point. The result in B&M (2009) can be understood as uncovering robust measurability as an additional restriction for RVI beyond ex post incentive compatibility. Recall that, given a class of environments indexed by different type spaces, robust measurability amounts to A&M measurability on every type space in the class. We shall provide here an argument of genericity of the Δ-TD assumption when the sets Q_i are finite.[21]

Recall that A is a finite set consisting of K alternatives, and recall our definition of the first-order interim utility $V_i(t_i) = (V_i^k(t_i))_{k=1,\dots,K}$. Let $K \geq 3$ for this subsection (if $K = 2$, a violation of Δ-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 $V_i^k(t_i) < V_i^{k'}(t_i)$, and choosing one such pair of alternatives with extreme values, normalize expected utilities so that $V_i^k(\theta_i, q_i) = 0$, $V_i^{k'}(\theta_i, q_i) = 1$, and $V_i(\theta_i, q_i) \in [0, 1]^K$.

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 Ω, the unit cube in $\mathbb{R}^{(K-2)S}$ with vertices at the points $(0, \dots, 0)$ and $(1, \dots, 1)$. Endow Ω with the uniform metric, and define open balls using this metric relative to Ω. Since the property of Δ-TD is defined by a finite number of inequalities, one can easily see that the set of points in Ω satisfying it is an open and dense subset of Ω. That is,

- for each environment in Ω that satisfies Δ-TD, there exists an open ball around it containing only the environments in which the property is maintained, and
- for each environment in Ω that violates Δ-TD and for each open ball around it, there always exists an environment satisfying Δ-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 Ω 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 Ω 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

[21] A similar genericity argument can be provided for the infinite case; see Appendix A and also Kunimoto and Serrano (2010).

and probabilities, agents stop after a finite number of decimals). It then follows from our Theorem 3 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 Δ-TD 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, Δ-TD 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 Δ-TD is much more compelling.

5 Sufficiency for Robust Virtual Implementation

So far, we have focused on necessary conditions for RVI. In the process, we have identified Δ-incentive compatibility and Δ-measurability as relevant conditions. In the previous section, we have argued that Δ-measurability is generically a trivial condition. In this section we will establish sufficiency results for RVI. Our proof is an extension to our environment of the corresponding proof in A&M (1992b). Both follow the same sequence of arguments; therefore we omit all formal details, which can be found in the online appendix to our paper. Instead, we outline the general proof technique. The argument for the environment where Δ-TD is satisfied is simple and a straightforward extension of the A&M mechanism and also borrows from Serrano and Vohra (2005). The proof for the general environment builds upon that simpler first argument. We present both arguments in sequence.

To establish our sufficiency results, we introduce two new assumptions. First, we assume that the set of first-order beliefs for every agent is finite. Second, we introduce the following assumption on environments:

Definition 9 (Quasi-Transferability) *An environment $\mathcal{E}$ satisfies **quasi-transferability** if there exists a collection of lotteries $\{\bar{a}_i\}_{i\in N}$ and $\{\underline{a}_i\}_{i\in N}$ in $\Delta(A)$ such that for any $\theta \in \Theta$,*

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

Remark: This is an exact analogue of A&M's (1992b) 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)).

To avoid unnecessary details, we will assume that f satisfies strict Δ-incentive compatibility (this assumption is not made in the online appendix). For the sufficiency argument, we construct a mechanism $\Gamma = (M, g)$, where $M = \times_{i \in N} M_i$,

$$M_i = M_i^0 \times M_i^1 \times \cdots \times M_i^J = \underbrace{T_i \times T_i \times \cdots \times T_i}_{J+1},$$

that is, each agent reports her first-order type $J+1$ times, and the outcome function $g(\cdot)$ consists of the following parts (this is exactly how A&M's (1992b) mechanism can also be described): for any $m \in M$,

$$\begin{aligned} g(m) &= \varepsilon \times \text{ separation function}(m^0) \\ &+ \varepsilon^2 \times \text{ punishment function}(m^0, \ldots, m^J) \\ &+ (1 - \varepsilon - \varepsilon^2) \times \frac{1}{J} \sum_{j=1}^{J} f(\hat{\theta}(m^j)) \end{aligned}$$

where $\hat{\theta}(m^j) \in \Theta$ denotes the payoff type component of m^j. The separation function, which only depends on the first report m^0, allows to distinguish different first-order types of each player. We can think of it as a menu of SCFs (lotteries in environments satisfying Δ-TD) such that an agent who has two first-order types with distinct payoff types will always select two different SCFs from the menu. Hence, if there were no other components, we would have had the truthful revelation of payoff types. Yet, as our goal is not to separate agents' payoff types, but to implement the SCF f, we add the two other components.

First observe that in quasi-transferable environments, for any $\eta > 0$, it is possible to construct two lotteries $\bar{a}_i[\eta]$ and $\underline{a}_i[\eta]$ so that for every agent $i \in N$ and $\theta \in \Theta$, $0 < u_i(\bar{a}_i[\eta]; \theta) - u_i(\underline{a}_i[\eta]; \theta) \leq \eta$ (the full formal details of this claim as well as the relevant discussion are provided in the online appendix). The second component punishes the agent for the inconsistency between her first report m^0 and a subsequent report $(m^1, \ldots, m^J)$; the punishment is only applied to the agent with the earliest inconsistent report. In that case only, any such agent i receives a lottery $\underline{a}_i[\eta]$ from the punishment component and all other agents i receive a lottery $\bar{a}_i[\eta]$. Since $\eta > 0$ can be chosen arbitrarily small, the size of the punishment can also be made very small. This is the way we construct the punishment function. This punishment alone could have changed the incentives of the agents to report truthfully in the separation component, but our weighting of these two components (ε and ε^2, respectively) guarantees that the incentives are preserved: the agent has higher utility from reporting truthfully in m^0 and taking the punishment than from changing m^0.

The last term is the SCF f that the planner wants to implement, which is split into J identical pieces (thus it depends on $m^1, \ldots, m^J$ messages). Consider one such piece $f(m^j)$. Assume that all agents report truthfully all the way to the $(j-1)$th message $(m^1, \ldots, m^{j-1})$ (this is our induction hypothesis). If all agents report truthfully at m^j, strict Δ-incentive compatibility guarantees that any deviation from truth-telling

is strictly worse for any agent. On the other hand, if agent $i^{'}$ deviates at the jth report m^j, some agent i (possibly $i = i^{'}$) may obtain a better outcome from this piece of f by deviating as well. By doing so, agent i will not only receive a better outcome, but also a punishment. The assumption of quasi-transferability is used in the punishment component, which allows to punish the deviator without inflicting "excessive" damage on the other agents.[22] Therefore, choosing a large J guarantees that punishment is costlier than the benefit from the deviation in the j-th small piece of f.

This mechanism is a close adaptation of the A&M (1992b) mechanism and the arguments are parallel to the ones thereof. However, our analysis also exhibits some differences. 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 since the SCF f only depends on payoff types and the misrepresentation of q_i into $q_i^{'}$ with the same payoff type θ_i is of no significance.

While the proof under Δ-TD is essentially identical to the proof in A&M (Section 3, 1992b), the mechanism under Δ-TD is very transparent and serves as a useful springboard to our further adaptation of the mechanism to the general environment. We only state the result here and make an argument required to apply A&M's proof. Those readers interested in the formal proof are referred to the online appendix.

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

Under Δ-TD, following the arguments in Lemma 1 of Serrano and Vohra (2005), we identify, for each $i \in N$, the family of lotteries $\{\ell_i(\theta_i, q_i)\}$. If an agent i of first-order type (θ_i, q_i) were asked to choose the most preferred lottery out of that set, each will uniquely choose the corresponding $\ell_i(\theta_i, q_i)$. These lotteries form the separation component of $g(m)$. The formal argument for the construction of these lotteries can be found in the online appendix. Note that these lotteries do not depend on the other agents' messages. Also, the menu of these lotteries separates all first-order types. Since all we need is to separate payoff types, multiplicity of Δ-rationalizable strategies in the mechanism is possible, where misrepresentations of the first-order belief may happen.

The proof for the general case that does not require Δ-TD has three key differences. First, each agent i's message space M_i^j becomes the partition of T_i, which is the finest partition generated by the measurability algorithm. Second, the separation function is a weighted sum of subcomponents constructed separately for each iteration of the same measurability algorithm. Third, to prove that the mechanism we construct indeed virtually implements an SCF f, we use double mathematical

[22] More specifically, quasi-transferability guarantees that for any $i, i^{'} \in N$ with $i \neq i^{'}$ and $\theta \in \Theta$, $u_i(\underline{a}_{i'}[\eta]; \theta) \geq u_i(\bar{a}_{i'}[\eta]; \theta)$.

induction, both on the message components (as in Section 3 of A&M (1992b) and as in our previous step) and on the steps of the measurability algorithm.

In the general environment, we need to rely on the messages of other players when we construct a separating function for agent i. Our construction of the separating function mirrors the steps of the measurability algorithm, which achieves further separation in step h by using "intermediate" separation in step $h-1$. Hence, the separating function is:

$$\text{separation function}(m^0) = \frac{1}{n}\sum_{i\in N}\left(\frac{1}{1+\delta+\delta^2+\cdots+\delta^L}\sum_{h=0}^{L}\delta^h x_i^h[\Psi_i^h(m_i^0)](\hat{\theta}(m^0))\right)$$

where $\delta > 0$ and a function $x_i^h[\Psi_i^h(m_i^0)] : \Theta \mapsto \Delta(A)$ is an SCF that separates first-order types of agent i given the separation that has already been achieved (hence the index $[\Psi_i^h(m_i^0)]$). Note also the increasing weights for each component; the choice of small δ guarantees that further separation does not interfere with the separation achieved in the earlier steps. This is formalized in the following claim, Claim 4.1 from the online appendix:

Claim 4.1: Suppose that $m \in S^{\tilde{\Gamma}}(\theta)$. 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, $m_i^0 = \Psi_i^*(\theta_i, q_i)$ with some $q_i \in Q_i$.

The other two components of the outcome function are as in A&M (1992b) and as in our previous step. The previous claim is key in establishing the following characterization:

Theorem 4 (A Characterization of RVI) *Suppose an environment $\mathcal{E}$ satisfies quasi-transferability and that the sets Q_i are finite. An SCF f is* ***robustly virtually implementable*** *if and only if it satisfies Δ-incentive compatibility and Δ-measurability.*

Remark: If the sets Q_i are not finite, an adaptation of the *maximally revealing mechanism* of B&M (2009) to the appropriately restricted notion of measurability would provide the proof of Theorem 4 for this case. In particular, we can use the adapted maximally revealing mechanism as the separation component of $g(m)$ and M_i^j is set as the partition over T_i generated by the appropriately restricted measurability algorithm. 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 Δ-TD. Of course, the general characterization of RVI so obtained, by means of Δ-incentive compatibility and Δ-measurability for any arbitrary collection of sets Q_i, boils down to the characterization theorem in B&M (2009), in terms of ex post incentive compatibility and robust measurability, when the set of first-order beliefs is unrestricted, i.e., $Q_i = \Delta(\Theta_{-i})$ for every $i \in N$.

6 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 RVI is often as powerful as it can possibly be. Indeed, with Δ-type diversity, the limits of implementation are given by Δ-incentive compatibility, but every Δ-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. Both are subject to it, 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 Subsections 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 (B&M (2011)) is often quite limiting, we have argued that a robust version of measurability is not. Indeed, in our settings, Δ-measurability – A&M measurability over the allowed type spaces – is a trivial condition if it is imposed over *almost* every type space.

Appendix A: Genericity of Δ-TD in Continuum Settings

We extend here our argument on the genericity of Δ-TD 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 $\mathcal{E}_\Delta = (A, \Theta_i, u_i)_{i \in N}$ 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 Θ. See also Section 4.3 for the argument.

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

Theorem 5 *Suppose that the payoff environment $\mathcal{E}_\Delta = (A, \Theta_i, u_i)_{i \in N}$ 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 Δ-TD holds.[23]

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

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

Then, we claim the following:

Claim A: For any $\theta_i^\ell, \theta_i^m \in \Theta_i$ with $\theta_i^\ell \neq \theta_i^m$ 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 $q_i^\epsilon \in V$, $V_i(\theta_i^\ell, q_i) \neq V_i(\theta_i^m, q_i^\epsilon)$. 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 $\{q_i^\lambda\}_{\lambda=1}^\infty$ as such a countable dense subset in $\Delta^0(\Theta_{-i})$. Define

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

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 Δ-TD 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 Δ-TD holds. Thus, this set Δ^* is a residual set in $\Delta^0(\Theta_{-i})$. This completes the proof.■

[23] 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.

References

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

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

Abreu, D. and A. Sen (1991): Virtual Implementation in Nash Equilibrium, Econometrica, 59, 997-1021.

Artemov, G., T. Kunimoto, and R. Serrano (2012): Online Appendix to "Robust Vitual Implementation: Toward a Reinterpretation of the Wilson Doctrine," Unpublished manuscript, available at `https://sites.google.com/site/takashikunimoto9/research`.

Battigalli, P. and M. Siniscalchi (2003): Rationalization and Incomplete Information, Advances in Theoretical Economics, 3 (1), Article 3.

Battigalli P., A. Di Tillio, E. Grillo and A. Penta (2011): Interactive Epistemology and Solution Concepts in Games with Asymmetric Information, *The BE Journal of Theoretical Economics*, vol. 11(Advances), Article 6.

Bergemann, D. and S. Morris (2011): Robust Implementation in General Mechanisms, Games and Economic Behavior 71, 261-281.

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

Bergemann, D. and S. Morris (2012): *Robust Mechanism Design,* World Scientific Press, forthcoming.

Chung, K. S. and J. Ely (2001): Efficient and Dominant Solvable Auctions with Interdependent Valuations, Discussion Paper, Northwestern University.

Cremer, J. R. McLean (1988): Full Extraction of the Surplus in Bayesian and Dominant Strategy Auctions, Econometrica, 56, 1247-1257.

Dekel, E., Fudenberg, D., and S. Morris (2007): Interim Correlated Rationalizability, Theoretical Economics, 2, 15-40.

Duggan, J. (1997): Virtual Bayesian Implementation, Econometrica, 65, 1175-1199.

Heifetz, A. and Z. Neeman (2006): On the Generic (Im)Possibility of Full Surplus Extraction in Mechanism Design, Econometrica, 74, 213-233.

Jackson, M. O. (2001): A Crash Course in Implementation Theory, Social Choice and Welfare, 18, 655-708.

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

Kunimoto, T. and R. Serrano (2010): Evaluating the Conditions for Robust Mechanism Design, Working Paper 2010-06, Department of Economics, Brown University.

Kunimoto, T. and R. Serrano (2011): A New Necessary Condition for Implementation in Iteratively Undominated Strategies, Journal of Economic Theory, 146, 2483-2495.

Matsushima, H. (1988): A New Approach to the Implementation Problem, Journal of Economic Theory 45, 128-144.

Neeman, Z. (2004): The Relevance of Private Information in Mechanism Design, Journal of Economic Theory, 117, 55-77.

Palfrey, T. R. and S. Srivastava (1993): Bayesian Implementation, Harwood Academic Publishers, New York.

Saijo, T., T. Sjöström and T. Yamato (2007): Secure Implementation, Theoretical Economics, 2, 203-229.

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

Serrano, R. and R. Vohra (2001): Some Limitations of Virtual Bayesian Implementation, Econometrica, 69, 785-792.

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

Wilson, R. (1987): Game Theoretic Analysis of Trading Processes, in Advances in Economic Theory, ed. by T. Bewley, Cambridge University Press.