Implementation with Transfers*

Yi-Chun Chen† Takashi Kunimoto‡ Yifei Sun§

March 23, 2016

Abstract

We say that a social choice rule is implementable with (small) transfers if one can design a mechanism whose set of equilibrium outcomes coincides with that specified by the rule but the mechanism allows for (small) ex post transfers among the players. We show in private-value environments that any incentive compatible rule is implementable with small transfers. We obtain this permissive implementation result by proposing a natural extension of Abreu and Matsushima (1994) to incomplete information environments. Furthermore, in order to showcase the applicability of our results, we relate them to the recent developments in implementation theory. Next we revisit the conjecture by Abreu and Matsushima (1994), who claim that the extension of Abreu and Matsushima (1994) may be possible by mimicking the argument of Abreu and Matsushima (1992b). To the extent that our mechanism is a natural extension of that of Abreu and Matsushima (1994), we show by example that their conjecture is not unconditionally warranted to cover fully interdependent-value environments. We therefore identify a condition under which our results can be extended to interdependent-value environments and tightly connect this identified condition to the notion of strategic distinguishability due to Bergemann and Morris (2009b).

JEL Classification: C72, D78, D82.

Keywords: Continuous implementation, full implementation, incentive compatibility, robustness, transfers

*We thank Atsushi Kajii, Michihiro Kandori, Hitoshi Matsushima, Rene Saran, Roberto Serrano, Satoru Takahashi, Olivier Tercieux, Rajiv Vohra, Takuma Wakayama and seminar participants at Bilkent University, Brown University, Deakin University, Hitotsubashi University, Hong Kong University of Science and Technology, Monash University, Singapore Management University, and University of Tokyo for very helpful comments. Financial support from the Singapore Ministry of Education Academic Research Fund Tier 1 (Chen and Sun) and from the Japan Society for the Promotion of Science (24330078, 25780128) (Kunimoto) is gratefully acknowledged. All remaining errors are our own.

†Department of Economics, National University of Singapore, Singapore 117570, ecsycc@nus.edu.sg

‡School of Economics, Singapore Management University, 90 Stamford Road, Singapore 178903, tkunimoto@smu.edu.sg

§School of International Trade and Economics, University of International Business and Economics, Beijing 100029, sunyifei@uibe.edu.cn

1 Introduction

The theory of *implementation* and *mechanism design* is mainly concerned with the following question: what is the set of outcomes that can be achieved by institutions (or mechanisms)? This institutional design problem is particularly relevant when a group of individuals with conflicting interests has to make a collective decision. The key question then becomes: when can individuals, acting in their own self-interest, arrive at the outcomes consistent with a given welfare criterion (or social choice rule)? To characterize the set of Pareto efficient allocations, for instance, we must know the preferences of those individuals, which is dispersed among the individuals involved. If Pareto efficiency is to be guaranteed, we must elicit this information from the individuals. In what follows, an individual's private information relevant to implementing some welfare criterion is referred to as the individual's *type*. Obviously, the difficulty of eliciting types lies in the fact that individuals need not tell the truth.

For this elicitation, we start our discussion from the notion of *partial implementation*. We say that a social choice rule is partially implementable if there exists (i) a mechanism, and (ii) an equilibrium whose outcome coincides with that specified by the rule. To understand the class of partially implementable rules, we often appeal to the *revelation principle*, which says that whenever partial implementation is possible, one can always duplicate the same equilibrium outcome by using the *truthful* equilibrium in the *direct revelation* mechanism. Thus, a necessary condition for the implementation of any welfare criterion is its *incentive compatibility*, which is simply the property that the best thing for each individual to do in the direct revelation mechanism is to report his true type as long as all other individuals truthfully announce their types.

Although the revelation principle can be adopted in many applications, it is important to realize that the direct revelation mechanism may possess other *untruthful* equilibria whose outcomes are not consistent with the welfare criterion. This problem of multiple equilibria is not merely hypothetical; rather, it has been found by researchers in numerous contexts to be a severe problem.[1] In order to take seriously the problems resulting from the multiplicity of equilibria, some researchers have turned to the question of *full implementation*, and explored the conditions under which the *set* of equilibrium outcomes coincides with a given welfare criterion. The literature of full implementation proposes a variety of mechanisms with the additional property that undesirable outcomes do not arise as equilibria.

The main objective of this paper is to advance the literature of (full) implementation theory in several dimensions. Before going into the detail of our results, we shall start by articulating the domain of problems to which our results apply. First, we consider environments in which monetary transfers among the players are available and all players have

[1] See Bassetto and Phelan (2008) in optimal income taxation, Demski and Sappington (1984) in incentive contracts, Postlewaite and Schmeidler (1986) and Palfrey and Srivastava (1987) in Bayesian implementation in exchange economies, and Repullo (1985) in dominant-strategy equilibrium implementation in social choice environments.

quasilinear utilities. We can justify this class of environments because most of the settings in the applications of mechanism design are in economies with money. Second, we employ the *stochastic* mechanisms in which lotteries are explicitly used. Therefore, we assume that each player has a von Neumann-Morgenstern expected utility. Third, we adopt *the iterative deletion of weakly dominated strategies* as our solution concept. We say that an action a_i is weakly dominated by another action a_i' if, no matter how other players play the game, a_i' cannot be worse than a_i and sometimes it can be strictly better. Fourth, we consider the case in which players' signals are correlated. Although this excludes the case of independent signals, we argue by example that this exclusion is indispensable in our construction. Note also that we can easily restore the correlation by slightly perturbing these signals. Finally, we adopt an approximate version of full implementation, which aims at achieving the socially optimal outcome together with some small ex post transfers. We say that a social choice rule is *implementable with arbitrarily small transfers* if one can design a mechanism whose set of equilibrium outcomes coincides with that specified by the rule, which allows for arbitrarily small ex post transfers among the players.

Given the qualification we have made clear above, we are ready to state one of our main results in private-value environments: a social choice rule is implementable with arbitrarily small transfers if and only if it is incentive compatible (Theorem 1). We also show in Section A.1 that the order of deletion of strategies does not matter. As we regard our mechanism as a natural extension of that of Abreu and Matsushima (1994) to incomplete information environments, we must stress that our mechanism is finite and uses no devices like "integer games" in which each player announces an integer and the player who announces the highest integer gets to be a dictator. This exhibits a clear contrast with Palfrey and Srivastava (1989), who employ an integer game-like device to establish a very similar result to ours. The finiteness of our mechanism is not only a nice property, as we need not employ integer games, but also an important instrument that enables us to derive all of our applications we discuss below. Although our mechanism exploits the power of ex post transfers, we can make these transfers arbitrarily small. Since small ex post transfers result in only an arbitrarily small cost for full implementation, we believe that all individuals would be willing to accept this small cost as a negligible entry fee to participate in the mechanism.

We now discuss how our results are closely related to the recent developments in implementation theory (Sections 1.1, 1.2, 1.3, and 1.4). Section 1.5 discusses how we can extend our result to interdependent-value environments. Finally, we provide the plan of the paper in Section 1.6.

1.1 No Transfers

While the use of small ex post transfers strikes us as being innocuous, it would still be interesting to know when we can avoid any ex post transfers "on the equilibrium." If there are no ex post transfers "on the equilibrium," a social choice rule is said to be *implementable with no transfers*. We propose two classes of environments in which we can achieve implementation

with no transfers. The first class of environments is the case of *nonexclusive-information (NEI)* structures (Theorem 2). NEI captures the situation in which any unilateral deception from the truth-telling in the direct revelation mechanism can be detected. The second class of environments is the case in which there are no consumption externalities among the players and each player only cares about his own consumption (Theorem 3). We can think of exchange economies as an example of this situation. In this environment, however, we need to strengthen the requirement of incentive compatibility into "strict" incentive compatibility.

1.2 Continuous Implementation

Oury and Tercieux (2012) recently shed light on the connection between partial and full implementation. They consider the following situation: The planner wants not only one equilibrium of his mechanism to yield a desired outcome in his initial model (i.e., partial implementation) but it to continue to do so in all models "close" to his initial model. This is what they call *continuous (partial) implementation*. They show that *Bayesian monotonicity*, which is a necessary condition for full implementation, becomes necessary for (strict) continuous implementation. Hence, continuous implementation can be a strong argument for full implementation.

As Bayesian monotonicity sometimes becomes a stringent constraint, we establish a very permissive continuous implementation result. That is, incentive compatibility is the only constraint we need to take into account so that our finite mechanism also achieves continuous implementation as long as the planner can allow for small ex post transfers (Theorem 4). We regard this as a significant finding because few positive continuous implementation results had been offered in the literature.[2]

1.3 $\overline{UNE}$-Implementation

If the planner wants all equilibria of his mechanism to yield a desired outcome under complete information, and entertains the possibility that players may have even the slightest uncertainty about payoffs, then the planner should insist on a solution concept hat has a closed graph in the limit of complete information. Chung and Ely (2003) add this closed-graph property to full implementation in undominated Nash equilibrium (i.e., Nash equilibrium where no players use weakly dominated actions) and call the corresponding concept "$\overline{UNE}$-implementation". They show that *Maskin monotonicity*, a necessary condition for Nash implementation, becomes a necessary condition for $\overline{UNE}$-implementation. For their proof, Chung and Ely need to construct a nearby interdependent-value environment around complete information, in which some players have superior information about the preferences of other players. Since we focus only on private-value environments, their result does not apply

[2]One notable exception is de Clippel, Saran, and Serrano (2014), who show that strict incentive compatibility is a sufficient condition for continuous implementation when the players are constrained by their reasoning ability.

to us. In complete information environments, we instead show that any social choice rule is $\overline{UNE}$-implementable with no transfers when there are at least three players.

1.4 Full Surplus Extraction

In a seminar paper, Crémer and McLean (1988) show that in a single object auction with generic correlated types, it is possible to design a mechanism that extracts all the surplus from the agents.[3] Although this is a surprisingly positive result, Brusco (1998) points out that the mechanism of Crémer and McLean (1988) might possess undesirable equilibria. We can resolve this multiplicity of equilibrium problem so that the full surplus extraction outcome is fully implementable with arbitrarily small transfers, as long as players do not use weakly dominated strategies (Corollary 4).

1.5 Interdependent-Value Environments

We consider our mechanism as a natural extension of the one proposed by Abreu and Matsushima (1994) to incomplete information environments. In fact, Abreu and Matsushima (1994) conjecture that their result under complete information can be extended in a similar manner of Abreu and Matsushima (1992b), who extend the result of *virtual* (as opposed to exact) implementation of Abreu and Matsushima (1992a) to incomplete information environments. By virtual implementation we mean a notion in which the planner contents himself with implementing the social choice rule with arbitrarily high probability. To the extent that our mechanism is a natural extension of that of Abreu and Matsushima (1994), we argue by example in Section 4.1 that one needs more conditions than those in Abreu and Matsushima (1992b) to extend the result to interdependent-value environments. Moreover, we identify a condition (called Assumption 2) under which our Theorem 1 can be extended to interdependent-value environments (Theorem 5). This exhibits a stark contrast with Palfrey and Srivastava (1989), who only provide an example (their Example 3) that illustrates the limitation of their result in interdependent-value environments.

We elaborate more on the issue of how to extend our result to interdependent-value environments by introducing the concept of *strategic (in)distinguishability* due to Bergemann and Morris (2009b): two payoff types are said to be strategically distinguishable if and only if there is a finite mechanism for which the two types do not share the intersection in terms of actions that survive iterated elimination of *ex post* (not interim) strictly dominated strategies.[4] We argue that the extension of our implementation result to interdependent-value environments is tightly connected to the notion of strategic distinguishability. First, in the example of Section 4.1, all the payoff types turn out to be strategically indistinguishable. Second, our Assumption 2 holds generically as long as each player has at least two distinct payoff types that are strategically distinguishable. Strategic distinguishability was originally

[3]See Section 3.4 and footnote 15 for the properties of the mechanism.

[4]See Bergemann and Morris (2009b) for more on this solution concept.

proposed in the context of *robust (virtual) implementation* whose results are made belief-free, whereas our paper is concerned with *interim implementation* that takes the players' interim beliefs as the primitive of the model. To the best of our knowledge, we are the first who find it essential to use the notion of robust implementation to obtain the results for interim implementation. In Section 5.2, we also detail exactly how Abreu and Matsushima (1992b) obtain permissive "virtual" implementation results in fully interdependent-value environments. We regard the difference between the permissive result of virtual implementation and our substantial restriction on the admissible class of interdependent-value environments as a revealing feature that was completely absent under complete information.

1.6 Plan of the Paper

The rest of the paper is organized as follows: In Section 2, we focus on private-value environments. More specifically, in Section 2.1, we introduce the preliminary notation and definitions. In Section 2.2, we discuss our solution concept and the concept of implementation. In Section 2.3, we introduce an assumption (Assumption 1) that we maintain throughout Sections 2 and 3 as well as provide some preliminary results. In Section 2.4, we establish our result in private-value environments (Theorem 1). Section 3 discusses four applications of our Theorem 1: we propose two classes of environments within which we can achieve implementation with "no" transfers (Section 3.1). We investigate the connection to continuous implementation (Section 3.2), to $\overline{UNE}$-implementation (Section 3.3), and to the full surplus extraction (Section 3.4). In Section 4, we extend Theorem 1 to interdependent-value environments. Specifically, in Section 4.1, we show by example that the most optimistic version of the extension to interdependent-value environments is not possible. In Section 4.2, we replace Assumption 1 with a new assumption (Assumption 2) under which our Theorem 1 can be extended to interdependent-value environments (Theorem 5). In Section 5, we discuss the role of honesty, with which we can strengthen our results to that of rationalizable implementation (Section 5.1); and we compare our results with virtual implementation results of Abreu and Matsushima (1992b) (Section 5.2). In the Appendix, we show that in private-value environments, the order of removal of strategies is irrelevant (Section A.1). In the rest of the Appendix, we provide all the proofs omitted from the main body of the paper (Sections A.2 through A.5).

2 Private-Value Environments

In this section, we first focus on the implementation problem in private-value environments. The implementation results in general interdependent-value environments will be provided in Section 4.

2.1 The Environment

Let I denote a finite set of players and with abuse of notation, we also denote by I the cardinality of I. The set of pure social alternatives is denoted by A, and $\Delta(A)$ denotes the set of all probability distributions over A with countable supports. In this context, $a \in A$ denotes a pure social alternative and $x \in \Delta(A)$ denotes a lottery on A.

The utility index of player i over the set A is denoted by $u_i : A \times \Theta_i \to \mathbb{R}$, where Θ_i is the countable set of payoff types and $u_i(a, \theta_i)$ specifies the bounded utility of player i from the social alternative a under $\theta_i \in \Theta_i$. Throughout the paper, we make the following mild assumptions on Θ_i and $u_i(\cdot)$: (i) every payoff type corresponds to a distinct preference, i.e., for any $i \in I$ and $\theta_i, \theta_i' \in \Theta_i$ with $\theta_i \neq \theta_i'$, $u_i(\cdot, \theta_i)$ is not a positive affine transformation of $u_i(\cdot, \theta_i')$; and (ii) every payoff type is never indifferent over outcomes, i.e., for every $i \in I$ and $\theta_i \in \Theta_i$, $u_i(\cdot, \theta_i)$ is not a constant function on A. Denote $\Theta = \Theta_1 \times \cdots \times \Theta_I$ and $\Theta_{-i} = \Theta_1 \times \cdots \times \Theta_{i-1} \times \Theta_{i+1} \times \cdots \times \Theta_I$.[5] We abuse notation to use $u_i(x, \theta_i)$ as player i's expected utility from a lottery $x \in \Delta(A)$ under θ_i. We also assume that player i's utility is quasilinear in transfers, denoted by $u_i(x, \theta_i) + \tau_i$ where $\tau_i \in \mathbb{R}$.

A *model* $\mathcal{T}$ is a triplet $(T_i, \hat{\theta}_i, \pi_i)_{i \in I}$, where T is a countable type space[6]; $\hat{\theta}_i : T_i \to \Theta_i$; and $\pi_i(t_i) \in \Delta(T_{-i})$ denotes the associated interim belief for each $t_i \in T_i$. We assume that each player of type t_i always knows his own type t_i. For each type profile $t = (t_i)_{i \in I}$, let $\hat{\theta}(t)$ denote the payoff type profile at t, i.e., $\hat{\theta}(t) \equiv (\hat{\theta}_i(t_i))_{i \in I}$. If T_i is a finite set for every player i, then we say $(T_i, \hat{\theta}_i, \pi_i)_{i \in I}$ is a *finite* model. Let $\pi_i(t_i)[E]$ denote the probability that $\pi_i(t_i)$ assigns to any set $E \subset T_{-i}$.

Given a model $(T_i, \hat{\theta}_i, \pi_i)_{i \in I}$ and a type $t_i \in T_i$, the *first-order belief* of t_i on Θ is computed as follows: for any $\theta \in \Theta$,

$$h_i^1(t_i)[\theta] = \pi_i(t_i)[\{t_{-i} \in T_{-i} : \hat{\theta}(t_i, t_{-i}) = \theta\}].$$

The *second-order belief* of t_i is his belief about t_{-i}^1, set as follows: for any measurable set $F \subset \Theta \times \Delta(\Theta)^{I-1}$,

$$h_i^2(t_i)[F] = \pi_i(t_i)\left[\{t_{-i} : (\hat{\theta}(t_i, t_{-i}), h_{-i}^1(t_{-i})) \in F\}\right].$$

An entire hierarchy of beliefs can be computed similarly. $\left(h_i^1(t_i), h_i^2(t_i), ..., h_i^\ell(t_i), ...\right)$ is an infinite hierarchy of beliefs induced by type t_i of player i. We assume the belief hierarchy is coherent, that is, for any l, any $X = \text{supp}\left(h_i^l(t_i)\right) \cap \text{supp}\left(h_i^{l-1}(t_i)\right)$,

$$\text{marg}_X h_i^l(t_i) = \text{marg}_X h_i^{l-1}(t_i).$$

Therefore, we assume it is common knowledge that each player of type t_i always knows his

[5] Similar notation will be used for other product sets.

[6] We distinguish payoff type space from type space for the discussion of the robustness property of our main results (see Section 3).

own payoff type and holds coherent belief hierarchy. We denote by T_i^* the set of player i's hierarchies of beliefs in this space and write $T^* = \prod_{i\in I} T_i^*$. T_i^* is endowed with the product topology so that we say a sequence of types $\{t_i[n]\}_{n=0}^{\infty}$ converges to a type t_i (denoted as $t_i[n] \to_p t_i$), if for every $\ell \in \mathbb{N}$, $h_i^\ell(t_i[n]) \to h_i^\ell(t_i)$ as $n \to \infty$. We write $t[n] \to_p t$ if $t_i[n] \to_p t_i$ for all i.

Throughout the paper, we consider a fixed environment $\mathcal{E}$ which is a triplet $\left(A, (u_i)_{i\in I}, \bar{\mathcal{T}}\right)$ with a finite model $\bar{\mathcal{T}} = \left(\bar{T}_i, \bar{\theta}_i, \bar{\pi}_i\right)_{i\in I}$ and a *planner* who aims to implement a *social choice function* (henceforth, SCF) $f : \bar{T} \to \Delta(A)$.[7]

2.2 Mechanisms, Solution Concepts, and Implementation

We assume that the planner can fine or reward any player by *side payments*. A *mechanism* $\mathcal{M}$ is a triplet $((M_i), g, (\tau_i))_{i\in I}$ where M_i is the nonempty "finite" *message space* for player i; $g : M \to \Delta(A)$ is an *outcome function*; and $\tau_i : M \to \mathbb{R}$ is a *transfer rule* from player $i \in I$ to the designer. For any $\alpha_i \in \Delta(M_i)$ and $\alpha_{-i} \in \Delta(M_{-i})$, we abuse the notation to denote by $g(\alpha_i, \alpha_{-i})$ the induced lottery in $\Delta(A)$ and by $\tau_i(\alpha_i, \alpha_{-i})$ the induced expected transfer. We say that a mechanism $\mathcal{M}$ has fines and rewards bounded by $\bar{\tau}$ if $|\tau_i(m)| \leq \bar{\tau}$ for every $i \in I$ and every $m \in M$. Note that there is a class of such mechanisms given $\bar{\tau}$. We denote one of the mechanisms by $(\mathcal{M}, \bar{\tau})$.

Given a mechanism $\mathcal{M}$, let $U(\mathcal{M}, \mathcal{T})$ denote an incomplete information game associated with a model $\mathcal{T}$. Fix a game $U(\mathcal{M}, \mathcal{T})$, player $i \in I$ and type $t_i \in T_i$. We say that $m_i \in S^0 W_i(t_i|\mathcal{M}, \mathcal{T})$ if and only if there does not exist $m_i' \in M_i$ such that

$$\begin{aligned} & \sum_{t_{-i}} \left[u_i(g(m_i', \nu_{-i}(t_{-i})), \hat{\theta}_i(t_i)) + \tau_i(m_i', \nu_{-i}(t_{-i}))\right] \pi_i(t_i)[t_{-i}] \\ \geq & \sum_{t_{-i}} \left[u_i(g(m_i, \nu_{-i}(t_{-i})), \hat{\theta}_i(t_i)) + \tau_i(m_i, \nu_{-i}(t_{-i}))\right] \pi_i(t_i)[t_{-i}] \end{aligned}$$

for all $\nu_{-i} : T_{-i} \to M_{-i}$ and a strict inequality holds for some $\nu_{-i} : T_{-i} \to M_{-i}$.[8] For any $l \geq 1$, we say that $m_i \in S^{l+1} W_i(t_i|\mathcal{M}, \mathcal{T})$ if and only if there does not exist $m_i' \in M_i$ such that

$$\begin{aligned} & \sum_{t_{-i}} \left[u_i(g(m_i', \nu_{-i}(t_{-i})), \hat{\theta}_i(t_i)) + \tau_i(m_i', \nu_{-i}(t_{-i}))\right] \pi_i(t_i)[t_{-i}] \\ > & \sum_{t_{-i}} \left[u_i(g(m_i, \nu_{-i}(t_{-i})), \hat{\theta}_i(t_i)) + \tau_i(m_i, \nu_{-i}(t_{-i}))\right] \pi_i(t_i)[t_{-i}] \end{aligned}$$

for all $\nu_{-i} : T_{-i} \to M_{-i}$ and for all t_{-i} and m_{-i} such that $\nu_{-i}(t_{-i}) \in S^l W_{-i}(t_{-i}|\mathcal{M}, \mathcal{T}) = \prod_{j\neq i} S^l W_j(t_j|\mathcal{M}, \mathcal{T})$. Let $S^\infty W$ denote the set of strategy profiles which survive one round

[7] We will consider a countable model when we define and study continuous implementation in Section 5.1.

[8] Our solution concept is equivalent to the one which allows for mixed strategy conjectures $\nu : T_{-i} \to \Delta(M_{-i})$.

of removal of weakly dominated strategies followed by iterative removal of strictly dominated strategies, i.e.,

$$S^{\infty}W_i\left(t_i|\mathcal{M},\mathcal{T}\right)=\bigcap_{l=1}^{\infty}S^{l}W_i\left(t_i|\mathcal{M},\mathcal{T}\right),$$

$$S^{\infty}W\left(t|\mathcal{M},\mathcal{T}\right)=\prod\nolimits_{i\in I}S^{\infty}W_i\left(t_i|\mathcal{M},\mathcal{T}\right).$$

In defining the solution, we require that the dominating strategies be pure strategies instead of mixed strategies. This makes the solution concept weaker and thus the implementation result stronger. We refer the reader to Börgers (1994) and Dekel and Fudenberg (1990) for the foundations of $S^{\infty}W$ in complete information games, and to Frick and Romm (2014) for its foundation in incomplete information games. The order of elimination of strategies in $S^{\infty}W$ generally matters, as WS^{∞} (the set of strategy profiles which survive iterative removal of strictly dominated strategies followed by one round of removal of weakly dominated strategies) may well be different from $S^{\infty}W$. In the appendix, we show that the iterative removal of weakly dominated strategy profiles in any order generates the same outcome as $S^{\infty}W$ in our mechanism. We can also define S^{∞} as the set of strategy profiles that survive the iterative removal of strictly dominated strategies. It is already well known that S^{∞} is order-independent and equivalent to the set of all rationalizable strategies in finite mechanisms. In Section 5.1, we will discuss the role of S^{∞} in our mechanism.

We now formally state the definition of implementability in $S^{\infty}W$. First, we allow the size of transfers to be arbitrarily small so that we propose the concept of implementation with arbitrarily small transfers.

Definition 1 (Implementation with Arbitrarily Small Transfers) *An SCF f is implementable in $S^{\infty}W$ with arbitrarily small transfers if for all $\bar{\tau}>0$, there is a mechanism $(\mathcal{M},\bar{\tau})$ such that for any $t\in\bar{T}$, and $m\in S^{\infty}W\left(t|\mathcal{M},\bar{\mathcal{T}}\right)$, we have $g\left(m\right)=f\left(t\right)$.*

2.3 An Assumption

We first follow Abreu and Matsushima (1992a), who show the following important result: it guarantees the existence of a function that can elicit each player's preference.

Lemma 1 *(Abreu and Matsushima (1992a)) For each $i\in I$, there exists a function $x_i:\bar{T}_i\to\Delta(A)$ such that for any $t_i,t_i'\in\bar{T}_i$, whenever $\hat{\theta}_i(t_i)\neq\hat{\theta}_i(t_i')$,*

$$u_i(x_i(t_i),\hat{\theta}_i(t_i))>u_i(x_i(t_i'),\hat{\theta}_i(t_i)) \tag{1}$$

Throughout until the end of Section 3, we make a single assumption on the environments.

Assumption 1 *An environment $\mathcal{E}$ satisfies Assumption 1 if, for all $i\in I$ and $t_i,t_i'\in\bar{T}_i$ with $t_i\neq t_i'$, it follows that $h_i^1(t_i)\neq h_i^1(t_i')$.*

Remark 1 *Assumption 1 states that two distinct types hold different beliefs over other players' preferences. Since $\bar{T}$ is finite, if there are at least two players and each player has at least two distinct preferences, Assumption 1 generically holds in the space of the probability distributions over $\bar{T}$. Note, however, that Assumption 1 fails to hold in the case of independent probability distributions.*

Remark 2 *Suppose that for any $i \in I$, and $t_i, t_i' \in \bar{T}_i$, whenever $t_i \neq t_i'$, we have $\hat{\theta}_i(t_i) \neq \hat{\theta}_i(t_i')$. In this case, Assumption 1 is equivalent to assuming any pair of distinct types hold distinct beliefs over the opponents' types. That is, the environment satisfies the beliefs-determine-preferences property on which Heifetz and Neeman (2006) has a detailed discussion (see also footnote 16).*

By Assumption 1, we can construct the following *transfer rule* $d_i^0 : T_i \times \Theta_{-i} \to \mathbb{R}$:

Lemma 2 *Suppose that an environment $\mathcal{E}$ satisfies Assumption 1. For all $i \in I$ with t_i and θ_{-i} such that there exists $t_{-i} \in T_{-i}$ such $\hat{\theta}_{-i}(t_{-i}) = \theta_{-i}$, define*

$$d_i^0(t_i, \theta_{-i}) = 2h_i^1(t_i)[(\hat{\theta}_i(t_i), \theta_{-i})] - h_i^1(t_i) \cdot h_i^1(t_i),$$

where $h_i^1(t_i) \cdot h_i^1(t_i)$ denotes its inner (or dot) product. Then, for all $i \in I$ and $t_i, t_i' \in \bar{T}_i$ with $t_i \neq t_i'$,

$$\sum_{\theta_{-i}} \left[d_i^0 \left(t_i, \theta_{-i} \right) - d_i^0 \left(t_i', \theta_{-i} \right) \right] h_i^1(t_i) \left[(\hat{\theta}_i(t_i), \theta_{-i}) \right] > 0. \tag{2}$$

Remark 3 *Lemma 2 guarantees the existence of a transfer rule in which each player will strictly prefer to tell the truth whenever he believes that every other one tells the truth. That is, such a transfer rule is strictly Bayesian incentive compatible. When there are more than two players, we can construct d_i^0 under a stronger (and yet still generic) version of Assumption 1, following d'Aspremont, Crémer, and Gérard-Varet (2003). Then, we can achieve budget balance for d_i^0 by allocating all the other transfers only across agents, we can achieve budget balance everywhere (both on and off the solution outcome). Assumption 1 is a condition which holds generically (see discussions in Johnson, Pratt, and Zeckhauser (1990) and d'Aspremont, Crémer, and Gérard-Varet (2003)).*

Proof. The construction of $d_i^0 \left(t_i, \theta_{-i} \right)$ makes itself a proper scoring rule. By Assumption 1, the strict inequality of (2) always holds. ■

2.4 The Results

Here we provide the main result of the section which characterizes implementation with arbitrarily small transfers. We shall show that an SCF f is implementable in $S^\infty W$ with

arbitrarily small transfers if and only if it is incentive compatible. First, we introduce the notation. For every $i \in I$, every $t_i, t_i' \in \bar{T}_i$, let

$$\sum_{t_{-i} \in \bar{T}_{-i}} u_i(f(t_{-i}, t_i'), \hat{\theta}_i(t_i)) \bar{\pi}_i(t_i)[t_{-i}]$$

denote the expected utility generated by the direct revelation mechanism for player i of type t_i when he announces t_i' and the other players all make truthful announcements.

Definition 2 *An SCF $f : \bar{T} \to \Delta(A)$ is **incentive compatible** if, for all $i \in I$ and all $t_i, t_i' \in \bar{T}_i$,*

$$\sum_{t_{-i} \in \bar{T}_{-i}} u_i(f(t_{-i}, t_i), \hat{\theta}_i(t_i)) \bar{\pi}_i(t_i)[t_{-i}] \geq \sum_{t_{-i} \in \bar{T}_{-i}} u_i(f(t_{-i}, t_i'), \hat{\theta}_i(t_i)) \bar{\pi}_i(t_i)[t_{-i}].$$

The theorem below shows that incentive compatibility is a necessary and sufficient condition for implementation with arbitrarily small transfers.

Theorem 1 *Suppose that $I \geq 2$ and the environment $\mathcal{E}$ satisfies Assumption 1. Then, an SCF f is implementable in $S^\infty W$ with arbitrarily small transfers if and only if it is incentive compatible.*

2.4.1 The Mechanism

We propose a natural extension of the mechanism proposed by Abreu and Matsushima (1994) from complete information to incomplete information environments. We define the mechanism as follows.

1. **The message space**:

   Each player i makes $(K+3)$ simultaneous announcements of his own type. We index each announcement by $-2, -1, 0, 1, \ldots, K$. That is, player i's message space is

   $$M_i = M_i^{-2} \times M_i^{-1} \times M_i^0 \times \cdots \times M_i^K = \underbrace{\bar{T}_i \times \cdots \times \bar{T}_i}_{K+3 \text{ times}},$$

   where K is an integer to be specified later. Denote

   $$m_i = \left(m_i^{-2}, ..., m_i^K\right) \in M_i,\ m_i^k \in M_i^k,\ k \in \{-2, -1, 0, ..., K\},$$

   and

   $$m = \left(m^{-2}, ..., m^K\right) \in M,\ m^k = \left(m_i^k\right)_{i \in I} \in M^k = \times_{i \in I} M_i^k.$$

   We use $m^k / \tilde{m}_i$ denote the message profile $\left(m_1^k, ..., m_{i-1}^k, \tilde{m}_i^k, m_{i+1}^k, ..., m_I^k\right)$.

2. **The outcome function**:

   Let $\epsilon \in (0,1)$ be a small positive number.

   Define $e : M^{-1} \times M^0 \to \mathbb{R}$ by

$$e(m^{-1}, m^0) = \begin{cases} \epsilon & \text{if } m_i^{-1} \neq m_i^0 \text{ for some } i \in I, \\ 0 & \text{otherwise.} \end{cases}$$

   The outcome function $g : M \to \Delta(A)$ is defined as follows: for each $m \in M$,

$$g(m) = e(m^{-1}, m^0) \frac{1}{I} \sum_{i \in I} x_i\left(m_i^{-2}\right) + \left\{1 - e(m^{-1}, m^0)\right\} \frac{1}{K} \sum_{k=1}^{K} f\left(m^k\right), \tag{3}$$

   The outcome function contains a "random dictator" component (recall the function x_i defined in (1)) which is triggered in the event that some player's -1th announcement does not equal his 0th announcement. When this event does not happen, only the nondictatorial component is triggered, which consists of K equally weighted lotteries the kth of which depends only on the I-tuple of kth announcements.

3. **The transfer rule**:

   Let λ, ξ and η be positive numbers. Player i is to pay:

   - $-\lambda d_i^0(m_i^{-1}, \hat{\theta}_{-i}(m_{-i}^{-2}))$;
   - $-\lambda d_i^0(m_i^0, \hat{\theta}_{-i}\left(m_{-i}^{-1}\right))$;[9]
   - ξ if he is the first player whose kth announcement ($k \geq 1$) differs from his own 0th announcement (All players who are the first to deviate are fined).

$$d_i\left(m^0, ..., m^K\right) = \begin{cases} \xi & \text{if there exists } k \in \{1, ..., K\} \text{ s.t. } m_i^k \neq m_i^0, \\ & \text{and } m_j^{k'} = m_j^0 \text{ for all } k' \in \{1, ..., k-1\} \text{ for all } j; \\ 0 & \text{otherwise.} \end{cases} \tag{4}$$

   - η if his kth announcement ($k \geq 1$) differs from his own 0th announcement.

$$d_i^k\left(m_i^0, m_i^k\right) = \begin{cases} \eta & \text{if } m_i^k \neq m_i^0; \\ 0 & \text{otherwise.} \end{cases} \tag{5}$$

[9]The design of the two scoring rules is needed for establishing the order independence result in the appendix. The other results in this paper still go through with only one scoring rule.

In total,

$$\tau_i(m) = -\lambda d_i^0(\hat{\theta}_{-i}(m_{-i}^{-2}), m_i^{-1}) - \lambda d_i^0(\hat{\theta}_{-i}(m_{-i}^{-1}), m_i^0) + d_i(m^0, ..., m^K) + \sum_{k=1}^{K} d_i^k\left(m_i^0, m_i^k\right). \tag{6}$$

4. Define $\bar{\Theta}_i = \{\theta_i \in \Theta_i |\ \hat{\theta}_i(\bar{t}_i) = \theta_i \text{ for some } \bar{t}_i \in \bar{T}_i\}$. We provide the summary of conditions on transfers:

   Let

$$E = \max_{m_i^{-2} \in M_i^{-2}, m^k \in M^k, \bar{\theta}_i \in \bar{\Theta}_i, i \in I} \left| \frac{1}{I} \sum_{j \in I} u_i\left(x_j(m_j^{-2}), \bar{\theta}_i\right) - u_i\left(f\left(m^k\right), \bar{\theta}_i\right) \right|; \tag{7}$$

$$D = \max_{\bar{m}_i^k \in M_i^k, m^k \in M^k, \bar{\theta}_i \in \bar{\Theta}_i, i \in I} \left\{ u_i\left(f\left(m^k\right), \bar{\theta}_i\right) - u_i(f(m_{-i}^k, \bar{m}_i^k), \bar{\theta}_i) \right\}, \tag{8}$$

   where E multiplied by ϵ is the upper bound of the gain for any player i, of triggering or not triggering the random dictatorial component; D is the maximum gain for player i from altering the kth announcement, where $k \geq 1$.

   We choose positive numbers λ, γ, K, ϵ, η, and ξ such that for every $i \in I$ and every $t_i, t_i' \in \bar{T}_i$ with $t_i \neq t_i'$,

$$\bar{\tau} > 2\lambda \bar{d}_i^0 + \xi + K\eta; \tag{9}$$

$$\lambda \sum_{\theta_{-i}} \left[d_i^0(t_i, \theta_{-i}) - d_i^0(t_i', \theta_{-i}) \right] h_i^1(t_i) \left[(\hat{\theta}_i(t_i), \theta_{-i}) \right] > \gamma; \tag{10}$$

$$\eta > \epsilon E; \tag{11}$$

$$\xi > \frac{1}{K} D; \tag{12}$$

$$\gamma > \epsilon E + \xi + K\eta, \tag{13}$$

   where $\bar{d}_i^0$ denotes an upper bound of $d_i^0(\cdot)$.[10]

2.4.2 The Proof

We use the following claims to prove the "if" part of Theorem 1.

Claim 1 *In the game $U\left(\mathcal{M}, \bar{\mathcal{T}}\right)$, for every $i \in I$, $\bar{t}_i, t_i \in \bar{T}_i$, and $m_i \in M_i$, if $m_i \in S^0 W_i\left(\bar{t}_i | \mathcal{M}, \bar{\mathcal{T}}\right)$, then $m_i^{-2} = t_i$ such that $\hat{\theta}_i(t_i) = \hat{\theta}_i(\bar{t}_i)$.*

[10] Given any $\bar{\tau} > 0$ exogenously, we first choose λ small enough so that $\lambda \bar{d}_i^0 < \frac{1}{4}\bar{\tau}$. Second, by (2), we can choose γ small enough so that (10) holds. Third, we choose K large enough so that $\frac{1}{K} D < \min\left\{\frac{1}{4}\bar{\tau}, \frac{1}{3}\gamma\right\}$. Fourth, we choose ε small enough so that $K\epsilon E < \min\left\{\frac{1}{4}\bar{\tau}, \frac{1}{3}\gamma\right\}$. Therefore, we have $\bar{\tau} > 2\lambda \bar{d}_i^0 + \frac{1}{K} D + K\epsilon E$ and $\gamma > \epsilon E + \frac{1}{K} D + K\epsilon E$. From these two inequalities, we can thus choose η and ξ such that (9), (11), (12) and (13) hold.

Proof. We show that for any $i \in I$, $\bar{t}_i, t_i \in \bar{T}_i$, and $m_i \in M_i$, if $m_i^{-2} = t_i$ and $\hat{\theta}_i(t_i) = \hat{\theta}_i(\bar{t}_i)$, then $m_i \notin S^0 W_i\left(\bar{t}_i | \mathcal{M}, \bar{\mathcal{T}}\right)$, i.e., m_i is weakly dominated by some m_i', which is constructed as follows:

$$m_i' = \left(\bar{t}_i, m_i^{-1}, ..., m_i^K\right).$$

We let $\hat{\theta}_i(\bar{t}_i) = \bar{\theta}_i$. and fix any conjecture $\nu_{-i} : \bar{T}_{-i} \to M_{-i}$ and we write $\nu_{-i}^k(t_{-i})$ for the kth round report of other player with type t_{-i}. The difference of the expected utilities between m_i' and m_i for player i of type $\bar{t}_i$ is shown as follows:

$$\begin{aligned}
& \sum_{t_{-i}} \left\{u_i\left(g\left(m_i', \nu_{-i}(t_{-i})\right), \bar{\theta}_i\right) + \tau_i\left(m_i', \nu_{-i}(t_{-i})\right)\right\} \pi_i(\bar{t}_i)[t_{-i}] \\
& - \sum_{t_{-i}} \left\{u_i\left(g\left(m_i, \nu_{-i}(t_{-i})\right), \bar{\theta}_i\right) + \tau_i\left(m_i, \nu_{-i}(t_{-i})\right)\right\} \pi_i(\bar{t}_i)[t_{-i}] \\
= & \sum_{t_{-i}} \frac{e\left(\left(m_i^{-1}, \nu_{-i}^{-1}(t_{-i})\right), m^0\right)}{I} \left\{u_i\left(x_i\left(\bar{t}_i\right), \bar{\theta}_i\right) - u_i\left(x_i\left(m_i^{-2}\right), \bar{\theta}_i\right)\right\} \pi_i(\bar{t}_i)[t_{-i}] \qquad (14) \\
= & \left\{\sum_{t_{-i}} \frac{e\left(\left(m_i^{-1}, \nu_{-i}^{-1}(t_{-i})\right), m^0\right)}{I} \pi_i(\bar{t}_i)[t_{-i}]\right\} \left\{u_i\left(x_i\left(\bar{t}_i\right), \bar{\theta}_i\right) - u_i\left(x_i\left(m_i^{-2}\right), \bar{\theta}_i\right)\right\} \\
\geq & \ 0,
\end{aligned}$$

where the first equality follows because the only difference lies in the function x_i when m_i' differs from m_i only in round -2 announcement, (see the definition of g in (3) and the definition of τ in (6)); by (1) the last inequality is strict whenever $e\left(\left(m_i^{-1}, \nu_{-i}^{-1}(t_{-i})\right), m^0\right) = \epsilon$ for some t_{-i}. ■

The next claim says that telling a lie in round -1 is strictly dominated by telling the truth, given the hypothesis that no players choose weakly dominated messages.

Claim 2 *In the game $U\left(\mathcal{M}, \bar{\mathcal{T}}\right)$, for every $i \in I$, $\bar{t}_i \in \bar{T}_i$, if $m_i \in S^1 W_i\left(\bar{t}_i | \mathcal{M}, \bar{\mathcal{T}}\right)$, then $m_i^{-1} = \bar{t}_i$.*

Proof. We show that for any $i \in I$, $\bar{t}_i \in \bar{T}_i$ with $\hat{\theta}_i(\bar{t}_i) = \bar{\theta}_i$, and $m_i \in S^0 W_i(\bar{t}_i | \mathcal{M}, \bar{\mathcal{T}})$, if $m_i^{-1} \neq \bar{t}_i$, then $m_i \notin S^1 W_i\left(\bar{t}_i | \mathcal{M}, \bar{\mathcal{T}}\right)$. We construct $\bar{m}_i$ as follows:

$$\bar{m}_i = \left(m_i^{-2}, \bar{t}_i, m_i^0, ..., m_i^K\right).$$

For any conjecture $\nu_{-i} : \bar{T}_{-i} \to M_{-i}$, we have that, for any t_{-i},

$$\nu_{-i}(t_{-i}) \in S^0 W_{-i}\left(t_{-i} | \mathcal{M}, \bar{\mathcal{T}}\right).$$

By Claim 1, we know that $m_{-i} \in W_{-i}^{\bar{l}+1}\left(t_{-i} | \mathcal{M}, \bar{\mathcal{T}}\right)$ implies $\hat{\theta}_{-i}\left(m_{-i}^{-2}\right) = \hat{\theta}_{-i}(t_{-i})$.

The difference of the expected values under $\bar{m}_i$ from m_i for player i of type $\bar{t}_i$ is shown as follows:

$$\sum_{t_{-i}} \left\{ u_i(g(\bar{m}_i, \nu_{-i}(t_{-i})), \bar{\theta}_i) + \tau_i \left(\bar{m}_i, \nu_{-i}(t_{-i})\right) \right\} \pi_i(\bar{t}_i)[t_{-i}]$$
$$- \sum_{t_{-i}} \left\{ u_i(g(m_i, \nu_{-i}(t_{-i})), \bar{\theta}_i) + \tau_i \left(m_i, \nu_{-i}(t_{-i})\right) \right\} \pi_i(\bar{t}_i)[t_{-i}]$$

$$\begin{aligned} = & \sum_{t_{-i}} \left\{ e\left(\left(\bar{t}_i, \nu_{-i}^{-1}(t_{-i})\right), m^0\right) - e\left(\left(m_i^{-1}, \nu_{-i}^{-1}(t_{-i})\right), m^0\right) \right\} \\ & \times \left\{ \frac{1}{I} \sum_{j \in I} u_i(x_j(\bar{t}_j), \bar{\theta}_i) - \frac{1}{K} \sum_{k=1}^{K} u_i(f((m_i^k, \nu_{-i}^k(t_{-i}))), \bar{\theta}_i) \right\} \pi_i(\bar{t}_i)[t_{-i}] \\ & + \sum_{t_{-i}} \left\{ \lambda d_i^0(\hat{\theta}_{-i}(t_{-i}), t_i) - \lambda d_i^0(\hat{\theta}_{-i}(t_{-i}), m_i^{-1}) \right\} \pi_i(\bar{t}_i)[t_{-i}] \end{aligned}$$

Observe that when $\bar{m}_i$ differs from m_i only in the -1th announcement, the difference in terms of $g(\cdot)$ (see the outcome function in (3)) lies in function $e(\cdot)$ and the difference in terms of transfer is summarized in functions d_i^0 (see the transfer rule in (6)). We observe the following points:

(i) In terms of outcomes, the possible expected gain of player i of type $\bar{t}_i$ by choosing m_i rather than $\bar{m}_i$ is

$$\sum_{t_{-i}} \left\{ e\left(\left(\bar{t}_i, \nu_{-i}^{-1}(t_{-i})\right), m^0\right) - e\left(\left(m_i^{-1}, \nu_{-i}^{-1}(t_{-i})\right), m^0\right) \right\}$$
$$\times \left\{ \frac{1}{I} \sum_{j \in I} u_i(x_j(\bar{t}_j), \bar{\theta}_i) - \frac{1}{K} \sum_{k=1}^{K} u_i(f((m_i^k, \nu_{-i}^k(t_{-i}))), \bar{\theta}_i) \right\} \pi_i(\bar{t}_i)[t_{-i}]$$

From (7), when playing m_i rather than $\bar{m}_i$, this possible gain is bounded above by ϵE.

(ii) In terms of payments, the expected loss by choosing m_i rather than $\bar{m}_i$ is

$$\begin{aligned} & \sum_{t_{-i}} \left\{ \lambda d_i^0(\hat{\theta}_{-i}(t_{-i}), \bar{t}_i) - \lambda d_i^0(\hat{\theta}_{-i}(t_{-i}), m_i^{-1}) \right\} \pi_i(\bar{t}_i)[t_{-i}] \\ = & \sum_{\theta_{-i}} \left[\lambda d_i^0\left(\theta_{-i}, \bar{t}_i\right) - \lambda d_i^0\left(\theta_{-i}, m_i^{-1}\right) \right] h_i^1(\bar{t}_i) \left[\left(\hat{\theta}_i(\bar{t}_i), \theta_{-i} \right) \right] \end{aligned}$$

Therefore, by (10), the loss is bounded below by γ. Note that $\gamma > \epsilon E$ by (13).

Therefore, m_i is strictly dominated by $\bar{m}_i$. ■

Claim 3 *In the game* $U\left(\mathcal{M},\bar{\mathcal{T}}\right)$*, for every* $i\in I$*,* $\bar{t}_i\in\bar{T}_i$*, if* $m_i\in S^2W_i\left(\bar{t}_i|\mathcal{M},\bar{\mathcal{T}}\right)$*, then* $m_i^0=\bar{t}_i$*.*

Proof. We show that for any $i\in I$, $\bar{t}_i\in\bar{T}_i$ with $\hat{\theta}_i(\bar{t}_i)=\bar{\theta}_i$, if $m_i^0\neq\bar{t}_i$, then $m_i\notin S^2W_i\left(\bar{t}_i|\mathcal{M},\bar{\mathcal{T}}\right)$. We construct $\bar{m}_i$ as follows:

$$\bar{m}_i=\left(m_i^{-2},m_i^{-1},\bar{t}_i,m_i^1,\ldots,m_i^K\right).$$

For any conjecture $\nu_{-i}:\bar{T}_{-i}\to M_{-i}$, we have that,

$$\nu_{-i}(t_{-i})\in S^1W_{-i}\left(t_{-i}|\mathcal{M},\bar{\mathcal{T}}\right).$$

From Claim 2, we know that for any $j\in I$, if $m_j\in S^1W_j\left(\bar{t}_j|\mathcal{M},\bar{\mathcal{T}}\right)$, then $m_j^{-1}=\bar{t}_j$.

The difference of the expected values under $\bar{m}_i$ from m_i for player i of type $\bar{t}_i$ is shown as follows:

$$\begin{aligned}
&\sum_{t_{-i}}\left\{u_i(g(\bar{m}_i,\nu_{-i}(t_{-i})),\bar{\theta}_i)+\tau_i\left(\bar{m}_i,\nu_{-i}(t_{-i})\right)\right\}\pi_i(\bar{t}_i)[t_{-i}]\\
&-\sum_{t_{-i}}\left\{u_i(g(m_i,\nu_{-i}(t_{-i})),\bar{\theta}_i)+\tau_i\left(m_i,\nu_{-i}(t_{-i})\right)\right\}\pi_i(\bar{t}_i)[t_{-i}]\\
=\;&\sum_{t_{-i}}\left\{e\left(\left(t_{-i},\bar{t}_i\right),\left(\nu_{-i}^0(t_{-i}),\bar{t}_i\right)\right)-e\left(\left(t_{-i},\bar{t}_i\right),\left(\nu_{-i}^0(t_{-i}),m_i^0\right)\right)\right\}\\
&\times\left\{\frac{1}{I}\sum_{j\in I}u_i(x_j(\bar{t}_j),\bar{\theta}_i)-\frac{1}{K}\sum_{k=1}^{K}u_i(f((m_i^k,\nu_{-i}^k(t_{-i}))),\bar{\theta}_i)\right\}\pi_i(\bar{t}_i)[t_{-i}]\\
&+\sum_{t_{-i}}\left\{\lambda d_i^0\left(\hat{\theta}_{-i}\left(t_{-i}\right),\bar{t}_i\right)-\lambda d_i^0\left(\hat{\theta}_{-i}\left(t_{-i}\right),m_i^0\right)\right\}\pi_i(\bar{t}_i)[t_{-i}]\\
&+\sum_{t_{-i}}\left\{d_i\left(\left(\nu_{-i}^0(t_{-i}),\bar{t}_i\right),\left(\nu_{-i}^1(t_{-i}),m_i^1\right),\ldots\right)-d_i\left(\left(\nu_{-i}^0(t_{-i}),m_i^0\right),\left(\nu_{-i}^1(t_{-i}),m_i^1\right),\ldots\right)\right\}\pi_i(\bar{t}_i)[t_{-i}]\\
&+\sum_{t_{-i}}\sum_{k=1}^{K}\{d_i^k\left(\bar{t}_i,m_i^k\right)-d_i^k\left(m_i^0,m_i^k\right)\}\pi_i(\bar{t}_i)[t_{-i}]\\
\geq\;&-\epsilon E+\gamma-\xi-K\eta\\
>\;&0
\end{aligned}$$

Observe that when $\bar{m}_i$ differs from m_i only in the 0th announcement, the difference in terms of $g(\cdot)$ (see the outcome function in (3)) lies in function $e(\cdot)$ and the difference in terms of transfer is summarized in functions d_i^0, d_i, and $\{d_i^k\}_{k=1,\ldots,K}$ (see the transfer rule in

(6)). In particular,

$$\begin{aligned} & \sum_{t_{-i}} \left\{ \lambda d_i^0 \left(\hat{\theta}_{-i}\left(t_{-i}\right), \bar{t}_i \right) - \lambda d_i^0 \left(\hat{\theta}_{-i}\left(t_{-i}\right), m_i^0 \right) \right\} \pi_i(\bar{t}_i)[t_{-i}] \\ = & \sum_{\theta} \left[\lambda d_i^0 \left(\theta_{-i}, t_i\right) - \lambda d_i^0 \left(\theta_{-i}, m_i^{-1}\right) \right] h_i^1(t_i) \left[\left(\hat{\theta}_i(t_i), \theta_{-i} \right) \right] \\ > & \gamma \end{aligned}$$

Therefore, m_i is strictly dominated by $\bar{m}_i$. ■

Claim 4 *Suppose that an SCF f is incentive compatible. Given $\bar{\tau} > 0$, let $\mathcal{M}$ be a mechanism associated with f as defined in Section 2.4.1. For each $k \geq 2$, $i \in I$, and $\bar{t}_i \in \bar{T}_i$, if $m_i \in S^k W_i(\bar{t}_i|\mathcal{M}, \bar{\mathcal{T}})$, then $m_i^{k-2} = \bar{t}_i$.*

Proof. Consider type $\bar{t}_i \in \bar{T}_i$ with $\hat{\theta}_i(\bar{t}_i) = \bar{\theta}_i$. When $k = 2$, the result follows from Claim 3. Fix $k \geq 2$. The induction hypothesis is that for every $i \in I$, $\bar{t}_i \in \bar{T}_i$, if $m_i \in S^k W_i \left(\bar{t}_i|\mathcal{M}, \bar{\mathcal{T}}\right)$, then $m_i^{k'} = \bar{t}_i$ for all $k' \leq k-2$.

Then, we show that if $m_i \in S^{k+1} W_i(\bar{t}_i|\mathcal{M}, \bar{\mathcal{T}})$, then $m_i^{k'} = \bar{t}_i$ for all $k' \leq k-1$. It suffices to prove $m_i^{k-1} = \bar{t}_i$. Suppose not, let $\tilde{m}_i$ be the dominating strategy defined as follows,

$$\tilde{m}_i \equiv \left(m_i^{-2}, ..., m_i^{k-2}, \bar{t}_i, m_i^k, ..., m_i^K\right).$$

We let $\widehat{M}_{-i} = \left\{ m_{-i} \in M_{-i} : m_{-i}^{k-1} = m_{-i}^0 \right\}$. Fix a conjecture $\nu_{-i} : \bar{T}_{-i} \to M_{-i}$. Note that, for each t_{-i},

$$\nu_{-i}(t_{-i}) \in S^k W_{-i} \left(t_{-i}|\mathcal{M}, \bar{\mathcal{T}}\right).$$

Thus, for any t, we obtain $e\left(m^{-1}, m^0\right) = 0$ for any $m \in S^k W\left(t|\mathcal{M}, \bar{\mathcal{T}}\right)$.

We will show that

$$\begin{aligned} & \sum_{t_{-i}} \left\{ u_i(g(\tilde{m}_i, \nu_{-i}(t_{-i})), \bar{\theta}_i) + \tau_i(\tilde{m}_i, \nu_{-i}(t_{-i})) \right\} \pi_i(\bar{t}_i)[t_{-i}] \\ & - \sum_{t_{-i}} \left\{ u_i(g(m_i, \nu_{-i}(t_{-i})), \bar{\theta}_i) + \tau_i(m_i, \nu_{-i}(t_{-i})) \right\} \pi_i(\bar{t}_i)[t_{-i}] \qquad (15) \\ > & \ 0. \end{aligned}$$

Note the left hand side of inequality is equal to

$$\begin{aligned} & \sum_{t_{-i}, \nu_{-i}(t_{-i}) \notin \widehat{M}_{-i}} \left\{ \begin{array}{c} \left\{ u_i(g(\tilde{m}_i, \nu_{-i}(t_{-i})), \bar{\theta}_i) + \tau_i(\tilde{m}_i, \nu_{-i}(t_{-i})) \right\} - \\ \left\{ u_i(g(m_i, \nu_{-i}(t_{-i})), \bar{\theta}_i) + \tau_i(m_i, \nu_{-i}(t_{-i})) \right\} \end{array} \right\} \pi_i(\bar{t}_i)[t_{-i}] \qquad (16) \\ + & \sum_{t_{-i}, \nu_{-i}(t_{-i}) \in \widehat{M}_{-i}} \left\{ \begin{array}{c} \left\{ u_i(g(\tilde{m}_i, \nu_{-i}(t_{-i})), \bar{\theta}_i) + \tau_i(\tilde{m}_i, \nu_{-i}(t_{-i})) \right\} - \\ \left\{ u_i(g(m_i, \nu_{-i}(t_{-i})), \bar{\theta}_i) + \tau_i(m_i, \nu_{-i}(t_{-i})) \right\} \end{array} \right\} \pi_i(\bar{t}_i)[t_{-i}]. \end{aligned}$$

Step 1:

$$\sum_{t_{-i},\nu_{-i}(t_{-i})\notin\widehat{M}_{-i}} \left\{ \begin{array}{c} \left\{u_i(g(\tilde{m}_i,\nu_{-i}(t_{-i})),\bar{\theta}_i)+\tau_i(\tilde{m}_i,\nu_{-i}(t_{-i}))\right\}- \\ \left\{u_i(g(m_i,\nu_{-i}(t_{-i})),\bar{\theta}_i)+\tau_i(m_i,\nu_{-i}(t_{-i}))\right\} \end{array} \right\} \pi_i(\bar{t}_i)[t_{-i}]>0.$$

From the induction hypothesis, for every $i\in I$ and $\bar{t}_i\in\bar{T}_i$, if $m_i\in S^kW_i(\bar{t}_i|\mathcal{M},\bar{\mathcal{T}})$, then $m_i^{k'}=\bar{t}_i$ for all $k'\leq k-2$. When $m_{-i}\notin\widehat{M}_{-i}$, there exists some $j\in I\backslash\{i\}$ such that $m_j^{k-1}\neq m_j^0$. We compute the expected loss in terms of payments for player i of type $\bar{t}_i$ when playing m_i rather than $\tilde{m}_i$:

$$\sum_{t_{-i},\nu_{-i}(t_{-i})\notin\widehat{M}_{-i}} \left\{\tau_i\left(\tilde{m}_i,\nu_{-i}(t_{-i})\right)-\tau_i\left(m_i,\nu_{-i}(t_{-i})\right)\right\}\pi_i(\bar{t}_i)[t_{-i}]$$

By choosing $\tilde{m}_i$ rather than m_i, player i will avoid the fine, η according to rule d_i^{k-1} (see (5) in Section 2.4.1) and ξ according to rule d_i (see (4)), that is,

$$\tau_i\left(\tilde{m}_i,\nu_{-i}(t_{-i})\right)-\tau_i\left(m_i,\nu_{-i}(t_{-i})\right)=\eta+\xi.$$

In terms of $g(\cdot)$ (see the outcome function in (3)), we have

$$\sum_{t_{-i},\nu_{-i}(t_{-i})\notin\widehat{M}_{-i}} \frac{1}{K}\left\{u_i(f(m_i^{k-1},\nu_{-i}^{k-1}(t_{-i})),\bar{\theta}_i)-u_i(f(\tilde{m}_i^{k-1},\nu_{-i}^{k-1}(t_{-i})),\bar{\theta}_i)\right\}\pi_i(\bar{t}_i)[t_{-i}]\leq\frac{1}{K}D. \tag{17}$$

This means that the possible gain from playing m_i rather than $\tilde{m}_i$ is bounded by D/K.

Since we have that $\xi>D/K$ (see (12) in Section 2.4.1), we have

$$\eta+\xi>\frac{1}{K}D. \tag{18}$$

This completes Step 1.

Step 2:

$$\sum_{t_{-i},\nu_{-i}(t_{-i})\in\widehat{M}_{-i}} \left\{ \begin{array}{c} \left\{u_i(g(\tilde{m}_i,\nu_{-i}^{k-1}(t_{-i})),\bar{\theta}_i)+\tau_i(\tilde{m}_i,\nu_{-i}^{k-1}(t_{-i}))\right\}- \\ \left\{u_i(g(m_i,\nu_{-i}^{k-1}(t_{-i})),\bar{\theta}_i)+\tau_i(m_i,\nu_{-i}^{k-1}(t_{-i}))\right\} \end{array} \right\} \pi_i(\bar{t}_i)[t_{-i}]>0$$

When $m_{-i}\in\widehat{M}_{-i}$, for any $j\in I\backslash\{i\}$, we have $m_j^{k-1}=m_j^0$. From the induction hypothesis, for every $i\in I$, $\bar{t}_i\in\bar{T}_i$, if $m_i\in S_i^k\left(t_i|\mathcal{M},\bar{\mathcal{T}}\right)$, then $m_i^{k'}=\bar{t}_i$, for all $k'\leq k-2$. We compute the expected loss in terms of payments for player i of type $\bar{t}_i$ when playing m_i rather than $\tilde{m}_i$:

$$\sum_{t_{-i},\nu_{-i}(t_{-i})\in\widehat{M}_{-i}} \left\{\tau_i\left(\tilde{m}_i,\nu_{-i}(t_{-i})\right)-\tau_i\left(m_i,\nu_{-i}(t_{-i})\right)\right\}\pi_i(\bar{t}_i)[t_{-i}]$$

For any t_{-i}, consider $\nu_{-i}(t_{-i}) = m_{-i}$.By choosing $\tilde{m}_i$ rather than m_i, player i will avoid the fine, η according to rule d_i^{k-1} (see (5) in Section 2.4.1), the expected loss in terms of payments from choosing m_i rather than $\tilde{m}_i$ in terms of $\tau(\cdot)$ (see (6) in Section 2.4.1) is

$$\begin{aligned} & \tau_i\left(\tilde{m}_i, m_{-i}\right) - \tau_i\left(m_i, m_{-i}\right) \\ = \; & \eta + \xi - d_i\left(m^0, ..., m^{k-1}, m^{k-2}/\tilde{m}_i ..., m^K\right) \\ \geq \; & \eta; \end{aligned}$$

Therefore, when playing m_i rather than $\tilde{m}_i$, the expected loss in terms of payments is bounded below by η.

In terms of $g(\cdot)$ (see the outcome function in (3)), the possible gain for player i to report m_i rather than $\tilde{m}_i$ is

$$\frac{1}{K} \sum_{t_{-i}, \nu_{-i}(t_{-i}) \in \widehat{M}_{-i}} \left\{ u_i(f(m_i^{k-1}, \nu_{-i}^{k-1}(t_{-i})), \bar{\theta}_i) - u_i(f(\tilde{m}_i^{k-1}, \nu_{-i}^{k-1}(t_{-i})), \bar{\theta}_i) \right\} \pi_i(\bar{t}_i)[t_{-i}],$$

Since $\tilde{m}_i$ differs from m_i only in the $(k-1)$th announcement.

That is, when playing m_i rather than $\tilde{m}_i$, the possible gain for player i of type $\bar{t}_i$ is which is bounded above by 0 from incentive compatibility of f. This completes Step 2. ■

The "only if" part of Theorem 1 is proved as follows.

Proof. Fix $\bar{\tau} > 0$ arbitrarily small. Let $\mathcal{M} = ((M_i), g, (\tau_i))_{i \in I}$ be a mechanism which implements f in $S^\infty W$ with transfers bounded $\bar{\tau}$. It is well known that a trembling-hand perfect equilibrium[11] is always contained in $S^\infty W$. Let σ be a trembling-hand perfect equilibrium in $U\left(\mathcal{M}, \bar{\tau}, \bar{\mathcal{T}}\right)$ such that $\sigma_i : T_i \to \Delta\left(M_i\right)$.

Since f is implementable in $S^\infty W$ by $(\mathcal{M}, \bar{\tau})$ and $m \in S^\infty W\left(t | \mathcal{M}, \bar{\mathcal{T}}\right)$ for every $m \in M$ with $\sigma\left(m|t\right) > 0$, it follows that for any $(t_i, t_{-i}) \in \bar{T}$,

$$f(t_i, t_{-i}) = g(\sigma_i\left(t_i\right), \sigma_{-i}\left(t_{-i}\right)). \tag{19}$$

We construct $(\bar{T}, f)$ as a direct revelation mechanism with the transfer rule

$$\tau_i(t_i, t_{-i}) = \tau_i\left(\sigma_i\left(t_i\right), \sigma_{-i}\left(t_{-i}\right)\right).$$

[11] Following Osborne and Rubinstein (1994), a strategy profile σ in a normal-form game is a trembling-hand perfect equilibrium if there exists a sequence $\left(\sigma^k\right)_{k=0}^\infty$ of completely mixed strategy profiles that converges to σ such that σ_i is a best response against σ_{-i}^k for every k. Here we consider the agent normal form of the incomplete information game $U(\mathcal{M}, \overline{\mathcal{T}})$ where each type t_i has the set of pure strategies M_i; moreover, for each pure strategy profile, t_i gets the expected payoff according to her belief. Note that a strategy is weakly dominated in mechanism $U(\mathcal{M}, \overline{\mathcal{T}})$ if and only if it is weakly dominated in its agent normal form. Also of course an NE in the agent form is an NE in $U(\mathcal{M}, \overline{\mathcal{T}})$.

Since σ is an equilibrium in $U\left(\mathcal{M},\bar{\tau},\bar{\mathcal{T}}\right)$, we have that for any $t_i \in \bar{T}_i$ and $m_i' \in M_i$,

$$\begin{aligned} & \sum_{t_{-i}} \pi_i(t_i)[t_{-i}] \left\{u_i\left(g(\sigma_i\left(t_i\right), \sigma_{-i}\left(t_{-i}\right))\right) + \tau_i\left(g(\sigma_i\left(t_i\right), \sigma_{-i}\left(t_{-i}\right))\right)\right\} \\ \geq & \sum_{t_{-i}} \pi_i(t_i)[t_{-i}] \left\{u_i\left(g(m_i', \sigma_{-i}\left(t_{-i}\right))\right) + \tau_i\left(m_i', \sigma_{-i}\left(t_{-i}\right)\right)\right\}. \end{aligned} \tag{20}$$

Then, by (19) and (20), the truth-telling is a Bayes Nash equilibrium in the incomplete information game induced by $(\bar{T}, f)$. That is, for any $t_i, t_i' \in \bar{T}_i$,

$$\begin{aligned} & \sum_{t_{-i}} \pi_i(t_i)[t_{-i}] \left\{u_i(f(t_i, t_{-i}), \hat{\theta}_i(t_i)) + \tau_i(t_i, t_{-i})\right\} \\ \geq & \sum_{t_{-i}} \pi_i(t_i)[t_{-i}] \left\{u_i(f(t_i', t_{-i}), \hat{\theta}_i(t_i)) + \tau_i(t_i', t_{-i})\right\}. \end{aligned} \tag{21}$$

Since f is implementable in $S^\infty W$ with arbitrarily small transfers, $\bar{\tau}$ can be arbitrarily small. Thus, we have

$$\sum_{t_{-i}} \pi_i(t_i)[t_{-i}] u_i(f(t_i, t_{-i}), \hat{\theta}_i(t_i)) \geq \sum_{t_{-i}} \pi_i(t_i)[t_{-i}] u_i(f(t_i', t_{-i}), \hat{\theta}_i(t_i)). \tag{22}$$

That is, f is incentive compatible. ■

If we impose no conditions on the size of transfers, *any* SCF is implementable with transfers. In this case, a very large size of transfers might be needed. We can get the following corollary by letting $K = 1$.

Corollary 1 *Suppose that $I \geq 2$ and the environment $\mathcal{E}$ satisfies Assumption 1. Any SCF is implementable in $S^\infty W$ with transfers.*

3 Applications

We now discuss the applications of our results in the previous section. First, in Section 3.1, we propose two classes of environments within which one can achieve implementation with no transfers, i.e., implementation with small transfers with the property that no transfers are required on the equilibrium. Next, in Section 3.2, we connect our results to *continuous* implementation, a concept proposed by Oury and Tercieux (2012). Specifically, we show that any incentive-compatible SCF is continuously implementable with arbitrarily small transfers. In Section 3.3, we discuss robust undominated Nash implementation, which Chung and Ely (2003) call $\overline{UNE}$-implementation. When there are at least three players, we then show that any SCF is $\overline{UNE}$-implementable with no transfers. In Section 3.4, with ex post small transfers, we obtain a full implementation result of the full surplus extraction in auction

environments.

3.1 Implementation with No Transfer

In Theorem 1, we use arbitrarily small transfers to achieve implementation of any incentive compatible SCF. In the mechanism, the ex post payment, although we can make it very small, is still necessary on the equilibrium. The concept of implementation with arbitrarily small transfers strikes us being rather innocuous. Still, it is sometimes impossible to assume that the planner can impose any transfers on the players in the equilibrium. Therefore, we propose the concept of implementation with no transfers.

Definition 3 (Implementation with No Transfers) *An SCF f is implementable in $S^\infty W$ with no transfers if for all $\bar{\tau} > 0$, it is implementable in $S^\infty W$ a mechanism $(\mathcal{M}, \bar{\tau})$ and moreover, for any $t \in \bar{T}$, and $m \in S^\infty W\left(t|\mathcal{M}, \bar{\mathcal{T}}\right)$, we have $\tau_i(m) = 0$ for each $i \in I$.*

Remark 4 *The concept of implementation with no transfers does not exclude a possibility that arbitrarily small transfers are made ex post out of the equilibrium. This concept of implementation is used by Abreu and Matsushima (1994) under complete information. We extend this to incomplete-information environments with private values.*

3.1.1 Non-Exclusive Information (NEI)

To discuss the result with no transfers, we need some extra assumptions. We first use *non-exclusive information structure* (NEI) for implementation with no transfers. To the best of our knowledge, NEI was first proposed by Postlewaite and Schmeidler (1986). We restate a version of its definition in Vohra (1999):

Definition 4 *The environment $\mathcal{E}$ satisfies the* ***non-exclusive information structure (NEI)*** *if, for each $t \in \bar{T}$, $i \in I$, and $t'_i \in \bar{T}_i$,*

$$\bar{\pi}_i(t'_i)[t_{-i}] = \begin{cases} 1 & \text{if } t'_i = t_i; \\ 0 & \text{otherwise.} \end{cases}$$

When $I = 2$, NEI is equivalent to complete information. NEI captures the idea that each agent is *informationally negligible* in the sense that any unilateral deception from the truth-telling in the direct revelation mechanism can be detected. Under NEI, we obtain the following result:

Theorem 2 *Suppose that $I \geq 2$ and the environment $\mathcal{E}$ satisfies NEI. Then, any incentive compatible SCF is implementable in $S^\infty W$ with no transfers.*

Proof. The mechanism is identical to the mechanism in Section 2.4.1 except that we replace $\lambda d_i^0\left(m_{-i}^{-2}, m_i^{-1}\right)$ and $\lambda d_i^0\left(m_{-i}^{-1}, m_i^0\right)$ with new transfer rules as follows:

$$\hat{d}_i^0(m_{-i}^{-2}, m_i^{-1}) = \begin{cases} \gamma & \text{if } \pi_i(m_i^{-1})[m_{-i}^{-2}] = 0; \\ 0 & \text{otherwise.} \end{cases}$$

$$\hat{d}_i^0(m_{-i}^{-1}, m_i^0) = \begin{cases} \gamma & \text{if } \pi_i(m_i^0)[m_{-i}^{-1}] = 0; \\ 0 & \text{otherwise.} \end{cases}$$

The proof then follows verbatim the proof of Theorem 1. ■

3.1.2 Strict Incentive Compatibility and Separability

Following Sjöström (1994), we introduce the following class of environments. We assume the outcome space has the product structure: $A = A_1 \times A_2 \times \cdots \times A_I$, and each player i's utility is defined as $u_i : A_i \times \Theta_i \to \mathbb{R}$. For each SCF f and type $t \in \bar{T}$, we denote $f(t) = (f_1(t), \ldots, f_I(t))$ where $f_i(t)$ denotes the marginal distribution of $f(t)$ on A_i where $A = A_1 \times A_2 \times \cdots \times A_I$. The reader is referred to Sjöström (1994) to see when this separable environment is valid. For example, we can consider an exchange economy where each player i has a consumption set A_i and cares only about his own consumption. We first introduce a stronger version of incentive compatibility.

Definition 5 *An SCF $f : \bar{T} \to \Delta(A)$ is **strictly incentive compatible** if, for all $i \in I$ and all $t_i, t_i' \in \bar{T}_i$ with $t_i \neq t_i'$,*

$$\sum_{t_{-i} \in \bar{T}_{-i}} u_i(f(t_i, t_{-i}), \hat{\theta}_i(t_i))\bar{\pi}_i(t_i)[t_{-i}] > \sum_{t_{-i} \in \bar{T}_{-i}} u_i(f(t_i', t_{-i}), \hat{\theta}_i(t_i))\bar{\pi}_i(t_i)[t_{-i}].$$

In the theorem below, we can drop Assumption 1 but instead, we need to strengthen incentive compatibility into strict incentive compatibility.

Theorem 3 *Let $\mathcal{E}$ be a separable environment with $I \geq 2$. Any strictly incentive compatible SCF is implementable in $S^\infty W$ with no transfers.*

The corresponding mechanism is provided as follows. Basically, in a separable environment, the strictly incentive compatible SCF replaces the role of scoring rule (d_i^0) in the previous discussion. Hence, we can handle any information structure. In particular, players' types can be independently distributed.

1. **The message space**:

   Each player i makes 4 simultaneous announcements of his own type. We index each announcement by $-2, -1, 0, 1$. That is, player i's message space is given as

   $$M_i = M_i^{-2} \times M_i^{-1} \times M_i^0 \times M_i^1 = \bar{T}_i \times \bar{T}_i \times \bar{T}_i \times \bar{T}_i.$$

Denote

$$m_i = \left(m_i^{-2}, m_i^{-1}, m_i^0, m_i^1\right) \in M_i,\ m_i^k \in M_i^k,\ k \in \{-2,-1,0,1\},$$

and

$$m = \left(m^{-2}, m^{-1}, m^0, m^1\right) \in M,\ m^k = \left(m_i^k\right)_{i\in I} \in M^k = \times_{i\in I} M_i^k.$$

We use $m^k/\tilde{m}_i$ to denote the strategy profile $\left(m_1^k, ..., m_{i-1}^k, \tilde{m}_i^k, m_{i+1}^k, ..., m_I^k\right)$.

2. **The outcome function**:

Let ϵ be a small positive number.

Define $e : M^{-1} \times M^0 \to \mathbb{R}$ by

$$e(m^{-1}, m^0) = \begin{cases} \epsilon & \text{if } m_i^{-1} \neq m_i^0 \text{ for some } i \in I, \\ 0 & \text{otherwise.} \end{cases}$$

The outcome function $g : M \to \Delta(A)$ is defined as follows: for each $m \in M$,

$$\begin{aligned} g(m) &= e\left(m^{-1}, m^0\right) \frac{1}{I} \sum_{i\in I} x_i\left(m_i^{-2}\right) \\ &+ \left\{1 - e\left(m^{-1}, m^0\right)\right\} \left\{\tilde{\lambda}_1 \tilde{f}(m^{-1}, m^{-2}) + \tilde{\lambda}_2 \tilde{f}(m^0, m^{-1}) + (1 - \tilde{\lambda}_1 - \tilde{\lambda}_2) f(m^1)\right\}, \end{aligned}$$

where $\tilde{f}(m^k, m^{k-1}) \equiv \times_{i\in I} f_i\left(m_i^k, m_{-i}^{k-1}\right)$ and $f_i(m_i^k, m_{-i}^{k-1})$ denotes the marginal distribution of $f(m_i^k, m_{-i}^{k-1})$ on A_i for $k \in \{-1, 0\}$.

3. **The transfer rule**:

Let η be positive numbers. Player i is to pay η if his 1st round announcement differs from his own 0th round announcement.

$$\tau_i\left(m_i^0, m_i^1\right) = \begin{cases} \eta & \text{if } m_i^1 \neq m_i^0; \\ 0 & \text{otherwise.} \end{cases} \tag{23}$$

The definitions of E and D are the same as in the previous section.

We choose positive numbers $\tilde{\lambda}_1$, $\tilde{\lambda}_2$, ϵ, and η such that for every t_i,$t_i' \in \bar{T}_i$ and every $i \in I$,

$$\bar{\tau}_i > \eta; \tag{24}$$

$$\min\left\{\tilde{\lambda}_1, \tilde{\lambda}_2\right\} \sum_{t_{-i}\in\bar{T}_{-i}} \left[u_i(f_i(t_i, t_{-i}), \hat{\theta}_i(t_i)) - u_i(f_i(t_i', t_{-i}), \hat{\theta}_i(t_i))\right] \bar{\pi}_i(t_i)\left[t_{-i}\right] > \gamma; \tag{25}$$

$$\eta > \epsilon E + (1 - \tilde{\lambda}_1 - \tilde{\lambda}_2) D; \tag{26}$$

and

$$\gamma > \epsilon E + (1 - \tilde{\lambda}_1 - \tilde{\lambda}_2)D + \eta. \tag{27}$$

Since f is strictly incentive compatible, the existence of γ is guaranteed in (25).

Remark 5 *In a separable environment, a proper adjustment of the weight between the* $0th$ *round report and the* $1st$ *round report can decrease the payment in a way that differs from that used in Abreu and Matsushima (1994). Specifically, given* $\bar{\tau}$*, the mechanism in Section 2.4.1 increases the number of rounds of reports to some large enough* K *to make sure each round report's effect on allocation is small enough (happens with probability* $1/K$*); instead, we can choose* $(1 - \tilde{\lambda}_1 - \tilde{\lambda}_2)$ *small enough to make the weight of the* $1st$ *round announcement small enough. Therefore,* η *can be chosen small enough to prevent the deviation in the* $1st$ *round.*

Remark 6 *We omit the proof of Theorem 3 and rather provide a heuristic argument of how the proof works. The first round deletion of weakly dominated strategies is the same as the procedure in the proof of Claim 1. Second, to elicit the true type profile in the* $-1th$ *and* $0th$ *rounds, the constructed SCF* $\tilde{f}$ *works in a similar way as the scoring rule* (d_i^0) *did in the proofs of Claims 2 and 3. Specifically, the function* $\tilde{f}$ *is constructed such that each player* i*'s payoff from* $\tilde{f}$ *is affected only by his own* $-1th$ *(resp.* $0th$*) round report and the other players'* $-2th$ *(resp.* $-1th$*) round report. By the strict incentive compatibility, each player will announce truthfully in the* $-1th$ *(resp.* $0th$*) round(given the truth telling in the* $-2th$ *(resp.* $-1th$*) reports for everyone). When all players tell the truth in every round, the constructed function* $\tilde{f}$ *coincides with the SCF* f*. This enables the mechanism to implement* f *without any ex post transfers. Finally, the last round of elimination of strictly dominated strategies works in a way that is parallel to the proof of Claim 4.*

3.2 Continuous Implementation

The mechanism design literature often deals with environments in which monetary payments are available, and the analyses are limited to partial implementation. Partial implementation is a notion that requires the planner to design a game in which only *some* equilibrium–but not necessarily *all equilibria*–yields the desired outcome. Then, appealing to the revelation principle, its analysis reduces to the characterization of incentive-compatible direct revelation mechanisms. This means that the mechanism design literature discounts the possibility that undesirable equilibria exist in the game. *Full*–as opposed to partial–implementation is a notion that requires that *all* equilibria deliver the desired outcome. Although it is unfortunate that the literature has thus far largely ignored the need to compare partial and full implementation, Oury and Tercieux (2012) have recently built a bridge between these two notions. They consider the following situation: The planner wants not only that the SCF be partially implementable, but also that it continue to be partially implementable in all the models *close* to his initial model. That is, the SCF is *continuously* (partial)

implemented. Oury and Tercieux (2012) show that Bayesian monotonicity (See definition on p. 1617 in Oury and Tercieux (2012)), which is a necessary condition for full implementation, becomes necessary even for continuous implementation; in light of this result, they argue that continuous implementation is tightly connected to full implementation.

We shall show that as long as the planner is willing to allow for small ex post transfers, any incentive-compatible SCF is continuously implementable in private-value environments. This stands in sharp contrast with Oury and Tercieux (2012) because our continuous implementation result does not need Bayesian monotonicity but only incentive compatibility, which is a necessary condition for partial implementation. Our result is consistent with Matsushima (1993), which shows that in Bayesian environments with side payments under strict incentive compatibility, Bayesian monotonicity holds generically. Therefore any incentive compatible SCF is fully implementable. Note that if one is willing to settle for allowing small ex post transfers, one can always transform any incentive-compatible SCF into a strict incentive-compatible one. However, the mechanism which can fully implement any incentive-compatible SCF employs either large transfers (See Matsushima (1991)) or infinite strategy spaces (See Jackson (1991)). We show that with arbitrarily small transfers, any incentive-compatible SCF is fully implementable by a finite mechanism, not only in the benchmark model but also in the nearby environment.

Given a mechanism $(\mathcal{M}, \bar{\tau})$ and a type space $\mathcal{T}$, we write $U(\mathcal{M}, \bar{\tau}, \mathcal{T})$ for the induced incomplete information game. In the game $U(\mathcal{M}, \bar{\tau}, \mathcal{T})$, a behavior strategy of a player i is $\sigma_i : T_i \to \Delta(M_i)$. We follow Oury and Tercieux (2012) to write down the following definitions. We define

$$V_i((m_i, \sigma_{-i}), t_i) = \sum_{t_{-i}} \pi_i(t_i)[t_{-i}] \left\{ u_i(g(m_i, \sigma_{-i}(t_{-i})), \theta_i(t_i)) + \tau_i(m_i, \sigma_{-i}(t_{-i})) \right\}.$$

Definition 6 *A profile of strategies $\sigma = (\sigma_1, ..., \sigma_I)$ is a **Bayes Nash equilibrium** in $U(\mathcal{M}, \bar{\tau}, \mathcal{T})$ if, for each $i \in I$ and each $t_i \in T_i$,*

$$m_i \in supp\left(\sigma_i\left(t_i\right)\right) \Rightarrow m_i \in argmax_{m_i' \in M_i} V_i\left(\left(m_i', \sigma_{-i}\right), t_i\right).$$

We write $\sigma_{|\bar{T}}$ for the strategy profile σ restricted to $\bar{T}$. For any $\mathcal{T} = (T_i, \hat{\theta}_i, \pi_i)_{i \in I}$, we will write $\mathcal{T} \supset \bar{\mathcal{T}}$ if $T \supset \bar{T}$ and for every $t_i \in \bar{T}_i$, we have $\pi_i\left(t_i\right)[E] = \bar{\pi}_i\left(t_i\right)\left[\bar{T}_{-i} \cap E\right]$ for any measurable $E \subset T_{-i}$.

Definition 7 *Fix a mechanism $(\mathcal{M}, \bar{\tau})$ and a model $\mathcal{T}$ such that $\bar{\mathcal{T}} \subset \mathcal{T}$. We say that a Bayes Nash equilibrium σ in $U(\mathcal{M}, \bar{\tau}, \mathcal{T})$ **(strictly) continuously implements** $f : \bar{T} \to \Delta(A)$ if the following two conditions hold: (i) $\sigma_{|\bar{T}}$ is a (strict) Bayes Nash equilibrium in $U\left(\mathcal{M}, \bar{\tau}, \bar{\mathcal{T}}\right)$; (ii) for any $\bar{t} \in \bar{T}$ and any sequence $t[n] \to_p \bar{t}$, whenever $t[n] \in T$ for each n, we have $(g \circ \sigma)(t[n]) \to f(\bar{t})$.*

We introduce two variants of continuous implementation:

Definition 8 *An SCF $f:\bar{T}\to\Delta(A)$ is continuously implementable with* ***transfers*** *if there exists a mechanism $(\mathcal{M},\bar{\tau})$ such that for each model $\mathcal{T}$ with $\bar{\mathcal{T}}\subset\mathcal{T}$, there is a Bayes Nash equilibrium σ in $U(\mathcal{M},\bar{\tau},\mathcal{T})$ that continuously implements f.*

Definition 9 *An SCF $f:\bar{T}\to\Delta(A)$ is continuously implementable with* ***arbitrarily small transfers*** *if for any $\bar{\tau}>0$, there exists a mechanism $(\mathcal{M},\bar{\tau})$ such that for each model $\mathcal{T}$ with $\bar{\mathcal{T}}\subset\mathcal{T}$, there is a Bayes Nash equilibrium σ in $U(\mathcal{M},\bar{\tau},\mathcal{T})$ that continuously implements f.*

Theorem 4 *Suppose that $I\geq 2$ and the environment $\mathcal{E}$ satisfies Assumption 1. Then, an SCF f is continuously implementable with arbitrarily small transfers if and only if it is incentive compatible.*

To prove Theorem 4, we establish the following important lemma.

Lemma 3 *Fix any model $\mathcal{T}$ such that $\bar{\mathcal{T}}\subset\mathcal{T}$ and a finite mechanism $\mathcal{M}$. Suppose that for each $t_i,t_i'\in T_i$, $W_i(t_i|\mathcal{M},\mathcal{T})=W_i(t_i'|\mathcal{M},\mathcal{T})$ whenever $\hat{\theta}_i(t_i)=\hat{\theta}_i(t_i')$. Then, for any $\bar{t}\in\bar{T}$ and any sequence $\{t[n]\}_{n=0}^{\infty}$ such that $t[n]\to_p\bar{t}$, we have $S^\infty W(t[n]|\mathcal{M},\mathcal{T})\subset S^\infty W(\bar{t}|\mathcal{M},\mathcal{T})$ for any large enough n.*

Proof. See Appendix A.2. ■

Now we prove Theorem 4.

Proof. We first prove "if" part. For any $\bar{\tau}>0$, we employ the mechanism $(\mathcal{M},\bar{\tau})$ constructed in Section 2.4.1. We employ the mechanism $(\mathcal{M},\bar{\tau})$ constructed in Section 2.4.1. Therefore, for all $\bar{t}\in\bar{T}$, $m\in S^\infty W\left(\bar{t}|\mathcal{M},\bar{\mathcal{T}}\right)\Rightarrow g(m)=f(\bar{t})$. Note that $S^\infty W\left(\bar{t}|\mathcal{M},\bar{\mathcal{T}}\right)=\{(\bar{t},...,\bar{t})\}$. We write σ^* such that $\sigma_i^*(\bar{t}_i)=(\bar{t}_i,...,\bar{t}_i)$ for all $\bar{t}_i\in\bar{T}_i$. Now pick any $\mathcal{T}$ such that $\bar{\mathcal{T}}\subset\mathcal{T}$. It is well known that a trembling hand perfect equilibrium is always contained in $S^\infty W$. Therefore, σ^* is a trembling hand perfect equilibrium in $U\left(\mathcal{M},\bar{\tau},\bar{\mathcal{T}}\right)$. We show that there exists an equilibrium that continuously implements f on $\bar{\mathcal{T}}$. For each player i and each type $\bar{t}_i\in\bar{T}_i$, restrict the space of strategies of player i by assuming that $\sigma_i(\bar{t}_i)=\sigma_i^*(\bar{t}_i)$ for each $\bar{t}_i\in\bar{T}_i$. Because M is finite and T is countable, standard arguments[12] show that there exists a trembling hand perfect equilibrium in $U(\mathcal{M},\bar{\tau},\mathcal{T})$, which is denoted by σ. Thus, σ is a trembling hand perfect equilibrium in $U(\mathcal{M},\bar{\tau},\mathcal{T})$ and $\sigma_{|\bar{T}}$ is a trembling hand perfect equilibrium in $U\left(\mathcal{M},\bar{\tau},\bar{\mathcal{T}}\right)$. Now, pick any sequence $\{t[n]\}_{n=0}^{\infty}$ such that $t[n]\to_p\bar{t}$. It is clear that, for each $n:\text{supp}(\sigma(t[n]))\subset S^\infty W(t[n]|\mathcal{M},\mathcal{T})$. Since $W_i(t_i|\mathcal{M},\mathcal{T})=W_i(t_i'|\mathcal{M},\mathcal{T})$ whenever $\hat{\theta}_i(t_i)=\hat{\theta}_i(t_i')$, for n large enough, we know by Lemma 3 that $S^\infty W(t[n]|\mathcal{M},\mathcal{T})\subset S^\infty W\left(\bar{t}|\mathcal{M},\bar{\mathcal{T}}\right)$. Thus, $(g\circ\sigma)(t[n])=f(\bar{t})$ as claimed.

[12]The existence of a trembling hand perfect equilibrium can be proved using Kakutani–Fan–Glicksberg's fixed point theorem. The space of strategy profiles is compact in the product topology. Using the fact that $u_i:A\times\Theta_i\to R$ is bounded, all the desired properties of the best-response correspondence (in particular upper hemicontinuity) can be established.

The "only if" part is proved as follows: Given f is continuously implementable with arbitrarily small transfers. Then, for any $\overline{\tau} > 0$, there is a Bayes Nash equilibrium σ in $U(\mathcal{M}, \bar{\mathcal{T}})$ such that $(g \circ \sigma)(\bar{t}) = f(\bar{t})$ for any $\bar{t} \in \bar{T}$ and $\tau(\sigma(\bar{t})) < \bar{\tau}$. By a similar argument in the proof of the "only if" part of Theorem 1, we conclude that f is incentive compatible. ■

If we do not impose any conditions on the size of ex post transfers, we obtain the following very permissive result.

Corollary 2 *Suppose that $I \geq 2$ and the environment $\mathcal{E}$ satisfies Assumption 1. Then,* ***any*** *SCF f is continuously implementable with transfers.*

The next result is one of the main results of Oury and Tercieux (2012).

Proposition 1 (Theorem 2 of Oury and Tercieux (2012)) *If an SCF f is* ***strictly*** *continuously implementable, it satisfies strict Bayesian monotonicity.*

Oury and Tercieux show that the condition for full implementation (i.e., Bayesian monotonicity) is necessary for "strict" continuous partial implementation. However, as long as we allow for ex post small transfers and consider private-value environments, we obtain yet another result that permits continuous implementation and our result is as permissive as it can be. The reader is referred to the discussion at the end of Section 3.3 where our Theorem 4 can accommodate slightly more general perturbations. We consider this as a significant finding because our Theorem 4 is one of few positive continuous implementation results in the literature and to the best of our knowledge, de Clippel, Saran, and Serrano (2014) is the only exception, which shows that any strict incentive compatible SCF is continuously implementable when the players are constrained by their reasoning ability in the spirit of level-k model and level-0 players are assumed to tell the truth in the direct mechanism.

To drop this "strictness," Oury and Tercieux assume instead that sending messages in the mechanism is slightly costly. To dispense with the assumption of costly messages, Oury (2015) rather proposes a stronger concept of continuous implementation that accommodates the local robustness with respect to payoff uncertainty. Recall that we assume that no players use weakly dominated actions. In fact, this weak dominance will be highly sensitive to payoff perturbations that can be induced by the cost of sending messages or local robustness of Oury (2015).

3.3 $\overline{UNE}$ Implementation

Chung and Ely (2003) contemplate the following situation: if a planner wants all equilibria of his mechanism yield a desired outcome under complete information, and if he entertains the possibility that players may have even the slightest uncertainty about payoffs, then the planner should insist on a solution concept that has a closed graph in the limit of complete information. Chung and Ely then adopt undominated Nash equilibrium (UNE) as a solution

concept and call the corresponding implementation concept "$\overline{UNE}$ implementation". In particular, Theorem 1 of Chung and Ely (2003) shows that Maskin monotonicity is a necessary condition for $\overline{UNE}$ implementation. For this proof, one needs to construct a near-complete information structure in which some players have superior information about the state, and consequently, about the preferences of other players. In their Section 6.2, Chung and Ely restrict their attention to private-value perturbations in which each type may be uncertain about the preferences of other players but always knows his own preferences.[13] Under such perturbations, they show that dominated strategies under complete information continue to be dominated.

In their footnote 7 Chung and Ely (2003) observe that the continuity of dominated strategies under private-value perturbations does not necessarily guarantee that UNE implementation suffices for $\overline{UNE}$-implementation. This leaves open the question as to whether $\overline{UNE}$-implementation can be achieved under private-value perturbation. Here we provide an affirmative answer. That is, our robustness argument can be adapted to prove that the mechanism provided in Abreu and Matsushima (1994) actually achieves $\overline{UNE}$ implementation. Following Chung and Ely (2003), we now rephrase their definition of $\overline{UNE}$-implementation.[14]

Definition 10 ($\overline{UNE}$ Implementation) *Fix a complete-information model $\bar{\mathcal{T}}$. We say that an SCF f is $\overline{UNE}$-implementable with no transfer if for any $\bar{\tau} > 0$, there exists a mechanism $(\mathcal{M}, \bar{\tau})$ such that the following three conditions hold: (i) there exists a strategy profile σ such that $\sigma_{|\bar{T}}$ is an undominated Nash equilibrium in $U\left(\mathcal{M}, \bar{\tau}, \bar{\mathcal{T}}\right)$; (ii) for any $\bar{t} \in \bar{T}$, any sequence $t[n] \to_p \bar{t}$, any model $\mathcal{T}$ with $\bar{\mathcal{T}} \subset \mathcal{T}$, and any sequence of undominated Bayes Nash equilibria $\{\sigma^n\}_{n=0}^{\infty}$ of the game $U(\mathcal{M}, \bar{\tau}, \mathcal{T})$, whenever $t[n] \in T$ for each n, we have $g(\sigma^n(t[n])) \to f(\bar{t})$; (iii) $\tau_i(\sigma(t)) = 0$ for any t, and any player i.*

Note that any complete-information model is a special case of an incomplete-information model. We therefore obtain the following permissive result:

Corollary 3 *Suppose that $I \geq 2$ and the model $\bar{\mathcal{T}}$ is a complete-information model. Then, any incentive-compatible SCF f is $\overline{UNE}$-implementable with no transfers.*

Proof. Note that complete-information environments trivially satisfy NEI (non-exclusive information) assumption. So, we modify the scoring rule d_i^0 as we did for Theorem 1. This allows us to dispense with Assumption 1. The rest of the proof is completed by Theorem 4. ■

Remark 7 *Assume that there are at least three players. In this case, under complete information, the planner can always detect any unilateral deviation from a truthful announcement.*

[13] The perturbation in Chung and Ely (2003) can be considered a special case of the perturbation defined in a universal type space that we formulate here.

[14] Our notion of $\overline{UNE}$-implementation is stronger than the one defined in Chung and Ely (2003) as we allow for a boarder class of perturbations. As a consequence, our positive result is stronger.

Therefore, we simply construct a new SCF that is the same as the original SCF, except that we simply ignore any such unilateral deviation and assign the same lottery as if there were no deviations. This new SCF is equivalent to the original SCF under the hypothesis of complete information so that we can make any SCF be incentive-compatible. So, when $I \geq 3$, we can drop incentive compatibility completely from Corollary 3. In fact, this is the main result of Abreu and Matsushima (1994). The novel contribution here is to observe that the result of Abreu and Matsushima (1994) can be adapted to establish $\overline{UNE}$-implementation.

Our result is consistent with Chung and Ely (2003). Theorem 1 of Chung and Ely (2003) shows that Maskin monotonicity is a necessary condition for $\overline{UNE}$-implementation. Specifically, for the proof of this theorem, one needs to exploit the interdependent values. It is also easy to show that Maskin monotonicity is still necessary for $\overline{UNE}$-implementation if the players are not very sure about their own payoff type, which is not the case of private values. In Section 4, we extend our implementation results to general interdependent-value environments. However, we no longer know that this extension exhibits the same robustness property as in Corollary 3.

In the following lines, we introduce a slightly more general class of perturbations than that we thus far considered. Specifically, each player holds a small uncertainty about his own payoff type, that is, each player almost knows his own payoff type. In addition, whether or not some player knows other players' type will not change his conjecture over his own payoff type. We write $\kappa(t_i)[\theta_i] = \text{marg}_{\Theta_i}\kappa(t_i)[\theta_i]$ for the belief on his own payoff type for player i with t_i and $\kappa(t_i)[\theta_i|t_{-i}] = \left(\text{marg}_{\Theta_i \times T_{-i}}\kappa(t_i)\right)[\theta_i|t_{-i}]$ for the belief conditional on some t_{-i}. Formally, it is captured by the following definition.

Definition 11 (convergence in private values) *Fix a model $\mathcal{T}$. We say a sequence of types $\{t_i[k]\}_{k=0}^{\infty}$* ***converges to a type*** *$t_i \in T_i$* ***in private values*** *if $t_i[k] \in T_i$ for each k and for any $t_{-i} \in T_{-i}$, whenever $\kappa(t_i[k])[t_{-i}] > 0$,*

$$\kappa(t_i[k])[\theta_i|t_{-i}] \to \kappa(t_i)[\theta_i] \quad \text{as } k \to \infty.$$

We write $t_i[k] \to_{pp} t_i$ for the class of convergent sequences which converge both in product topology and in private values.

Our robustness results in Sections 3.2 and 3.3 hold even when the nearby perturbation admits convergence in private values. No essential changes are needed in the proof.

3.4 Full Surplus Extraction

In a seminal paper, Crémer and McLean (1988) show that in a single object auction with generic correlated types, it is possible to design a mechanism (which we call a CM mechanism) in such a way that (i) each bidder earns an expected surplus of zero in a Bayes Nash equilibrium and (ii) the object is allocated to the agent with the highest valuation. This

outcome is referred to as the *full surplus extraction (henceforth, FSE)* outcome. Although this is a surprisingly positive result, an FSE outcome is rarely observed in reality. Many explanations have been proposed to resolve this discrepancy between theory and reality, including risk neutrality, unlimited liability, the absence of collusion among agents, a lack of competition among sellers, and the restrictiveness of a fixed finite type space. Although these are important issues, we rather follow Brusco (1998) who points out another weakness of the FSE result. In particular, Brusco provides an example in which every mechanism that has the FSE property as a Bayes Nash equilibrium must have another Bayes Nash equilibrium which is weakly Pareto superior for the agents. This implies that the multiplicity of equilibria might be a reason why the FSE outcome is not observed in reality, despite the fact that the FSE outcome is an equilibrium in dominant strategies. Brusco shows that one can devise a two-stage sequential mechanism that implements the FSE outcome in all perfect Bayesian equilibria. Chen and Xiong (2013) show that the FSE outcome is virtually Bayesian fully implemented.

We can establish a similar result, by adopting a static mechanism to achieve full implementation, as long as players do not use weakly dominated strategies. First, we include the range of payment schemes of the CM mechanism as part of A (the set of pure outcomes). Second, following Crémer and McLean (1988), we observe that the social choice function that achieves the FSE outcome is Bayesian incentive compatible, i.e., incentive compatible.[15] So, by Theorem 1, we obtain the following:

Corollary 4 *Suppose that $I \geq 2$ and the environment $\mathcal{E}$ satisfies Assumption 1. The FSE outcome is implementable in $S^{\infty}W$ with arbitrarily small transfers.*

Therefore, we still obtain the FSE property even when we insist on full implementation with small transfers. Note that we achieve full implementation in a finite mechanism, whereas the mechanisms in Brusco (1998) and Chen and Xiong (2013) are infinite and involve either integer games or an "open set trick." One crucial assumption that we adopt for this result is that no players use weakly dominated actions.[16]

4 Interdependent-Value Environments

We now deal with the case of interdependent-value environments where each player i's utility function is defined as $u_i : A \times \Theta \to \mathbb{R}$ with $\Theta = \Theta_1 \times \cdots \times \Theta_I$. In Section 4.1, we construct

[15] Crémer and McLean (1988) show two main results: their Theorem 1 achieves FSE in dominant-strategy incentive-compatibility when agents' beliefs satisfy a full-rank condition, whereas their Theorem 2 achieves FSE in Bayesian incentive-compatibility when agents' beliefs satisfy a weaker spanning condition. Corollary 4 therefore strengthens only their Theorem 2, while the results in Brusco (1998) and Chen and Xiong (2013) apply to their Theorem 1 as well.

[16] The FSE outcome hinges on the assumption that players' beliefs satisfy BDP property. A prior satisfies the BDP property if it assigns probability 1 to a set of type profiles in which no distinct types have the same belief. (see discussions of the genericity of BDP in Heifetz and Neeman (2006), Chen and Xiong (2011) and Chen and Xiong (2013))

two examples to demonstrate an essential difference between virtual implementation and implementation with transfers in the incomplete information setups. Despite the difference and the interim perspective that we take throughout, we show in Section 4.2 that we can employ the *maximally revealing mechanism* due to Bergemann and Morris (2009b) to extend our result to interdependent-value environments.

4.1 The Difficulty of Extending Abreu and Matsushima (1994) to Incomplete Information Environments

In this section, we will first recall the notion of *measurability* (henceforth, AM-measurability) due to Abreu and Matsushima (1992b), who extend their result of Abreu and Matsushima (1992a) to incomplete information environments. Specifically, they show that any incentive compatible and AM-measurable SCF is virtually implementable in S^{∞}. As conjectured in Abreu and Matsushima (1994), one may speculate that by suitably adapting the argument of Abreu and Matsushima (1992b) to the result in Abreu and Matsushima (1994), any incentive compatible and AM-measurable SCF is implementable with arbitrarily small transfers. We show by two examples that the conjecture is not unconditionally warranted. In Example 1, we show that our mechanism does not implement an incentive compatible and AM-measurable social choice function in a private-value environment which violates Assumption 1. In Example 2, we show that even when Assumption 1 holds, a natural extension of our previous mechanism to interdependent-value environments still does not implement an incentive compatible and AM-measurable social choice function.

We first define AM-measurability. Let Π_{-i} be a partition of $\bar{T}_{-i}$. Say that t_i is equivalent to t_i' with respect to Π_{-i} if player i's interim expected payoff under type t_i is exactly the same as under type t_i' when evaluating any allocation function $y : \bar{T} \to \Delta(A) \times \mathbb{R}^I$ which is measurable with respect to $\bar{T}_i \times \Pi_{-i}$. Let $\rho_i(t_i, \Pi_{-i})$ be the set of all elements of $\bar{T}_i$ that are equivalent to t_i with respect to Π_{-i}, and let

$$R_i(\Pi_{-i}) = \left\{ \rho_i(t_i, \Pi_{-i}) \subset \bar{T}_i | t_i \in \bar{T}_i \right\}.$$

We define an infinite sequence of I-tuples of partitions, $\{\Pi^h\}$ in the following way: $\Pi_i^0 = \{\bar{T}_i\}$, and recursively, for every i and $h \geq 1$, $\Pi_i^h = R_i\left(\Pi_{-i}^{h-1}\right)$. Note that for every $h \geq 0$, Π_i^{h+1} is the same as, or finer than, Π_i^h. Since $\overline{T}_i$ is finite, there is some $\overline{h}$ and partition Π_i^* such that $\Pi_i^h = \Pi_i^*$ for every $h \geq \overline{h}$.

Definition 12 *An SCF f satisfies* ***AM-measurability*** *if it is measurable with respect to* Π^*.

In Example 1 below, we will construct an environment in which (1) players' values are private; and (2) Assumption 1 is violated. We show that there is an incentive compatible and AM-measurable social choice function that cannot be implemented by our mechanism.

Example 1 *$A = \{a_1, a_2\}$; $I = \{1, 2, 3\}$; $\bar{T}_i = \{t_i^1, t_i^2\}$ for all $i \in I$. Define $a_1 \equiv (1, 0); a_2 \equiv (0, 1)$; $t_i^1 \equiv (1, 0)$; and $t_i^2 \equiv (0, 1)$. Let $3 + 1 \equiv 1$. Let $\pi_i : \bar{T}_i \to \Delta\left(\bar{T}_{-i}\right)$ be player i's interim belief map from $\bar{T}_i \to \Delta(\bar{T}_{-i})$:*

$$\pi_i(t_i^1)[t_{-i}] = \pi_i(t_i^2)[t_{-i}] \ = \begin{cases} 1/2 & \text{if } t_{i+1} = t_{i+2} = a_1 \\ 1/2 & \text{if } t_{i+1} = t_{i+2} = a_2 \\ 0 & \text{otherwise.} \end{cases}$$

That is, in player i's view, player $(i+1)$'s type and player $(i+2)$'s type are perfectly correlated but they are independent of player i's type. Each player i has the following preferences: for any $a \in A$ and $t \in \bar{T}$,

$$u_i(a, t) = a \cdot t_i,$$

where $a \cdot t_i$ denotes the dot (or, inner) product of the two vectors a and t_i.

Consider the following incentive-compatible social choice function $f^* : \bar{T} \to A$: for any $t \in \bar{T}$, $f^*(t) = a$ if and only if $\#\{i \in I : t_i = a\} \geq 2$. We can interpret this f^* as the majority rule.

We construct a set of allocation rules $\left\{x_i : \bar{T}_i \to A\right\}_{i \in I}$ such that $x_i\left(t_i\right) = t_i$ for each $i \in I$ and $t_i \in \bar{T}_i$. It is easy to see that for all $t_i, t_i' \in \bar{T}_i$ with $t_i \neq t_i'$,

$$u_i(x_i(t_i), t_i) > u_i(x_i(t_i'), t_i). \tag{28}$$

Thus, for each $i \in I$, we obtain $\Pi_i^* = \{\{t_i^1\}, \{t_i^2\}\}$, the finest possible partition over $\bar{T}_i$. Therefore, every social choice function is AM-measurable in the environment. Therefore, by Abreu and Matsushima (1992b), f^* is virtually implementable in S^∞. However, one cannot obtain exact implementation only by replacing ϵ in the mechanism in Abreu and Matsushima (1992b) with the indicator function $e(\cdot)$ which takes either ϵ or 0.[17] That is why we turn to the mechanism we constructed in Section 2.4.1.

Consider player i of type t_i and a strategy $\sigma_i : \bar{T}_i \to M_i$ such that $\sigma_i\left(t_i\right) = (t_i, t_i', \ldots, t_i')$ where $t_i' \neq t_i$. That is, player i tells the truth in the first round but consistently misrepresent his type in the rest of rounds. By inequality (28), every player will tells the truth in the first round report. However, whatever his type is, each player holds the same belief over others' payoff types. Therefore, under the scoring rule, any report in the second round can be rationalized. Due to the property of the majority rule f^*, $\sigma_i(t_i)$ survives $S^\infty W$ but induces an outcome which is "not" consistent with the one specified by the SCF f^*. Thus, to the extent that our mechanism is a natural extension of that of Abreu and Matsushima (1994), the conjecture in Abreu and Matsushima (1994) fails in private-value environments. This

[17] To see this, consider a complete information setup. Suppose that for some fixed t_{-i} and two types t_i and t_i', t_i regards $f\left(t_i', t_{-i}\right)$ as the best outcome among all. Then, for type t_i, reporting t_i' all the way is strictly better than reporting t_i in the first round and t_i' in all subsequent rounds. This is because (1) the former avoids the penalty for being inconsistent; (2) the former entails no loss in changing the epsilon portion of the allocation no matter how we choose the dictator lotteries for the first round.

justifies our use of Assumption 1.

In Example 2 below, we will construct an interdependent-value environment in which (1) Assumption 1 is satisfied; and (2) some message profile in $S^\infty W$ induces an outcome different from the one specified by an incentive compatible and AM-measurable social choice function. Thus, even with Assumption 1, interdependent value environments present new challenges for extending our result.

Example 2 *We build upon Example 1. The new elements we add to Example 1 are summarized as follows: Let* $\pi_i : \bar{T}_i \to \Delta\left(\bar{T}_{-i}\right)$ *be player* i*'s interim belief map from* $\bar{T}_i \to \Delta(\bar{T}_{-i})$ *:*

$$\pi_i(t_i)[t_{-i}] = \begin{cases} p & \text{if } t_{i+1} = t_{i+2} = t_i; \\ 1-p & \text{if } t_{i+1} = t_{i+2} \neq t_i. \end{cases}$$

where $\frac{1}{2} < p < 1$*. That is, in player* i*'s view, player* $(i+1)$*'s type and player* $(i+2)$*'s type are perfectly correlated but they are only partially correlated with player* i*'s type. Each player* i *has the following preferences: for any* $a \in A$ *and* $t \in \bar{T}$*,*

$$u_i(a, t) = a \cdot t_{i+1}.$$

That is, player i*'s preference is determined by player* $(i+1)$*'s type.*

Consider a set of allocation rules $\left\{x_i : \bar{T}_i \to A\right\}_{i \in I}$ where x_i such that $x_i\left(t_i\right) = t_i$ for each t_i. It is easy to see that for all $t_i, t_i' \in \bar{T}_i$ with $t_i \neq t_i'$,

$$\sum_{t_{-i}} u_i(x_i(t_i), t_{-i})\pi_i(t_i)[t_{-i}] > \sum_{t_{-i}} u_i(x_i(t_i'), t_{-i})\pi_i(t_i)[t_{-i}].$$

Thus, we obtain $\Pi_i^* = \{\{t_i^1\}, \{t_i^2\}\}$, the finest possible partition, for each player i. Therefore, every social choice function is AM-measurable in the environment.[18]

For any $\bar{\tau} > 0$, we can adopt the corresponding mechanism $(\mathcal{M}, \bar{\tau})$ defined in Section 2.4.1. We claim that the mechanism generates a strategy profile which survives $S^\infty W$ but induces an outcome which is *not* consistent with the one specified by the SCF f^*. We formally state this result in the following claim:

Claim 5 *Consider Example 2. Fix any mechanism* $(\mathcal{M}, \bar{\tau})$ *defined in Section 2.4.1. For any* $i \in I$ *and any* $t_i \in \bar{T}_i$*, we have that* $(t_i', \ldots, t_i') \in S_i^\infty W_i(t_i | \mathcal{M}, \bar{\mathcal{T}})$ *for some* $t_i' \neq t_i$*.*

Proof. See Appendix A.3. ■

Remark 8 *Note that the SCF* f^* *in Example 2 also satisfies the notion of strict (Maskin) monotonicity defined in Chung and Ely (2003). In particular, Chung and Ely (2003) show*

[18] In fact, this environment satisfies no-total-indifference and type diversity, under which AM-measurability is automatically satisfied. See Serrano and Vohra (2005) for a detailed discussion about these notions.

that under strict monotonicity and a no-veto power condition, there exists an infinite mechanism that achieves $\overline{UNE}$-implementation. In Section 3.3, we show that the mechanism in Abreu and Matsushima (1994) can also achieve $\overline{UNE}$-implementation in private-value environments. However, we argue by Claim 5 (which holds for any p close to 1) that the mechanism fails to achieve $\overline{UNE}$-implementation when the information structure differs slightly from complete information. This shows that our finite mechanism cannot $\overline{UNE}$-implement some strictly monotone SCF. This shows that if we are to achieve $\overline{UNE}$-implementation in general interdependent-value environments, we need to either appeal to infinite mechanisms as in Chung and Ely (2003) or construct a finite mechanism different from that of Abreu and Matsushima.

4.2 Mechanisms, Solution Concepts, and Implementation

We formulate the extension of our results to interdependent-value environments. First, we strengthen the solution concept from $S^{\infty}W$ to *the iterative elimination of weakly dominated strategies*, W^{∞} defined as follows. Set $W_i^0(t_i|\mathcal{M},\mathcal{T}) = M_i$. For any $l \geq 1$, we say that $m_i \in W_i^{l+1}(t_i|\mathcal{M},\mathcal{T})$ if and only if there does not exist $\alpha_i \in \Delta(M_i)$ such that[19]

$$\begin{aligned} & \sum_{t_{-i}} \left[u_i(g(\alpha_i, \nu_{-i}(t_{-i})), \hat{\theta}(t)) + \tau_i(\alpha_i, \nu_{-i}(t_{-i})) \right] \pi_i(t_i)[t_{-i}] \\ \geq & \sum_{t_{-i}} \left[u_i(g(m_i, \nu_{-i}(t_{-i})), \hat{\theta}(t)) + \tau_i(m_i, \nu_{-i}(t_{-i})) \right] \pi_i(t_i)[t_{-i}] \end{aligned}$$

for all $\nu_{-i} : T_{-i} \to M_{-i}$ such that $\nu_{-i}(t_{-i}) \in W_{-i}^l(t_{-i}|\mathcal{M},\mathcal{T})$ and a strict inequality holds for some $\nu_{-i} : T_{-i} \to M_{-i}$. Let W^{∞} denote the set of strategy profiles which survive iterative removal of weakly dominated strategies, i.e.,

$$W_i^{\infty}(t_i|\mathcal{M},\mathcal{T}) = \bigcap_{l=1}^{\infty} W_i^l(t_i|\mathcal{M},\mathcal{T}),$$

$$W^{\infty}(t|\mathcal{M},\mathcal{T}) = \prod\nolimits_{i\in I} W_i^{\infty}(t_i|\mathcal{M},\mathcal{T}).$$

Definition 13 *An SCF f is **implementable in** W^{∞} with arbitrarily small transfers if, for all $\bar{\tau} > 0$, there is a mechanism $(\mathcal{M}, \bar{\tau})$ such that for any $t \in \bar{T}$, and $m \in W^{\infty}\left(t|\mathcal{M},\bar{\mathcal{T}}\right)$, we have $g(m) = f(t)$.*

Example 2 shows that in order to extend our result to interdependent-value environments, we need to strengthen Assumption 1. To do so, we introduce the concept of strategic distinguishability of players' payoff types, which is proposed by Bergemann and Morris (2009b).

[19]Here we instead allow for a mixed strategy to be a dominator because later we would like to employ a result in Bergemann and Morris (2009b) where they adopt ex post rationalizability as their solution concept.

Given a mechanism $\mathcal{M}=(M,g)$, we first define the process of iterative elimination of *ex post* never best responses, which makes no assumptions on each player's belief about other players' payoff types. We set $S_i^0(\theta_i|\mathcal{M})=M_i$ and for each $k=0,\ldots$, we inductively define

$$S_i^{k+1}(\theta_i|\mathcal{M}) = \left\{ m_i \in S_i^k(\theta_i|\mathcal{M}) \;\middle|\; \begin{array}{l} \exists \mu_i \in (\Theta_{-i}\times M_{-i}) \text{ s.t.} \\ (1)\ \mu_i(\theta_{-i},m_{-i})>0 \Rightarrow m_{-i}\in S_{-i}^k(\theta_{-i}|\mathcal{M}) \\ (2)\ m_i \in \arg\max_{m_i'} \sum_{\theta_{-i},m_{-i}} \mu_i(\theta_{-i},m_{-i})\, u_i\left(g\left(m_i',m_{-i}\right),(\theta_i,\theta_{-i})\right) \end{array} \right\}.$$

Finally, we let $S_i^\infty(\theta_i|\mathcal{M})=\bigcap_{k\geq 0} S_i^k(\theta_i|\mathcal{M})$ and call it the set of *ex post (as opposed to interim) rationalizable strategies* for payoff type θ_i.

Definition 14 *Payoff types θ_i and θ_i' are* ***strategically indistinguishable***[20] *(we denote it by $\theta_i \sim \theta_i'$) then $S_i^\infty(\theta_i|\mathcal{M})\cap S_i^\infty(\theta_i'|\mathcal{M})\neq\varnothing$ in any mechanism $\mathcal{M}$.*

In their Proposition 2, Bergemann and Morris (2009b) construct a (finite) mechanism $\mathcal{M}^*=(M^*,g^*)$ with the property that if $\theta_i \not\sim \theta_i'$, then $S_i^\infty(\theta_i|\mathcal{M}^*)\cap S_i^\infty(\theta_i'|\mathcal{M}^*)=\varnothing$. Following Bergemann and Morris (2009b), we refer to $\mathcal{M}^*$ as the *maximally revealing mechanism*. Since M^* is finite, there exists some positive number $\bar{l}$ such that, for any $l\geq\bar{l}$, $i\in I$, and $\theta_i\in\Theta_i$, $S_i^l(\theta_i|\mathcal{M}^*)=S_i^\infty(\theta_i|\mathcal{M}^*)$.

Let $\sim^*$ be the transitive closure of the binary relation $\sim$. For each player i of type θ_i, we define $\psi_i^*(\theta_i)=\{\theta_i'\in\Theta_i|\theta_i'\sim^*\theta_i\}$. Since $\sim^*$ is transitive, it follows that $\{\psi_i^*(\theta_i)\}_{\theta_i\in\Theta_i}$ forms a partition over Θ_i, which we denote by Ξ_i^*. For each $\psi_i\in\Xi_i^*$, we write

$$M_i^*(\psi_i)=\bigcup_{\theta_i\in\psi_i}\{m_i\in M_i^*|m_i\in S_i^\infty(\theta_i|\mathcal{M}^*)\}.$$

In words, $M_i^*(\psi_i)$ describes the set of ex post rationalizable strategies for any payoff type that belongs to ψ_i. For each $\psi_{-i}\in\Xi_{-i}^*$, we also define

$$M_{-i}^*(\psi_{-i})=\times_{j\neq i}M_j^*(\psi_j).$$

For any $\theta_i,\theta_i'\in\Theta_i$, whenever $\theta_i\not\sim\theta_i'$, we have that $S_i^\infty(\theta_i|\mathcal{M}^*)\cap S_i^\infty(\theta_i'|\mathcal{M}^*)=\varnothing$. Thus, the maximally revealing mechanism $\mathcal{M}^*$ can elicit the true payoff type of player i modulo the binary relation $\sim^*$. Specifically, for any $\theta_i\in\Theta_i$ and $m_i\in S_i^\infty(\theta_i|\mathcal{M}^*)$, we are able to identify a unique $\psi_i\in\Xi_i^*$ such that $m_i\in M_i^*(\psi_i)$.

We define type t_i's belief over the partition of his opponents' payoff types by $\chi_{t_i}\in\Delta(\Xi_{-i}^*)$. We define type t_i's belief over the payoff types of other players modulo ψ_{-i} as

[20](Bergemann and Morris, 2009b, Theorem 1) show that strategic indistinguishability is equivalent to a primitive notion called pairwise inseparability.

follows: for any $\psi_{-i} \in \Xi^*_{-i}$,

$$\chi_{t_i}(\psi_{-i}) = \sum_{t_{-i}:\hat{\theta}_{-i}(t_{-i})\in\psi_{-i}} \pi_i(t_i)[t_{-i}].$$

To deal with interdependent-value environments, we introduce the following condition on environments.

Assumption 2 *An environment $\mathcal{E}$ satisfies Assumption 2 if, if, for all $i \in I$ and $t_i, t'_i \in \bar{T}_i$ with $t_i \neq t'_i$, it follows that $\chi_{t_i} \neq \chi_{t'_i}$.*

Remark 9 *Note that in Example 2, Ξ^*_i is trivial, i.e., it constitutes the coarsest possible partition over Θ_i and hence it is ruled out by Assumption 2. Provided that Ξ^*_i is nontrivial, Assumption 2 generically holds over the space of probability distributions over $\bar{T}$. In particular, Bergemann and Morris (2009a) show that each Ξ^*_i becomes the finest possible partition $\{\{\theta_i\}|\theta_i \in \Theta_i\}$ in a class of interdependent-value environments satisfying three conditions: (1) there is a strictly ex post incentive compatible SCF; (2) agents have single-crossing preferences; (3) the agents preferences satisfy a condition called the contraction property, which demands that value interdependence is not too large. In this case, Assumption 2 is equivalent to Assumption 1.*

Given the maximally revealing mechanism $\mathcal{M}^*$, we can construct a new scoring rule in this environment: for any $t_i \in \bar{T}_i$, $\psi_{-i} \in \Xi^*_{-i}$, and $m_{-i} \in M^*_{-i}$,

$$\tilde{d}^0_i(t_i, m_{-i}) = \begin{cases} 2\chi_{t_i}(\psi_{-i}) - \sum_{\psi'_{-i}\in\Xi^*_{-i}} \{\chi_{t_i}(\psi'_{-i})\}^2 & \text{if } m_{-i} \in M^*_{-i}(\psi_{-i}) \\ \text{arbitrary} & \text{if } m_{-i} \notin M^*_{-i}(\psi_{-i}). \end{cases} \tag{29}$$

The next lemma shows that the scoring rule $\tilde{d}^0_i$ gives each agent i a strict incentive to announce the true type, provided that all other agents choose ex post rationalizable strategies in the maximally revealing mechanism.

Lemma 4 *Suppose that an environment $\mathcal{E}$ satisfies Assumption 2. For all $i \in I$, any $t_i, t'_i \in \bar{T}_i$ with $t_i \neq t'_i$, and $\sigma_i : \bar{T}_{-i} \to M^*_{-i}$, whenever $\sigma_i(t_{-i}) \in S^\infty_{-i}(\theta_{-i}(t_{-i})|\mathcal{M}^*)$, we have*

$$\sum_{t_{-i}\in\bar{T}_{-i}} \left[\tilde{d}^0_i(t_i, \sigma_i(t_{-i})) - \tilde{d}^0_i(t'_i, \sigma_i(t_{-i}))\right] \pi_i(t_i)[t_{-i}] > 0.$$

Proof. See Appendix A.4. ■

Given this scoring rule and the hypothesis that player i's opponents choose their ex post rationalizable strategy in the maximally revealing mechanism, we can elicit each player i's interim beliefs, which further enables us to identify each player's true (not payoff) type by Assumption 2 (see the proof of Theorem 5 for details). We are now ready to state the main result of this section.

Theorem 5 *Suppose that $I \geq 2$ and the environment $\mathcal{E}$ satisfies Assumption 2. Then, an SCF f is implementable in W^∞ with arbitrarily small transfers if and only if it is incentive compatible.*

Remark 10 *Although we adopt W^∞ as our solution concept, we only perform $\bar{l}$ rounds of iterative deletion of weakly dominated strategies to induce the truth-telling (recall $S_i^l(\theta_i|\mathcal{M}^*) = S_i^\infty(\theta_i|\mathcal{M}^*)$ for any $l \geq \bar{l}$). Hence, when $\bar{l} = 1$ (e.g., the case of private values), we can replace W^∞ by $S^\infty W$.*

Proof of the "if" part of Theorem 5: We modify the mechanism in Section 2.4.1 in the following ways: (i) we adopt the maximally revealing mechanism constructed in Bergemann and Morris (2009b) for the first round report; (ii) we increase the number of each player's reports from $K+3$ to $K+3+\bar{l}$; (iii) we increase the number of scoring rules from 2 (one between round -2 and -1, and the other between round -1 and 0) to $\bar{l}+1$ (each scoring rule is defined between round l and $l+1$ where l runs from $-(\bar{l}+2)$ to 0). While the mechanism appears to be a natural generalization from the mechanism proposed in the previous section, the argument requires several subtle steps summarized in the proof of Claim 6 in Appendix A.5.

We detail the modification of the mechanism below:

1. **The message space**:

   Each player i simultaneously makes an announcement in the maximally revealing mechanism $\mathcal{M}^* = (M^*, g^*)$ and $(K+\bar{l}+2)$ announcements of his own type. We index each announcement by $-(\bar{l}+2), -(\bar{l}+1), ..., 0, 1, \ldots, K$. That is, player i's message space is

   $$M_i = M_i^{-(\bar{l}+2)} \times \cdots \times M_i^0 \times \cdots \times M_i^K = M_i^* \times \underbrace{\bar{T}_i \times \cdots \times \bar{T}_i}_{K+\bar{l}+2 \text{ times}},$$

   where K is an integer to be specified later.

2. **The outcome function**:

   Let $\epsilon \in (0,1)$ be a small positive number. Define $e : M^{-(\bar{l}+1)} \times \cdots \times M^0 \to \mathbb{R}$ by

   $$e((m^{-l})_{l=0}^{\bar{l}+1}) = \begin{cases} \epsilon & \text{if } m_i^{-l} \neq m_i^0 \text{ for some } i \in I \text{ and some } l \in \{1, \ldots, \bar{l}+1\}, \\ 0 & \text{otherwise.} \end{cases}$$

   Based on the outcome function g^* in the maximally revealing mechanism[21], our outcome function $g : M \to \Delta(A)$ is defined as follows: for each $m \in M$,

   $$g(m) = e((m^{-l})_{l=0}^{\bar{l}+1}) g^*(m^{-(\bar{l}+2)}) + \left\{1 - e(\left(m^{-l}\right)_{l=0}^{\bar{l}+1})\right\} \frac{1}{K} \sum_{k=1}^{K} f\left(m^k\right). \tag{30}$$

[21] Here g^* needs to be generic in the sense defined in Claim 11 in the appendix.

3. **The transfer rule**:

We abuse notation to use E and D as

$$E = \max_{m_i^{-(\bar{l}+2)} \in M_i^*, m^k \in M^k, \theta \in \Theta, i \in I} \left| u_i \left(g^*(m^{-(\bar{l}+2)}), \theta \right) - u_i \left(f \left(m^k \right), \theta \right) \right|; \tag{31}$$

$$D = \max_{\bar{m}_i^k \in M_i^k, m^k \in M^k, \theta \in \Theta, i \in I} \left\{ u_i \left(f \left(m^k \right), \theta \right) - u_i(f(m_{-i}^k, \bar{m}_i^k), \theta) \right\}, \tag{32}$$

Now, in addition to the transfers defined in Section 2.4.1, player i is to pay $-\lambda \tilde{d}_i^0(m_i^{-(\bar{l}+1)}, m_{-i}^{-(\bar{l}+2)})$, and pay $-\lambda d_i^0(m_i^{-l}, \hat{\theta}_{-i}(m_{-i}^{-(l+1)}))$ for any $l \in \{0, \ldots, \bar{l}\}$.

In total,

$$\tau_i(m) = \lambda \tilde{d}_i^0(m_i^{-(\bar{l}+1)}, m_{-i}^{-(\bar{l}+2)}) + \lambda \sum_{l=0}^{\bar{l}} d_i^0(m_i^{-l}, \hat{\theta}_{-i}(m_{-i}^{-(l+1)})) + d_i(m^0, ..., m^K) + \sum_{k=1}^{K} d_i^k(m_i^0, m_i^k).$$

We choose positive numbers λ, γ, K, ϵ, η, and ξ such that (9), (10), (11), (12), and (13) hold; moreover, by Lemma 4, for every $i \in I$, every $t_i, t_i' \in \bar{T}_i$ with $t_i \neq t_i'$, and $\sigma_i : \bar{T}_{-i} \to M_{-i}^*$ such that $\sigma_i(t_{-i}) \in S_{-i}^{\infty}(\theta_{-i}(t_{-i}) | \mathcal{M}^*)$ for any t_{-i},

$$\lambda \sum_{t_{-i} \in \bar{T}_{-i}} \left[\tilde{d}_i^0(t_i, \sigma_i(t_{-i})) - \tilde{d}_i^0(t_i', \sigma_i(t_{-i})) \right] \pi_i(t_i)[t_{-i}] > \gamma. \tag{33}$$

The rest of the mechanism is the same as the one in Section 2.4.1. We shall establish the following claim to prove the theorem. Claim 6 says that any message that survives iterated weak dominance in our mechanism must entail a message that was ex post rationalizable in the maximally revealing mechanism.

Claim 6 *For every* $m_i \in W_i^{\infty}(t_i | \mathcal{M}, \bar{T})$, $m_i^{-(\bar{l}+2)} \in S_i^{\infty}(\hat{\theta}_i(t_i) | \mathcal{M}^*)$.

We relegate the proof of Claim 6 to Appendix A.5. By inequality (33), we can show that any message that survives the iterated weak dominance in our mechanism must entail the truth-telling in round $-(\bar{l}+1)$ of announcement. Then, the rest of the proof then follows verbatim the "if" part of proof of Theorem 1.■

Proof of the "only-if" part of Theorem 5: Fix an arbitrary $\bar{\tau} > 0$. Let $\mathcal{M} = ((M_i), g, (\tau_i))_{i \in I}$ be a mechanism which implements f in W^{∞} with transfers bounded by $\bar{\tau}$. We exploit the concept of stable set introduced by Kohlberg and Mertens (1986).[22] Specifically, by Corollary

[22]Following Kohlberg and Mertens (1986) and (van Damme, 1987, p. 265), a set E of Nash equilibria of a normal-form game Γ is stable if it is a minimal set with the following property: E is a closed set of equilibria and for every $\varepsilon > 0$ there exists some $\hat{\eta} > 0$ such that, if $\eta \in (0, \hat{\eta})$, every η-perturbed game associated with the reduced normal form of Γ has an equilibrium that is ε-close to E. To apply the result, we consider the agent normal form of the incomplete information game $U(\mathcal{M}, \bar{\mathcal{T}})$ as in footnote 11.

10.3.2 in van Damme (1987), every finite game has a stable set which is contained in the set of Nash equilibria; by Theorem 10.3.3 in van Damme (1987), a stable set contains a stable set of any game obtained by eliminating a weakly dominated strategy. Therefore, a stable set contains a nonempty subset in W^∞. In other words, there is a Bayes Nash equilibrium σ of $U(\mathcal{M}, \bar{\mathcal{T}})$ such that for any $t \in \bar{T}$ and $m \in M$, $m \in W^\infty\left(t|\mathcal{M}, \bar{\mathcal{T}}\right)$ implies $\sigma(m|t) > 0$. The rest of the proof follows verbatim the proof of the "only if" part of Theorem 1.■

Remark 11 *We do not know whether the order of deletion of weakly dominated strategies matters for Theorem 5.*

5 Discussion

In Section 5.1, we introduce the concept of partial honesty and propose a way of making the dominance "strict." This allows us to connect our results to *rationalizable* implementation. Section 5.2 discusses the difference between implementation with small transfers and virtual implementation.

5.1 The Role of Honesty and Rationalizable Implementation

Following Matsushima (2008) and Dutta and Sen (2012), we depart from the assumption that all players are motivated solely by their self-interest and instead assume that they all have a small intrinsic preference for honesty. This implies that such players have preferences not just on outcomes but also directly on the *messages* that they are required to send to the planner.

Fix the mechanism $\Gamma = (\mathcal{M}, \bar{\tau})$ that we constructed in Section 4. First, recall that each player i's preferences are given by $u_i : \Delta(A) \times \Theta \to \mathbb{R}$. Following the setup of Dutta and Sen (2012), we extend this $u_i(\cdot)$ to $v_i : M \times \Theta \to \mathbb{R}$ satisfying the following two properties: for all $i \in I,\ t = (t_i, t_{-i}) \in \bar{T}$, $m_i, \tilde{m}_i, \in M_i$, and $m_{-i} \in M_{-i}$:

1. If $u_i(g(m_i, m_{-i}), \hat{\theta}(t_i, t_{-i})) \geq u_i(g(\tilde{m}_i, m_{-i}), \hat{\theta}(t_i, t_{-i}))$, $m_i^k = t_i$, and $\tilde{m}_i^k \neq t_i$ for some $k = -\bar{l} - 2, \ldots, 0$ (or, we have $\bar{l} = 0$ in the case of private values), then

$$v_i((m_i, m_{-i}), \hat{\theta}(t_i, t_{-i})) > v_i((\tilde{m}_i, m_{-i}), \hat{\theta}(t_i, t_{-i})).$$

2. In all other cases, $v_i((m_i, m_{-i}), \hat{\theta}(t_i, t_{-i})) \geq v_i((\tilde{m}_i, m_{-i}), \hat{\theta}(t_i, t_{-i}))$ if and only if

$$u_i(g(m_i, m_{-i}), \hat{\theta}(t_i, t_{-i})) \geq u_i(g(\tilde{m}_i, m_{-i}), \hat{\theta}(t_i, t_{-i})).$$

The first part of the definition captures an individual's preference for *partial* honesty. That is, he strictly prefers (m_i, m_{-i}) to $(\tilde{m}_i, m_{-i})$ *only if* he thinks $g(m_i, m_{-i})$ is at least as good as $g(\tilde{m}_i, m_{-i})$. We consider this to be a very weak assumption, and this

weakness makes the concept of partial honesty particularly compelling. If all players are partially honest in this sense, we can conclude that in private-value environments, for any $k = -2, -1, 0$, any message containing truth-telling in round k announcement *strictly* dominates the same message except that round k announcement involves non-truth telling. Similarly, in interdependent-value environments, for each $k \in \{-\bar{l} - 2, \ldots, 0\}$, any message containing truth-telling in round k announcement strictly dominates the same message except that round k announcement involves non-truth telling. Hence, given partial honesty, every dominance becomes *strict* in our mechanism. This means that we can improve upon our previous results by replacing both $S^\infty W$ (for private-value environments) and W^∞ (for interdependent-value environments) with S^∞, which is the (interim correlated) *rationalizability* correspondence, which maps each type profile to the set of message profiles that survive the iterated deletion of never best responses.[23] By Dekel, Fudenberg, and Morris (2006), we know that this rationalizability correspondence is upper hemi-continuous. Hence, we obtain the following result:

Proposition 2 *Suppose that $I \geq 2$ and that all agents are partially honest. Then,*

1. *(Private-Value) If the environment $\mathcal{E}$ satisfies Assumption 1, any incentive-compatible SCF is implementable in S^∞ with arbitrarily small transfers. Moreover, any incentive-compatible SCF is "strictly continuously" implementable with arbitrarily small transfers.*

2. *(Interdependent-Value) If the environment $\mathcal{E}$ satisfies Assumption 2, any incentive-compatible SCF f is implementable in S^∞ with arbitrarily small transfers and it is strictly continuously implementable with arbitrarily small transfers.*

Proof. We simply combine all the arguments we made above with Theorem 1 for private value environments and Theorem 6 for interdependent-value environments, respectively together with the fact that the interim correlated rationalizable correspondence is upper-hemicontinuous in finite mechanisms. This completes the proof. ■

Assuming that sending messages is slightly costly, Oury and Tercieux (2012) show in their Theorem 4 that an SCF f is continuously implementable by a finite mechanism if and only if it is implementable in rationalizable strategies by a finite mechanism. Although they do not need ex post payments or partial honesty, without either of these we know of no exact rationalizable implementation result with finite mechanisms.

Matsushima (2008) imposes more stringent structures on the players' cost function of sending messages than our partial honesty so that he can take care of fully interdependent values without imposing any conditions. We believe that one of the strongest assumptions he made was that the cost of sending messages depends on the *proportion* of a player's dishonest

[23] In finite games, it is well known that an action is strictly dominated if and only if it is never a best response.

announcements. This assumption is very specific to the construction of our mechanism and that in Matsushima (2008) (and thus, to basically any mechanism that resembles the Abreu-Matsushima type of construction) in the sense that each player is required to make a number of announcements of his type in the mechanism. In other words, Matsushima's assumption no longer makes sense once we adopt a different construction of the mechanism, according to which all players are not necessarily required to report their types many times. Nevertheless, the concept of partial honesty can still be valid as long as the messages in the mechanism contain the players' types. The lesson we draw here is that there seems to be a clear trade-off between the permissiveness of implementation results and more structures in regard to the cost function of sending messages.

5.2 Implementation with Arbitrarily Small Transfers and Virtual Implementation

Virtual implementation means that the planner contents himself with implementing a social choice function with arbitrarily high probability. In complete information environments with at least three players, Abreu and Sen (1991), Abreu and Matsushima (1992a), and Matsushima (1988) all show that essentially any SCF is virtually implementable. In incomplete information environments with side-payments, Abreu and Matsushima (1992b) show that an SCF is virtually implementable in S^∞ by a finite mechanism if and only if it satisfies incentive compatibility and AM measurability. Example 2 we discussed in Section 4 satisfies type diversity. Under type diversity, we know that every social choice function satisfies AM measurability (see Serrano and Vohra (2005)). This means that the difficulty we encounter in Example 2 has nothing to do with AM measurability. In this section, we elaborate on the comparison between Abreu and Matsushima (1992b) and our mechanism. To accommodate the case of interdependent values, we introduce the following condition:

Condition (S) : We say that an environment $\mathcal{E}$ satisfies *Condition (S)* if, for each $i \in I$, there exist a function $x_i : \bar{T}_i \to \Delta(A)$ and $\zeta > 0$ such that for all $t_i, t_i' \in \bar{T}_i$ with $t_i \neq t_i'$ and $t_{-i} \in \bar{T}_{-i}$,

$$u_i(x_i(t_i), (\hat{\theta}_i(t_i), t_{-i})) - u_i(x_i(t_i'), (\hat{\theta}_i(t_i), t_{-i})) > \zeta. \tag{34}$$

Condition (S) is admittedly a strong requirement but, for the sake of our argument, we assume that Condition (S) holds. Under this condition, we can dispense with the maximally revealing mechanism that we employed for our construction in Section 4 so that we rewrite the payoff difference between m_i' and m_i for type $\bar{t}_i$ in (14) in the proof of Claim 1: for any $(m^{-1}, m^0) \in M^{-1} \times M^0$,

$$\sum_{t_{-i}} e(m^{-1}, m^0) \left\{ u_i(x_i(\bar{t}_i), \hat{\theta}(\bar{t}_i, t_{-i})) - u_i(x_i(m_i^{-2}), \hat{\theta}(\bar{t}_i, t_{-i})) \right\} \pi_i(\bar{t}_i)[t_{-i}] \geq 0.$$

In the case of virtual implementation, we are allowed to achieve the wrong outcome

with small probability so that we can set $e(m^{-1}, m^0) = \epsilon$ for any $(m^{-1}, m^0) \in M^{-1} \times M^0$. In this case, under Condition (S), we reduce the above payoff difference for type $\bar{t}_i$ to the following:

$$\epsilon \sum_{t_{-i}} \left\{ u_i(x_i(\bar{t}_i), \hat{\theta}(\bar{t}_i, t_{-i})) - u_i(x_i(m_i^{-2}), \hat{\theta}(\bar{t}_i, t_{-i})) \right\} \pi_i(\bar{t}_i)[t_{-i}] > 0.$$

This specification of $e(\cdot)$ allows us to dispense with all the announcements between round -1 and 0. This further implies that we need neither Assumption 1 nor Assumption 2. Therefore, in the case of virtual implementation, we can handle "independent" beliefs as well as correlated beliefs and the associated mechanism works in fully interdependent-value environments. Furthermore, by this specification of $e(\cdot)$, we always put positive probability to the outcome, $\sum_{i \in I} x_i(m_i^{-2})/I$, determined by the announcement in round -2. Therefore, the strategies deleted in the first round of the iterative process change from weakly dominated strategies into "strictly dominated" one. That is, we obtain virtual implementation in S^∞ (under iterated strict dominance).

Even without Condition (S), Abreu and Matsushima (1992b) constructed an SCF $x : \bar{T} \to \Delta(A)$ such that modulo AM-measurability, only the true type profile survives S^∞ in the direct revelation mechanism associated with x. Indeed, Abreu and Matsushima (1992b) show that AM-measurability is a necessary condition for virtual implementation in S^∞. This explains why virtual implementation can handle fully interdependent-value environments, as long as the SCF to be implemented satisfies incentive compatibility and AM-measurability.

While virtual implementation provides for an impressive conclusion, it comes at the expense of some assumptions. In virtual implementation, the planner is willing to settle for implementing something that is ϵ-close to the SCF. This implies that the planner is considered capable of committing to any mechanism, which might assign a very bad outcome with probability ϵ. In order for this argument to work, players must take these small probabilities seriously and base decisions on them, with the rational expectation that these outcomes will be enforced if they happen to be selected by the mechanism.

We propose the concept of implementation with arbitrarily small transfers; this is another concept of approximate implementation, very much like virtual implementation. The key feature of our mechanism, however, is that undesirable outcomes never occur with positive probability. Indeed, we need ex post transfers but we can make them arbitrarily small. We therefore believe that implementation with small transfers becomes a more appropriate candidate than virtual implementation when the planner's commitment power to the mechanism is in question.

A Appendix

There are five subsections in the appendix. In Section A.1, we show that in private-value environments, our mechanism also works under iterative deletion of *weakly* dominated strate-

gies, i.e., W^∞ and moreover, the order of removal of strategies in W^∞ is irrelevant in our mechanism. Section A.2 provides the proof of Lemma 3. Section A.3 provides the proof of Claim 5 for Example 2 we discussed in Section 4.1. We provide the proof of Claim 4 in Section A.4 and the proof of Claim 6 in Section A.5, respectively. These claims are needed for proving Theorem 5.

A.1 Order Independence

We now define the process of iterative removal of weakly dominated strategies. We seek to define mechanisms for which the order of removal of weakly dominated strategies is irrelevant, that is, given an arbitrary type profile, any message profile in the set of iteratively weakly undominated strategies can implement the socially desired outcome at that type profile. Given a mechanism $\mathcal{M}$, let $U(\mathcal{M}, \bar{\mathcal{T}})$ denote an incomplete information game associated with a model $\bar{\mathcal{T}}$. Fix a game $U(\mathcal{M}, \bar{\mathcal{T}})$, player $i \in I$ and type $\bar{t}_i \in \bar{T}_i$. Let H be a profile of correspondences $(H_i)_{i \in I}$ where H_i is a mapping from $\bar{T}_i$ to a subset of M_i. A message $m_i \in H_i(\bar{t}_i)$ is *weakly dominated with respect to* H for player i of type $\bar{t}_i \in \bar{T}_i$ if there exists $m_i' \in M_i$ such that

$$\begin{aligned} & \sum_{t_{-i}} \left[u_i(g(m_i', \nu_{-i}(t_{-i})), \hat{\theta}_i(t_i)) + \tau_i(m_i', \nu_{-i}(t_{-i})) \right] \pi_i(t_i)[t_{-i}] \\ \geq & \sum_{t_{-i}} \left[u_i(g(m_i, \nu_{-i}(t_{-i})), \hat{\theta}_i(t_i)) + \tau_i(m_i, \nu_{-i}(t_{-i})) \right] \pi_i(t_i)[t_{-i}] \end{aligned}$$

for all $\nu_{-i} : \bar{T}_{-i} \to M_{-i}$ such that $\nu_{-i}(t_{-i}) \in H_{-i}(t_{-i})$ for all $t_{-i} \in \bar{T}_{-i}$ and a strict inequality holds for some ν_{-i}.[24]

Let $\left\{W^k\right\}_{k=0}^\infty$ be a sequence of profiles of correspondences with the following three properties: for each $i \in I$, $\bar{t}_i \in \bar{T}_i$, and $m_i \in M_i$, (i) $W_i^0\left(\bar{t}_i | \mathcal{M}, \bar{\mathcal{T}}\right) = M_i$; (ii) if $m_i \in W_i^{k+1}\left(\bar{t}_i | \mathcal{M}, \bar{\mathcal{T}}\right) \setminus W_i^k\left(\bar{t}_i | \mathcal{M}, \bar{\mathcal{T}}\right)$, it is weakly dominated with respect to W^k for player i of type $\bar{t}_i$; and (iii) if $m_i \in W_i^\infty\left(\bar{t}_i | \mathcal{M}, \bar{\mathcal{T}}\right)$, it is weakly undominated with respect to W^∞ for player i of type $\bar{t}_i$ where $W_i^\infty(\bar{t}_i | \mathcal{M}, \bar{\mathcal{T}}) \equiv \bigcap_{l=1}^\infty W_i^l(\bar{t}_i | \mathcal{M}, \bar{\mathcal{T}})$.

For any $\bar{t} \in \bar{T}$, we let $W^\infty\left(\bar{t} | \mathcal{M}, \bar{\mathcal{T}}\right) = \prod_{i \in I} W_i^\infty\left(\bar{t} | \mathcal{M}, \bar{\mathcal{T}}\right)$. Since M is finite, $W_i^k\left(\bar{t}_i | \mathcal{M}, \bar{\mathcal{T}}\right)$ is nonempty for any k. Thus, W^∞ is nonempty-valued. Note that $W^\infty\left(\bar{t} | \mathcal{M}, \bar{\mathcal{T}}\right)$ depends on the sequence $\left\{W^k\right\}_{k=0}^\infty$. However, we will show that for any $t \in \bar{T}$ and $m \in W^\infty\left(t | \mathcal{M}, \bar{\mathcal{T}}\right)$, we have $g(m) = f(t)$. That is, the socially desired outcome achieved in W^∞ is obtained by any elimination order. We first establish the following claim.

Claim 7 *Assume that the environment $\mathcal{E}$ satisfies Assumption 1. For any $\gamma' > 0$, there exist $\lambda > 0$ and a proper scoring rule d_i^0 such that for any $t_i, t_i', t_i'' \in \bar{T}_i$ with $t_i' \neq t_i''$ and any*

[24] We consider player i's belief over other players' *pure* strategies. However, this formulation is equivalent to taking player i's belief as a conjecture over other players' (correlated) mixed strategies, i.e., $\nu_i : \bar{T}_{-i} \to \Delta(M_{-i})$ such that $\nu_i(t_{-i})[H_{-i}(t_{-i})] = 1$.

$\sigma_{-i} : \bar{T}_{-i} \to \bar{T}_{-i}$, *we have that*

$$\lambda \left| \sum_{t_{-i} \in \bar{T}_{-i}} \left[d_i^0 \left(\sigma_{-i} \left(t_{-i} \right), t_i' \right) - d_i^0 \left(\sigma_{-i} \left(t_{-i} \right), t_i'' \right) \right] \pi_i \left(t_i \right) \left[t_{-i} \right] \right| > \gamma'. \tag{35}$$

Proof. Fix any i. Let

$$C_i = \left\{ d_i^0 \in \mathbb{R}^{\bar{T}} : \sum_{t_{-i} \in \bar{T}_{-i}} \left[d_i^0 \left(t_{-i}, t_i \right) - d_i^0 \left(t_{-i}, t_i' \right) \right] \bar{\pi}_i \left(t_i \right) \left[t_{-i} \right] > 0, \forall t_i \neq t_i' \right\}.$$

C_i is the set of proper scoring rules in $\mathbb{R}^{\bar{T}}$. By Lemma 2, C_i is a nonempty open set. Let

$$C_i' = \left\{ d_i^0 \in \mathbb{R}^{\bar{T}} : \sum_{t_{-i} \in \bar{T}_{-i}} \left[d_i^0 \left(\sigma_{-i} \left(t_{-i} \right), t_i' \right) - d_i^0 \left(\sigma_{-i} \left(t_{-i} \right), t_i'' \right) \right] \bar{\pi}_i \left(t_i \right) \left[t_{-i} \right] \neq 0, \forall t_i, t_i', t_i'' : t_i' \neq t_i'', \forall \sigma_{-i} \right\}.$$

Since $\bar{T}$ is finite, the complement of C_i' has measure zero in $\mathbb{R}^{\bar{T}}$. Therefore, $C_i \cap C_i'$ has a positive measure in $\mathbb{R}^{\bar{T}}$. Thus we can find a proper scoring rule d_i^0 such that for any $\sigma_{-i} : \bar{T}_{-i} \to \bar{T}_{-i}$ and $t_i, t_i', t_i'' \in \bar{T}_i$ with $t_i' \neq t_i''$,

$$\sum_{t_{-i} \in \bar{T}_{-i}} \left[d_i^0 \left(\sigma_{-i} \left(t_{-i} \right), t_i' \right) - d_i^0 \left(\sigma_{-i} \left(t_{-i} \right), t_i'' \right) \right] \pi_i \left(t_i \right) \left[t_{-i} \right] \neq 0.$$

Finally, since $\bar{T}$ is finite, for any $\gamma' > 0$, we can find some $\lambda > 0$ such that inequality (35) holds for any $\sigma_{-i} : \bar{T}_{-i} \to \bar{T}_{-i}$ and $t_i, t_i', t_i'' \in \bar{T}_i$ with $t_i' \neq t_i''$. ■

We introduce one mild condition on the type space.

Condition 1 *A type space $\bar{\mathcal{T}}$ satisfies **Condition 1** if $\pi_i \left(t_i \right) \left[t_{-i} \right] > 0$ for some $\left(t_j \right)_{j \neq i} = t_{-i} \in \bar{T}_{-i}$, then $\pi_j \left(t_j \right) \left[t_i \right] > 0$ for any $j \neq i$, where $\pi_j(t_j)[t_i] = \sum_{(t_i, \tilde{t}_{-i-j}) \in \bar{T}_{-j}} \pi_j(t_j)[t_i, \tilde{t}_{-i-j}]$.*

For instance, Condition 1 is adopted in Vohra (1999) and Jackson (1991). It automatically holds when $\bar{T}$ admits the common support for all players' priors. We are ready to state our main result in this section.

Proposition 3 *Suppose that $I \geq 2$, the environment $\mathcal{E}$ satisfies Assumption 1, and $\bar{\mathcal{T}}$ satisfies Condition 1. Then, for any incentive compatible SCF f and any $\bar{\tau} > 0$, there exists a mechanism $(\mathcal{M}, \bar{\tau})$ such that for any $t \in \bar{T}$ and $m \in W^{\infty}\left(t | \mathcal{M}, \bar{\mathcal{T}} \right)$, we have $g(m) = f(t)$.*

Proof of Proposition 3: Fix $\bar{\tau} > 0$. Choose the mechanism $(\mathcal{M}, \bar{\tau})$ defined in Section 2.4.1, with the proper scoring rule d_i^0 given in Claim 7, and λ under $\gamma' = \gamma$ (which is defined in Section 2.4.1). Note how we can choose λ arbitrarily small by choosing γ small enough.

To prove Proposition 3, we need to show that for any $i \in I$, $\bar{t}_i \in \bar{T}_i$, and $m_i \in M_i$, if $m_i \in W_i^\infty\left(\bar{t}_i|\mathcal{M},\bar{\mathcal{T}}\right)$, then $m_i^{-1} = \bar{t}_i$. This will be established in the three claims (Claims 9, 10, and 11) below.

The claim below says that if a message m_i survives the iterated weak dominance, the message that keeps the same announcement of round -2 as m_i but replaces everywhere else with truth-telling also survives the iterated weak dominance.

Claim 8 *For any $i \in I$, $\bar{t}_i \in \bar{T}_i$, and $m_i \in M_i$, if $m_i \in W_i^\infty\left(\bar{t}_i|\mathcal{M},\bar{\mathcal{T}}\right)$, then $\left(m_i^{-2},\bar{t}_i,...,\bar{t}_i\right) \in W_i^\infty\left(\bar{t}_i|\mathcal{M},\bar{\mathcal{T}}\right)$.*

Proof. For any player $i \in I$, let σ_i be i's strategy such that $\sigma_i\left(\bar{t}_i\right) = \left(\bar{t}_i,...,\bar{t}_i\right)$ for every type $\bar{t}_i \in \bar{T}_i$. Note that we use this notation throughout Section A.1. We prove this claim in two steps.

Step 1: $\sigma_i(\bar{t}_i) \in W_i^\infty(\bar{t}_i|\mathcal{M},\bar{\mathcal{T}})$ *for any $i \in I$ and $\bar{t}_i \in \bar{T}_i$.*

We prove this by induction. Note first that for all $\bar{t} \in \bar{T}$, we have $\sigma(\bar{t}) \in W^0(\bar{t}|\mathcal{M},\bar{\mathcal{T}})$ trivially. For any $n \geq 0$, assume by our induction hypothesis that $\sigma(\bar{t}) \in W^n(\bar{t}|\mathcal{M},\bar{\mathcal{T}})$. Then, we shall show that $\sigma(\bar{t}) \in W^{n+1}(\bar{t}|\mathcal{M},\bar{\mathcal{T}})$ for all $\bar{t} \in \bar{T}$. This is equivalent to showing the following: for any $i \in I$, $\bar{t}_i \in \bar{T}_i$ and $\tilde{m}_i \in W_i^n(\bar{t}_i|\mathcal{M},\bar{\mathcal{T}})$, either $\sigma_i(\bar{t}_i)$ is always at least as good as $\tilde{m}_i$ or there exists $\nu_{-i} : \bar{T}_{-i} \to M_{-i}$ such that $\nu_{-i}(\bar{t}_{-i}) \in W_{-i}^n(\bar{t}_{-i}|\mathcal{M},\mathcal{T})$ for all $\bar{t}_{-i} \in \bar{T}_{-i}$ and $\sigma_i(\bar{t}_i)$ is a strictly better reply to ν_{-i} than $\tilde{m}_i$. Fix $i \in I$, $\bar{t}_i \in \bar{T}_i$, and $\tilde{m}_i \in W_i^n(\bar{t}_i|\mathcal{M},\bar{\mathcal{T}})$. We verify this by considering the following two cases of $\tilde{m}_i$: (i) $\tilde{m}_i^{-2} \neq \sigma_i^{-2}(\bar{t}_i)$ and $\tilde{m}_i^k = \sigma_i^k(\bar{t}_i)$ for all $k \geq -1$; and (ii) $\tilde{m}_i^k \neq \sigma_i^k(\bar{t}_i)$ for some $k \geq -1$. In Case (i), due to the construction of the mechanism, $\sigma_i(\bar{t}_i)$ is at least as good as $\tilde{m}_i$ for any $\nu_{-i} : \bar{T}_{-i} \to M_{-i}$ by inequality (14). In Case (ii), against the conjecture $\nu_{-i}(\cdot) = \sigma_{-i}(\cdot)$, $\sigma_i(\bar{t}_i)$ is a strictly better message than $\tilde{m}_i$ by the argument in Claims 2, 3 and 4. Therefore, there is no $\tilde{m}_i$ that weakly dominates $\sigma_i(\bar{t}_i)$. Thus, $\sigma(\bar{t}) \in W^{k+1}(\bar{t}|\mathcal{M},\bar{\mathcal{T}})$. This completes the proof of Step 1.

Step 2: *for any $i \in I$, $\bar{t}_i \in \bar{T}_i$, and $m_i \in M_i$, if $m_i \in W_i^\infty\left(\bar{t}_i|\mathcal{M},\bar{\mathcal{T}}\right)$, then $\left(m_i^{-2},\bar{t}_i,...,\bar{t}_i\right) \in W_i^\infty\left(\bar{t}_i|\mathcal{M},\bar{\mathcal{T}}\right)$.*

By Step 1, it suffices to show $\left(m_i^{-2},\bar{t}_i,...,\bar{t}_i\right) \in W_i^\infty\left(\bar{t}_i|\mathcal{M},\bar{\mathcal{T}}\right)$ even when $m_i^{-2} \neq \bar{t}_i$. We shall show that no $\tilde{m}_i$ can weakly dominate $\left(m_i^{-2},\bar{t}_i,...,\bar{t}_i\right)$. Specifically, we do so by considering the following two cases of $\tilde{m}_i$: (i) $\tilde{m}_i^{-2} \neq \sigma_i^{-2}(\bar{t}_i)$ and $\tilde{m}_i^k = \sigma_i^k(\bar{t}_i)$ for all $k \geq -1$; (ii) $\tilde{m}_i^k \neq \sigma_i^k(\bar{t}_i)$ for some $k \geq -1$. In Case (i), due to the construction of the mechanism, $\left(m_i^{-2},\bar{t}_i,\ldots,\bar{t}_i\right)$ is at least as good as $\tilde{m}_i$ for any $\nu_{-i} : \bar{T}_{-i} \to M_{-i}$ by inequality (14). In Case (ii), $\left(m_i^{-2},\bar{t}_i,...,\bar{t}_i\right)$ is a strictly better message than $\tilde{m}_i$ against conjecture $\nu_{-i}(\cdot) = \sigma_{-i}(\cdot)$ by the argument in Case (ii) of Step 1. Thus, no $\tilde{m}_i$ can weakly dominate $\left(m_i^{-2},\bar{t}_i,...,\bar{t}_i\right)$. This completes the proof. ■

The next claim says that if a message m_i survives the iterated weak dominance, the

message that keeps the same announcement of round -1 as m_i but replaces everywhere else with truth-telling also survives the iterated weak dominance.

Claim 9 *For any $i \in I$, $\bar{t}_i \in \bar{T}_i$, and $m_i \in M_i$, if $m_i \in W_i^\infty(\bar{t}_i|\mathcal{M}, \bar{\mathcal{T}})$, then $(\bar{t}_i, m_i^{-1}, \bar{t}_i, ..., \bar{t}_i) \in W_i^\infty(\bar{t}_i|\mathcal{M}, \bar{\mathcal{T}})$.*

Proof. By Step 1 in the proof of Claim 8, it suffices to consider the case that $m_i^{-1} \neq \bar{t}_i$. By considering the following two cases, we shall show that no $\tilde{m}_i$ can weakly dominate $(\bar{t}_i, m_i^{-1}, \bar{t}_i, ..., \bar{t}_i)$: (i) $\tilde{m}_i^{-1} \neq m_i^{-1}$ and $\tilde{m}_i^k = \bar{t}_i$ for all $k \neq -1$; (ii) $\tilde{m}_i^k \neq \bar{t}_i$ for some $k \neq -1$. In Case (i), we proceed in two steps.

Step 1: *for any $\tilde{m}_i$, if $\tilde{m}_i^{-1} \neq m_i^{-1}$ and $\tilde{m}_i^k = m_i^k$ for all $k \neq -1$, m_i is strictly better than $\tilde{m}_i$ against some conjecture ν_{-i} such that $\nu_{-i}(\bar{t}_{-i}) \in W_{-i}^\infty\left(\bar{t}_{-i}|\mathcal{M}, \bar{\mathcal{T}}\right)$ for all $\bar{t}_{-i} \in \bar{T}_{-i}$.*

Since $m_i \in W_i^\infty(\bar{t}_i|\mathcal{M}, \bar{\mathcal{T}})$, one of the following two cases must hold: (1) player i of type $\bar{t}_i$ is indifferent between $\tilde{m}_i$ and m_i against any conjecture ν_{-i} such that $\nu_{-i}(\bar{t}_{-i}) \in W_{-i}^\infty(\bar{t}_{-i}|\mathcal{M}, \bar{\mathcal{T}})$ for all $\bar{t}_{-i}$; or (2) m_i is strictly better than $\tilde{m}_i$ for player i of type $\bar{t}_i$ against some conjecture ν_{-i} such that $\nu_{-i}(\bar{t}_{-i}) \in W_{-i}^\infty(\bar{t}_{-i}|\mathcal{M}, \bar{\mathcal{T}})$ for all $\bar{t}_{-i} \in \bar{T}_{-i}$.

By Claim 7, Case (1) is impossible. Thus, we must have Case (2). Since m_i and $\tilde{m}_i$ only differ in round -1, the utility gain for player i of type $\bar{t}_i$ by using m_i rather than $\tilde{m}_i$ is concentrated in the payment rule λd_i^0, which is larger than γ by inequality (35). Next, the utility loss comes from the random dictator component of the outcome function, which is bounded above from ϵE. By inequality (13), we know $\gamma - \epsilon E > 0$. Thus, m_i is strictly better than $\tilde{m}_i$.

Step 2: *for any $\tilde{m}_i$, if $\tilde{m}_i^{-1} \neq m_i^{-1}$ and $\tilde{m}_i^k = \bar{t}_i$ for all $k \neq -1$, then $(\bar{t}_i, m_i^{-1}, \bar{t}_i, ..., \bar{t}_i)$ is strictly better than $\tilde{m}_i$ against some conjecture ν_{-i} such that $\nu_{-i}(\bar{t}_{-i}) \in W_{-i}^\infty(\bar{t}_{-i}|\mathcal{M}, \bar{\mathcal{T}})$ for all $\bar{t}_{-i} \in \bar{T}_{-i}$.*

Since $m_i^{-1} \neq \bar{t}_i$ and $m_i \in W_i^\infty(\bar{t}_i|\mathcal{M}, \bar{\mathcal{T}})$, by Claim 7, there exist a nonempty set of players $J \subset I \backslash \{i\}$, $\{\bar{t}_j\}_{j \in J}$, and a collection of strategies $\{\hat{\sigma}_j\}_{j \in J}$ such that $\hat{\sigma}_j(\bar{t}_j) \in W_j^\infty(\bar{t}_j|\mathcal{M}, \bar{\mathcal{T}})$ and $\hat{\sigma}_j^{-2}(\bar{t}_j) \neq \bar{t}_j$ for all $j \in J$. From Claim 8, we know that $(\hat{\sigma}_j^{-2}(\bar{t}_j), \bar{t}_j, ..., \bar{t}_j) \in W_j^\infty(\bar{t}_j|\mathcal{M}, \bar{\mathcal{T}})$ for all $j \in J$. Let $\tilde{\sigma}_{-i}$ be defined such that $\tilde{\sigma}_{-i}^{-2}(\bar{t}_{-i}) = \hat{\sigma}_{-i}^{-2}(\bar{t}_{-i})$ and $\tilde{\sigma}_{-i}^k(\bar{t}_{-i}) = \sigma_{-i}(\bar{t}_{-i})$ for all $\bar{t}_{-i} \in \bar{T}_{-i}$ and $k \geq -1$. Thus, $\tilde{\sigma}_{-i}(\bar{t}_{-i}) \in W_{-i}^\infty(\bar{t}_{-i}|\mathcal{M}, \bar{\mathcal{T}})$ for all $\bar{t}_{-i} \in \bar{T}_{-i}$.

Fix such conjecture $\nu_{-i}(\cdot) = \tilde{\sigma}_{-i}(\cdot)$. Since $(\bar{t}_i, m_i^{-1}, \bar{t}_i, ..., \bar{t}_i)$ and $\tilde{m}_i$ only differ in round -1, the utility gain for player i of type $\bar{t}_i$ by using $(\bar{t}_i, m_i^{-1}, \bar{t}_i, ..., \bar{t}_i)$ rather than $\tilde{m}_i$ is concentrated in the payment rule λd_i^0, which is larger than γ. Next, the utility loss through the random dictator component of the outcome function, which is bounded above from ϵE. Since we know that $\gamma - \epsilon E > 0$ from the proof of Step 1, $(\bar{t}_i, m_i^{-1}, \bar{t}_i, ..., \bar{t}_i)$ is strictly better than $\tilde{m}_i$ against conjecture $\tilde{\sigma}_{-i}$.

In Case (ii), $(\bar{t}_i, m_i^{-1}, \bar{t}_i, ..., \bar{t}_i)$ is strictly better than $\tilde{m}_i$ against some conjecture, as we can make an argument parallel to Step 2 in the proof of Claim 8. Thus, no $\tilde{m}_i$ can weakly dominate $(\bar{t}_i, m_i^{-1}, \bar{t}_i, ..., \bar{t}_i)$. This completes the proof. ■

The next claim says that if a message survives the iterated weak dominance, it must contain the truth telling in the announcement of round -1.

Claim 10 *For any $i \in I$, $\bar{t}_i \in \bar{T}_i$, and $m_i \in M_i$, if $m_i \in W_i^{\infty}(\bar{t}_i|\mathcal{M}, \bar{\mathcal{T}})$, then $m_i^{-1} = \bar{t}_i$.*

Proof. Suppose not, that is, there exists some $m_i \in W_i^{\infty}(\bar{t}_i|\mathcal{M}, \bar{\mathcal{T}})$ with $m_i^{-1} \neq \bar{t}_i$. Then by Claim 9, $(\bar{t}_i, m_i^{-1}, \bar{t}_i, ..., \bar{t}_i) \in W_i^{\infty}(\bar{t}_i|\mathcal{M}, \bar{\mathcal{T}})$. Since the indicator function $e(\cdot)$ has a positive weight in this case, by inequality (14), we conclude that for any $j \in I \backslash \{i\}$ and $\bar{t}_j \in \bar{T}_j$, if $m_j \in W_j^{\infty}(\bar{t}_j|\mathcal{M}, \bar{\mathcal{T}})$, then $m_j^{-2} = \bar{t}_j$. Since $m_i \in W_i^{\infty}(\bar{t}_i|\mathcal{M}, \bar{\mathcal{T}})$, by Claim 2, whenever $m_i^{-1} \neq \bar{t}_i$, m_i is weakly dominated by $(m_i^{-2}, \bar{t}_i, m_i^0, \ldots, m_i^K)$. This is a contradiction. ■

The rest of the proof of Proposition 3 is completed by adapting the proof of Theorem 1.■

A.2 Proof of Lemma 3

Proof. Since $\mathcal{M}$ is finite, there is k such that $S^k W(\bar{t}|\mathcal{M}, \mathcal{T}) = S^{\infty} W(\bar{t}|\mathcal{M}, \mathcal{T})$ for every $\bar{t} \in \overline{T}$. Thus, it suffices to show that for each k, each $\bar{t} \in \bar{T}$, and sequence $\{t[n]\}_{n=0}^{\infty}$ in T such that $t[n] \to_p \bar{t}$ as $n \to \infty$, there exists a natural number $N_k \in \mathbb{N}$ such that, for any $n \geq N_k$, we have $S^k W(t[n]|\mathcal{M}, \mathcal{T}) \subset S^k W(\bar{t}|\mathcal{M}, \mathcal{T})$. We prove this by induction. Since $S^0 W_i(t_i|\mathcal{M}, \mathcal{T}) = S^0 W_i(t_i'|\mathcal{M}, \mathcal{T})$ whenever $\hat{\theta}_i(t_i) = \hat{\theta}_i(t_i')$ and $\hat{\theta}_i(t_i[n]) = \hat{\theta}_i(\bar{t}_i)$ for some sufficiently large n. The claim is true for $k = 0$. Now suppose that the claim holds for $k \geq 0$ and we show that the claim is also valid for $k+1$.

Fix $m_i \in S^{k+1} W_i(t_i[n]|\mathcal{M}, \mathcal{T})$. Recall the notation in Section 2.2. Then, for any m_i', there exists some $\nu_{-i}^{[n]} : T_{-i} \to S^k W_{-i}(t_{-i}|\mathcal{M}, \mathcal{T})$ such that

$$\begin{aligned} & \sum_{t_{-i}} \left[u_i(g(m_i, \nu_{-i}^{[n]}(t_{-i})), \bar{\theta}_i) + \tau_i\left(m_i, \nu_{-i}^{[n]}(t_{-i})\right) \right] \pi_i(t_i[n])[t_{-i}] \qquad (36) \\ \geq & \sum_{t_{-i}} \left[u_i(g(m_i', \nu_{-i}^{[n]}(t_{-i})), \bar{\theta}_i) + \tau_i\left(m_i', \nu_{-i}^{[n]}(t_{-i})\right) \right] \pi_i(t_i[n])[t_{-i}]. \end{aligned}$$

Let

$$V_i\left(m_i, \nu_{-i}^{[n]}\right) \equiv \sum_{t_{-i}} \left[u_i(g(m_i, \nu_{-i}^{[n]}(t_{-i})), \bar{\theta}_i) + \tau_i\left(m_i, \nu_{-i}^{[n]}(t_{-i})\right) \right] \pi_i(t_i[n])[t_{-i}].$$

For any m_i and m_i', we define $\beta^{m_i, m_i'} : T_{-i} \to M_{-i}$ such that, for any t_{-i},

$$\beta^{m_i, m_i'}(t_{-i}) = \arg \max_{\nu_{-i}^{[n]}(t_{-i}) \in S_{-i}^k(t_{-i}|\mathcal{M}, \mathcal{T})} \left\{ V_i\left(m_i, \nu_{-i}^{[n]}(t_{-i})\right) - V_i\left(m_i', \nu_{-i}^{[n]}(t_{-i})\right) \right\}.$$

We can interpret $\beta^{m_i, m_i'}$ as player i's belief about the best possible scenario for the choice of m_i against m_i' where other players use k-times iteratively undominated strategies. Thus, we

have

$$\begin{aligned}
&\sum_{m_{-i}} \left[u_i(g(m_i, m_{-i}), \bar{\theta}_i) + \tau_i(m_i, m_{-i})\right] \pi_i(t_i[n]) \left[\{t_{-i} \in T_{-i} : \beta^{m_i, m_i'}(t_{-i}) = m_{-i}\}\right] \\
\geq &\sum_{m_{-i}} \left[u_i(g(m_i', m_{-i}), \bar{\theta}_i) + \tau_i(m_i', m_{-i})\right] \pi_i(t_i[n]) \left[\{t_{-i} \in T_{-i} : \beta^{m_i, m_i'}(t_{-i}) = m_{-i}\}\right].
\end{aligned}$$

Note that this is where the assumption of private values becomes crucial. Since $t[n] \to_p \bar{t}$, for any $n > 0$, there exists $\varepsilon_n > 0$ such that

$$\pi_i(t_i[n]) [(\bar{t}_{-i})^{\varepsilon_n}] \to \pi_i(\bar{t}_i) [\bar{t}_{-i}], \text{ as } n \to \infty,$$

where $(\bar{t}_{-i})^{\varepsilon_n}$ denotes an open ball consisting of the set of types t_{-i} whose $(k-1)$-order beliefs are ε_n-close to those of types $\bar{t}_{-i}$.[25] It follows that the following probability is well defined.

For any $\bar{t}_{-i} \in \bar{T}_{-i}$ such that $\pi_i(\bar{t}_i)[\bar{t}_{-i}] > 0$, and m_{-i}, we define the following:

$$\beta_{-i}(\bar{t}_{-i})[m_{-i}] \equiv \lim_{n \to \infty} \frac{\pi_i(t_i[n]) \left[\left\{t_{-i} \in (\bar{t}_{-i})^{\varepsilon_n} : \beta^{m_i, m_i'}(t_{-i}) = m_{-i}\right\}\right]}{\pi_i(\bar{t}_i)[\bar{t}_{-i}]}.$$

Now we construct a conjecture $\nu_{-i} : \bar{T}_{-i} \to \Delta(M_{-i})$ for type $\bar{t}_i$. For any $(\bar{t}_{-i}, m_{-i})$, we set $\nu_{-i}(m_{-i}|\bar{t}_{-i}) = \beta_{-i}(\bar{t}_{-i})[m_{-i}]$. From the inequality above we have

$$\begin{aligned}
&\sum_{m_{-i}} \left[u_i(g(m_i, m_{-i}), \bar{\theta}_i) + \tau_i(m_i, m_{-i})\right] \sum_{\bar{t}_{-i} \in T} \beta_{-i}(\bar{t}_{-i})[m_{-i}] \pi_i(\bar{t}_i)[\bar{t}_{-i}] \\
\geq &\sum_{m_{-i}} \left[u_i(g(m_i', m_{-i}), \bar{\theta}_i) + \tau_i(m_i', m_{-i})\right] \sum_{\bar{t}_{-i} \in T} \beta_{-i}(\bar{t}_{-i})[m_{-i}] \pi_i(\bar{t}_i)[\bar{t}_{-i}].
\end{aligned}$$

Therefore,

$$\begin{aligned}
&\sum_{\bar{t}_{-i}, m_{-i}} \left[u_i(g(m_i, m_{-i}), \bar{\theta}_i) + \tau_i(m_i, m_{-i})\right] \nu_{-i}(m_{-i}|\bar{t}_{-i}) \pi_i(\bar{t}_i)[\bar{t}_{-i}] \\
\geq &\sum_{\bar{t}_{-i}, m_{-i}} \left[u_i(g(m_i', m_{-i}), \bar{\theta}_i) + \tau_i(m_i, m_{-i})\right] \nu_{-i}(m_{-i}|\bar{t}_{-i}) \pi_i(\bar{t}_i)[\bar{t}_{-i}]
\end{aligned}$$

By construction, $\nu_{-i}(m_{-i}|\bar{t}_{-i})\pi_i(\bar{t}_i)[\bar{t}_{-i}] > 0$ implies that $m_{-i} \in S^k W_{-i}(t_{-i}[n]|\mathcal{M}, \mathcal{T})$. By our induction hypothesis, $S^k W_{-i}(t_{-i}[n]|\mathcal{M}, \mathcal{T}) \subset S^k W_{-i}(\bar{t}_{-i}|\mathcal{M}, \mathcal{T})$. Thus, we have $m_{-i} \in S^k W_{-i}(\bar{t}_{-i}|\mathcal{M}, \mathcal{T})$. Since the choice of m_i' is arbitrary, so this completes the proof. ■

[25] This follows from the fact that the Prohorov distance between $t_i[n]$ and $\bar{t}_i$ converges to 0 due to the finiteness of $\bar{T}_{-i}$. See Dudley (2002, pp. 398 and 411).

A.3 Proof of Claim 5

Recall that $\bar{T}_i = \{t_i^1, t_i^2\} = \{(1,0),(0,1)\}$ for each $i \in I$ and $A = \{(1,0),(0,1)\}$. Recall also that player i's preferences only depend on player $i+1$'s type. To simplify the notation, we write player i's preferences as follows: $u_i(a,t) \equiv u_i(a,t_{-i}) = a \cdot t_{i+1}$, for any $a \in A$ and $t \in \bar{T}$.

Let σ' be a strategy profile such that for each $i \in I$ and $t_i \in \bar{T}_i$, $\sigma_i'(t_i) = (t_i', ..., t_i')$ where $t_i' \in \bar{T}_i \backslash \{t_i\}$. Then we show that $\sigma_i'(t_i) \in S_i^\infty W_i(t_i|\mathcal{M}, \bar{\mathcal{T}})$ by the following lemmas. For each $i \in I$, we define $\alpha_i : \bar{T}_i \to \bar{T}_i$ such that $\alpha_i(t_i) \neq t_i$ for all $t_i \in \bar{T}_i$.

We show that a non-truthful announcement by all players constitutes a Bayes Nash equilibrium in the direct-revelation mechanism $(\bar{T}, f^*)$ in Lemma 5.

Lemma 5 *For any player i of type t_i,*

$$\sum_{t_{-i} \in \bar{T}_{-i}} u_i(f^*(t_i', \alpha_{-i}(t_{-i})), t_{-i}) \pi_i(t_i)[t_{-i}] \geq \sum_{t_{-i} \in \bar{T}_{-i}} u_i(f^*(t_i, \alpha_{-i}(t_{-i})), t_{-i}) \pi_i(t_i)[t_{-i}]. \quad (37)$$

Proof. In player i's view, other players' types are perfectly correlated. Besides, f^* is a majority rule. Therefore, in player i's view, player i cannot change the outcome by his unilateral deviation when the other players are making a consistent (false) announcement. Thus, we complete the proof. ■

The next lemma says that if player i's type is different from that of player $i+1$, he has a better outcome by matching his announcement to his neighbors' than that by telling the true type.

Lemma 6 *For any player i of type t_i, $u_i(x_i(t_i'), t_{i+1}') - u_i(x_i(t_i), t_{i+1}') > 0$ if $t_i \neq t_i' = t_{i+1}'$.*

Proof. Fix any outcome $a \in A$. Player i of type t_i's interim utility is given as follows:

$$\sum_{t_{-i} \in \bar{T}_{-i}} u_i(a, t_{-i}) \pi_i(t_i)[t_{-i}] = pa \cdot t_i + (1-p)\, a \cdot t_i',$$

where $t_i \neq t_i'$. Therefore, player i of type t_i strictly prefers a to the other outcome if and only if $a = t_i$. Since $u_i(a, t_{-i}) = a \cdot t_{i+1}$, $u_i(x_i(t_i'), t_{i+1}') - u_i(x_i(t_i), t_{i+1}') > 0$ if $t_i \neq t_i' = t_{i+1}'$. ■

The lemma below says that the message that has the consistent misrepresentation of types survives $S^\infty W$.

Lemma 7 *For every $i \in I$ and $t_i \in \bar{T}_i$, we have $\sigma_i'(t_i) \in S_i^\infty W_i(t_i|\mathcal{M}, \bar{\mathcal{T}})$.*

Proof. We prove Lemma 7 in the following three steps.

Step 1: *For every $i \in I$ and $t_i \in \bar{T}_i$, against conjecture σ'_{-i}, $\sigma_i'(t_i)$ is a strictly better message than $\tilde{m}_i$ if $\tilde{m}_i^k = t_i'$ for any $k \geq -1$.*

Fix any $\tilde{m}_i$. First, consider the case that $\tilde{m}_i^k \neq t_i'$ for some $k \in \{-1, 0\}$. The utility gain in payment rule λd_i^0 from using $\sigma_i'(t_i)$ rather than $\tilde{m}_i$ is

$$\begin{aligned} & \lambda \sum_{t_{-i} \in \bar{T}_{-i}} \left[d_i^0(\sigma_{-i}'^{-1}(t_{-i}), t_i') - d_i^0(\sigma_{-i}'^{-1}(t_{-i}), t_i) \right] \pi_i(t_i)[t_{-i}] \\ = & \lambda \sum_{t_{-i}' \in \bar{T}_{-i}} \left[d_i^0\left(t_{-i}', t_i'\right) - d_i^0\left(t_{-i}', t_i\right) \right] \pi_i\left(t_i'\right)\left[t_{-i}'\right] \\ > & \gamma, \end{aligned}$$

where $t_{i+1} = t_{i+2} = t_i \neq t_i' = t_{i+1}' = t_{i+2}'$ and the first equality follows from that $\pi_i(t_i)[t_{-i}] = \pi_i(t_i')[t_{-i}']$ in this example; the last inequality follows from inequality (10). All the possible loss (from using $\sigma_i'(t_i)$ rather than $\tilde{m}_i$) consists of (i) the utility loss in the random dictatorial component of the outcome function weighted by $e(\cdot)$ function, which is bounded above from ϵE; (ii) the utility loss in d_i, which is bounded above from ξ; (iii) the utility loss in d_i^k for all $k \geq 1$. The total loss is bounded above from $\epsilon E + \xi + K\eta$.

For any outcome that depends on kth message profile, if $\tilde{m}_i^k \neq t_i'$, $\sigma_i'(t_i)$ is at least as good as $\tilde{m}_i$ by inequality (37).

By inequality (13), we know $\gamma > \epsilon E + \xi + K\eta$. Therefore, $\sigma_i'(t_i)$ is a strictly better reply to σ_{-i}' than any such $\tilde{m}_i$.

Finally, consider the case that $\tilde{m}_i^{-1} = \tilde{m}_i^0 = t_i'$ and $\tilde{m}_i^k \neq t_i'$ for some $k \geq 1$. For any $k \geq 1$, in terms of the outcome that depends on the kth message profile, if $\tilde{m}_i^k \neq t_i'$, $\sigma_i'(t_i)$ is at least as good as $\tilde{m}_i$ by inequality (37). In terms of payments, since $\sigma_i'(t_i) = (t_i', ..., t_i')$ is a consistent message, the utility gain (from using $\sigma_i'(t_i)$ rather than $\tilde{m}_i$) in the payment rules d_i and d_i^k for all $k \geq 1$ is bounded below by $\xi + \eta$. Therefore, $\sigma_i'(t_i)$ is a strictly better reply to σ_{-i}' than any such $\tilde{m}_i$. This completes the proof of Step 1.

Step 2: *For every $i \in I$ and $t_i \in \bar{T}_i$, $\sigma_i'(t_i) \in W_i^1(t_i|\mathcal{M}, \bar{\mathcal{T}})$.*

Fix any player i of type t_i and $\tilde{m}_i \neq \sigma_i'(t_i)$. Then, it suffices to show that no $\tilde{m}_i$ can weakly dominate $\sigma_i'(t_i)$. More specifically, Taking the previous step into account, we can decompose our argument into the following two cases of $\tilde{m}_i$:

Case (i) $\tilde{m}_i^{-2} \neq t_i'$ and $\tilde{m}_i^k = t_i'$ for all $k \geq -1$.

Let $\bar{m}_{-i} \in M_{-i}$ be defined such that $\bar{m}_j^{-1} = \bar{m}_j^0$ for all $j \neq i$. Therefore, we have $e((m_i^{-1}, \bar{m}_{-i}^{-1}), (m_i^0, \bar{m}_{-i}^0)) = 0$ when $m_i^{-1} = m_i^0$. Let $\tilde{m}_{-i} \in M_{-i}$ be defined such that $\tilde{m}_j^{-1} \neq \tilde{m}_j^0$ for some $j \neq i$. Then, we have $e((m_i^{-1}, \tilde{m}_{-i}^{-1}), (m_i^0, \tilde{m}_{-i}^0)) = \epsilon$ for all m_i. Let ν_{-i} be a conjecture of type t_i such that $\nu_{-i}(\bar{m}_{-i}|t_{-i}) = 1$ and $\nu_{-i}(\tilde{m}_{-i}|t_{-i}') = 1$ where $t_{i+1} = t_{i+2} = t_i \neq t_i' = t_{i+1}' = t_{i+2}'$. Then, the utility net gain for player i of type t_i from

choosing $\sigma_i'(t_i)$ rather than $\tilde{m}_i$ is given:

$$\begin{aligned} & \left\{0 \times u_i(x_i(t_i'), t_{-i})\pi_i(t_i)[t_{-i}] + \epsilon \times u_i(x_i(t_i'), t'_{-i})\pi_i(t_i)[t'_{-i}]\right\} \\ - & \left\{0 \times u_i(x_i(t_i), t_{-i})\pi_i(t_i)[t_{-i}] + \epsilon \times u_i(x_i(t_i), t'_{-i})\pi_i(t_i)[t'_{-i}]\right\} \\ = & \epsilon\left\{u_i(x_i(t_i'), t'_{-i}) - u_i(x_i(t_i), t'_{-i})\right\}\pi_i(t_i)[t'_{-i}] \\ > & 0, \end{aligned}$$

where the last inequality follows from Lemma 6. Therefore, $\sigma_i'(t_i)$ is a strictly better reply to ν_{-i} than any such $\tilde{m}_i$.

Case (ii) $\tilde{m}_i^k \neq t_i'$ for some $k \geq -1$.

By Step 1, we conclude that $\sigma_i'(t_i)$ is a strictly better message to conjecture σ'_{-i} than any such $\tilde{m}_i$. Thus, no $\tilde{m}_i$ can weakly dominate $\sigma_i'(t_i)$ so that $\sigma_i'(t_i) \in W_i^1(t_i|\mathcal{M}, \bar{\mathcal{T}})$. This completes the proof of Step 2.

Step 3: *For every $i \in I$ and $t_i \in \bar{T}_i$, we have $\sigma_i'(t_i) \in S_i^\infty W_i(t_i|\mathcal{M}, \bar{T})$.*

Fix conjecture σ'_{-i} and any $\tilde{m}_i$. We first show that for each player i of type t_i, $\sigma_i'(t_i)$ is a best response to σ'_{-i} by considering the following two cases: (i) $\tilde{m}_i^{-2} \neq t_i'$ and $\tilde{m}_i^k = t_i'$ for all $k \geq -1$; (ii) $\tilde{m}_i^k \neq t_i'$ for some $k \geq -1$. In Case (i), player i of type t_i is indifferent between $\tilde{m}_i$ and $\sigma_i'(t_i)$ since the indicator function $e(\cdot)$ has a value of 0. In Case (ii), it follows immediately from Step 1. Thus, for every $i \in I$ and $t_i \in \bar{T}_i$, we have $\sigma_i'(t_i) \in S_i^2(t_i|\mathcal{M}, \bar{\mathcal{T}})$. Fix $i \in I$ and $t_i \in \bar{T}_i$. For each $k \geq 2$, we assume by our inductive hypothesis that $\sigma_i'(t_i) \in S_i^k(t_i|\mathcal{M}, \bar{\mathcal{T}})$. Then, we can conclude that $\sigma_i'(t_i) \in S_i^{k+1}(t_i|\mathcal{M}, \bar{\mathcal{T}})$, since we can always fix σ'_{-i} as a conjecture of player i of type t_i. This completes the proof of Step 3. ■

A.4 Proof of Lemma 4

Proof. Fix $i \in I$ and $t_i, t_i' \in \bar{T}_i$ with $t_i \neq t_i'$. Let $\sigma_{-i} : \bar{T}_{-i} \to M_{-i}^*$ be type t_i's conjecture in the maximally revealing mechanism such that $\sigma_{-i}(t_{-i}) \in S_{-i}^\infty(\hat{\theta}_{-i}(t_{-i})|\mathcal{M}^*)$ for each $t_{-i} \in \bar{T}_{-i}$. Thus, for any $t_{-i} \in \bar{T}_{-i}$ and $\psi_{-i} \in \Xi_{-i}^*$, we have $\sigma_{-i}(t_{-i}) \in M_{-i}^*(\psi_{-i})$ if and only if $\hat{\theta}_{-i}(t_{-i}) \in \psi_{-i}$. Consider any $\psi_{-i} \in \Xi_{-i}^*$ such that $\sigma_{-i}(t_{-i}) \in M_{-i}^*(\psi_{-i})$ for each $t_{-i} \in \bar{T}_{-i}$. The expected payoff of player i of type t_i over ψ_{-i} in $\tilde{d}_i^0$ from announcing t_i' is computed as

follows:

$$\begin{aligned}
&\sum_{t_{-i}\in\bar{T}_{-i}:\hat{\theta}_{-i}(t_{-i})\in\psi_{-i}} \lambda\tilde{d}_i^0\left(t_i', \sigma_{-i}(t_{-i})\right)\pi_i(t_i)[t_{-i}] \\
= &\sum_{t_{-i}\in\bar{T}_{-i}:\hat{\theta}_{-i}(t_{-i})\in\psi_{-i}} \lambda\left(2\chi_{t_i'}\left(\psi_{-i}\right) - \sum_{\psi'_{-i}\in\Xi^*_{-i}}\chi^2_{t_i'}\left(\psi'_{-i}\right)\right)\pi_i(t_i)[t_{-i}] \\
= &\lambda\left(2\chi_{t_i'}\left(\psi_{-i}\right) - \sum_{\psi'_{-i}\in\Xi^*_{-i}}\chi^2_{t_i'}\left(\psi'_{-i}\right)\right)\sum_{t_{-i}\in\bar{T}_{-i}:\hat{\theta}_{-i}(t_{-i})\in\psi_{-i}}\pi_i(t_i)[t_{-i}] \\
= &\lambda\left(2\chi_{t_i'}\left(\psi_{-i}\right) - \sum_{\psi'_{-i}\in\Xi^*_{-i}}\chi^2_{t_i'}\left(\psi'_{-i}\right)\right)\chi_{t_i}(\psi_{-i}), \qquad (38)
\end{aligned}$$

where the first equality follows from the definition of $\tilde{d}_i^0$ (see (29)) and the fact that $\sigma_{-i}(t_{-i}) \in M^*_{-i}\left(\psi_{-i}\right)$ if and only if $\hat{\theta}_{-i}(t_{-i}) \in \psi_{-i}$. By Assumption 2, we obtain

$$\{t_i\} = \arg\max_{t_i'\in\bar{T}_i}\sum_{\psi_{-i}\in\Xi^*_{-i}}\lambda\left(2\chi_{t_i'}\left(\psi_{-i}\right) - \sum_{\psi'_{-i}\in\Xi^*_{-i}}\chi^2_{t_i'}\left(\psi'_{-i}\right)\right)\chi_{t_i}(\psi_{-i}).$$

This implies that given the construction of the scoring rule $\tilde{d}_i^0$ and the hypothesis that all other players choose their ex post rationalizable strategies in the maximally revealing mechanism, telling the true type is strictly better than telling any other types. This completes the proof. ■

A.5 Proof of Claim 6

The proof of Claim 6 is reduced to establishing Lemma 8: if a message m_i in the game $U\left(\mathcal{M}, \bar{\mathcal{T}}\right)$ entails a message m_i^* that is not ex post rationalizable in the maximally revealing mechanism, then m_i is weakly dominated in the game $U\left(\mathcal{M}, \bar{\mathcal{T}}\right)$. Claims 11, 12 and 13 all constitute the building blocks for proving Lemma 8. First, we prove Claims 11, 12 and 13 and thereafter, we prove Lemma 8.

Claim 11 below shows that we can have a maximally revealing mechanism with the property that any two distinct messages of every type result in two different payoffs against any pure strategy profile of the other players. We call such a mechanism a "generic" maximally revealing mechanism.

Claim 11 *We can construct a maximally revealing mechanism $\mathcal{M}^* = (M^*, g^*)$ with the following property: for any $t_i \in \bar{T}_i$, $\sigma^*_{-i} : T_{-i} \to M^*_{-i}$, and $m_i, m_i' \in M_i^*$ with $m_i \neq m_i'$, we*

have

$$\begin{aligned}&\sum_{t_{-i}\in\bar{T}_{-i}} u_i(g^*(m_i,\sigma^*_{-i}(t_{-i})),\hat{\theta}_i(t_i),\hat{\theta}_{-i}(t_{-i}))\pi_i(t_i)[t_{-i}]\\ \neq &\sum_{t_{-i}\in\bar{T}_{-i}} u_i(g^*(m'_i,\sigma^*_{-i}(t_{-i})),\hat{\theta}_i(t_i),\hat{\theta}_{-i}(t_{-i}))\pi_i(t_i)[t_{-i}].\end{aligned}$$

Proof. Let $\mathcal{M}=(M^*,g)$ be a finite maximally revealing mechanism proposed in Bergemann and Morris (2009b). We define $\mathcal{M}^\tau=(M^*,g,\tau)$ as the mechanism which augments $\mathcal{M}=(M^*,g)$ with a transfer rule $\tau=(\tau_i)_{i\in I}$ such that $\tau_i:M^*\to\mathbb{R}$ for each $i\in I$. Fix any i. Let

$$C_i=\left\{\tau_i\in\mathbb{R}^{\bar{T}}:S^\infty(\theta_i|\mathcal{M}^\tau)\cap S^\infty(\theta'_i|\mathcal{M}^\tau)=\varnothing,\forall\theta_i,\theta'_i\text{ with }\theta_i\not\sim\theta'_i\right\}.$$

Note that C_i is a nonempty open set. Let

$$C'_i=\left\{\tau_i\in\mathbb{R}^{\bar{T}}:\begin{array}{r}\sum_{t_{-i}\in\bar{T}_{-i}}\left\{u_i(g^*(m_i,\sigma^*_{-i}(t_{-i})),\hat{\theta}_i(t_i),\hat{\theta}_{-i}(t_{-i}))+\tau_i(m_i,\sigma_{-i}(t_{-i}))\right\}\pi_i(t_i)[t_{-i}]\\ \neq\sum_{t_{-i}\in\bar{T}_{-i}}\left\{u_i(g^*(m'_i,\sigma^*_{-i}(t_{-i})),\hat{\theta}_i(t_i),\hat{\theta}_{-i}(t_{-i}))+\tau_i(m'_i,\sigma_{-i}(t_{-i}))\right\}\pi_i(t_i)[t_{-i}],\\ \forall t_i\in\bar{T}_i,\ \forall\sigma^*_{-i}:T_{-i}\to M_{-i},\text{ and }\forall m_i,m'_i\in M^*_i\text{ with }m_i\neq m'_i\end{array}\right\}.$$

Since $\bar{T}$ is finite, the complement of C'_i has measure zero in $\mathbb{R}^{\bar{T}}$. Therefore, $C_i\cap C'_i$ has a positive measure in $\mathbb{R}^{\bar{T}}$. Thus we can find a transfer rule $\tau^*_i\in C_i\cap C'_i$. Thus, we set $\mathcal{M}^*=(M^*,g^*)=(M^*,g,\tau^*)$ as a maximally revealing mechanism as we desire. ■

In what follows, we assume without loss of generality that the maximally revealing mechanism we use in the proof of Lemma 8 is generic. Claims 12 shows that the scoring rule is also generic in the sense that two distinct announcements result in two different payoffs, as we show in Claim 7 of Section A.1.

Claim 12 *Assume that the environment $\mathcal{E}$ satisfies Assumption 2. For any $i\in I$, $t_i\in\bar{T}_i$, and $\gamma'>0$, there exist $\lambda>0$ and a proper scoring rule $\tilde{d}^0_i$ such that for any t'_i, $t''_i\in\bar{T}_i$ with $t'_i\neq t''_i$ and any $\hat{\sigma}_{-i}:\bar{T}_{-i}\to M^*_{-i}$, we have that*

$$\lambda\left|\sum_{t_{-i}\in\bar{T}_{-i}}\left[\tilde{d}^0_i(\hat{\sigma}_{-i}(t_{-i}),t'_i)-\tilde{d}^0_i(\hat{\sigma}_{-i}(t_{-i}),t''_i)\right]\pi_i(t_i)[t_{-i}]\right|>\gamma'. \tag{39}$$

Proof. Fix any i. Let $\Sigma_{-i}=\left\{\sigma_{-i}:\overline{T}_{-i}\to M^*_{-i}|\sigma_{-i}(t_{-i})\in S^\infty_{-i}(\theta_{-i}(t_{-i})|\mathcal{M}^*),\forall t_{-i}\right\}$. For each $\sigma_{-i}\in\Sigma_{-i}$, let

$$C^{\sigma_{-i}}_i=\left\{\tilde{d}^0_i\in\mathbb{R}^{\bar{T}}:\sum_{t_{-i}\in\bar{T}_{-i}}\left[\tilde{d}^0_i(\sigma_{-i}(t_{-i}),t_i)-\tilde{d}^0_i(\sigma_{-i}(t_{-i}),t'_i)\right]\bar{\pi}_i(t_i)[t_{-i}]>0,\forall t'_i\neq t_i\right\}.$$

By Lemma 4, $C_i \equiv \bigcap_{\sigma_{-i} \in \Sigma_{-i}} C_i^{\sigma_{-i}}$ is a nonempty open set. The rest of the proof is identical to the argument in the proof of Claim 7. ■

It immediately follows from Claim 12 that inequality (39) holds if we choose $\gamma' = \gamma$ chosen in (10) and (13) in Section 2.4.1.

We introduce the following claim to show that under the scoring rule, whenever player i of type t_i misrepresents his type t_i as t_i', there always exists a conjecture that rationalizes this misrepresentation. This constitutes a key result to show that for any $l \in \{1, \ldots, \bar{l}\}$, there always exists a message profile m that survives W^l (i.e., l-times iterative deletion of weakly dominated strategies) such that $e((m^{-l})_{l=0}^{\bar{l}+1}) = \epsilon$.

Claim 13 *For any $t_i, t_i' \in \bar{T}_i$ with $t_i \neq t_i'$, there exists $\sigma_{-i} : \bar{T}_{-i} \to \Delta\left(\bar{T}_{-i}\right)$, such that*

$$\{t_i'\} = \arg\max_{\tilde{t}_i \in \bar{T}_i} \sum_{t_{-i} \in \bar{T}_{-i}} \pi_i\left(t_i\right)\left[t_{-i}\right] \sum_{\tilde{t}_{-i} \in \bar{T}_{-i}} d_i^0\left(\tilde{t}_{-i}, \tilde{t}_i\right) \sigma_{-i}\left(\tilde{t}_{-i} | t_{-i}\right),$$

where we denote by $\sigma_{-i}\left(\tilde{t}_{-i} | t_{-i}\right)$ the probability that $\tilde{t}_{-i}$ is realized given that $\sigma_{-i}(t_{-i})$ is played.

Proof. First note that

$$\{t_i'\} = \arg\max_{\tilde{t}_i \in \bar{T}_i} \sum_{\tilde{t}_{-i} \in \bar{T}_{-i}} d_i^0\left(\tilde{t}_{-i}, \tilde{t}_i\right) \pi_i\left(t_i'\right)\left[\tilde{t}_{-i}\right],$$

since the scoring rule is strictly incentive compatible. We construct type t_i's conjecture denoted by $\sigma_{-i} : \bar{T}_{-i} \to \Delta\left(\bar{T}_{-i}\right)$ such that for any $t_{-i}, \tilde{t}_{-i} \in \bar{T}_{-i}$,

$$\sigma_{-i}\left(\tilde{t}_{-i} | t_{-i}\right) = \pi_i\left(t_i'\right)\left[\tilde{t}_{-i}\right]. \tag{40}$$

We consider player i of type t_i and compute type t_i's expected utility from d_i^0:

$$\begin{aligned}
& \sum_{t_{-i} \in \bar{T}_{-i}} \pi_i\left(t_i\right)\left[t_{-i}\right] \sum_{\tilde{t}_{-i} \in \bar{T}_{-i}} d_i^0\left(\tilde{t}_{-i}, \tilde{t}_i\right) \sigma_{-i}\left(\tilde{t}_{-i} | t_{-i}\right) \\
= & \sum_{t_{-i} \in \bar{T}_{-i}} \pi_i\left(t_i\right)\left[t_{-i}\right] \sum_{\tilde{t}_{-i} \in \bar{T}_{-i}} d_i^0\left(\tilde{t}_{-i}, \tilde{t}_i\right) \pi_i\left(t_i'\right)\left[\tilde{t}_{-i}\right] \\
= & \sum_{\tilde{t}_{-i} \in \bar{T}_{-i}} d_i^0\left(\tilde{t}_{-i}, \tilde{t}_i\right) \pi_i\left(t_i'\right)\left[\tilde{t}_{-i}\right],
\end{aligned}$$

where the first equality follows from (40), and the second equality follows from the fact that $d_i^0\left(\tilde{t}_{-i}, \tilde{t}_i\right) \pi_i\left(t_i'\right)\left[\tilde{t}_{-i}\right]$ does not depend on t_{-i}. Thus, we complete the proof. ■

Note that Claim 6 immediately follows from Lemma 8, which shows that for any t_i and $m_i \in W_i^{\infty}(t_i | \mathcal{M}, \bar{\mathcal{T}})$, we have $m_i^{-\bar{l}-2} \in S_i^{\infty}(\hat{\theta}_i(t_i) | \mathcal{M}^*)$.

Lemma 8 *For any $t_i \in \bar{T}_i$ and $l = 0, 1, ..., \bar{l}$, the following two statements, each of which is denoted by $P^1(l)$ and $P^2(l)$, respectively, hold:*
$P^1(l)$: for any $\hat{m}_i \in M_i$, whenever $\hat{m}_i^{-\bar{l}-2} \notin S_i^l(\hat{\theta}_i(t_i)|\mathcal{M}^*)$, then $\hat{m}_i \notin W_i^l(t_i|\mathcal{M}, \bar{\mathcal{T}})$;
$P^2(l)$: there is some $\left(m_i^{-\bar{l}-2}, ..., m_i^{-\bar{l}+l-1}\right) \in \times_{k=-\bar{l}-2}^{-\bar{l}+l-2} M_i^k$ such that for every $t_i' \in \bar{T}_i$,

$$\left(m_i^{-\bar{l}-2}, ..., m_i^{-\bar{l}+l-2}, t_i', t_i, ..., t_i\right) \in W_i^l\left(t_i|\mathcal{M}, \bar{\mathcal{T}}\right). \tag{41}$$

Proof. We prove Lemma 8 by induction. We observe that $P^1(0)$ and $P^2(0)$ hold trivially. Next, for each $l \in \{0, \ldots, \bar{l}-1\}$, we assume that $P^1(l)$ and $P^2(l)$ hold and prove $P^1(l+1)$ and $P^2(l+1)$.

We first prove $P^1(l+1)$. Consider player i of type t_i. Assume that there is $m_i^* \in M_i^*$ such that $m_i^* \in S_i^l(\hat{\theta}_i(t_i)|\mathcal{M}^*)$ and $m_i^* \notin S_i^{l+1}(\hat{\theta}_i(t_i)|\mathcal{M}^*)$.[26] This implies that there exists some $\alpha_i^* \in \Delta(M_i^*)$ such that

$$u_i(g^*(\alpha_i^*, m_{-i}^*), (\hat{\theta}_i(t_i), \theta_{-i})) > u_i(g^*(m_i^*, m_{-i}^*), (\hat{\theta}_i(t_i), \theta_{-i})), \tag{42}$$

for all $\theta_{-i} \in \Theta_{-i}$ and $m_{-i}^* \in S_{-i}^l(\theta_{-i}|\mathcal{M}^*)$.

For any $t_{-i} \in$ supp $\pi_i(t_i)$, $\hat{m}_{-i} \in M_{-i}$, and $j \neq i$, if $\hat{m}_j \in W_j^l\left(t_j|\mathcal{M}, \bar{\mathcal{T}}\right)$, then $\hat{m}_j^{-\bar{l}-2} \in S_j^l(\hat{\theta}_j(t_j)|\mathcal{M}^*)$ by $P^1(l)$. Fix $m_i \in M_i$ such that $m_i^{-\bar{l}-2} = m_i^*$. Let $\alpha_i \in \Delta(M_i)$ such that $\alpha_i^{-\bar{l}-2} = \alpha_i^*$ and $\alpha_i^k\left(m_i^k\right) = 1$ for any $k \neq -\bar{l}-2$. Thus, for any $\nu_{-i} : \bar{T}_{-i} \to M_{-i}$, whenever $\nu_{-i}(t_{-i}) \in W_{-i}^l(t_{-i}|\mathcal{M}, \bar{\mathcal{T}})$ for each $t_{-i} \in \bar{T}_{-i}$,

$$\begin{aligned}
& \sum_{t_{-i}} \left\{u_i(g(\alpha_i, \nu_{-i}(t_{-i})), \hat{\theta}(t_i, t_{-i})) + \tau_i\left(\tilde{m}_i, \nu_{-i}(t_{-i})\right)\right\} \pi_i(t_i)[t_{-i}] \\
& - \sum_{t_{-i}} \left\{u_i(g(m_i, \nu_{-i}(t_{-i})), \hat{\theta}(t_i, t_{-i})) + \tau_i(m_i, \nu_{-i}(t_{-i}))\right\} \pi_i(t_i)[t_{-i}] \qquad (43) \\
= & \sum_{t_{-i}} e(\left(m_i^{-l}, \nu_{-i}^{-l}(t_{-i})\right)_{l=0}^{\bar{l}+1}) \pi_i(t_i)[t_{-i}] \\
& \times \left\{u_i(g^*(\alpha_i, \nu_{-i}^{-\bar{l}-2}(t_{-i})), \hat{\theta}(t_i, t_{-i})) - u_i(g^*(m_i^*, \nu_{-i}^{-\bar{l}-2}(t_{-i})), \hat{\theta}(t_i, t_{-i}))\right\} \\
\geq & 0
\end{aligned}$$

where the equality follows from the fact that the only difference lies in the function g^* when α_i differs from m_i only in round $-(\bar{l}+2)$ announcement; the inequality follows from (42).

In addition, by $P^2(l)$, for any $t_{-i} \in \bar{T}_{-i}$, there exists some $\tilde{m}_{-i} \in W_{-i}^l(t_{-i}|\mathcal{M}, \mathcal{T})$ such that $\tilde{m}_{-i}^{-\bar{l}+l-1} \neq \tilde{m}_{-i}^0$. Thus, $e(\left(m^{-l}\right)_{l=0}^{\bar{l}+1}) = \epsilon$ when $m_{-i} = \tilde{m}_{-i}$. Let ν_{-i} be a conjecture such that $\nu_{-i}(t_{-i}) = \tilde{m}_{-i}$ for some t_{-i}. Against such conjecture ν_{-i}, the inequality in (43) becomes strict. Thus, m_i is weakly dominated by α_i so that $m_i \notin W_i^{l+1}(t_i|\mathcal{M}, \mathcal{T})$. So, $P^1(l+1)$ holds.

Second, we shall prove $P^2(l+1)$. In the proof, for any conjecture $\nu_{-i} : \bar{T}_{-i} \to \Delta(M_{-i})$,

[26] Throughout this section, we use m_i^* to denote a generic element in M_i^*.

we write $\nu_{-i}^{k}\left(t_{-i}\right)$ as the marginal distribution of ν_{-i} on M_{-i}^{k}. Moreover, we write the "coordinate-wise" best replies as follows:

$$\begin{aligned} b_{i}^{*}(\nu_{-i}^{-\bar{l}-2}) &\in \arg\max_{m_{i}^{*}\in M_{i}^{*}} \sum_{t_{-i}\in\bar{T}_{-i}} \sum_{\tilde{t}_{-i}\in\bar{T}_{-i}} u_{i}(g^{*}\left(m_{i}^{*},\tilde{t}_{-i}\right),\hat{\theta}\left(t_{i},t_{-i}\right))\nu_{-i}^{-\bar{l}-2}\left(\tilde{t}_{-i}|t_{-i}\right)\pi_{i}\left(t_{i}\right)\left[t_{-i}\right]; \\ b_{i}(\nu_{-i}^{-\bar{l}-2}) &\in \arg\max_{t_{i}'\in\bar{T}_{i}} \sum_{t_{-i}\in\bar{T}_{-i}} \sum_{\tilde{t}_{-i}\in\bar{T}_{-i}} \tilde{d}_{i}^{0}\left(\tilde{t}_{-i},t_{i}'\right)\nu_{-i}^{-\bar{l}-2}\left(\tilde{t}_{-i}|t_{-i}\right)\pi_{i}\left(t_{i}\right)\left[t_{-i}\right]; \qquad (44) \\ b_{i}\left(\nu_{-i}^{k}\right) &\in \arg\max_{t_{i}'\in\bar{T}_{i}} \sum_{t_{-i}\in\bar{T}_{-i}} \sum_{\tilde{t}_{-i}\in\bar{T}_{-i}} d_{i}^{0}\left(\tilde{t}_{-i},t_{i}'\right)\nu_{-i}^{k}\left(\tilde{t}_{-i}|t_{-i}\right)\pi_{i}\left(t_{i}\right)\left[t_{-i}\right] \qquad (45) \end{aligned}$$

for all $k=-\bar{l}-1,\ldots,K$. Observe that by Claims 11, 12 and 7, the best reply is unique if $\nu_{-i}^{k}\left(t_{-i}\right)$ is a point mass for every t_{-i}. By $P^{2}\left(l\right)$, there is a mapping $\widetilde{\nu}_{-i}:\bar{T}_{-i}\to\times_{k=-\bar{l}-2}^{-\bar{l}+l-2}M_{-i}^{k}$ such that for every $t_{-i},t_{-i}'\in\bar{T}_{-i}$,

$$\hat{\nu}_{-i}(t_{-i},t_{-i}')\equiv(\widetilde{\nu}_{-i}^{-\bar{l}-2}(t_{-i}),...,\widetilde{\nu}_{-i}^{-\bar{l}+l-2}(t_{-i}),t_{-i}',t_{-i},...,t_{-i})\in W_{-i}^{l}\left(t_{-i}|\mathcal{M},\bar{\mathcal{T}}\right).$$

We now prove $P^{2}\left(l+1\right)$ in the following two steps.

Step 1. *If* $t_{i}'\neq t_{i}$, *then* $\bar{m}_{i}\equiv(b_{i}^{*}(\widetilde{\nu}_{-i}^{-\bar{l}-2}),b_{i}(\widetilde{\nu}_{-i}^{-\bar{l}-2}),b_{i}(\widetilde{\nu}_{-i}^{-\bar{l}-1}),\ldots,b_{i}(\widetilde{\nu}_{-i}^{-\bar{l}+l-2}),t_{i}',t_{i},...,t_{i})\in W_{i}^{l+1}\left(t_{i}|\mathcal{M},\bar{\mathcal{T}}\right)$.

By Claim 13, there exists a mapping $\tilde{\sigma}_{-i}:\bar{T}_{-i}\to\Delta\left(\bar{T}_{-i}\right)$ such that

$$\{t_{i}'\}=\arg\max_{\tilde{t}_{i}\in\bar{T}_{i}} \sum_{t_{-i}\in\bar{T}_{-i}} \sum_{\tilde{t}_{-i}\in\bar{T}_{-i}} d_{i}^{0}\left(\tilde{t}_{-i},\tilde{t}_{i}\right)\tilde{\sigma}_{-i}\left(\tilde{t}_{-i}|t_{-i}\right)\pi_{i}\left(t_{i}\right)\left[t_{-i}\right],$$

Let $\nu_{-i}:\bar{T}_{-i}\to\Delta\left(M_{-i}\right)$ be type t_{i}'s conjecture such that

$$\nu_{-i}\left(\hat{\nu}_{-i}(t_{-i},t_{-i}')|t_{-i}\right)=\tilde{\sigma}_{-i}\left(t_{-i}'|t_{-i}\right),\forall t_{-i},t_{-i}'\in\bar{T}_{-i}.$$

Note that $b_{i}(\nu_{-i}^{-\bar{l}+l-1})=t_{i}'$ and by construction, ν_{-i} is a valid conjecture. We show that $\bar{m}_{i}$ is a strict better reply than any other $\tilde{m}_{i}$ against ν_{-i}. Fix $\tilde{m}_{i}\neq\bar{m}_{i}$. This is proved by considering the following two cases: (A) $\tilde{m}_{i}^{-\bar{l}-2}\neq\bar{m}_{i}^{-\bar{l}-2}$ and $\tilde{m}_{i}^{k}=\bar{m}_{i}^{k}$ for any $k\neq-\bar{l}-2$; (B) $\tilde{m}_{i}^{k}\neq\bar{m}_{i}^{k}$ for some $k\neq-\bar{l}-2$.

Case A: $\tilde{m}_{i}^{-\bar{l}-2}\neq\bar{m}_{i}^{-\bar{l}-2}$ and $\tilde{m}_{i}^{k}=\bar{m}_{i}^{k}$ for any $k\neq-\bar{l}-2$

Note that $e(\left(\bar{m}_{i}^{-l},\nu_{-i}^{-l}\left(t_{-i}\right)\right)_{l=0}^{\bar{l}+1})=\epsilon$ for any $t_{-i}\in\bar{T}_{-i}$ since $\bar{m}_{i}^{-\bar{l}+l}=t_{i}'\neq t_{i}=\bar{m}_{i}^{0}$. Then, the payoff difference from changing $\tilde{m}_{i}$ into $\bar{m}_{i}$ is

$$\begin{aligned} & \epsilon\sum_{t_{-i}} u_{i}(g^{*}(\bar{m}_{i}^{-\bar{l}-2},\nu_{-i}^{-\bar{l}-2}\left(t_{-i}\right)),\hat{\theta}\left(t_{i},t_{-i}\right))\pi_{i}(t_{i})[t_{-i}] \\ - & \epsilon\sum_{t_{-i}} u_{i}(g^{*}(\tilde{m}_{i}^{-\bar{l}-2},\nu_{-i}^{-\bar{l}-2}\left(t_{-i}\right)),\hat{\theta}\left(t_{i},t_{-i}\right))\pi_{i}(t_{i})[t_{-i}] \\ > & 0 \qquad (46) \end{aligned}$$

where the inequality follows because $\bar{m}_i^{-\bar{l}-2} = b_i^*(\nu_{-i}^{-\bar{l}-2})$ which is the unique best reply against $\nu_{-i}^{-\bar{l}-2}$.

Case B: $\tilde{m}_i^k \neq \bar{m}_i^k$ for some $k \neq -\bar{l}-2$.

We prove Case B by considering the following subcases.

B1: $\tilde{m}_i^k \neq \bar{m}_i^k = b_i\left(\nu_{-i}^{k-1}\right)$for some $k = -\bar{l}-1, ..., -\bar{l}+l$

By (39) and (44), we have

$$\lambda \sum_{t_{-i}\in\bar{T}_{-i}} \left[d_i^0\left(\nu_{-i}^{k-1}(t_{-i}), b_i\left(\nu_{-i}^{k-1}\right)\right) - d_i^0\left(\nu_{-i}^{k-1}\left(t_{-i}\right), \tilde{m}_i^k\right)\right] \pi_i\left(t_i\right)\left[t_{-i}\right] > \gamma.$$

Given the conjecture ν_{-i}, the gain from changing $\tilde{m}_i^k$ to $\bar{m}_i^k$ is at least γ, while the potential loss is at most ϵE. Since $\gamma > \varepsilon E + \xi + K\eta$ by (13), $\bar{m}_i$ is strictly better than $\tilde{m}_i$ against ν_{-i}.

B2: $\tilde{m}_i^k \neq \bar{m}_i^k = t_i$ for some $k = -\bar{l}+l+1, ..., 0$.

Note that by (39), given the conjecture ν_i, the gain from changing $\tilde{m}_i^k$ to $\bar{m}_i^k$ is at least γ; while the potential loss from changing $\tilde{m}_i$ into $\bar{m}_i$ is at most $\epsilon E + \xi + K\eta$. Since $\gamma > \varepsilon E + \xi + K\eta$ by (13), $\bar{m}_i$ is strictly better than $\tilde{m}_i$ against ν_{-i}.

B3: $\tilde{m}_i^k \neq \bar{m}_i^k = t_i$ for some $k = 1..., K$.

By **B1** and **B2**, it suffices to show $\bar{m}_i$ is strictly better than $\tilde{m}_i$ when $\tilde{m}_i^k = t_i$ for any $k = -\bar{l}-1, ..., 0$. Against the belief ν_{-i}, the gain from changing $\tilde{m}_i^k$ to m_i^k is at least η, while there is no loss incurred. Therefore, $\bar{m}_i$ is strictly better than $\tilde{m}_i$ against ν_{-i}.

This completes the proof for Case B. Thus, $\bar{m}_i \in W_i^{l+1}\left(t_i|\mathcal{M}, \bar{\mathcal{T}}\right)$.

Step 2. *If* $t_i' = t_i$, *then* $\bar{m}_i \equiv (b_i^*(\widetilde{\nu}_{-i}^{-\bar{l}-2}), b_i(\widetilde{\nu}_{-i}^{-\bar{l}-2}), b_i(\widetilde{\nu}_{-i}^{-\bar{l}-1}), \ldots, b_i(\widetilde{\nu}_{-i}^{-\bar{l}+l-2}), t_i', t_i, ..., t_i) \in W_i^{l+1}\left(t_i|\mathcal{M}, \bar{\mathcal{T}}\right)$.

Consider $\bar{\sigma}_{-i} : \bar{T}_{-i} \to \Delta\left(\bar{T}_{-i}\right)$ such that for any t_{-i} we have $\bar{\sigma}_{-i}\left(t_{-i}|t_{-i}\right) = 1-\varsigma$ and $\bar{\sigma}_{-i}\left(t_{-i}'|t_{-i}\right) = \varsigma$ for some $t_{-i}' \neq t_{-i}$. Since d_i^0 is a proper scoring rule, we can choose $\varsigma > 0$ sufficiently small such that

$$\{t_i\} = \arg\max_{\tilde{t}_i\in\bar{T}_i} \sum_{t_{-i}\in\bar{T}_{-i}} \sum_{\tilde{t}_{-i}\in\bar{T}_{-i}} d_i^0\left(\tilde{t}_{-i}, \tilde{t}_i\right) \bar{\sigma}_{-i}\left(\tilde{t}_{-i}|t_{-i}\right) \pi_i\left(t_i\right)\left[t_{-i}\right]$$

and meanwhile by Claim 7, given $\gamma > 0$ satisfying inequality (13), we can choose $\lambda > 0$ such that

$$\lambda \sum_{t_{-i}\in\bar{T}_{-i}} \sum_{\tilde{t}_{-i}\in\bar{T}_{-i}} \left[d_i^0\left(\tilde{t}_{-i}, t_i\right) - d_i^0\left(\tilde{t}_{-i}, t_i'\right)\right] \bar{\sigma}_{-i}\left(\tilde{t}_{-i}|t_{-i}\right) \pi_i\left(t_i\right)\left[t_{-i}\right] > \gamma. \tag{47}$$

Let $\nu_{-i} : \bar{T}_{-i} \to \Delta\left(M_{-i}\right)$ be type t_i's conjecture such that $\forall t_{-i}, t_{-i}' \in \bar{T}_{-i}$

$$\nu_{-i}\left(\hat{\nu}_{-i}(t_{-i}, t_{-i}')|t_{-i}\right) = \bar{\sigma}_{-i}\left(t_{-i}'|t_{-i}\right),$$

Note that $b_i(\nu_{-i}^{-\bar{l}+l-1}) = t_i$ and by construction, ν_{-i} is a valid conjecture. We will show that $\bar{m}_i$ is a strict better reply than any other $\tilde{m}_i$ against ν_{-i}. Fix $\tilde{m}_i \neq \bar{m}_i$. This is proved in the following two cases: (A') $\tilde{m}_i^{-\bar{l}-2} \neq \bar{m}_i^{-\bar{l}-2}$ and $\tilde{m}_i^k = \bar{m}_i^k$ for any $k \neq -\bar{l}-2$; (B') $\tilde{m}_i^k \neq \bar{m}_i^k$

for some $k \neq -\bar{l}-2$.

Case A': $\tilde{m}_i^{-\bar{l}-2} \neq \bar{m}_i^{-\bar{l}-2}$ and $\tilde{m}_i^k = \bar{m}_i^k$ for any $k \neq -\bar{l}-2$

The payoff difference from changing $\tilde{m}_i$ into $\bar{m}_i$ is

$$\begin{aligned} & e(\left(\bar{m}_i^{-l}, \nu_{-i}^{-l}\left(t_{-i}\right)\right)_{l=0}^{\bar{l}+1})\pi_i(t_i)[t_{-i}] \\ & \times \sum_{t_{-i}} \left\{ u_i(g^*\left(\bar{m}_i^{-\bar{l}-2}, \nu_{-i}^{-\bar{l}-2}\left(t_{-i}\right)\right), \hat{\theta}\left(t_i, t_{-i}\right)) - u_i(g^*\left(\tilde{m}_i^{-\bar{l}-2}, \nu_{-i}^{-\bar{l}-2}\left(t_{-i}\right)\right), \hat{\theta}\left(t_i, t_{-i}\right)) \right\} \\ \geq \; & 0, \end{aligned} \tag{48}$$

where the inequality follows because $\bar{m}_i^{-\bar{l}-2} = b_i^*\left(\nu_{-i}^{-\bar{l}-2}\right)$ which is the unique best reply against $\nu_{-i}^{-\bar{l}-2}$. Inequality (48) is strict since for any t_{-i} we have that $\nu_{-i}^{-\bar{l}+l-1}\left(t'_{-i}|t_{-i}\right) = \bar{\sigma}_{-i}\left(t'_{-i}|t_{-i}\right)$ and $\bar{\sigma}_{-i}\left(t'_{-i}|t_{-i}\right) > 0$ for some $t'_{-i} \neq t_{-i}$.

Case B': $\tilde{m}_i^k \neq \bar{m}_i^k$ for some $k \neq -\bar{l}-2$.

The proof of Case B' is identical to that of Case B in Step 1 (where in Case B'1 we use (47)). Thus, $\bar{m}_i \in W_i^{l+1}\left(t_i|\mathcal{M}, \bar{\mathcal{T}}\right)$. This completes the proof of Lemma 8. ■

References

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

——— (1992b): "Virtual implementation in iteratively undominated strategies: incomplete information," *mimeo*.

——— (1994): "Exact Implementation," *Journal of Economic Theory*, 64, 1–19.

Abreu, D., and A. Sen (1991): "Virtual implementation in Nash equilibrium," *Econometrica: Journal of the Econometric Society*, pp. 997–1021.

Bassetto, M., and C. Phelan (2008): "Tax riots," *The Review of Economic Studies*, 75(3), 649–669.

Bergemann, D., and S. Morris (2009a): "Robust implementation in direct mechanisms," *The Review of Economic Studies*, 76(4), 1175–1204.

——— (2009b): "Robust virtual implementation," *Theoretical Economics*, 4, 45–88.

Börgers, T. (1994): "Weak dominance and approximate common knowledge," *Journal of Economic Theory*, 64(1), 265–276.

Brusco, S. (1998): "Unique implementation of the full surplus extraction outcome in auctions with correlated types," *Journal of Economic Theory*, 80(2), 185–200.

Chen, Y.-C., and S. Xiong (2011): "The genericity of beliefs-determine-preferences models revisited," *Journal of Economic Theory*, 146(2), 751–761.

——— (2013): "Genericity and robustness of full surplus extraction," *Econometrica*, 81(2), 825–847.

Chung, K.-S., and J. C. Ely (2003): "Implementation with Near-Complete Information," *Econometrica*, 71(3), 857–871.

Crémer, J., and R. P. McLean (1988): "Full Extraction of the Surplus in Bayesian and Dominant Strategy Auctions," *Econometrica*, 56(6), 1247–57.

d'Aspremont, C., J. Crémer, and L.-A. Gérard-Varet (2003): "Correlation, independence, and Bayesian incentives," *Social Choice and Welfare*, 21(2), 281–310.

de Clippel, G., R. Saran, and R. Serrano (2014): "Mechanism Design with Bounded Depth of Reasoning and Small Modeling Mistakes," *working paper*.

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

Dekel, E., D. Fudenberg, and S. Morris (2006): "Topologies on types," *Theoretical Economics*, 1, 275–309.

Demski, J. S., and D. Sappington (1984): "Optimal incentive contracts with multiple agents," *Journal of Economic Theory*, 33(1), 152–171.

Dudley, R. M. (2002): *Real analysis and probability*, vol. 74. Cambridge University Press.

Dutta, B., and A. Sen (2012): "Nash implementation with partially honest individuals," *Games and Economic Behavior*, 74(1), 154–169.

Frick, M., and A. Romm (2014): "Rational Behavior under Correlated Uncertainty," *working paper*.

Heifetz, A., and Z. Neeman (2006): "On the generic (im) possibility of full surplus extraction in mechanism design," *Econometrica*, pp. 213–233.

Jackson, M. O. (1991): "Bayesian Implementation," *Econometrica*, 59(2), 461–477.

Johnson, S., J. W. Pratt, and R. J. Zeckhauser (1990): "Efficiency despite mutually payoff-relevant private information: The finite case," *Econometrica: Journal of the Econometric Society*, pp. 873–900.

Kohlberg, E., and J.-F. Mertens (1986): "On the strategic stability of equilibria," *Econometrica: Journal of the Econometric Society*, pp. 1003–1037.

Matsushima, H. (1988): "A new approach to the implementation problem," *Journal of Economic Theory*, 45(1), 128–144.

——— (1991): "Incentive compatible mechanisms with full transferability," *Journal of Economic Theory*, 54(1), 198–203.

——— (1993): "Bayesian monotonicity with side payments," *Journal of Economic Theory*, 59(1), 107–121.

——— (2008): "Role of honesty in full implementation," *Journal of Economic Theory*, 139(1), 353–359.

Osborne, M., and A. Rubinstein (1994): *A Course in Game Theory*. Cambridge, MA: MIT Press.

Oury, M. (2015): "Continuous implementation with local payoff uncertainty," *Journal of Economic Theory*, 159, 656–677.

Oury, M., and O. Tercieux (2012): "Continuous implementation," *Econometrica*, 80(4), 1605–1637.

Palfrey, T. R., and S. Srivastava (1987): "On Bayesian implementable allocations," *The Review of Economic Studies*, 54(2), 193–208.

——— (1989): "Mechanism design with incomplete information: A solution to the implementation problem," *Journal of Political Economy*, pp. 668–691.

Postlewaite, A., and D. Schmeidler (1986): "Implementation in differential information economies," *Journal of Economic Theory*, 39(1), 14–33.

Repullo, R. (1985): "Implementation in dominant strategies under complete and incomplete information," *The Review of Economic Studies*, 52(2), 223–229.

Serrano, R., and R. Vohra (2005): "A characterization of virtual Bayesian implementation," *Games and Economic Behavior*, 50(2), 312–331.

Sjöström, T. (1994): "Implementation in undominated Nash equilibria without integer games," *Games and Economic Behavior*, 6(3), 502–511.

van Damme, E. (1987): *Stability and perfection of Nash equilibria*. Springer-Verlag.

Vohra, R. (1999): "Incomplete information, incentive compatibility, and the core," *Journal of Economic Theory*, 86(1), 123–147.