WORKING PAPERS

N° TSE-514

April 2014

"Strategic and structural uncertainties in robust implementation"

Takuro Yamashita

Toulouse School of Economics

Strategic and structural uncertainties
in robust implementation[*]

Takuro Yamashita[†]

April 25, 2014

Abstract

This paper discusses some connections among several robustness concepts of mechanisms in terms of agents' behaviors. Specifically, under certain conditions such as private values and "rich" interdependent values, we show that implementation in (one-round or iterative) undominated strategies, a solution concept robust to strategic uncertainty, is equivalent to Bayesian implementation with arbitrary type spaces, a solution concept robust to structural uncertainty.

[†]Toulouse School of Economics. takuro.yamashita@tse-fr.eu

1

1 Introduction

In mechanism design or implementation theory, one often models the strategic interactions of agents in a given mechanism as a Bayesian game by introducing a *type space* as in Harsanyi (1967-68), which induces the agents' (high-order) beliefs over their payoff-relevant private information.[1]

Some recent studies examine "robust" mechanism design, by investigating which social objectives are implementable if the mechanism designer has little information about the agents' high-order beliefs. Roughly, an objective is said to be "robustly implementable" if, given whatever type space is given, a mechanism Bayesian implements the objective (e.g., Bergemann and Morris (2005) and Bergemann and Morris (2011)). In this approach, the agents can have arbitrary high-order beliefs about the states of the world, while they are still assumed to play a Bayesian equilibrium given each type space. In this sense, this approach considers robustness with respect to the agents' *structural uncertainties*.

Another robustness notion often studied in the literature is with respect to *strategic uncertainties* of the agents.[2] In this approach, each agent is assumed to be "rational" in the sense that he does not play any "dominated" strategy, but he may play any "undominated" strategy. In particular, the agents may play some non-equilibrium strategy profiles, and in this sense, this strategic uncertainty approach considers a different aspect of the robustness of mechanisms.

The objective of this paper is to contribute to the literature on the relationships between those different kinds of robustness in the context of implementation, by showing that some of the solution concepts in these two different approaches are "close" to each other in terms of the implementable

---

[1]A typical additional assumption is that the agents have a common prior over their payoff-relevant private information, and so there is no higher-order uncertainty. In this paper, when we refer to a type space, it does not necessarily have a common prior.

[2]The terms "structural uncertainty" and "strategic uncertainty" appear in, for example, Morris and Shin (2002).

objectives.

To be more specific, we first compare *Bayesian implementation given any type space* as an implementation concept in the structural uncertainty approach, and *implementation in undominated strategies* as an implementation concept in the strategic uncertainty approach. Bayesian implementation requires that a desirable outcome is induced given whatever Bayesian equilibrium is played.[3] Implementation in undominated strategies requires that a desirable outcome is induced whenever each agent plays any strategy that is not strictly dominated.[4] Note that we do not allow the iterative elimination of dominated strategies, and in this sense, do not assume mutual or common knowledge of rationality among the agents.

We show that, if a social choice correspondence is implementable in undominated strategies, then it is Bayesian implementable given any type space. Conversely, in a private-value environment (i.e., one where each agent knows his preference over allocations), the other direction is also true: if a social choice correspondence is Bayesian implementable given any type space, then it is implementable in undominated strategies (Theorem 1). Therefore, even though we consider different uncertainties, robustness to one kind of uncertainty implies robustness to the other kind. Although the private-value assumption is restrictive, it may be considered reasonable in various contexts, including certain private-goods auctions, cost sharing for public good provision, preference aggregation in voting, and some two-sided matching

---

[3]Because the concept of structural uncertainty refers to the designer's assumption about the agents' possible high-order beliefs, it is not necessarily solely for Bayesian implementation. Indeed, several papers in game theory investigate the implications of structural uncertainty with solution concepts other than Bayesian equilibria, such as rationalizability (see, for example, Battigalli and Siniscalchi (2003) and Bergemann and Morris (2007)). In this paper, we consider Bayesian implementation as a solution concept that does not exhibit strategic uncertainty, in order to compare with implementation robust to strategic uncertainty.

[4]Although we consider strict dominance to define the "rational" behaviors of the agents, analogous results would hold for some different notions of rationality, such as that based on weak dominance. See Remark 3 in Section 4.

problems.

In Section 4, we generalize some of the results to an environment with interdependent values. With interdependent values, we have the equivalence between implementation in iteratively undominated strategies and Bayesian implementation given any type space, as in Bergemann and Morris (2011). However, implementation in one-round undominated strategies is in general more demanding than these concepts. Nevertheless, we obtain an analogous result as in Theorem 1 if each agent's signal space is sufficiently "rich".

In Section 5, we examine *partial* Bayesian implementation given any type space as a less demanding implementation concept in the structural uncertainty approach, which requires that, for each given type space, there is *at least one* Bayesian equilibrium that induces a desirable outcome. With private values and certain technical conditions, we show that, if a social choice correspondence is partially Bayesian implementable given any type space, then it is *virtually* implementable in undominated strategies, where the term "virtual" is in the sense of Abreu and Matsushima (1992) and Abreu and Matsushima (1991), i.e., a correspondence that is arbitrarily close to the desirable correspondence is implementable in undominated strategies (Theorem 3).[5] In this sense, implementation in (one-round/iteratively) undominated strategies, Bayesian implementation, and partial Bayesian implementation are all "close" to each other in the private-value environment. Thus, Bayesian incentive compatibility conditions given any type space have "virtually" as strong implications as implementation in undominated strategies. However,

---

[5]An analogous result is established by Bergemann and Morris (2009b) for *social choice functions with interdependent values*. Our result is a complement to their findings in establishing an analogous result for *social choice correspondences with private values*. From a technical viewpoint, the proof of our result trivializes the issue of their *strategic distinguishability* by the private-value assumption. On the other hand, the proof for their result uses the property that Bayesian incentive compatibility given any type space implies ex post incentive compatibility (Bergemann and Morris (2005)), which does not hold for social choice correspondences. Because of these differences, neither of these proofs implies the other.

4

this virtual equivalence result does not generally hold with interdependent values.

2 Private-value model

The set of agents is denoted by $I = \{1, \ldots, N\}$. An allocation is denoted by $x \in X$. Each agent $i \in I$ has payoff-relevant private information $\theta_i \in \Theta_i$, which we call i's *payoff type*. We assume that each Θ_i is finite. We denote the profile of payoff types by $\theta = (\theta_i)_{i \in I}$, and the set of all payoff type profiles by $\Theta = \prod_{i \in I} \Theta_i$.

Agent i's preference for each x is represented by a utility function $u_i(x, \theta_i)$, which does not depend on θ_{-i}, and hence is a *private-value* environment. In Section 5, we discuss how the results would change with interdependent values, i.e., when i's preference also depends on θ_{-i}.

The objective of the mechanism designer is represented by a social choice correspondence $F : \Theta \to 2^X$, which assigns a subset of X for each $\theta \in \Theta$. We interpret $F(\theta) \subseteq X$ as the set of desirable outcomes in state $\theta \in \Theta$.

A mechanism is denoted by $\Gamma = \langle M, g \rangle$, where $M = \prod_i M_i$, each M_i is a finite set of messages for agent i, and $g : M \to X$ is an outcome function.[6]

2.1 Implementation based on dominance

We first introduce implementation in undominated strategies as an implementation concept that is robust to the agents' strategic uncertainty. In mechanism $\Gamma = \langle M, g \rangle$, for each i and θ_i, we say that $m_i \in M_i$ is *strictly*

---

[6]Throughout the paper, we only consider finite mechanisms. The author believes that some of the results could hold for a class of infinite mechanisms, such as those with compact message spaces and a continuous outcome function.

dominated for θ_i if there exists $\mu_i \in \Delta(M_i)$ such that,[7] for any $m_{-i} \in M_{-i}$,

$$\sum_{m_i' \in M_i} u_i(g(m_i', m_{-i}), \theta_i)\mu_i(m_i') > u_i(g(m_i, m_{-i}), \theta_i).$$

We say that m_i is *undominated for* θ_i if it is not strictly dominated for θ_i. The set of undominated messages for θ_i is denoted by $M_i^U(\theta_i)$. Hence,

$$
\begin{aligned}
M_i^U(\theta_i) \;=\; \{ m_i \in M_i | \nexists \mu_i \in \Delta(M_i); \; \forall m_{-i} \in M_{-i}, \\
\sum_{m_i' \in M_i} u_i(g(m_i', m_{-i}), \theta_i)\mu_i(m_i') > u_i(g(m_i, m_{-i}), \theta_i).\}.
\end{aligned}
$$

Definition 1. A mechanism Γ *implements F in undominated strategies* if for each θ and $m \in M^U(\theta)$, we have $g(m) \in F(\theta)$.

We say that F is *implementable in undominated strategies* if some mechanism implements F in undominated strategies. Implementation in undominated strategies requires that, given whatever undominated actions the agents play in any state, the induced outcome is desirable.[8]

The concept of implementation in undominated strategies would be relevant in a situation where the designer assumes that each agent is "rational" in the sense of not playing strictly dominated strategies, but does not assume that the agents mutually or commonly know their rationality.

If we additionally assume common knowledge of rationality, then the iterative elimination of strictly dominated strategies would be possible. We introduce implementation based on the iterative elimination as follows.

---

[7]For a finite set S, $\Delta(S)$ denotes the set of all probability mass functions over S, so that, for each $\alpha \in \Delta(S)$, $\alpha(s) \in [0, 1]$ denotes the probability that α assigns to $s \in S$, and $\sum_{s \in S} \alpha(s) = 1$. We treat $\Delta(S)$ as a subspace of an $|S|$-dimensional Euclidean space.

[8]Our implementation concept is sometimes called *weak* implementation (in undominated strategies). This is more demanding than *partial* implementation, where at least one strategy profile of the solution concept yields a desirable outcome. On the other hand, it is less demanding than *full* implementation, which requires that $\{g(m)|m \in M^U(\theta)\} = F(\theta)$ (while the weak implementation only requires $\{g(m)|m \in M^U(\theta)\} \subseteq F(\theta)$).

In a mechanism $\Gamma = \langle M, g \rangle$, for each i, θ_i, let $M_i^0(\theta_i) = M_i^U(\theta_i)$. Then, for each $k = 1, 2, \ldots$, given $(M_j^{k-1}(\theta_j))_{\theta_j \in \Theta_j, j \in I}$, we define $M_i^k(\theta_i)$ as follows.

$$
\begin{aligned}
M_i^k(\theta_i) \;=\; & \{m_i \in M_i^{k-1}(\theta_i) | \nexists \mu_i \in \Delta(M_i^{k-1}(\theta_i)); \; \forall (\theta_{-i}, m_{-i}) \text{ s.t. } m_{-i} \in M_{-i}^{k-1}(\theta_{-i}), \\
& \sum_{m_i'} \mu_i(m_i') u_i(g(m_i', m_{-i}), \theta_i) > u_i(g(m_i, m_{-i}), \theta_i) \}.
\end{aligned}
$$

We define $M_i^\infty(\theta_i) = \bigcap_{k=0}^\infty M_i^k(\theta_i)$, and $M^\infty(\theta) = \prod_{i \in I} M_i^\infty(\theta_i)$.

Definition 2. A mechanism Γ *implements F in iteratively undominated strategies* if for each θ and $m \in M^\infty(\theta)$, we have $g(m) \in F(\theta)$.

2.2   Bayesian implementation with arbitrary type spaces

Another implementation concept that considers robustness to the agents' structural uncertainties assumes that the agents play a Bayesian equilibrium, which means that they predict each other's strategy correctly (and so there is no strategic uncertainty), but they can have arbitrary high-order beliefs about each other's payoff type. Such high-order beliefs are modeled as "types" as follows.

For each i, let T_i be a countable set, and we call $t_i \in T_i$ his *type*. There exist two mappings (τ_i, β_i) such that, for each $t_i \in T_i$, his payoff type is given by $\tau_i(t_i) \in \Theta_i$, and his *belief type* is given by $\beta_i(t_i) \in \Delta(T_{-i})$. A tuple $\mathcal{T} = (T_i, \tau_i, \beta_i)_{i=1}^N$ is called a (countable) *type space*.[9]

A mechanism $\Gamma = \langle M, g \rangle$ induces a Bayesian game with type space $\mathcal{T}$, and each agent i's (mixed) strategy is given by a mapping $\sigma_i : T_i \to \Delta(M_i)$.

Definition 3. In a mechanism $\Gamma = \langle M, g \rangle$, a strategy profile $\sigma = (\sigma_i)_{i=1}^N$ is

---

[9]Throughout the paper, unless explicitly mentioned, a type space means a countable type space. One merit of this approach over uncountably infinite type spaces is notational simplicity and existence of Bayesian equilibria in finite games. With uncountably infinitely many types, even in a finite game, a Bayesian equilibrium may fail to exist (see, for example, Brandenburger, Friedenberg, and Keisler (2008)).

a *Bayesian equilibrium given type space* $\mathcal{T}$ if for each i, t_i, m_i,

$$
\sum_{t_{-i}} \sum_{m'_{-i}} \sum_{m'_i} u_i(g(m'_i, m'_{-i}), \tau_i(t_i)) \sigma_i(m'_i | t_i) \sigma_{-i}(m'_{-i} | t_{-i}) \beta_i(t_{-i} | t_i)
$$
$$
\geq \sum_{t_{-i}} \sum_{m'_{-i}} u_i(g(m_i, m'_{-i}), \tau_i(t_i)) \sigma_{-i}(m'_{-i} | t_{-i}) \beta_i(t_{-i} | t_i).
$$

Definition 4. We say that a mechanism Γ *Bayesian implements F given type space* $\mathcal{T}$ if, for each Bayesian equilibrium σ^* given $\mathcal{T}$, for each $t \in T$ and $m \in M$ with $\sigma^*(m|t) > 0$, we have $g(m) \in F(\tau(t))$.

We say that F is *Bayesian implementable* given $\mathcal{T}$ if some mechanism Bayesian implements F given $\mathcal{T}$.

When the designer does not know which $\mathcal{T}$ is the true type space, then he may aim to design a mechanism so that the mechanism Bayesian implements F given *any* type space $\mathcal{T}$.

3 Equivalence

We observe the equivalence of the three concepts in terms of implementable social choice correspondences.

Theorem 1. The following three statements are equivalent.

1. F is implementable in undominated strategies.

2. F is implementable in iteratively undominated strategies.

3. F is Bayesian implementable given any type space $\mathcal{T}$.

The equivalence between statements 2 and 3 is based on some epistemic results established by Bergemann and Morris (2011) for social choice func-

tions.[10] We can straightforwardly extend their proof for social choice correspondences, which is provided in the appendix for the sake of completeness.[11]

Proof. For the equivalence between statements 2 and 3, see the proof of Proposition 1 in the appendix. We prove that statement 2 implies statement 1. The other direction is obvious.

Suppose that F is implemented in iteratively undominated strategies by a mechanism $\Gamma = \langle M, g \rangle$. That is, for each θ and $m \in M^\infty(\theta)$, we have $g(m) \in F(\theta)$. It suffices to show that F is in fact implementable in (one-round) undominated strategies. The following lemma that holds true in private-value environments is crucial.

Lemma 1. For each i, θ_i, and $m_i \notin M_i^\infty(\theta_i)$, there exists $\mu_i \in \Delta(M_i^\infty(\theta_i))$ such that, for any $m_{-i} \in M_{-i}^\infty = \prod_{j \neq i}(\bigcup_{\theta_j} M_j^\infty(\theta_j))$,

$$\sum_{m_i} \mu_i(m_i) u_i(g(m_i, m_{-i}), \theta_i) > u_i(g(m_i, m_{-i}), \theta_i).$$

Proof. (of the lemma)

Fix i, θ_i, and $m_i \notin M_i^\infty(\theta_i)$. By finiteness of Γ, there is an integer K such that $M_i^K(\theta_i) = M_i^\infty(\theta_i)$ for any i, θ_i, and thus,

$$M_i^K(\theta_i) = \{\tilde{m}_i \in M_i^K(\theta_i) | \nexists \mu_i \in \Delta(M_i^K(\theta_i)); \ \forall(\theta_{-i}, m_{-i}) \text{ s.t. } m_{-i} \in M_{-i}^K(\theta_{-i}),$$
$$\sum_{m_i'} \mu_i(m_i') u_i(g(m_i', m_{-i}), \theta_i) > u_i(g(\tilde{m}_i, m_{-i}), \theta_i)\}.$$

Thus, for $m_i \notin M_i^K(\theta_i)$, there exists $\mu_i \in \Delta(M_i^K(\theta_i))$ such that, for any θ_{-i} and $m_{-i} \in M_{-i}^K(\theta_{-i})$,

$$\sum_{m_i'} \mu_i(m_i') u_i(g(m_i', m_{-i}), \theta_i) > u_i(g(m_i, m_{-i}), \theta_i).$$

---

[10] The proof is based on the equivalence between the collection of Bayesian equilibria of all type spaces and the collection of *iteratively* undominated strategy profiles in any (finite) game. See Battigalli and Siniscalchi (2003) and Bergemann and Morris (2007) who show such an equivalence result in game theory.

[11] The relationship between the statements 2 and 3 becomes less clear if we allow uncountably infinite type spaces (e.g., a universal type space), because of the existence issue of Bayesian equilibria with such type spaces. See also Footnote 9.

Fix an arbitrary $m_{-i} \in M_{-i}^K = \bigcup_{\theta'_{-i} \in \Theta_{-i}} M_{-i}^K(\theta'_{-i})$. Then there exists θ_{-i} such that $m_{-i} \in M_{-i}^K(\theta_{-i})$. Then,

$$\sum_{m'_i} \mu_i(m'_i) u_i(g(m'_i, m^*_{-i}), \theta_i) > u_i(g(m_i, m^*_{-i}), \theta_i),$$

which establishes the lemma. □

Define a mechanism $\Gamma' = \langle M', g' \rangle$ so that $M'_i = M_i^\infty = \bigcup_{\theta_i} M_i^\infty(\theta_i)$ for each i, and $g' : M' \to X$ is the restriction of $g : M \to X$ on $M'(\subseteq M)$. Then by Lemma 1, for each i, θ_i, if $m_i \notin M_i^\infty(\theta_i)$, then m_i is (one-round) strictly dominated. Thus, $M_i'^U(\theta_i)$, the set of undominated messages for θ_i in mechanism Γ', is a subset of M_i^∞, which implies that Γ' implements F in undominated strategies. □

Remark 1. It may be worth noting that, in general, social choice correspondences that are implementable in either of these approaches are not necessarily dominant-strategy incentive compatible (or more precisely, they do not necessarily have subcorrespondences that are dominant-strategy incentive compatible). In this sense, the set of implementable objectives in these approaches could be strictly larger than the set of dominant-strategy implementable objectives, one of the classical "robust" approaches in the literature.[12]

Remark 2. Our equivalence result also implies that the findings in one approach can apply to the other approach. For example, as shown by Bergemann and Morris (2005), Chung and Ely (2007), and Börgers and Smith (2012) in various contexts, under certain conditions, Bayesian implementability given large type spaces implies dominant-strategy incentive compatibility. Such results would hold true as well for implementation in undominated strategies.[13] Conversely, Yamashita (2012) shows that a condition called the

---

[12]See, for example, Bergemann and Morris (2005) and Yamashita (2012).

[13]Indeed, for example, Jackson (1992) shows that a social choice function that is implementable in undominated strategies must be dominant-strategy incentive compatible,

"chain dominance property" is necessary for implementation in undominated strategies, and this would also hold true for Bayesian implementation given any type space. Such a result might be useful as a "detour" for studying robust mechanisms in terms of structural uncertainties, where the Bayesian incentive compatibility constraints given large type spaces could be highly multidimensional and complicated.

4 Interdependent values

For some economic problems, it may be more natural to model the situation as one with interdependent values, i.e., agent i's preference given θ_i is represented by a utility function $u_i(x, \theta)$ that can vary with θ_{-i}. Even with this general specification of the agents' preferences, some of the results obtained in this paper still hold true.

In the following, we redefine each implementation concept with interdependent values. We say that, for each i, θ_i, a message $m_i \in M_i$ in mechanism $\Gamma = \langle M, g \rangle$ is strictly dominated, if there is $\mu_i' \in \Delta(M_i)$ such that, given any θ_{-i} and $m_{-i} \in M_{-i}$,

$$\sum_{m_i'} u_i(g(m_i', m_{-i}), \theta) \mu_i(m_i') > u_i(g(m_i, m_{-i}), \theta).$$

That is, μ_i' is strictly a better response than m_i, given whatever payoff types the others have, and whatever actions the others play. Let $M_i^U(\theta_i)$ be the set of all messages that are not strictly dominated for i with θ_i, i.e.,

$$M_i^U(\theta_i) = \{m_i \in M_i | \nexists \mu_i \in \Delta(M_i); \ \forall \theta_{-i} \in \Theta_{-i}, m_{-i} \in M_{-i},$$
$$\sum_{m_i'} u_i(g(m_i', m_{-i}), \theta_i) \mu_i(m_i') > u_i(g(m_i, m_{-i}), \theta_i).\}.$$

while Bergemann and Morris (2005) show that Bayesian implementation of a social choice function (or more generally, a *separable* correspondence) implies its dominant-strategy incentive compatibility.

We say that mechanism Γ *implements F in undominated strategies* if for each θ and $m \in M^U(\theta) = \prod_i M_i^U(\theta_i)$, we have $g(m) \in F(\theta)$.

Iterative elimination is defined in an analogous way. Let $M_i^0(\theta_i) = M_i^U(\theta_i)$ for each i, θ_i. Then, for each $k = 1, 2, \ldots$, given $(M_j^{k-1}(\theta_j))_{\theta_j \in \Theta_j, j \in I}$, we define $M_i^k(\theta_i)$ as follows.

$$M_i^k(\theta_i) = \{m_i \in M_i^{k-1}(\theta_i) | \nexists \mu_i \in \Delta(M_i^{k-1}(\theta_i)); \; \forall (\theta_{-i}, m_{-i}) \text{ s.t. } m_{-i} \in M_{-i}^{k-1}(\theta_{-i}),$$
$$\sum_{m_i'} \mu_i(m_i') u_i(g(m_i', m_{-i}), \theta) > u_i(g(m_i, m_{-i}), \theta)\}.$$

We define $M_i^\infty(\theta_i) = \bigcap_{k=0}^\infty M_i^k(\theta_i)$, and $M^\infty(\theta) = \prod_{i \in I} M_i^\infty(\theta_i)$. We say that mechanism Γ *implements F in iteratively undominated strategies* if for each θ and $m \in M^\infty(\theta)$, we have $g(m) \in F(\theta)$.

Bayesian implementation is also defined in an analogous fashion. In a mechanism $\Gamma = \langle M, g \rangle$, a strategy profile $\sigma^* = (\sigma_i^*)_{i=1}^N$ is a Bayesian equilibrium given type space $\mathcal{T}$ if, for any i, m_i, t_i,

$$\sum_{t_{-i}} \sum_{m_{-i}'} \sum_{m_i'} u_i(g(m_i', m_{-i}'), \tau(t)) \sigma_i^*(m_i'|t_i) \sigma_{-i}^*(m_{-i}'|t_{-i}) \beta_i(t_{-i}|t_i)$$
$$\geq \sum_{t_{-i}} \sum_{m_{-i}'} u_i(g(m_i, m_{-i}'), \tau(t)) \sigma_{-i}^*(m_{-i}'|t_{-i}) \beta_i(t_{-i}|t_i).$$

We say that a mechanism Γ *Bayesian implements F given type space $\mathcal{T}$* if for any Bayesian equilibrium σ^* given $\mathcal{T}$, for each $t \in T$ and $m \in M$ with $\sigma^*(m|t) > 0$, we have $g(m) \in F(\tau(t))$. Similarly, Γ *partially Bayesian implements F given type space $\mathcal{T}$* if there exists a Bayesian equilibrium σ^* given $\mathcal{T}$ such that, for each $t \in T$ and $m \in M$ with $\sigma^*(m|t) > 0$, we have $g(m) \in F(\tau(t))$.

Proposition 1. F is Bayesian implementable given any type space if and only if it is implementable in iteratively undominated strategies.

We provide the proof in the appendix for completeness, but it is a straightforward extension of that in Bergemann and Morris (2011).

Next, we examine the relationship between one-round and iteratively un-dominated strategies. In general, there are some social choice correspon-dences that are implementable in iteratively undominated strategies but not in (one-round) undominated strategies, as in the following example.

Example 1. Let $I = 2$, $\Theta_1 = \Theta_2 = \{-1, 1\}$, and $X = \{-2, -1, 1, 2\}$. For agent 1, $u_1(x, \theta) = x\theta_1$, and for agent 2, $u_2(x, \theta) = x\theta_1\theta_2$. Let $F(\theta) = \{2\theta_1 + \theta_1\theta_2\}$.

The following "revelation" mechanism $\Gamma = \langle \Theta, g \rangle$ implements F in itera-tively undominated strategies: $g(\theta) = 2\theta_1 + \theta_1\theta_2$ for each θ. First, for agent 1, truth-telling is strictly dominant. Given this, for agent 2, truth-telling becomes strictly dominant.[14]

However, no mechanism can implement F in (one-round) undominated strategies. Specifically, for agent 2, we have $u_2(x, \theta_1, \theta_2) = u_2(x, -\theta_1, -\theta_2)$ for any x. Therefore, we have $M_2^U(1) = M_2^U(-1)$ for any mechanism, which implies that F is not implementable in undominated strategies.

This is different to the previous case because, with iterative elimination, the first round of elimination (based on strict dominance for agent 1) restricts the set of agent 2's joint belief about θ_1 and agent 1's play. Because the value of θ_1 completely changes agent 2's preference, such a restriction could make a significant difference in agent 2's behavior.

Nevertheless, there are some environments with interdependence where an analogous result to Theorem 1 holds. One such environment is where the agents' payoff-relevant information has a multidimensional and rich structure such that mutual knowledge of rationality would not have any additional behavioral implications compared with rationality only.

Specifically, let $\Theta_i = C_i \times D_i$ for each i, where, for each $\theta_i = (c_i, d_i)$, we call $c_i \in C_i$ the *common component* and D_i the *private component* of θ_i.[15]

---

[14]For $\theta_2 = -1$, $u_2(g(\theta_1, -1), \theta_1, -1) = -\theta_1^2 > -3\theta_1^2 = u_2(g(\theta_1, 1), \theta_1, -1)$.

For $\theta_2 = 1$, $u_2(g(\theta_1, -1), \theta_1, 1) = \theta_1^2 < 3\theta_1^2 = u_2(g(\theta_1, 1), \theta_1, 1)$.

[15]These terms appear in, for example, Jehiel, Moldovanu, Meyer-ter-Vehn, and Zame (2006).

We assume that i's preference depends on $\theta_i = (c_i, d_i)$ and $c_{-i} \in C_{-i}$, but not on $d_{-i} \in D_{-i}$ (and hence, $u_i(x, \theta) = u_i(x, \theta_i, c_{-i})$). We impose the following "richness" condition.

Assumption 1. For each i, $(c_i, d_i) \in \Theta_i$, and $c_i' \in C_i$, there exist $d_i' \in D_i$ and $\psi_i : C_{-i} \to C_{-i}$ such that, for any $c_{-i} \in C_{-i}$, $u_i(\cdot, c_i, d_i, c_{-i})$ is an affine transformation of $u_i(\cdot, c_i', d_i', \psi_i(c_{-i}))$ (i.e., they exhibit the same preferences over the lotteries).

To interpret this assumption, imagine that one could observe agent i's choices, which are the best choices if i's payoff type is (c_i, d_i) (with some belief over C_{-i} and the others' choices). In this sense, the observed choices do not falsify that i's payoff type is (c_i, d_i). The assumption says that, in such a case, one cannot falsify either that i's common component is any c_i' because i would behave exactly in the same way if his private component is d_i' and his belief over C_{-i} is the one that is appropriately translated by ψ_i. Therefore, the assumption means that agent i's behavior implies no information about his common component.

For example, consider an interdependent-value auction of an oil tract, where each bidder has two sorts of private information, a noisy estimate of the amount of oil in the tract (common component c_i), and his cost of digging the well (private component d_i). Let $v_i(c, d_i) = c_i + \delta_i(\sum_{j \neq i} c_j) + d_i$ be i's willingness to pay for the tract, and $v_i(c, d_i) - p_i$ be i's utility if he wins the auction (and zero otherwise) where $p_i \in \mathbb{R}$ is his payment to the seller. Let $C_i \subseteq \mathbb{R} = D_i$.[16] In this case, the assumption is satisfied by letting $d_i' = c_i - c_i' + d_i$ and ψ_i be an identity map. To understand the implication of the assumption, suppose that agent i bids $c_i + d_i$ in a second-price auction. His bid of $c_i + d_i$ may reflect (i) that i's payoff type is (c_i, d_i) and he believes that the others' common components are all zero, or (ii) that i's payoff type is $(c_i', c_i - c_i' + d_i)$ with the same belief over C_{-i}. Therefore, i's bid does

---

[16]We adopt a continuous payoff-type space in this example to simplify the argument, but a similar conclusion is obtained even with finite Θ_i.

not induce any information about his common component.[17] Intuitively, if bidder i makes an aggressive bid, it may be because he has a high estimate of the amount of oil (i.e., c_i is high), or because he has a low estimate but his cost of digging the well is low (i.e., d_i is high). Assumption 1 says that the bidders' signal spaces are so rich that the bidders' behaviors do not reject either possibility.[18]

Theorem 2. Under Assumption 1, the following three statements are equivalent.

1. F is implementable in undominated strategies.

2. F is implementable in iteratively undominated strategies.

3. F is Bayesian implementable given any type space.

---

[17]Although i's bid induces no information about the common component (in any mechanism, as we show in Theorem 2), his bid could be informative about possible *combinations* of the private and common components, and in this sense, the design of desirable mechanisms is still nontrivial. See Yamashita (2013b) for revenue maximization in such interdependent-value auction settings when bidders may play any strategies that are not weakly dominated.

[18]The assumption can also be satisfied with a more general valuation function. For example, let $d_i = (d_{i1}, d_{i2}) \in D_i$ be two-dimensional, and $v_i = \pi_i(c)d_{i1} + d_{i2}$ be bidder i's willingness to pay, where $c_i \in C_i = [\underline{c}, \overline{c}]$, $d_{i1} \in [\underline{d}_1, \overline{d}_1] \subseteq \mathbb{R}_+$, and $d_{i2} \in [\underline{d}_2, \overline{d}_2] \subseteq \mathbb{R}$. We interpret $\pi_i(c) > 0$ as the estimated amount of oil in the tract given c, d_{i1} is i's (constant) marginal revenue from the sale of the oil, and d_{i2} is the cost of digging the well. We assume π_i is continuous and strictly increasing in every argument. Assumption 1 is satisfied if the intervals $[\underline{d}_1, \overline{d}_1]$ and $[\underline{d}_2, \overline{d}_2]$ are sufficiently large, because then we can let $d_i' = (d_{i1}', d_{i2}')$ be such that

$$\pi_i(\underline{c}, \ldots, \underline{c})d_{i1} + d_{i2} = \pi_i(\underline{c}, \ldots, \underline{c})d_{i1}' + d_{i2}',$$
$$\pi_i(\overline{c}, \ldots, \overline{c})d_{i1} + d_{i2} = \pi_i(\overline{c}, \ldots, \overline{c})d_{i1}' + d_{i2}',$$

and ψ_i be such that $\pi_i(c_i, c_{-i})d_{i1} + d_{i2} = \pi_i(c_i', \psi_i(c_{-i}))d_{i1}' + d_{i2}'$ for each c_{-i}. Note that such a ψ_i exists by continuity of π_i. For example, for each $x \in \mathbb{R}$, let $\gamma(x) \in C_i$ be such that $x = \pi_i(c_i', \gamma(x), \ldots, \gamma(x))d_{i1}' + d_{i2}'$, which exists by continuity of π_i. Then, we let $\psi_i(c_{-i}) = \gamma(\pi_i(c_i, c_{-i})d_{i1} + d_{i2})$.

Proof. The equivalence between statements 2 and 3 is Proposition 1. We show that statement 2 implies statement 1. The other statement that statement 1 implies statement 2 is obvious.

Suppose that F is implemented in iteratively undominated strategies by a mechanism $\Gamma = \langle M, g \rangle$. That is, for each θ and $m \in M^\infty(\theta)$, we have $g(m) \in F(\theta)$.

First, we observe that Assumption 1 implies the following lemma.

Lemma 2. For each i, $(c_i, d_i) \in \Theta_i$, and $c_i' \in C_i$, there exists $d_i' \in D_i$ such that, for each k, $M_i^k(c_i, d_i) \subseteq M_i^k(c_i', d_i')$.

Proof. (of the lemma) Fix i, c_i, d_i, c_i'. First, let $k = 0$.

Let $m_i \in M_i^0(c_i, d_i)$. Then, there is no $\mu_i \in \Delta(M_i)$ such that for all c_{-i} and m_{-i},

$$\sum_{m_i'} \mu_i(m_i') u_i(g(m_i', m_{-i}), c_i, d_i, c_{-i}) > u_i(g(m_i, m_{-i}), c_i, d_i, c_{-i}).$$

By Assumption 1, there exist $d_i' \in D_i$ and $\psi_i : C_{-i} \to C_{-i}$ such that, for any $c_{-i} \in C_{-i}$, $u_i(\cdot, c_i, d_i, c_{-i})$ is an affine transformation of $u_i(\cdot, c_i', d_i', \psi_i(c_{-i}))$. This implies that there is no $\mu_i \in \Delta(M_i)$ such that for all c_{-i} and m_{-i},

$$\sum_{m_i'} \mu_i(m_i') u_i(g(m_i', m_{-i}), c_i', d_i', \psi_i(c_{-i})) > u_i(g(m_i, m_{-i}), c_i', d_i', \psi_i(c_{-i})).$$

Therefore, $m_i \in M_i^0(c_i', d_i')$. In the following, we denote such d_i' by $\delta_i(c_i, d_i, c_i')$.

Now, suppose that, for every $j \in I$, $(\tilde{c}_j, \tilde{d}_j) \in \Theta_j$, and $\hat{c}_j \in C_j$, we have $M_j^{k-1}(\tilde{c}_j, \tilde{d}_j) \subseteq M_j^{k-1}(\hat{c}_j, \delta_j(\tilde{c}_j, \tilde{d}_j, \hat{c}_j))$.

Let $m_i \in M_i^k(c_i, d_i)$. Then, by the separation theorem, there exists $\lambda_i \in \Delta(\Theta_{-i} \times M_{-i})$ such that (i) $\lambda_i(\theta_{-i}, m_{-i}) > 0$ implies $m_{-i} \in M_{-i}^{k-1}(\theta_{-i})$, and (ii) for each $m_i' \in M_i$,

$$\sum_{c_{-i}, d_{-i}, m_{-i}} \lambda_i(c_{-i}, d_{-i}, m_{-i}) u_i(g(m_i', m_{-i}), c_i, d_i, c_{-i})$$
$$\leq \sum_{c_{-i}, d_{-i}, m_{-i}} \lambda_i(c_{-i}, d_{-i}, m_{-i}) u_i(g(m_i, m_{-i}), c_i, d_i, c_{-i}).$$

16

By Assumption 1, there exists $\psi_i : C_{-i} \to C_{-i}$ such that, for any $c_{-i} \in C_{-i}$, $u_i(\cdot, c_i, d_i, c_{-i})$ is an affine transformation of $u_i(\cdot, c_i', d_i', \psi_i(c_{-i}))$. Define $\lambda_i' \in \Delta(\Theta_{-i} \times M_{-i})$ so that, for each $(c_{-i}', d_{-i}', m_{-i}) \in \Theta_{-i} \times M_{-i}$,

$$\lambda_i'(c_{-i}', d_{-i}', m_{-i}) = \sum_{c_{-i}, d_{-i}} \lambda_i(c_{-i}, d_{-i}, m_{-i}) 1\{c_{-i}' = \psi_i(c_{-i})\} 1\{d_{-i}' = \delta_{-i}(c_{-i}, d_{-i}, c_{-i}')\}.$$

Note that, for each c_{-i}', d_{-i}', and m_{-i}, $\lambda_i(c_{-i}', d_{-i}', m_{-i}) > 0$ implies $m_{-i} \in M_{-i}^{k-1}(c_{-i}', d_{-i}')$. To see this, suppose contrarily that, for some c_{-i}', d_{-i}', m_{-i}, we have $m_{-i} \notin M_{-i}^{k-1}(c_{-i}', d_{-i}')$ but $\lambda_i'(c_{-i}', d_{-i}', m_{-i}) > 0$. Then, there exist c_{-i}, d_{-i} such that $c_{-i}' = \psi_i(c_{-i})$, $d_{-i}' = \delta_{-i}(c_{-i}, d_{-i}, c_{-i}')$, and $\lambda_i(c_{-i}, d_{-i}, m_{-i}) > 0$. Moreover, by assumption, $M_{-i}^{k-1}(c_{-i}, d_{-i}) \subseteq M_{-i}^{k-1}(c_{-i}', d_{-i}')$, and hence, $m_{-i} \notin M_{-i}^{k-1}(c_{-i}, d_{-i})$. This contradicts that $\lambda_i(c_{-i}, d_{-i}, m_{-i}) > 0$ implies $m_{-i} \in M_{-i}^{k-1}(c_{-i}, d_{-i})$.

This implies that there is no $\mu_i \in \Delta(M_i)$ such that, for any (c_{-i}, d_{-i}) and $m_{-i} \in M_{-i}^{k-1}(c_{-i}, d_{-i})$,

$$\sum_{m_i'} \mu_i(m_i') u_i(g(m_i', m_{-i}), c_i', \delta_i(c_i, d_i, c_i'), c_{-i}) > u_i(g(m_i, m_{-i}), c_i', \delta_i(c_i, d_i, c_i'), c_{-i}),$$

and therefore, $m_i \in M_i^k(c_i', \delta_i(c_i, d_i, c_i'))$. $\qquad\square$

Define a mechanism $\Gamma' = \langle M', g' \rangle$ so that $M_i' = M_i^\infty = \bigcup_{\theta_i} M_i^\infty(\theta_i)$ for each i, and $g' : M' \to X$ is the restriction of $g : M \to X$ on $M'(\subseteq M)$. For some i, θ_i, suppose that $m_i \notin M_i^\infty(\theta_i)$. Then, there exists $\mu_i \in \Delta(M_i^\infty(\theta_i))$ such that, for any (c_{-i}, d_{-i}) and $m_{-i} \in M_{-i}^\infty(c_{-i}, d_{-i})$,

$$\sum_{m_i'} \mu_i(m_i') u_i(g(m_i', m_{-i}), \theta_i, c_{-i}) > u_i(g(m_i, m_{-i}), \theta_i, c_{-i}).$$

Now, fix an arbitrary $m_{-i}' \in M_{-i}^\infty$ and $c_{-i}' \in C_{-i}$. By Lemma 2, there exists d_{-i}' such that $m_{-i}' \in M_{-i}^\infty(c_{-i}', d_{-i}')$. Thus,

$$\sum_{m_i'} \mu_i(m_i') u_i(g(m_i', m_{-i}'), \theta_i, c_{-i}') > u_i(g(m_i, m_{-i}'), \theta_i, c_{-i}').$$

Because this inequality is true for any $m_{-i}' \in M_{-i}'$ and $c_{-i}' \in C_{-i}$, we have $m_i \notin M_i'^U(\theta_i)$ in the new mechanism Γ'. This implies that Γ' implements F in undominated strategies. $\qquad\square$

5 Partial Bayesian implementation

In the literature on robust mechanism design with respect to the agents' structural uncertainties, several studies, such as Bergemann and Morris (2005), Chung and Ely (2007), Smith (2011), and Börgers and Smith (2012), examine the implications of Bayesian incentive compatibility conditions given arbitrary type spaces. Bayesian incentive compatibility is equivalent to the *partial* Bayesian implementation of an objective, which requires that there exists *at least one* Bayesian equilibrium that induces a desirable outcome. In this section, we examine the relationships between the partial Bayesian implementation and the other implementation concepts discussed in the previous sections. We first consider the private-value environment, and then consider the interdependent-value environment.

5.1 Private values

Definition 5. We say that a mechanism Γ *partially Bayesian implements F given type space* $\mathcal{T}$ if there exists a Bayesian equilibrium σ^* given $\mathcal{T}$ such that, for each $t \in T$ and $m \in M$ with $\sigma^*(m|t) > 0$, we have $g(m) \in F(\tau(t))$.

Obviously, partial Bayesian implementation is a less demanding concept than Bayesian implementation (and implementation in one-round or iteratively undominated strategies), but not necessarily vice versa. Nevertheless, under certain conditions, together with the private-value assumption, we show that these concepts are "close" to each other in terms of implementable objectives.

To explain the additional conditions we need, the following example illustrates that, *without any additional condition*, partial Bayesian implementation is *strictly* less demanding than Bayesian implementation, even in a single-agent setting.[19]

---

[19]This distinction between partial implementation (i.e., requiring that one of the possible outcomes is desirable) and weak (or full) implementation (i.e., requiring that every possible outcome is desirable) is well known in the literature. See, for example, Plott (1976).

Example 2. Let $N = 1$, $X = \{a, a'\}$, and $\Theta_1 = \{t, t'\}$. The agent is always indifferent between a and a', regardless of θ_1, i.e., $u_1(a, \theta_1) = u_1(a', \theta_1)$ for each $\theta_1 = t, t'$. However, the designer aims to implement the following state-contingent outcomes: $F(t) = \{a\}$ and $F(t') = \{a'\}$.

F is partially Bayesian implementable given any type space, by a "revelation mechanism" $\Gamma = \langle \Theta_1, g \rangle$ such that $g(t) = a$ and $g(t') = a'$. However, Γ cannot Bayesian implement F, because $M_1^U(t) = M_1^U(t') = \{t, t'\}$. Indeed, in any mechanism Γ, we have $M_1^U(t) = M_1^U(t')$, and therefore, F is not Bayesian implementable.

This distinction occurs because, for Γ to implement F, the agent needs to send different messages in different states, even though he is indifferent. Such F is never implementable in undominated strategies (and thus, nor Bayesian implementable) by any mechanism, because the two types of the agent have the same set of undominated messages.

Nevertheless, we show that, if F is partially Bayesian implementable given any type space, then with some additional assumptions, it is "virtually" implementable in undominated strategies. To introduce the notion of virtual implementation, we assume that there exists a finite set of "pure allocations" Y such that $X = \Delta(Y)$ is the set of all lotteries over Y. For each y, i's utility given θ_i is denoted by $v_i(y, \theta_i)$, and we assume $u_i(x, \theta_i) = \sum_y x(y) v_i(y, \theta_i)$, where $x(y) \in [0, 1]$ is the probability that lottery x assigns to y. We also assume that F is convex-valued, i.e., for each θ, if $x, x' \in F(\theta)$, then $\alpha x + (1-\alpha)x' \in F(\theta)$ for any $\alpha \in (0, 1)$. That is, if two lotteries x, x' are considered to be desirable in state θ, then any compound lottery that randomly selects x or x' is desirable in θ as well.

Definition 6. Given $\varepsilon \in (0, 1)$ and a social choice correspondence F, let F_ε be a social choice correspondence such that, for each θ,

$$F_\varepsilon(\theta) = \{x | \exists x' \in F(\theta), \ \max_y |x(y) - x'(y)| < \varepsilon\}.$$

We say that F is *virtually implementable in undominated strategies* if, for any $\varepsilon \in (0, 1)$, F_ε is implementable in undominated strategies.

We introduce an assumption where an agent's indifference can always be broken by some perturbation.

Assumption 2. There exists an allocation rule $f^U : \Theta \to X$ such that, in a revelation mechanism $\Gamma^U = \langle \Theta, f^U \rangle$, θ_i is the unique undominated message for each i and θ_i (i.e., $M_i^U(\theta_i) = \{\theta_i\}$).[20]

Theorem 3. Under Assumption 2, if F is partially Bayesian implementable given any $\mathcal{T}$, then it is virtually implementable in undominated strategies.[21]

The proof comprises the following two steps. We first introduce another implementation concept, which we refer to as *implementation in (some) best-response correspondences*. We show that (i) if F is partially Bayesian implementable, then it is implementable in best-response correspondences (Proposition 2).[22] Then we show that (ii) under Assumption 2, if F is implementable in best-response correspondences, then it is virtually implementable in undominated strategies (Proposition 3).

Definition 7. For each i, θ_i, let $M_i^R(\theta_i) \subseteq M_i$. A mechanism Γ *implements F in best-response correspondences* $(M_i^R(\theta_i))_{i,\theta_i}$ if (i) for each θ and $m \in M^R(\theta) = \prod_i M_i^R(\theta_i)$, we have $g(m) \in F(\theta)$, and (ii) for each i, θ_i, and $\mu_{-i} \in \Delta(M_{-i})$, there exists $m_i \in M_i^R(\theta_i)$ that is a best response for θ_i against μ_{-i}, i.e., for any $m_i' \in M_i$,

$$\sum_{m_{-i}} \mu_{-i}(m_{-i}) u_i(g(m_i, m_{-i}), \theta_i) \geq \sum_{m_{-i}} \mu_{-i}(m_{-i}) u_i(g(m_i', m_{-i}), \theta_i).$$

---

[20]Assumption 2 may not be very demanding if the feasible set of allocations does not depend on θ. For example, it is sufficient if for each $i, \theta_i \neq \theta_i'$, there exist x, x' such that $u_i(x, \theta_i) > u_i(x', \theta_i)$ and $u_i(x, \theta_i) < u_i(x', \theta_i)$.

[21]Obviously, if F is virtually implementable in undominated strategies, then it is virtually partially implementable given any $\mathcal{T}$ (with or without Assumption 2). On the other hand, it is an open question whether virtual implementation in undominated strategies implies partial Bayesian implementation given any $\mathcal{T}$.

[22]In fact, the other direction also holds. The proof is available upon request.

Proposition 2. If F is partially Bayesian implementable given any $\mathcal{T}$, then there exists $M_i^R(\theta_i) \subseteq M_i$ for each i and θ_i such that F is implementable in best-response correspondences $(M_i^R(\theta_i))_{i,\theta_i}$.

The formal proof is in the appendix, and here, we explain the idea of the proof for a special case to provide some intuition. Suppose that, given a type space $\mathcal{T}$, $\Gamma = \langle M, g \rangle$ has a pure-strategy Bayesian equilibrium σ^* such that, for each t with $\tau(t) = \theta$, we have $g(\sigma^*(t)) \in F(\theta)$ (so Γ partially Bayesian implements F). Suppose also that, for each i and $m_i \in M_i$, there is some t_i such that $\sigma_i^*(t_i) = m_i$. Such a type t_i is denoted by $t_i(m_i)$ (fixed arbitrarily if there are multiple of such types).

Imagine that $\mathcal{T}$ is indeed a universal type space of Mertens and Zamir (1985) or Brandenburger and Dekel (1993).[23] Then, for each i, θ_i, and any probability distribution μ_{-i} over M_{-i}, there is some t_i such that $\tau_i(t_i) = \theta_i$ and $\beta_i(t_{-i}(m_{-i})|t_i) = \mu_{-i}(m_{-i})$ for each $m_{-i} \in M_{-i}$.[24] Then, because $\sigma_i^*(t_i)$ is the equilibrium action for t_i, we have, for any $m_i' \in M_i$,

$$\sum_{m_{-i}} \mu_{-i}(m_{-i}) u_i(g(m_i, m_{-i}), \theta_i) \geq \sum_{m_{-i}} \mu_{-i}(m_{-i}) u_i(g(m_i', m_{-i}), \theta_i).$$

Now, we define $M_i^R(\theta_i) = \{\sigma_i^*(t_i)|\tau_i(t_i) = \theta_i\}$ for each i, θ_i. Then, Γ implements F in best-response correspondences $(M_i^R(\theta_i))_{i,\theta_i}$ because (i) for each θ and $m \in M^R(\theta)$, we have $g(m) \in F(\theta)$ because Γ partially Bayesian implements F given $\mathcal{T}$ (at the equilibrium σ^*), and (ii) for each $i, \theta_i, \mu_{-i} \in \Delta(M_{-i})$, there exists $m_i \in M_i^R(\theta_i)$ that is a best response for θ_i against μ_{-i}, by the construction of $M_i^R(\theta_i)$.

In general, of course, σ^* may be in mixed strategies and may assign probabilities less than one (or even zero) for some of the messages. Furthermore, $\mathcal{T}$ is restricted to be at most countably infinite, and thus, the argument based

---

[23]Because we only allow for at most countably infinite type spaces, and so do not allow uncountably infinite type spaces such as the universal type space, in the formal proof, we find another type space that is at most countably infinite, and still sufficiently rich to obtain the result.

[24]This is an implication of the *completeness* of the universal type space.

on the completeness of the universal type space is not applicable. The formal proof treats these complicated issues, while the main intuition is already explained above.

Now we show that, under Assumption 2, if F is implementable in best-response correspondences, then it is virtually implementable in undominated strategies.

Proposition 3. Under Assumption 2, suppose that there exists $M_i^R(\theta_i) \subseteq M_i$ for each i, θ_i such that F is implementable in best-response correspondences $(M_i^R(\theta_i))_{i,\theta_i}$. Then F is virtually implementable in undominated strategies.

Remark 3. Although this paper considers strict dominance to define the agents' "rational" behaviors, an alternative approach in the literature is based on (one-round) elimination of weakly dominated strategies.[25] Obviously, implementation based on weak dominance is less demanding than implementation in (strict) undominated strategies. Conversely, partial Bayesian implementation given any type space is less demanding than implementation based on weak dominance. Therefore, as long as Assumption 2 is satisfied, the distinction between the weak and strict dominance concepts is "virtually" nonessential.

More generally, as long as Assumption 2 is satisfied, implementation based on any solution concept that is (i) more demanding than partial Bayesian implementation given any type space and (ii) less demanding than implementation in undominated strategies is "virtually" equivalent to each other.[26]

---

[25]For example, see Börgers (1991) and Jackson (1992). In this remark, we discuss only the one-round elimination of weakly dominated strategies, but Yamashita (2013a) shows that the same argument applies to an iterative elimination concept, which is based on the one-round elimination of weakly dominated strategies followed by the iterative elimination of strictly dominated strategies, as in Dekel and Fudenberg (1990) and Börgers (1994).

[26]For example, (appropriately defined) "perfect Bayesian" implementation given any type space using dynamic mechanisms would be in this class of solution concepts.

5.2 Interdependent values

With interdependent values, there can be some F that is *partially* Bayesian implementable, but is not implementable in (even virtually) iterative undominated strategies, as in the following example.

Example 3. Let $I = 2$, $\Theta_1 = \Theta_2 = \{-1, 1\}$, and $X = \Delta(\{1, -1\})$. For each agent i, $u_i(x, \theta) = x\theta_1\theta_2$. Let $F(\theta) = \{\theta_1\theta_2\}$ for each θ.

The following "revelation" mechanism $\Gamma = \langle \Theta, g \rangle$ partially Bayesian implements F given any type space: $g(\theta) = \theta_1\theta_2$ for each θ.

However, no mechanism can (even virtually) implement F in iteratively undominated strategies, because in any mechanism, for each i, $M_i^U(1) = M_i^U(-1)$, and thus, no iterative elimination is possible.

In this example, in the truth-telling equilibrium, no agent is made better off by any deviation *if the other agent is truthful*. However, the mechanism has another untruthful Bayesian equilibrium. Indeed, we can only Bayesian implement constant correspondences, i.e., such F that $F(\theta)$ does not vary with θ.

Identifying general conditions for any partially Bayesian implementable correspondence to be virtually implementable in (one-round) undominated strategies is left as a future research topic. However, such "general conditions" could be somewhat restrictive. Bergemann and Morris (2009a) obtain ex post incentive compatibility and *robust measurability* as two necessary conditions (and also sufficient with some additional "technical" conditions) for social choice functions to be virtually implementable in iteratively undominated strategies. Recall that a social choice function is ex post incentive compatible if and only if it is partially Bayesian implementable given any type space (Bergemann and Morris (2005)), and therefore, robust measurability is the additional condition that makes the function virtually implementable in iteratively undominated strategies. However, robust measurability could be a substantial restriction on social choice functions. For example, in their private-goods allocation environment, robust measurability requires that in-

terdependence is "sufficiently small" with respect to a certain measure.

6 Conclusion

In this paper, in certain environments, we have established equivalence in terms of implementable social choice correspondences between (i) implementation in (one-round or iterative) undominated strategies, a solution concept that is robust to strategic uncertainties, and (ii) Bayesian implementation with arbitrary type spaces, a solution concept that is robust to structural uncertainties. The class of environments that admit this equivalence includes private-value environments and interdependent-value environments with certain richness conditions.

Furthermore, we have established virtual equivalence of those concepts with partial Bayesian implementation in the private-value environment. The results suggest that, in some economically important environments, assuming the agents' (Bayesian) equilibrium plays does not enlarge the set of implementable objectives unless we make additional assumptions on their (high-order) beliefs, compared to the case where the agents are just assumed to be rational.

We also believe that the results are useful in suggesting that some theoretical tools developed in one approach would be useful in another approach, as we have discussed in Remark 2.

A Proofs

B Proof of Proposition 1

Suppose that F is implemented in iteratively undominated strategies by a mechanism $\Gamma = \langle M, g \rangle$. Given any type space $\mathcal{T}$, let σ^* be any Bayesian equilibrium in Γ. First, for each i and t_i such that $\tau_i(t_i) = \theta_i$, if $\sigma_i^*(m_i|t_i) > 0$, then $m_i \in M_i^0(\theta_i)$.

For $k = 1, 2, \ldots$, suppose the following holds: for each i and t_i such that $\tau_i(t_i) = \theta_i$, if $\sigma_i^*(m_i|t_i) > 0$, then $m_i \in M_i^{k-1}(\theta_i)$. Now, suppose that for some i, t_i such that $\tau_i(t_i) = \theta_i$, there is $m_i \in M_i^{k-1}(\theta_i) \setminus M_i^k(\theta_i)$ such that $\sigma_i^*(m_i|t_i) > 0$. Then, for such m_i, there exists $\mu_i \in M_i^{k-1}(\theta_i)$ such that, for any θ_{-i} and $m_{-i} \in M_{-i}^{k-1}(\theta_{-i})$,

$$\sum_{m_i'} \mu_i(m_i') u_i(g(m_i', m_{-i}), \theta) > u_i(g(m_i, m_{-i}), \theta).$$

Because the support of $\sigma_{-i}^*(\cdot|t_{-i})$ is a subset of $M_{-i}^{k-1}(\tau_{-i}(t_{-i}))$ for each t_{-i}, we have

$$\sum_{t_{-i}} \sum_{m_{-i}} \sum_{m_i'} \beta_i(t_{-i}|t_i) \sigma_{-i}^*(m_{-i}|t_{-i}) \mu_i(m_i') u_i(g(m_i', m_{-i}), \tau(t))$$
$$> \sum_{t_{-i}} \sum_{m_{-i}} \beta_i(t_{-i}|t_i) \sigma_{-i}^*(m_{-i}|t_{-i}) u_i(g(m_i, m_{-i}), \tau(t)).$$

Instead of playing $\sigma_i^*(t_i)$, suppose that t_i plays (i) each $m_i' \in M_i$ such that $m_i' \neq m_i$ with probability $\sigma_i^*(m_i'|t_i)$, and (ii) μ_i with probability $\sigma_i^*(m_i|t_i)$. This is strictly a better response for t_i than $\sigma_i^*(t_i)$, which contradicts that σ^* is a Bayesian equilibrium. Therefore, if $\sigma_i^*(m_i|t_i) > 0$, then $m_i \in M_i^k(\theta_i)$.

Iterating this argument for all k, we conclude that, for each i, t_i such that $\tau_i(t_i) = \theta_i$, if $\sigma_i^*(m_i|t_i) > 0$, then $m_i \in M_i^\infty(\theta_i)$. This implies that Γ Bayesian implements F given any type space.

Next, we show that Bayesian implementation given any type space implies implementation in iteratively undominated strategies. Let $\Gamma = \langle M, g \rangle$ be a mechanism that Bayesian implements F given any type space. Let

$$\Lambda_{-i} = \{\lambda_{-i} \in \Delta(\Theta_{-i} \times M_{-i}) | \forall \theta_{-i}, \ [\lambda_{-i}(\theta_{-i}, m_{-i}) > 0 \Rightarrow m_{-i} \in M_{-i}^\infty(\theta_{-i})]\}.$$

$\lambda_{-i} \in \Lambda_{-i}$ is agent i's (joint) conjecture about the others' payoff types and their messages that survive the iterative elimination procedure. Note that Λ_{-i} is convex and closed.

Lemma 3. For each $m_i \in M_i^\infty(\theta_i)$, there exists $\lambda_{-i} \in \Lambda_{-i}$ such that, for any $m_i' \in M_i$,

$$\sum_{\theta_{-i}} \sum_{m_{-i}} \lambda_{-i}(\theta_{-i}, m_{-i}) u_i(g(m_i, m_{-i}), \theta) \geq \sum_{\theta_{-i}} \sum_{m_{-i}} \lambda_{-i}(\theta_{-i}, m_{-i}) u_i(g(m_i', m_{-i}), \theta).$$

We denote such λ_{-i} by $\lambda_{-i}^{\theta_i, m_i}$.

Proof. Suppose that there exist i, θ_i, and $m_i \in M_i^\infty(\theta_i)$ such that, for any $\lambda_{-i} \in \Lambda_{-i}$, there is some $m_i' \in M_i$ such that

$$\sum_{\theta_{-i}} \sum_{m_{-i}} \lambda_{-i}(\theta_{-i}, m_{-i}) u_i(g(m_i, m_{-i}), \theta) < \sum_{\theta_{-i}} \sum_{m_{-i}} \lambda_{-i}(\theta_{-i}, m_{-i}) u_i(g(m_i', m_{-i}), \theta).$$

We can find such m_i' in $M_i^\infty(\theta_i)$. Then, by the separation theorem, we can find $\mu_i' \in \Delta(M_i^\infty(\theta_i))$ such that, for any $\lambda_{-i} \in \Lambda_{-i}$,

$$\sum_{\theta_{-i}} \sum_{m_{-i}} \lambda_{-i}(\theta_{-i}, m_{-i}) u_i(g(m_i, m_{-i}), \theta) < \sum_{\theta_{-i}} \sum_{m_{-i}} \sum_{m_i'} \lambda_{-i}(\theta_{-i}, m_{-i}) \mu_i(m_i') u_i(g(m_i', m_{-i}), \theta).$$

In particular, this implies that, for any θ_{-i} and $m_{-i} \in M_{-i}^\infty(\theta_{-i})$, we have

$$u_i(g(m_i, m_{-i}), \theta) < \sum_{m_i'} \mu_i(m_i') u_i(g(m_i', m_{-i}), \theta),$$

which contradicts that $m_i \in M_i^\infty(\theta_i)$. $\qquad\square$

We construct a type space $\mathcal{T}$ as follows. For each i, let $T_i = \{(\theta_i, m_i) | \theta_i \in \Theta_i,\ m_i \in M_i^\infty(\theta_i)\}$. For each $t_i = (\theta_i, m_i)$ and $t_{-i} = (\theta_{-i}, m_{-i})$, we define $\tau_i(t_i) = \theta_i$, and $\beta_i(t_{-i}|t_i) = \lambda_{-i}^{\theta_i, m_i}(\theta_{-i}, m_{-i})$.

Consider σ^* such that, for each i and $t_i = (\theta_i, m_i)$, we have $\sigma_i^*(m_i|t_i) = 1$. It is a Bayesian equilibrium because, for each i and $t_i = (\theta_i, m_i)$, given that the other agents $-i$ follow σ_{-i}^*, m_i satisfies that, for each $m_i' \in M_i$,

$$
\sum_{t_{-i}} \sum_{m_{-i}} \sigma_{-i}^*(m_{-i}|t_{-i}) \beta_i(t_{-i}|t_i) u_i(g(m_i, m_{-i}), \tau(t))
$$
$$
= \sum_{\theta_{-i}} \sum_{m_{-i}} \lambda_{-i}^{\theta_i, m_i}(\theta_{-i}, m_{-i}) u_i(g(m_i, m_{-i}), \theta)
$$
$$
\geq \sum_{\theta_{-i}} \sum_{m_{-i}} \lambda_{-i}^{\theta_i, m_i}(\theta_{-i}, m_{-i}) u_i(g(m_i', m_{-i}), \theta)
$$
$$
= \sum_{t_{-i}} \sum_{m_{-i}} \sigma_{-i}^*(m_{-i}|t_{-i}) \beta_i(t_{-i}|t_i) u_i(g(m_i', m_{-i}), \tau(t)).
$$

Because Γ Bayesian implements F, for each $\theta, m \in M^\infty(\theta)$, we have $g(m) \in F(\theta)$. This implies that the mechanism Γ implements F in iteratively undominated strategies.

B.1 Proof of Proposition 2

Let Γ be a mechanism that partially Bayesian implements F given any type space.

Consider the following type space $\mathcal{T}$. Let $L_i = |M_{-i}| + 1 (< \infty)$, where $|M_{-i}|$ is the number of message profiles for agent i's opponents.

For each i, we define

$$
T_i = \{(\theta_i; q_{-i}^1, \ldots, q_{-i}^{L_i}; p^1, \ldots, p^{L_i}) | \forall l,\ q_{-i}^l \in \mathbb{Q}^{N-1},\ p^l \in \mathbb{Q} \cap [0,1];\ \sum_{l=1}^{L_i} p^l = 1\}.
$$

Because T_i is at most countably infinite, there is a one-to-one mapping $\rho_i : T_i \to \mathbb{Q}$. In the following, each rational number $q_j \in \mathbb{Q}$ corresponds to one type of j, $\rho_j(q_j) \in T_j$. Each type $t_i = (\theta_i; q_{-i}^1, \ldots, q_{-i}^{L_i}; p^1, \ldots, p^{L_i})$ of

27

agent i has (i) a payoff type θ_i, and (ii) assigns positive probabilities for at most $L_i(<\infty)$ type profiles of the opponents, and (iii) each probability is a rational number. More precisely, we define τ_i and β_i as follows.

$$\tau_i(\theta_i; q_{-i}^1, \ldots, q_{-i}^{L_i}; p^1, \ldots, p^{L_i}) = \theta_i,$$

$$\beta_i(t_{-i}|\theta_i; q_{-i}^1, \ldots, q_{-i}^{L_i}; p^1, \ldots, p^{L_i}) = \begin{cases} p^l & \text{if} \quad t_{-i} = \rho_{-i}^{-1}(q_{-i}^l), \\ 0 & \text{otherwise.} \end{cases}$$

By assumption, in mechanism $\Gamma = \langle M, g \rangle$ and given type space $\mathcal{T}$, there exists a Bayesian equilibrium σ^* such that, for each t and m with $\sigma^*(m|t) > 0$, we have $g(m) \in F(\tau(t))$. For each i, θ_i, let $\Sigma_i^*(\theta_i) = \{\sigma_i^*(t_i) \in \Delta(M_i)|\tau_i(t_i) = \theta_i\}$. Let $\Sigma_i^* = \bigcup_{\theta_i} \Sigma_i^*(\theta_i)$, and let $\Lambda_{-i} = co(\Sigma_{-i}^*)$ denote the convex hull of Σ_{-i}^*.

For each i, t_i, his belief about the opponents' messages in the equilibrium is given by $\mu_{-i}^{t_i} \in \Lambda_{-i}$ such that, for each m_{-i},

$$\mu_{-i}^{t_i}(m_{-i}) = \sum_{t_{-i}} \beta_i(t_{-i}|t_i)\sigma_{-i}^*(m_{-i}|t_{-i}),$$

which is a convex combination of $|M_{-i}|$ elements of Σ_{-i}^* where each weight is a rational number. Let $\tilde{\Lambda}_{-i} = \{\mu_{-i}^{t_i}|t_i \in T_i\}$.

The following technical result will be used later.

Lemma 4. $\tilde{\Lambda}_{-i}$ is a dense subset of Λ_{-i}.

Proof. First, by Carathéodory's theorem,[27] each $\mu_{-i} \in \Lambda_{-i}$ is given by a convex combination of L_i elements in Σ_{-i}^*, say $\sigma_{-i}^*(t_{-i}^1), \ldots, \sigma_{-i}^*(t_{-i}^{L_i})$, with the weight vector $(b^1, \ldots, b^{L_i})$, i.e., $(b^1, \ldots, b^{L_i}) \in \mathbb{R}_+^{L_i}$, $\sum_l b^l = 1$, and $\mu_{-i}(m_{-i}) = \sum_l b^l \sigma_{-i}^*(m_{-i}|t_{-i}^l)$ for each $m_{-i} \in M_{-i}$.

By the denseness of rational numbers, for any $\varepsilon > 0$, we can find a rational vector, $(\beta^1, \ldots, \beta^{L_i}) \in \mathbb{Q}_+^{L_i}$, such that $\sum_l \beta^l = 1$ and $|b^l - \beta^l|$ is less than ε

---

[27]Specifically, let A be an L-dimensional subset of a Euclidean space, and let $co(A)$ be the convex hull of A. For each $a \in co(A)$, there exist $L+1$ elements in A, say $a_1, \ldots, a_{L+1} \in A$, such that a is a convex combination of them.

for each l. By the construction of the type space, there exists t_i such that $\beta_i(t^l_{-i}|t_i) = \beta^l$ for each l. Therefore, for each $\varepsilon > 0$ and each $\mu_{-i} \in \Lambda_{-i}$, we can find $t_i \in T_i$ such that

$$\max_{m_{-i} \in M_{-i}} |\mu^{t_i}_{-i}(m_{-i}) - \mu_{-i}(m_{-i})| < \varepsilon,$$

which implies that $\tilde{\Lambda}_{-i}$ is a dense subset of Λ_{-i}. $\qquad\square$

The following lemma is the key step of the proof of the theorem.

Lemma 5. For each i, θ_i, there exists a finite subset $\Sigma^R_i(\theta_i)$ of $\Sigma^*_i(\theta_i)$ such that, for any $\mu_{-i} \in \Lambda_{-i}$, there exists $\mu_i \in \Sigma^R_i(\theta_i)$ that is a best response for θ_i against μ_{-i}, i.e., for any $m'_i \in M_i$,

$$\sum_{\theta_{-i}} \sum_{m_{-i}} \sum_{m_i} \mu_{-i}(m_{-i}) \mu_i(m_i) u_i(g(m_i, m_{-i}), \theta_i) \geq \sum_{\theta_{-i}} \sum_{m_{-i}} \mu_{-i}(m_{-i}) u_i(g(m'_i, m_{-i}), \theta_i).$$

Proof. For each i, θ_i, and $\mu_{-i} \in \tilde{\Lambda}_{-i}$, let $R_i(\theta_i, \mu_{-i}) \subseteq M_i$ be the set of the "pure" best responses for θ_i to μ_{-i}, i.e., $m_i \in R_i(\theta_i, \mu_{-i})$ if and only if, for any $m'_i \in M_i$,

$$\sum_{\theta_{-i}} \sum_{m_{-i}} \mu_{-i}(m_{-i}) u_i(g(m_i, m_{-i}), \theta_i) \geq \sum_{\theta_{-i}} \sum_{m_{-i}} \mu_{-i}(m_{-i}) u_i(g(m'_i, m_{-i}), \theta_i).$$

For each i, θ_i, and $\mu_{-i} \in \tilde{\Lambda}_{-i}$, let $t_i(\theta_i, \mu_{-i})$ denote agent i's type such that $\mu^{t_i}_{-i} = \mu_{-i}$ (fixed arbitrarily if there are multiple of such types), i.e., $t_i(\theta_i, \mu_{-i})$ is agent i's type whose payoff type is θ_i and his belief about the opponents' messages in the equilibrium is μ_{-i}.

We partition $\tilde{\Lambda}_{-i}$ so that μ'_{-i}, μ''_{-i} are in the same partition element if and only if $R_i(\theta_i, \mu'_{-i}) = R_i(\theta_i, \mu''_{-i})$ for any θ_i. Then $\tilde{\Lambda}_{-i}$ is finitely partitioned, and we denote the partition by $\{B^1, B^2, \ldots, B^{K_i}\}$. For each $B^k \subseteq \tilde{\Lambda}_{-i}$ in the partition, fix any element $\mu^k_{-i} \in B^k$, and we define $\Sigma^R_i(\theta_i)$ as $\Sigma^R_i(\theta_i) = \{\sigma^*_i(t_i(\theta_i, \mu^k_{-i})) | k = 1, \ldots, K_i\}$.

By upper hemi-continuity of best responses, $\sigma^*_i(t_i(\theta_i, \mu^k_{-i}))$ is a best response to any element in $\overline{B^k}$, the closure of B^k. Because $\tilde{\Lambda}_{-i}$ is dense in Λ_{-i}, we have $\Lambda_{-i} \subseteq \bigcup_{k=1}^{K_i} \overline{B^k}$. Therefore, for each $\mu_{-i} \in \Lambda_{-i}$, there is $\mu_i \in \Sigma^R_i(\theta_i)$ that is a best response for θ_i against μ_{-i}. $\qquad\square$

We define a mechanism $\Gamma' = \langle M', g' \rangle$ as follows. For each i, $M'_i = \bigcup_{\theta_i} \Sigma_i^R(\theta_i)$, and for each $\mu \in M'$, $g'(\mu) = \sum_{m \in M} \mu(m) g(m)$. The next lemma completes the proof of the theorem.

Lemma 6. Γ' implements F in best-response correspondences $(\Sigma_i^R(\theta_i))_{i,\theta_i}$.

Proof. For each i and $\mu_{-i} \in \Delta(M'_{-i})$, agent i with θ_i can find a best response against μ_{-i} in $\Sigma_i^R(\theta_i)$.

Recall that, in mechanism Γ, for each θ and $\mu \in \prod_i (\Sigma_i^R(\theta_i))$, μ was the equilibrium message profile of some t such that $\tau(t) = \theta$, and therefore, for any m such that $\mu(m) > 0$, we have $g(m) \in F(\theta)$. Now, in mechanism Γ', we have $g'(\mu) = \sum_{m \in M} \mu(m) g(m)$, and it is an element in $F(\theta)$ as well because F is convex-valued. □

C Proof of Proposition 3

Let Γ be a mechanism that implements F in best-response correspondences $(M_i^R(\theta_i))_{i,\theta_i}$. Without loss of generality, we assume that $M_i = \bigcup_{\theta_i} M_i^R(\theta_i)$ for each i.[28]

We define another mechanism $\Gamma^\varepsilon = \langle M^\varepsilon, g^\varepsilon \rangle$ as follows. For each i, let $M_i^\varepsilon = \{(m_i, \theta_i) | m_i \in M_i^R(\theta_i), \; \theta_i \in \Theta_i\}$, and for each $(m, \theta) \in M^\varepsilon$, let $g^\varepsilon(m, \theta) = (1 - \varepsilon) g(m) + \varepsilon f^U(\theta)$.

Fix arbitrary i, θ_i, and $m_i \notin M_i^R(\theta_i)$. Because $M_i = \bigcup_{\theta_i} M_i^R(\theta_i)$, we have $m_i \in M_i^R(\theta'_i)$ for some $\theta'_i \neq \theta_i$. Because m_i cannot be the unique best response to any $\mu_{-i} \in \Delta(M_{-i})$, by the separation theorem, there exists $\mu_i \in \Delta(M_i^R(\theta_i))$ such that, for any $m_{-i} \in M_{-i}$,

$$\sum_{m'_i} \mu_i(m'_i) u_i(g(m'_i, m_{-i}), \theta_i) \geq u_i(g(m_i, m_{-i}), \theta_i).$$

In mechanism Γ^ε, observe that, for this agent i with θ_i, (θ'_i, m_i) is strictly dominated if $\theta'_i \neq \theta_i$. Specifically, consider a mixed action μ'_i such that, for

---

[28]If not, then redefine the mechanism by eliminating all messages in $M_i \setminus (\bigcup_{\theta_i} M_i^R(\theta_i))$. Then the new mechanism implements F in the same best-response correspondences.

each $m_i' \in M_i^R(\theta_i)$, $\mu_i'(\theta_i, m_i') = \mu_i(m_i')$ (so $\sum_{m_i' \in M_i^R(\theta_i)} \mu_i'(\theta_i, m_i') = 1$). Then, for each (m_{-i}, θ_{-i}),

$$
\begin{aligned}
&\sum_{m_i'} \mu_i'(m_i', \theta_i) u_i(g^\varepsilon((m_i', \theta_i), (m_{-i}, \theta_{-i})), \theta_i) \\
=\ &\sum_{m_i'} \mu_i(m_i')(1-\varepsilon) u_i(g(m_i', m_{-i}), \theta_i) + \varepsilon u_i(f^U(\theta_i, \theta_{-i}), \theta_i) \\
>\ &(1-\varepsilon) u_i(g(m_i, m_{-i}), \theta_i) + \varepsilon u_i(f^U(\theta_i', \theta_{-i}), \theta_i) \\
=\ &u_i(g^\varepsilon((m_i, \theta_i'), (m_{-i}, \theta_{-i})), \theta_i),
\end{aligned}
$$

where the strict inequality is implied by the definition of f^U. Thus, (θ_i', m_i) is strictly dominated.

Therefore, in mechanism Γ^ε, if a message is undominated for θ_i, then it must take the form (θ_i, m_i'), where $m_i' \in M_i^R(\theta_i)$. This implies that, for each θ, g' assigns $g(m')$ for some $m' \in M^R(\theta)$ with a probability of at least $1 - \varepsilon$. In conclusion, F is virtually implemented in undominated strategies.

References

ABREU, D., AND H. MATSUSHIMA (1991): "Virtual Implementation in Iteratively Undominated Strategies: Incomplete Information," Discussion paper.

ABREU, D., AND H. MATSUSHIMA (1992): "Virtual Implementation in Iteratively Undominated Strategies: Complete Information," *Econometrica*, 60(5), 993–1008.

BATTIGALLI, P., AND M. SINISCALCHI (2003): "Rationalization and Incomplete Information," *The B.E. Journal of Theoretical Economics*, advances.3(1), 3.

BERGEMANN, D., AND S. MORRIS (2005): "Robust Mechanism Design," *Econometrica*, 73(6), 1771–1813.

———— (2007): "Belief Free Incomplete Information Games," Cowles Foundation Discussion Papers 1629, Cowles Foundation for Research in Economics, Yale University.

———— (2009a): "Robust Implementation in Direct Mechanisms," *Review of Economic Studies*, 76(4), 1175–1204.

———— (2009b): "Robust virtual implementation," *Theoretical Economics*, 4(1).

———— (2011): "Robust implementation in general mechanisms," *Games and Economic Behavior*, 71(2), 261–281.

BÖRGERS, T. (1991): "Undominated strategies and corordination in normalform games," *Social Choice and Welfare*, 8, 65–78.

———— (1994): "Weak Dominance and Approximate Common Knowledge," *Journal of Economic Theory*, 64(1), 265–276.

BÖRGERS, T., AND D. SMITH (2012): "Robust Mechanism Design and Dominant Strategy Voting Rules," forthcoming, *Theoretical Economics*.

BRANDENBURGER, A., AND E. DEKEL (1993): "Hierarchies of Beliefs and Common Knowledge," *Journal of Economic Theory*, 59, 189–198.

BRANDENBURGER, A., A. FRIEDENBERG, AND H. J. KEISLER (2008): "Admissibility in Games," *Econometrica*, 76(2), 307–352.

CHUNG, K.-S., AND J. ELY (2007): "Foundations of Dominant-Strategy Mechanisms," *Review of Economic Studies*, 74(2), 447–476.

DEKEL, E., AND D. FUDENBERG (1990): "Rational behavior with payoff uncertainty," *Journal of Economic Theory*, 52(2), 243–267.

HARSANYI, J. (1967-68): "Games with incomplete information played by Bayesian players," *Management Science*, 14, 159–182,320–334,486–502.

JACKSON, M. O. (1992): "Implementation in undominated strategies: a look at bounded mechanisms," *Review of Economic Studies*, 59, 757–775.

JEHIEL, P., B. MOLDOVANU, M. MEYER-TER-VEHN, AND W. ZAME (2006): "The limits of ex post implementation," *Econometrica*, 74, 585–610.

MERTENS, J. F., AND S. ZAMIR (1985): "Formulation of Bayesian analysis for games with incomplete information," *International Journal of Game Theory*, 14, 1–29.

MORRIS, S., AND H. S. SHIN (2002): "Measuring Strategic Uncertainty," mimeo.

PLOTT, C. R. (1976): "Axiomatic Social Choice Theory: An Overview and Interpretation," *American Journal of Political Science*, 20(3), 511–596.

SMITH, D. (2011): "A Prior Free Efficiency Comparison of Mechanisms for the Public Good Problem," mimeo, University of Michigan.

YAMASHITA, T. (2012): "A necessary condition for implementation in undominated strategies, with applications to robustly optimal trading mechanisms," mimeo, Toulouse School of Economics.

——— (2013a): "A Note on strategic and structural uncertainties in mechanism design: weak dominance," mimeo, Toulouse School of Economics.

——— (2013b): "Robust auctions with interdependent values," mimeo, Toulouse School of Economics.