Discussion Paper No, 2015-04

Implementation with Transfers*

Yi-Chun Chen† Takashi Kunimoto‡ Yifei Sun§

March 6, 2015

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 then show in private-value environments that any incentive compatible rule is implementable with small transfers. Therefore, our mechanism only needs small ex post transfers to make our implementation results completely free from the multiple equilibrium problem. In addition, our mechanism possesses the unique equilibrium that is robust to higher-order belief perturbations. We also identify a class of interdependent-value environments to which our results can be extended.

JEL Classification: C72, D78, D82.

Keywords: Full implementation, universal type space, robustness, transfers

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

*We thank Atsushi Kajii, Michihiro Kandori, Hitoshi Matsushima, Rene Saran, Satoru Takahashi, Olivier Tercieux, Takuma Wakayama and seminar participants at Bilkent University, Hitotsubashi 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

‡Department of Economics, Hitotsubashi University, 2-1 Naka, Kunitachi, Tokyo 186-8601, Japan, takashi_kunimoto@econ.hit-u.ac.jp

§Department of Economics, National University of Singapore, Singapore 117570, sunyifei@nus.edu.sg

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 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 such 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. This fundamental insight allows us to transform *any* implementation problem into the planner's problem of maximizing a given social welfare, subject to incentive compatibility-constraints. This is the standard constrained-optimization problem. Due to its tractability, this approach turns out to be powerful enough to produce many applications–in auctions, bargaining, organizational economics, monetary economics, and many others domains

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, as demonstrated by 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. 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. These proposed mechanisms originally looked promising as a way to fix the direct revelation mechanism. However, many of these mechanisms share one serious drawback: undesirable equilibria are eliminated by triggering the "integer games" in which each player announces an integer and the player who announces the highest integer gets to be a dictator. For example, Palfrey and Srivastava (1989) establish a very permissive implementation result in private-value environments: *any* incentive compatible social rule can be fully implementable in undominated Bayes Nash equilibrium. However, their mechanism also employ the integer games. Many researchers consider the integer game or any variant of it as an unrealistic device, presumably relying on the argument that the truthful equilibrium is cognitively simple and can be a strong focal point among the individuals involved; those researchers confine themselves to characterizing incentive-compatible rules. Thus, there is a clear divide between those who are content with partial implementation and those who work on full implementation; moreover, there is unfortunately little interaction between them.

The main objective of this paper is to build a bridge between partial and full implementation. 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 quasilinear utilities. We focus on 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 von Neumann-Morgenstern expected utility. Third, we focus on *private-value* environments. That is, each player's utility depends only upon his own payoff type (but not the other players' payoff types) as well as upon the lottery chosen and his monetary payment (or subsidy). Fourth, we assume that no players use *weakly dominated* actions in the game. 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. We consider eliminating weakly dominated actions as a minor qualification on the players' strategic behavior because most refinements of Nash equilibrium do not involve weakly dominated actions. 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 preparation we have made thus far, we are ready to state our main result: a social choice rule is implementable with arbitrarily small transfers if and only if it is incentive compatible (Theorem 2). This is quite consistent with the idea of partial implementation because if the planner is content with small ex post transfers, the only constraint for full implementation is incentive compatibility. However, the mechanism we employ here is *not* the direct revelation mechanism. Rather, our mechanism is based on the mechanism in Abreu and Matsushima (1994), but we extend it to an incomplete-information environment. We must also stress that our mechanism is finite and uses no devices like integer games. Recall that Palfrey and Srivastava (1989) use the integer games to show a similar permissive result. Although our mechanism, unlike Palfrey and Srivastava (1989), 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 will show that all these features of our mechanism are valuable ones, which remove it from the scope of the criticisms usually made of full 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*. Oury and Tercieux show that when sending messages in the mechanism is slightly costly, *Bayesian monotonicity*, which is a necessary condition for full implementation, becomes necessary for continuous implementation. Hence, continuous implementation can be a strong argument for full implementation.

Like Oury and Tercieux (2012), we also show that our mechanism achieves continuous implementation as long as the planner can allow for small ex post transfers (Theorems 5 and 6). Recall that we assume that no players use weakly dominated actions. In fact, this weak dominance will be highly sensitive to payoff perturbations induced by the cost of sending messages. It is for this reason that our continuous implementation result does not follow from Oury and Tercieux (2012).

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 is 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 3). NEI captures the situation in which any unilateral deception from the truth-telling in the direct revelation mechanism can be detected. Furthermore, since complete-information environments can be considered a special case of NEI, our Theorem 3 can be considered an extension of the result of Abreu and Matsushima (1994) to incomplete-information environments. 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 4). We can think of exchange economies as an example of this situation. In this environment, however, we need to strengthen incentive compatibility.

If the planner wants all equilibria of his mechanism yield a desired outcome, and entertains the possibility that players may have even the slightest uncertainty about payoffs, then the planner should insist on a solution concept with a closed graph. 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 complete information environment nearby, 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 to us. Instead, we show that any incentive-compatible social choice rule is $\overline{UNE}$-implementable with no transfers (Corollary 2).

The rest of the paper is organized as follows: In Section 2, we introduce the preliminary notation and definitions as well as two assumptions (Assumptions 1 and 2) that we maintain throughout the paper. In Section 3, we construct a mechanism and discuss some of its basic properties. Section 4 provides our main results. More specifically, we establish Theorem 1 for implementation with transfers (Section 4.1), Theorem 2 for implementation with arbitrarily small transfers (Section 4.2), and Theorems 3 and 4 for implementation with no transfers (Section 4.3). Section 5 discusses three applications of our results: we investigate the connection to continuous implementation (Section 5.1), to $\overline{UNE}$-implementation (Section 5.2), and to the full surplus extraction (Section 5.3). In Section 6, we provide some extensions of our results and also discuss the limitations of our results. In particular, we discuss the role of honesty and rationalizable implementation (Section 6.1); we identify a class of interdependent-value environments to which our permissive results can be extended (Section 6.2); we propose a way of achieving budget balance when there are at least three individuals (Section 6.3); and finally, we compare our results with those of *virtual implementation*, a process in which the planner contents himself with implementing the social choice rule with arbitrarily high probability.

2 Preliminaries

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 $\theta_i \in \Theta_i$. Denote $\Theta = \Theta_1 \times \cdots \times \Theta_I$ and $\Theta_{-i} = \Theta_1 \times \cdots \times \Theta_{i-1} \times \Theta_{i+1} \times \cdots \times \Theta_I$.[1] 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; $\hat{\theta}_i : T_i \to \Theta_i$; and $\pi_i(t_i) \in \Delta(T_{-i})$ denotes the associated 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 and supp $\pi_i(t_i)$ (i.e., the support of $\pi_i(t_i)$) is also finite for each $t_i \in T_i$, 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 measurable 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 $\theta \in \Theta$,

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

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 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) $f : \bar{T} \to \Delta(A)$.[2]

2.2 Mechanisms, Solution Concepts, and Implementation

We assume that the planner can fine or reward a player $i \in I$ by *side payments*. A *mechanism* $\mathcal{M}$ is a triplet $((M_i), g, (\tau_i))_{i\in I}$ where M_i is the nonempty countable *message space* for player i; $g : M \to \Delta(A)$ is an *outcome function*; and $\tau_i(m) : M \to \mathbb{R}$ is a *transfer rule* for player

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

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

$i \in I$. A mechanism $\mathcal{M}$ is *finite* if M_i is finite for every player $i \in I$. 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 $m \in M$. We denote such an mechanism 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 $m_i \in W_i(t_i|\mathcal{M}, \mathcal{T})$ if and only if there does not exist $m_i' \in M_i$ such that

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

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

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

for all $\nu : T_{-i} \to \Delta(M_{-i})$ and for all t_{-i} and m_{-i}, $\nu(m_{-i}|t_{-i})\pi_i(t_i)[t_{-i}] > 0$ implies that $m_{-i} \in S_{-i}^l(t_{-i}|\mathcal{M}, \mathcal{T}) = \prod_{j \neq i} S_j^l(t_j|\mathcal{M}, \mathcal{T})$. Let $S^\infty W$ denote the set of strategy profiles which survive one round of removal of weakly dominated strategies followed by iterative removal of strictly dominated strategies, i.e.,

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

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

Here we restrict attention to pure strategies, but without loss of generality. In the mechanism we construct below, we have $S^\infty W$ as a singleton; this constitutes a unique, undominated Bayesian Nash equilibrium in pure strategies. Moreover, this undominated Bayesian Nash equilibrium remains the unique equilibrium in the mechanism even when mixed strategies are allowed. Several foundations for $S^\infty W$ in normal-form games are known in the literature. We refer the reader to Börgers (1994) and Dekel and Fudenberg (1990) for its foundations 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 W^{∞} generates the same outcome as $S^{\infty}W$ in our mechanism, regardless of the order of removal of strategies, where W^{∞} denotes the set of strategies that survive the iterative removal of dominated strategy profiles. 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 6.1, we will discuss the role of S^{∞} in our mechanism.

We introduce the following definition:

Definition 1 *Fix a model $\bar{\mathcal{T}}$. We say that a mechanism $(\mathcal{M}, \bar{\tau})$ implements an SCF f in $S^{\infty}W$* ***with transfers*** *if, for any $t \in \bar{T}$ and $m \in S^{\infty}W\left(t|\mathcal{M}, \bar{\mathcal{T}}\right)$, we have $g(m) = f(t)$.*

We now formally state the definition of implementation in $S^{\infty}W$. First, we impose no conditions on the magnitude of transfers and propose the concept of implementation with transfers.

Definition 2 (Implementation with Transfers) *An SCF f is implementable in $S^{\infty}W$ with transfers if there exists a mechanism $(\mathcal{M}, \bar{\tau})$ which implements f in $S^{\infty}W$ with transfers.*

It is often unrealistic to assume that the planner can impose large transfers on the players. Hence, we only allow for arbitrarily small transfers and propose the following concept.

Definition 3 (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 exists a mechanism $(\mathcal{M}, \bar{\tau})$ which implements f in $S^{\infty}W$ with transfers.*

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 4 (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 1 *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.*

2.3 Assumptions

Throughout the paper we make two assumptions on the environments. First, we follow Abreu and Matsushima (1992a) and propose the following assumption.

Assumption 1 *An environment $\mathcal{E} = (A, (u_i)_{i \in I}, \bar{\mathcal{T}})$ satisfies Assumption 1 if the following two conditions hold:*

1. *for each $t_i \in \bar{T}_i$, $u_i(\cdot, \hat{\theta}_i(t_i))$ is not a constant function on A;*
2. *for any $t_i, t_i' \in \bar{T}_i$ with $t_i \neq t_i'$, $u_i(\cdot, \hat{\theta}_i(t_i))$ is not a positive affine transformation of $u_i(\cdot, \hat{\theta}_i(t_i'))$.*

Under Assumption 1, Abreu and Matsushima (1992a) show the following important result. Lemma 1 guarantees the existence of a function that can elicit each player's type.

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

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

We next introduce the following assumption.

Assumption 2 *An environment $\mathcal{E}$ satisfies Assumption 2 if, for all $i \in I$ and $t_i, t_i' \in \bar{T}_i$ with $t_i \neq t_i'$, $\pi_i(t_i) \neq \pi_i(t_i')$.*

Remark 2 *Since $\bar{T}$ is finite, Assumption 2 generically holds in the space of the probability distributions over $\bar{T}$. Note, however, that Assumption 2 fails to hold in the case of independent probability distributions.*

By Assumption 2, we can construct the following *scoring rule* $d_i^0 : T \to \mathbb{R}$:

Lemma 2 *Suppose that an environment $\mathcal{E}$ satisfies Assumption 2. For all $i \in I$ and $(t_i, t_{-i}) \in \bar{T}$, define*

$$d_i^0\left(t_i, t_{-i}\right) = 2\bar{\pi}_i\left(t_i\right)\left[t_{-i}\right] - \bar{\pi}_i\left(t_i\right) \cdot \bar{\pi}_i\left(t_i\right),$$

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

$$\sum_{t_{-i} \in T_{-i}} \left[d_i^0\left(t_i, t_{-i}\right) - d_i^0\left(t_i', t_{-i}\right)\right] \bar{\pi}_i\left(t_i\right)\left[t_{-i}\right] > 0. \tag{2}$$

Remark 3 *Lemma 2 guarantees the existence of a proper scoring rule in which each player will tell the truth whenever he believes that every other one tells the truth. Such a constructed scoring rule is strictly Bayesian incentive compatible. When there are more than two players, we can achieve budget balance. (See the discussion in Section 6)*

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

3 The Mechanism and its Basic Properties

3.1 The Mechanism

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 $\left(m_1^k, ..., m_{i-1}^k, \tilde{m}_i^k, m_{i+1}^k, ..., m_I\right)$.

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:

   - $-\lambda d_i^0\left(m_{-i}^{-2}, m_i^{-1}\right)$ (if $d_i^0\left(m_{-i}^{-2}, m_i^{-1}\right)$ is positive, it means player i is paid);
   - $-\lambda d_i^0(m_{-i}^{-1}, m_i^0)$ (if $d_i^0(m_{-i}^{-1}, m_i^0)$ is positive, it means player i is paid);[3]
   - 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.

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

   In total,

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

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

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

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

[3] The design of the two scoring rules is needed for establishing the order independence of W^∞ in the Appendix. The results in the main body of the paper still go through with one scoring rule.

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 $t_i, t_i' \in \bar{T}_i$ and every $i \in I$,

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

$$\sum_{t_{-i} \in \bar{T}_{-i}} \left[\lambda d_i^0 \left(t_{-i}, t_i'\right) - \lambda d_i^0 \left(t_{-i}, t_i\right)\right] \bar{\pi}_i \left(t_i\right) \left[t_{-i}\right] > \gamma; \tag{10}$$

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

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

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

where $\bar{d}_i^0$ denotes an upper bound of $d_i^0(t)$ over $t \in \bar{T}$.[4]

3.2 Basic Properties of the Mechanism

In this section, we exploit some basic properties of the mechanism constructed in the previous section. These properties play an important role in the rest of the paper.

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

Proof. We show that for any $i \in I$, $\bar{t}_i \in \bar{T}_i$, and $m_i \in M_i$, if $m_i^{-2} \neq \bar{t}_i$, then $m_i \notin S_i^1\left(\bar{t}_i | \mathcal{M}, \bar{\mathcal{T}}\right)$, i.e., m_i is weakly dominated by some m_i'. We construct m_i' as follows:

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

Fix any conjecture $\nu : \bar{T}_{-i} \to \Delta(M_{-i})$.

[4] Given any $\bar{\tau} > 0$, 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.

The difference of the expected values between m_i' and m_i for player i of type $\bar{t}_i$ is shown as follows:

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

where the first equality follows because the only difference lies in function x_i when m_i' differs from m_i only in the first announcement, (see the definition of g in (3) and the definition of τ in (6)); by (1) the last inequality is strict whenever $e\left(m^{-1}, m^0\right) = \epsilon$ for some m_{-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 $m_i \in S_i^2\left(\bar{t}_i|\mathcal{M}, \bar{\mathcal{T}}\right)$, 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 $m_i \in S_i^1(\bar{t}_i|\mathcal{M}, \bar{\mathcal{T}})$, if $m_i^0 \neq \bar{t}_i$, then $m_i \notin S_i^2(\bar{t}_i|\mathcal{M}, \bar{\mathcal{T}})$. We construct $\bar{m}_i$ as follows:

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

Then, for any conjecture $\nu : \bar{T}_{-i} \to \Delta(M_{-i})$, we have that, for each (t_{-i}, m_{-i}),

$$\nu(m_{-i}|t_{-i})\pi_i(\bar{t}_i)[t_{-i}] > 0 \Rightarrow m_{-i} \in S_{-i}^1\left(t_{-i}|\mathcal{M}, \bar{\mathcal{T}}\right).$$

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:

$$\begin{aligned}
& \sum_{t_{-i},m_{-i}} \left\{u_i(g(\bar{m}_i, m_{-i}), \bar{\theta}_i) + \tau_i\left(\bar{m}_i, m_{-i}\right)\right\} \nu(m_{-i}|t_{-i})\pi_i(\bar{t}_i)(t_{-i}) \\
& - \sum_{t_{-i},m_{-i}} \left\{u_i(g(m_i, m_{-i}), \bar{\theta}_i) + \tau_i\left(m_i, m_{-i}\right)\right\} \nu(m_{-i}|t_{-i})\pi_i(\bar{t}_i)[t_{-i}]
\end{aligned}$$

$$
\begin{aligned}
= & \sum_{t_{-i},m_{-i}} \left\{e\left(m^{-1}/\bar{m}_i, m^0\right) - e\left(m^{-1}, m^0\right)\right\} \\
& \times \left\{\frac{1}{I}\sum_{j\in I} u_i(x_j(\bar{t}_j),\bar{\theta}_i) - \frac{1}{K}\sum_{k=1}^{K} u_i(f(m^k),\bar{\theta}_i)\right\} \nu(m_{-i}|t_{-i})\pi_i(\bar{t}_i)[t_{-i}] \\
& + \sum_{t_{-i},m_{-i}} \left\{d_i^0\left(m_{-i}^{-2},\bar{t}_i\right) - d_i^0\left(m_{-i}^{-2}, m_i^{-1}\right)\right\} \nu(m_{-i}|t_{-i})\pi_i(\bar{t}_i)[t_{-i}]
\end{aligned}
$$

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)).

Note that

(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

$$
\begin{aligned}
& \sum_{t_{-i},m_{-i}} \left\{e\left(m^{-1}/\bar{m}_i, m^0\right) - e\left(m^{-1}, m^0\right)\right\} \\
& \times \left\{\frac{1}{I}\sum_{j\in I} u_i(x_j(\bar{t}_j),\bar{\theta}_i) - \frac{1}{K}\sum_{k=1}^{K} u_i(f(m^k),\bar{\theta}_i)\right\} \nu(m_{-i}|t_{-i})\pi_i(\bar{t}_i)[t_{-i}]
\end{aligned}
$$

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

$$
\sum_{t_{-i},m_{-i}} \left[\lambda d_i^0\left(m_{-i}^{-2},\bar{t}_i\right) - \lambda d_i^0\left(m_{-i}^{-2}, m_i^{-1}\right)\right] \nu(m_{-i}|t_{-i})\pi_i(\bar{t}_i)[t_{-i}].
$$

By Claim 1, we know that $m_{-i} \in S_{-i}^1\left(\bar{t}_{-i}|\mathcal{M},\bar{\mathcal{T}}\right)$ implies $m_{-i}^{-2} = \bar{t}_{-i}$. Therefore, by (10), we obtain

$$
\begin{aligned}
& \sum_{t_{-i},m_{-i}} \left[\lambda d_i^0\left(m_{-i}^{-2},\bar{t}_i\right) - \lambda d_i^0\left(m_{-i}^{-2}, m_i^{-1}\right)\right] \nu(m_{-i}|t_{-i})\pi_i(\bar{t}_i)[t_{-i}] \\
= & \sum_{\bar{t}_{-i}\in\bar{T}_{-i}} \left[\lambda d_i^0\left(\bar{t}_{-i},\bar{t}_i\right) - \lambda d_i^0\left(\bar{t}_{-i}, m_i^{-1}\right)\right] \bar{\pi}_i\left(\bar{t}_i\right)\left[\bar{t}_{-i}\right] \\
> & \gamma,
\end{aligned}
$$

where γ is chosen such 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* $m_i \in S_i^3\left(\bar{t}_i|\mathcal{M},\bar{\mathcal{T}}\right)$, *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 $m_i \notin S_i^3(\bar{t}_i|\mathcal{M},\bar{\mathcal{T}})$. We construct $\bar{m}_i$ as follows:

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

Then, for any conjecture $\nu : \bar{T}_{-i} \to \Delta(M_{-i})$, we have that, for each (t_{-i}, m_{-i}),

$$\nu(m_{-i}|t_{-i})\pi_i(\bar{t}_i)[t_{-i}] > 0 \Rightarrow m_{-i} \in S_{-i}^2\left(t_{-i}|\mathcal{M},\bar{\mathcal{T}}\right).$$

From Claim 1, we know that for any $j \in I$, if $m_j \in S_j^2\left(\bar{t}_j|\mathcal{M},\bar{\mathcal{T}}\right)$, 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:

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

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)).

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

4 Main Results

There are three subsections here. In Section 4.1, we provide a result of implementation with transfers where very large transfers are allowed. In Section 4.2, we make the size of transfers arbitrarily small and establish a characterization of implementation with arbitrarily small transfers. Here, incentive compatibility is an important condition. Finally, in Section 4.3, we propose two classes of environments in each of which we need no transfers on the equilibrium in the mechanism.

4.1 Implementation with Transfers

The following theorem shows that 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 even on the equilibrium.

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

We use the following claim to prove Theorem 1.

Claim 4 *Let $K = 1$. In the game $U\left(\mathcal{M}, \bar{\mathcal{T}}\right)$ for every $i \in I$, $\bar{t}_i \in \bar{T}_i$, if $m_i \in S_i^4\left(\bar{t}_i|\mathcal{M}, \bar{\mathcal{T}}\right)$, then $m_i^1 = \bar{t}_i$.*

Proof. Fix $i \in N$, $\bar{t}_i \in \bar{T}_i$ with $\hat{\theta}_i(\bar{t}_i) = \bar{\theta}_i$. We shall show that

$$m_i^1 \neq \bar{t}_i \Rightarrow m_i \notin S_i^4\left(\bar{t}_i|\mathcal{M}, \bar{\mathcal{T}}\right).$$

That is, we shall show that m_i is strictly dominated. Let $\tilde{m}_i$ be the dominating strategy defined as follows,

$$\tilde{m}_i = \left(m_i^{-2}, m_i^{-1}, m_i^0, \bar{t}_i\right).$$

Then, for any conjecture $\nu : \bar{T}_{-i} \to \Delta(M_{-i})$, we have that, for each (t_{-i}, m_{-i}),

$$\nu(m_{-i}|t_{-i})\pi_i(\bar{t}_i)[t_{-i}] > 0 \Rightarrow m_{-i} \in S_{-i}^3\left(t_{-i}|\mathcal{M}, \bar{\mathcal{T}}\right).$$

From Claim 3, we know that for any $j \in I$, if $m_j \in S_j^3\left(\bar{t}_j|\mathcal{M}, \bar{\mathcal{T}}\right)$, then $m_j^0 = \bar{t}_j$.

By choosing m_i rather than $\tilde{m}_i$, in terms of transfer rule, one possible loss from reporting is

$$\sum_{t_{-i}, m_{-i}} \left\{\tau_i\left(\tilde{m}_i, m_{-i}\right) - \tau_i\left(m\right)\right\} \nu(m_{-i}|t_{-i})\pi_i(\bar{t}_i)[t_{-i}] = \eta + \xi, \tag{15}$$

where player i of type $\bar{t}_i$ will get punished by η according to rule d_i^1 (by (5)) and ξ according to rule d_i (by (4)).

Note that $e\left(m^{-1}, m^0\right) = 0$. In terms of outcome function $g(\cdot)$ (defined in (3)): the possible gain from playing m_i rather than $\tilde{m}_i$ is

$$\sum_{t_{-i},m_{-i}} \left\{u_i(f(m^1), \bar{\theta}_i) - u_i(f(\bar{t}_i, m^1_{-i}), \bar{\theta}_i)\right\} \nu(m_{-i}|t_{-i})\pi_i(\bar{t}_i)[t_{-i}].$$

From (8), we also have the following inequality on the expected gain of type t_i when playing m_i rather than $\tilde{m}_i$:

$$\sum_{t_{-i},m_{-i}} \left\{u_i(f(m^1), \bar{\theta}_i) - u_i(f(\bar{t}_i, m^1_{-i}), \bar{\theta}_i)\right\} \nu(m_{-i}|t_{-i})\pi_i(\bar{t}_i)[t_{-i}] \leq D. \tag{16}$$

When $K = 1$, we know from Section 3.1 that $\xi > D$ (see (12)).[5] So, we obtain

$$\eta + \xi > D. \tag{17}$$

To sum up, we have

$$\begin{aligned}
& \sum_{t_{-i},m_{-i}} \left\{u_i(g(\tilde{m}_i, m_{-i}), \bar{\theta}_i) + \tau_i\left(\tilde{m}_i, m_{-i}\right)\right\} \nu(m_{-i}|t_{-i})\pi_i(\bar{t}_i)[t_{-i}] \\
& - \sum_{t_{-i},m_{-i}} \left\{u_i(g(m_i, m_{-i}), \bar{\theta}_i) + \tau_i\left(m_i, m_{-i}\right)\right\} \nu(m_{-i}|t_{-i})\pi_i(\bar{t}_i)[t_{-i}] \\
= & \sum_{t_{-i},m_{-i}} \left\{u_i(f(m^1_{-i}, \bar{t}_i), \bar{\theta}_i) - u_i(f(m^1), \bar{\theta}_i) + \xi + \eta\right\} \nu(m_{-i}|t_{-i})\pi_i(\bar{t}_i)[t_{-i}] \\
\geq & \sum_{t_{-i},m_{-i}} \left\{\eta + \xi - D\right\} \nu(m_{-i}|t_{-i})\pi_i(\bar{t}_i)[t_{-i}] \\
> & \ 0.
\end{aligned}$$

The first equality follows from the outcome function (3) and the transfer rule (6); the second inequality follows from (16); the last inequality follows from (17). Therefore, player i of type $\bar{t}_i$ will report $\bar{t}_i$ rather than m_i^1. ■

[5] When $K = 1$, we can appropriately choose λ, γ, ϵ, ξ, and η to satisfy those conditions on transfers and utilities in Section 3.1. This means that ξ can be a very large number. Since we now impose no restrictions on the size of transfers, by choosing $\lambda > 0$ large enough, we can choose γ arbitrarily large to satisfy $\gamma > \epsilon E + \xi + \eta$ (inequality (13)). Hence, ξ can be chosen large enough to satisfy $\xi > D$ (inequality (12)).

4.2 Implementation with Arbitrarily Small Transfers

We shall show that if an SCF f is *incentive compatible*, our mechanism can implement f in $S^\infty W$ with arbitrarily small transfers. 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 $(\bar{T}, f)$ for player i of type t_i when he announces t_i' and the other players all make truthful announcements.

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

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

We are now ready to state the main result of this section. The theorem below shows that incentive compatibility is a necessary and sufficient condition for implementation with arbitrarily small transfers.

Theorem 2 *Suppose that the environment $\mathcal{E}$ satisfies Assumptions 1 and 2. Assume $I \geq 2$. An SCF f is implementable in $S^\infty W$ with arbitrarily small transfers where $S^\infty W\left(t|\mathcal{M}, \bar{\mathcal{T}}\right)$ is a singleton if and only if f is incentive compatible.*

Remark 4 *Palfrey and Srivastava (1989) establish a very similar implementation result in their Theorem 2: any incentive compatible social choice function is fully implementable in undominated Bayes Nash equilibrium. We clarify a few differences between our result and that of Palfrey and Srivastava (1989). Although Palfrey and Srivastava (1989) do not need ex post small transfers, they use the integer games as part of their mechanism. On the other hand, although our mechanism does not use any devices such as the integer games, it exploits the power of ex post small transfers. In addition, our solution concept of $S^\infty W$ is more robust (or permissive) than undominated Bayes Nash equilibrium. Although Theorem 2 of Palfrey and Srivastava (1989) needs at least three players, our result works even for the case of two players. One common feature these two papers share is the difficulty of extending the results to interdependent-value environments. The reader is referred to both Section 6.2 of our paper and Section 4 of Palfrey and Srivastava (1989) for appreciating this difficulty.*

We use the following claim to prove the "if" part of Theorem 2.

Claim 5 *Suppose that an SCF f is incentive compatible. For each $k \geq 3, i \in I$, and $\bar{t}_i \in \bar{T}_i$, if $m_i \in S_i^k(\bar{t}_i|\mathcal{M}, \bar{\mathcal{T}})$, then $m_i^{k-3} = \bar{t}_i$.*

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

Then, we show that if $m_i \in S_i^{k+1}(\bar{t}_i|\mathcal{M}, \bar{\mathcal{T}})$, then $m_i^{k'} = \bar{t}_i$ for all $k' \leq k-2$. It suffices to prove $m_i^{k-2} = \bar{t}_i$. Suppose not, let $\tilde{m}_i$ be the dominating strategy defined as follows,

$$\tilde{m}_i \equiv \left(m_i^{-2}, ..., m_i^{k-3}, \bar{t}_i, m_i^{k-1}..., m_i^K\right).$$

We let $M_{-i}^* = \left\{m_{-i} \in M_{-i} : m_{-i}^{k-2} = m_{-i}^0\right\}$. Fix a conjecture $\nu : \bar{T}_{-i} \to \Delta(M_{-i})$. Note that, for each (t_{-i}, m_{-i}),

$$\nu(m_{-i}|t_{-i})\pi_i(\bar{t}_i)[t_{-i}] > 0 \Rightarrow m_{-i} \in S_{-i}^k\left(t_{-i}|\mathcal{M}, \bar{\mathcal{T}}\right).$$

Thus, we obtain $e\left(m^{-1}, m^0\right) = 0$.

We will show that

$$\begin{aligned} & \sum_{t_{-i}, m_{-i}} \left\{u_i(g(\tilde{m}_i, m_{-i}), \bar{\theta}_i) + \tau_i(\tilde{m}_i, m_{-i})\right\} \nu(m_{-i}|t_{-i})\pi_i(\bar{t}_i)[t_{-i}] \\ & - \sum_{t_{-i}, m_{-i}} \left\{u_i(g(m_i, m_{-i}), \bar{\theta}_i) + \tau_i(m_i, m_{-i})\right\} \nu(m_{-i}|t_{-i})\pi_i(\bar{t}_i)[t_{-i}] \\ > \; & 0. \end{aligned} \tag{18}$$

Note the left hand side of inequality is equal to

$$\begin{aligned} & \sum_{t_{-i}, m_{-i} \notin M_{-i}^*} \left\{ \begin{array}{c} \left\{u_i(g(\tilde{m}_i, m_{-i}), \bar{\theta}_i) + \tau_i(\tilde{m}_i, m_{-i})\right\} - \\ \left\{u_i(g(m_i, m_{-i}), \bar{\theta}_i) + \tau_i(m_i, m_{-i})\right\} \end{array} \right\} \nu(m_{-i}|t_{-i})\pi_i(\bar{t}_i)[t_{-i}] \\ & + \sum_{t_{-i}, m_{-i} \in M_{-i}^*} \left\{ \begin{array}{c} \left\{u_i(g(\tilde{m}_i, m_{-i}), \bar{\theta}_i) + \tau_i(\tilde{m}_i, m_{-i})\right\} - \\ \left\{u_i(g(m_i, m_{-i}), \bar{\theta}_i) + \tau_i(m_i, m_{-i})\right\} \end{array} \right\} \nu(m_{-i}|t_{-i})\pi_i(\bar{t}_i)[t_{-i}]. \end{aligned} \tag{19}$$

Step 1:

$$\sum_{t_{-i}, m_{-i} \notin M_{-i}^*} \left\{ \begin{array}{c} \left\{u_i(g(\tilde{m}_i, m_{-i}), \bar{\theta}_i) + \tau_i(\tilde{m}_i, m_{-i})\right\} - \\ \left\{u_i(g(m_i, m_{-i}), \bar{\theta}_i) + \tau_i(m_i, m_{-i})\right\} \end{array} \right\} \nu(m_{-i}|t_{-i})\pi_i(\bar{t}_i)[t_{-i}] > 0.$$

From the induction hypothesis, for every $i \in I$ and $\bar{t}_i \in \bar{T}_i$, if $m_i \in S_i^k(\bar{t}_i|\mathcal{M}, \bar{\mathcal{T}})$, then $m_i^{k'} = \bar{t}_i$ for all $k' \leq k-3$. When $m_{-i} \notin M_{-i}^*$, there exists some $j \in I\backslash\{i\}$ such that $m_j^{k-1} = m_j^0$. We compute the expected loss in terms of payments for player i of type $\bar{t}_i$ when playing m_i rather than $\tilde{m}_i$:

$$\sum_{t_{-i}, m_{-i} \notin M^*_{-i}} \left\{ \tau_i\left(\tilde{m}_i, m_{-i}\right) - \tau_i\left(m\right) \right\} \nu(m_{-i}|t_{-i}) \pi_i(\bar{t}_i)[t_{-i}]$$

By choosing $\tilde{m}_i$ rather than m_i, player i will avoid the fine, η according to rule d_i^{k-2} (see (5) in Section 3.1) and ξ according to rule d_i (see (4)), that is,

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

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

$$\sum_{t_{-i}, m_{-i} \notin M^*_{-i}} \frac{1}{K} \left\{ u_i(f(m^{k-1}), \bar{\theta}_i) - u_i(f(\tilde{m}_i^{k-1}, m_{-i}^{k-1}), \bar{\theta}_i) \right\} \nu(m_{-i}|t_{-i}) \pi_i(\bar{t}_i)[t_{-i}] \leq \frac{1}{K} D. \quad (20)$$

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 3.1), we have

$$\eta + \xi > \frac{1}{K} D. \quad (21)$$

This completes Step 1.

Step 2:

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

When $m_{-i} \in M^*_{-i}$, 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 $k' \leq k-3$. We compute the expected loss in terms of payments for player i of type $\bar{t}_i$ when playing m_i rather than $\tilde{m}_i$:

$$\sum_{t_{-i}, m_{-i} \in M^*_{-i}} \left\{ \tau_i\left(\tilde{m}_i, m_{-i}\right) - \tau_i\left(m\right) \right\} \nu(m_{-i}|t_{-i}) \pi_i(\bar{t}_i)[t_{-i}]$$

By choosing $\tilde{m}_i$ rather than m_i, player i will avoid the fine, η according to rule d_i^{k-2} (see (5) in Section 3.1), 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 3.1) is

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

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

$$\sum_{t_{-i}} \{\tau_i(\tilde{m}_i, m_{-i}) - \tau_i(m)\} \pi_i(\bar{t}_i)[t_{-i}] \geq \eta.$$

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

$$\frac{1}{K}\sum_{m_{-i}} \left\{u_i(f(m^{k-2}), \bar{\theta}_i) - u_i(f(m^{k-2}/\tilde{m}_i), \bar{\theta}_i)\right\} \pi_i(\bar{t}_i)[t_{-i}],$$

Since $\tilde{m}_i$ differs from m_i only in the $(k-2)$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 2 is proved as follows.

Proof. Fix $\bar{\tau} > 0$ arbitrarily small. Given $f : \bar{T} \to \Delta(A)$ is implementable in $S^\infty W$ with arbitrarily small transfers by a mechanism $(\mathcal{M}, \bar{\tau})$, then for any $t \in \bar{T}$ and $m \in S^\infty W\left(t|\mathcal{M}, \bar{\mathcal{T}}\right)$, we have $g(m) = f(t)$ and $\tau(m) < \bar{\tau}$. Since $S^\infty W\left(t|\mathcal{M}, \bar{\mathcal{T}}\right)$ is a singleton, we know that $S^\infty W$ is a pure Bayesian Nash Equilibrium in $U(\mathcal{M}, \bar{\tau}, \bar{\mathcal{T}})$. Then, we have for all $m_i' \in M_i$,

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

Let $(\bar{T}, f)$ be a direct revelation mechanism. Then truth telling must be a Bayesian nash equilibrium. That is, for any $t_i, t_i' \in \bar{T}_i$,

$$\begin{aligned} & \sum_{t'_{-i}} \pi_i(t_i)[t'_{-i}] \left\{u_i(f(t_i, t'_{-i}), \hat{\theta}_i(t_i)) + \tau_i(t_i, t'_{-i})\right\} \\ \geq & \sum_{t'_{-i}} \pi_i(t_i)[t'_{-i}] \left\{u_i(f(t'_i, t'_{-i}), \hat{\theta}_i(t_i)) + \tau_i(t'_i, t'_{-i})\right\} \end{aligned} \tag{22}$$

Note that (22) holds for any $\tau(\sigma(t)) < \bar{\tau}$. Since $\bar{\tau}$ can be arbitrarily close to 0, we must have

$$\sum_{t'_{-i}} \pi_i(t_i)[t'_{-i}] u_i(f(t_i, t'_{-i}), \hat{\theta}_i(t_i)) \geq \sum_{t'_{-i}} \pi_i(t_i)[t'_{-i}] u_i(f(t'_i, t'_{-i}), \hat{\theta}_i(t_i)) \tag{23}$$

That is, f is incentive compatible. ■

4.3 Implementation with No Transfer

In Theorem 2, 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. We will show that under some condition, the ex post payment is not required on the equilibrium.

4.3.1 Non-Exclusive Information (NEI)

Recall the following definition: an SCF $f : \bar{T} \to \Delta(A)$ is implementable in $S^\infty W$ with *no transfers* if it is implementable in $S^\infty W$ with arbitrarily small transfers by a mechanism $(\mathcal{M}, \bar{\tau})$ such that for any $t \in \bar{T}$ and $m \in S^\infty W(t|\mathcal{M}, \bar{\mathcal{T}})$, $\tau_i(m) = 0$ for each $i \in I$. 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 is first proposed by Postlewaite and Schmeidler (1986). We provide a version of its definition as follows:

Definition 6 *The environment $\mathcal{E}$ satisfies the* ***non-exclusive information structure (NEI)*** *if, for each $\bar{t} \in \bar{T}$, $i, j \in I$, and $t_j \in \bar{T}_j$,*

$$\bar{\pi}_i(\bar{t}_i)[t_j, \bar{t}_{-ij}] = \begin{cases} 1 & \textit{if } t_j = \bar{t}_j \\ 0 & \textit{otherwise} \end{cases}$$

where $\bar{t}_{-ij}$ denotes a type profile that is obtained from $\bar{t}$ after eliminating $\bar{t}_i$ and $\bar{t}_j$.

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 3 *Suppose that the environment $\mathcal{E}$ satisfies Assumptions 1 and NEI. Assume $I \geq 2$. Any incentive compatible SCF is implementable in $S^\infty W$ with no transfers.*

Proof. The mechanism is identical to the mechanism in Section 3.1 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 2. ■

4.3.2 Strict Incentive Compatibility and Separability

Following Sjöström (1994), we introduce the following class of environments. 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 $A = A_1 \times A_2 \times ... \times A_I$. 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 7 *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 $t_i \neq t_i'$,*

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

In the theorem below, we can drop Assumption 2 but instead, we need to strengthen incentive compatibility into strict incentive compatibility.

Theorem 4 *Suppose that a separable environment $\mathcal{E}$ satisfies Assumptions 1. Assume $I \geq 2$. 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. We can drop the assumption on information structure, that is, players' information can be independent.

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 $\left(m_1^k, ..., m_{i-1}^k, \tilde{m}_i^k, m_{i+1}^k, ..., m_I\right)$.

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.

$$\tau_i\left(m_i^0, m_i^1\right) = \begin{cases} \eta & \text{if } m_i^1 \neq m_i^0; \\ 0 & \text{otherwise.} \end{cases} \tag{24}$$

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

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

$$\bar{\tau}_i > \eta; \tag{25}$$

$$\tilde{\lambda}_q \sum_{t_{-i} \in \bar{T}_{-i}} \left[u_i(f_i(t_i, t_{-i}), \hat{\theta}_i(t_i)) - u_i(f_i(t_i', t_{-i}), \hat{\theta}_i(t_i))\right] \bar{\pi}_i(t_i)\left[t_{-i}\right] > \gamma, \text{ for } q \in \{1, 2\}; \tag{26}$$

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

   and

$$\gamma > \epsilon E + (1 - \tilde{\lambda}_1 - \tilde{\lambda}_2) D + \eta. \tag{28}$$

   Since f is strictly incentive compatible, the existence of γ is guaranteed in (26).

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 $\bar{\tau}$, we can choose $(1-\tilde{\lambda}_1-\tilde{\lambda}_2)$ small enough to make the weight of the* 1*st round announcement small enough. Therefore, η can be chosen small enough to prevent the deviation in the* 1*st round.*

Remark 6 *We omit the proof of Theorem 4 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* -1*th and* 0*th 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* -1*th (resp.* 0*th) round report and the other players'* -2*th (resp.* -1*th) round report. By the strict incentive compatibility, each player will announce truthfully in the* -1*th (resp.* 0*th) round(given the truth telling in the* -2*th (resp.* -1*th) 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..*

5 Applications

We now discuss the applications of our results. First, we connect our results to *continuous* implementation, a concept proposed by Oury and Tercieux (2012). In Section 5.1, we show that any incentive-compatible SCF is continuously implementable with arbitrarily small transfers. Second, we discuss robust undominated Nash implementation, which Chung and Ely (2003) call $\overline{UNE}$-implementation. Chung and Ely show that when $\overline{UNE}$-implementation is defined to be robust to perturbations accommodating interdependent values, Maskin monotonicity is a necessary condition. In contrast, when we require $\overline{UNE}$-implementation to be robust only to private-value perturbations, we establish a very permissive result. That is, as long as we allow for a tiny number of transfers out of equilibrium, any incentive-compatible SCF is shown to be $\overline{UNE}$-implementable. Finally, with ex post small transfers, we obtain a full implementation result of the full surplus extraction in auctions environments.

5.1 Continuous Implementation

The mechanism design literature often deals with environments in which monetary payments are available, and they are content to limit their analyses 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-values 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 (Matsushima (1991)) or infinite strategy space (In the Bayesian environments with side payments, the set of allocation rules is infinite in 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 any measurable function $\sigma_i : T_i \to \Delta(M_i)$. We follow Oury and Tercieux (2012) to write down the following definitions. We define

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

Definition 8 *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 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 9 *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* $U\left(\mathcal{M}, \bar{\tau}, \mathcal{T}\right)$ ***(strictly) continuously implements*** $f : \bar{T} \rightarrow \Delta(A)$ *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* $t[n] \rightarrow_p \bar{t}$*, whenever* $t[n] \in T$ *for each* n*, we have* $(g \circ \sigma)(t[n]) \rightarrow f(\bar{t})$.

We introduce two variants of continuous implementation:

Definition 10 *An SCF* $f : \bar{T} \rightarrow \Delta\left(A\right)$ *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 11 *An SCF* $f : \bar{T} \rightarrow \Delta\left(A\right)$ *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* $U\left(\mathcal{M}, \bar{\tau}, \mathcal{T}\right)$ *that continuously implements* f.

First, we establish the following important lemma.

Lemma 3 *Fix any model* $\mathcal{T}$ *such that* $\bar{\mathcal{T}} \subset \mathcal{T}$*. There exists a finite mechanism* $\mathcal{M}$*. For any* $\bar{t} \in \bar{T}$ *and any sequence* $\{t\,[n]\}_{n=0}^{\infty}$ *in* T*, if* $t\,[n] \rightarrow_p \bar{t}$*, then, for each* n *large enough, we have* $S^{\infty}W(t[n]|\mathcal{M}, \mathcal{T}) \subset S^{\infty}W(\bar{t}|\mathcal{M}, \mathcal{T})$.

Let $\mathcal{M}$ be any one of the mechanisms used in Section 4. The proof of Lemma 3 builds upon the following claims.

Claim 6 *Fix any model* $\mathcal{T}$ *such that* $\bar{\mathcal{T}} \subset \mathcal{T}$*. For any* $\bar{t} \in \bar{T}$ *and any sequence* $\{t[n]\}_{n=0}^{\infty}$ *such that* $t\,[n] \rightarrow_p \bar{t}$*, there exists* $N_1 \in \mathbb{N}$ *such that for any* $n \geq N_1$*, we have if* $m_i \in W_i^1\left(t_i\,[n]\,|\mathcal{M}, \mathcal{T}\right)$*, then* $m_i^{-2} = \bar{t}_i$.

Proof. Fix $\bar{t} \in \bar{T}$. Let $\{t[n]\}_{n=0}^{\infty}$ be such that $t\,[n] \rightarrow_p \bar{t}$. There exists a natural number $N_1 \in \mathbb{N}$ such that for each $n > N_1$, we have $\hat{\theta}_i\left(t_i\,[n]\right) = \hat{\theta}_i(\bar{t}_i) = \bar{\theta}_i$ for some $\bar{\theta}_i \in \Theta_i$.

This is due to the fact that Θ_i is finite and endowed with the discrete topology. It follows immediately from Claim 1 that if $m_i^{-2} \neq \bar{t}_i$, then $m_i \notin W_i^1(t_i[n]|\mathcal{M}, \mathcal{T})$. ■

Fix a mechanism $(\mathcal{M}, \bar{\tau})$ and a type space $\bar{\mathcal{T}}$. For any $\bar{t} \in \bar{T}$, we define a new iteration process. We say that $m_i \in \tilde{W}_i(\bar{t}_i|\mathcal{M}, \bar{\mathcal{T}})$ if and only if $m_i^{-2} = \bar{t}_i$. We set $S_i^1(\bar{t}_i|\mathcal{M}, \bar{\mathcal{T}}) = \tilde{W}_i(\bar{t}_i|\mathcal{M}, \bar{\mathcal{T}})$. $S_i^{l+1}(\bar{t}_i|\mathcal{M}, \bar{\mathcal{T}})$ is defined in the same way as in Section 2.2 for all $l \geq 1$.

$$S_i^{\infty}\tilde{W}_i(\bar{t}_i|\mathcal{M}, \bar{\mathcal{T}}) = \bigcap_{l=1}^{\infty} S_i^l(\bar{t}_i|\mathcal{M}, \bar{\mathcal{T}}),$$

$$S^{\infty}\tilde{W}(\bar{t}_i|\mathcal{M}, \bar{\mathcal{T}}) = \prod\nolimits_{i\in I} S_i^{\infty}\tilde{W}_i(\bar{t}_i|\mathcal{M}, \bar{\mathcal{T}}).$$

Fix any model $\mathcal{T}$ such that $\bar{\mathcal{T}} \subset \mathcal{T}$, and a finite mechanism $\mathcal{M}$, for any $\bar{t} \in \bar{T}$ and any sequence $\{t[n]\}_{n=0}^{\infty}$ in T, if $t[n] \to_p \bar{t}$, then, for any $n > N_1$, $S^{\infty}W(t[n]|\mathcal{M}, \mathcal{T}) \subset S^{\infty}\tilde{W}(t[n]|\mathcal{M}, \bar{\mathcal{T}})$ by Claim 6.

Claim 7 *Fix any model $\mathcal{T}$ such that $\bar{\mathcal{T}} \subset \mathcal{T}$, there exists a finite mechanism $\mathcal{M}$. For any $\bar{t} \in \bar{T}$ and any sequence $\{t[n]\}_{n=0}^{\infty}$ in T, if $t[n] \to_p \bar{t}$, then, for each n large enough, we have $S^{\infty}W(t[n]|\mathcal{M}, \mathcal{T}) \subset S^{\infty}W(\bar{t}|\mathcal{M}, \mathcal{T})$.*

Proof. From Claim 2, 3, and 5 in Section 3, we know that for any $\bar{t} \in \bar{T}$, $S^{\infty}\tilde{W}(\bar{t}|\mathcal{M}, \bar{\mathcal{T}}) = \{(\bar{t}, ..., \bar{t})\}$. Therefore, $S^{\infty}W(\bar{t}|\mathcal{M}, \bar{\mathcal{T}}) = S^{\infty}\tilde{W}(\bar{t}|\mathcal{M}, \bar{\mathcal{T}})$. So it suffices to show for each n large enough, $S^{\infty}W(t[n]|\mathcal{M}, \mathcal{T}) \subset S^{\infty}\tilde{W}(t[n]|\mathcal{M}, \mathcal{T})$.

That is equivalent to show that for 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 $S^k(t[n]|\mathcal{M}, \mathcal{T}) \subset S^k(\bar{t}|\mathcal{M}, \mathcal{T})$, for all k. We prove this by induction. From Claim 6, we know that for any large enough n, $\hat{\theta}_i(t_i[n]) = \hat{\theta}_i(\bar{t}_i) = \bar{\theta}_i$ for some $\bar{\theta}_i \in \Theta_i$. We fix such large n. By definition, $m_i \in \tilde{W}_i(t_i[n]|\mathcal{M}, \mathcal{T})$ then $m_i^{-2} = \bar{t}_i$. Thus, $S^1(t[n]|\mathcal{M}, \mathcal{T}) \subset \tilde{W}^1(\bar{t}|\mathcal{M}, \mathcal{T})$. Suppose the claim is true for any $k > 1$. We then show that it is also valid for $k+1$.

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

$$\begin{aligned} &\sum_{t_{-i},m_{-i}} \left[u_i(g(m_i, m_{-i}), \bar{\theta}_i) + \tau_i(m_i, m_{-i})\right] \nu^{[n]}(m_{-i}|t_{-i})\pi_i(t_i[n])[t_{-i}] \\ \geq &\sum_{t_{-i},m_{-i}} \left[u_i(g(m_i', m_{-i}), \bar{\theta}_i) + \tau_i(m_i', m_{-i})\right] \nu^{[n]}(m_{-i}|t_{-i})\pi_i(t_i[n])[t_{-i}], \end{aligned} \tag{29}$$

where $\nu^{[n]}(m_{-i}|t_{-i})\pi_i(t_i)[t_{-i}] > 0$ implies that $m_{-i} \in S_{-i}^k(t_{-i}[n]|\mathcal{M}, \mathcal{T})$. Let

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

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},

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

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 become crucial. Since $t[n] \to_p \bar{t}$,

$$\pi_i(t_i[n])\left[(\bar{t}_{-i})^{\varepsilon_n}\right] \to \pi_i(\bar{t}_i)[\bar{t}_{-i}]$$

as $n \to \infty$ where $\varepsilon_n > 0$ and $(\bar{t}_{-i})^{\varepsilon_n}$ 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 $\bar{t}_{-i}$.[6] 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 $\nu : \bar{T}_{-i} \to \Delta(M_{-i})$ for type $\bar{t}_i$. For any $(\bar{t}_{-i}, m_{-i})$, we set $\nu(m_{-i}|\bar{t}_{-i}) = \beta_{-i}(\bar{t}_{-i})[m_{-i}]$. From the inequality above we have

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

[6]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).

Therefore,

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

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

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

Theorem 5 *Suppose that the environment $\mathcal{E}$ satisfies Assumptions 1 and 2. Assume $I\geq 2$.* ***Any*** *SCF f is continuously implementable with transfers.*

Proof. We employ the mechanism $(\mathcal{M},\bar{\tau})$ constructed in Section 2.1 and let $K=1$. Therefore, for all $\bar{t}\in\bar{T}$, $m\in S^\infty W\left(\bar{t}|\mathcal{M},\bar{\mathcal{T}}\right)\Rightarrow g\left(m\right)=f\left(\bar{t}\right)$. Note that $S^\infty W\left(\bar{t}|\mathcal{M},\bar{\mathcal{T}}\right)=\left\{\left(\bar{t},...,\bar{t}\right)\right\}$. We write σ^* such that $\sigma_i^*\left(\bar{t}_i\right)=\left(\bar{t}_i,...,\bar{t}_i\right)$ 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[7] 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 $\sigma_i\left(\bar{t}_i\right)=\sigma_i^*\left(\bar{t}_i\right)$ for each $\bar{t}_i\in\bar{T}_i$. Because M is finite and T is countable, standard arguments (see footnote 1 of online appendix of Oury and Tercieux (2012)) show that there exists a Bayes Nash equilibrium in $U(\mathcal{M},\bar{\tau},\mathcal{T})$, which is denoted by σ. Thus, σ is a Bayes Nash equilibrium in $U(\mathcal{M},\bar{\tau},\mathcal{T})$ and $\sigma_{|\bar{T}}$ is a Bayes Nash equilibrium in $U\left(\mathcal{M},\bar{\tau},\bar{\mathcal{T}}\right)$. Now, pick any sequence $\{t\left[n\right]\}_{n=0}^\infty$ such that $t\left[n\right]\to_p\bar{t}$. It is clear that, for each n : $\mathrm{Supp}(\sigma\left(t\left[n\right]\right))\subset S^\infty W\left(t\left[n\right]|\mathcal{M},\mathcal{T}\right)$. In addition, for n large enough, we know by Lemma 3 that $S^\infty W\left(t\left[n\right]|\mathcal{M},\mathcal{T}\right)\subset S^\infty W\left(\bar{t}|\mathcal{M},\bar{\mathcal{T}}\right)$ and so, $(g\circ\sigma)(t[n])=f(\bar{t})$ as claimed. ■

It is often unrealistic to assume that the mechanism can induce very large transfers even out of equilibrium. Therefore, we obtain the following characterization of continuous implementation with arbitrarily small transfers.

[7] We follow Osborne and Rubinstein (1994) and provide a version in our context. A profile of strategies $\sigma=(\sigma_1,...,\sigma_I)$ is a trembling hand perfect equilibrium in $U\left(\mathcal{M},\bar{\tau},\mathcal{T}\right)$ if, for each $i\in I$ and each $t_i\in T_i$, there exists a sequence $\left(\sigma^k\right)_{k=0}^\infty$ of completely mixed strategy profiles that converges to σ such that, $m_i\in\mathrm{supp}(\sigma_i\left(t_i\right))\Rightarrow m_i\in\mathrm{argmax}_{m_i'\in M_i}V_i\left(\left(m_i',\sigma_{-i}^k\right),t_i\right)$, for every k.

Theorem 6 *Suppose that the environment $\mathcal{E}$ satisfies Assumptions 1 and 2. Assume $I \geq 2$. An SCF f is continuously implementable with arbitrarily small transfers if and only if f is incentive compatible.*

Proof. For any $\bar{\tau} > 0$, we employ the mechanism $(\mathcal{M}, \bar{\tau})$ constructed in Section 2.1. The proof for "if" part is parallel to the proof of 5.

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 2, we conclude that f is incentive compatible. ■

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. To drop this "strictness," they assume instead that sending messages in the mechanism is slightly costly. Recall that our mechanism exploits the weak dominance in round -2 announcement. This weak dominance will be highly sensitive to payoff perturbations that are induced by the cost of sending messages. Therefore, Oury and Tercieux's argument cannot apply here; as a result the relation between Bayesian monotonicity and continuous implementation disappears. However, as long as we allow for ex post small transfers and consider private-values environments, we obtain yet another result that permits continuous implementation and our result is as permissive as it can be. Oury and Tercieux's result also holds in any interdependent-value environments, while our result can be extended to a particular class of interdependent-value environments (see the discussion in Section 6.2).

5.2 $\overline{UNE}$ Implementation

Chung and Ely (2003) contemplate the following situation: if a planner wants all equilibria of his mechanism yield a desired outcome, 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 with a closed graph. Chung and Ely then adopt undominated Nash equilibrium 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[8] in which each type may be uncertain about the preferences of other players but always knows his own preferences. 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. In fact, we provide an affirmative answer to Chung and Ely's question. That is, our robustness argument can be adapted to prove that the mechanism provided in Abreu and Matsushima (1994) actually achieves $\overline{UNE}$ implementation. Thus, if we consider private-value environments and allow for small ex post transfers, we provide a permissive result for $\overline{UNE}$-implementation.

Following Chung and Ely (2003), we now rephrase their definition of $\overline{UNE}$-implementation.

Definition 12 *Fix a mechanism $(\mathcal{M}, \bar{\tau})$ and a complete-information model $\bar{\mathcal{T}}$. We say that $(\mathcal{M}, \bar{\tau})$ $\overline{UNE}$-implements $f : \bar{T} \to \Delta(A)$ if the following two 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 $g(\sigma^n(t[n])) \to f(\bar{t})$.*

Note that any complete-information model is a special case of an incomplete-information model. By Theorem 5, we record the following result:

Corollary 1 *Suppose that the environment $\mathcal{E}$ satisfies Assumptions 1 and 2. Assume $I \geq 2$.* ***Any*** *SCF f is $\overline{UNE}$-implementable with transfers.*

More importantly, we obtain the following permissive result:

Corollary 2 *Suppose that the environment $\mathcal{E}$ satisfies Assumptions 1 and 2. Assume $I \geq 2$. Any incentive-compatible SCF f is $\overline{UNE}$-implementable with no transfers.*

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. 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*

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

information so that we can make any SCF be incentive-compatible. So, when $I \geq 3$, we can drop incentive compatibility completely from Corollary 2. 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.

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 3. The rest of the proof is completed by Theorem 6. ■

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 players are not very sure about their own payoff type in the case of private values. In the present paper, we assume private values and it is also possible to extend our continuous implementation result to a particular class of interdependent-value environments. In Section 6.2 below, we elaborate more on the difficulty of extending our results to general interdependent-value environments.

5.3 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 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.[9] So, by Theorem 2, we obtain the following:

Corollary 3 *Suppose that the environment $\mathcal{E}$ satisfies Assumption 1 and 2. Assume $I \geq 2$. 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.

6 Discussion

Throughout our argument, the dominance is always strict except in round -2. In Section 6.1, we introduce the concept of partial honesty and propose a way of making the dominance in round -2 "strict." This allows us to connect our results to *rationalizable* implementation. In Section 6.2, we provide a sufficient condition for our results in interdependent-value environments.

6.1 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 3. First, recall that each player i's preferences are given by $u_i : \Delta(A) \times \Theta_i \to \mathbb{R}$. Following the setup of Dutta and

[9] 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 3 therefore strengthens only their Theorem 2, while the results in Brusco (1998) and Chen and Xiong (2013) apply to their Theorem 1 as well.

Sen (2012), we extend this $u_i(\cdot)$ to $v_i : M \times \Theta_i \to \mathbb{R}$ satisfying the following two properties: for all $\bar{\mathcal{T}} = (\bar{T}_i, \hat{\theta}_i, \pi_i)_{i \in I},\ i \in I,\ t = (t_i, t_{-i}) \in \bar{T}$, $m_i, \tilde{m}_i, \in M_i$, and $m_{-i} \in M_{-i}$:

1. If $u_i(g(m_i, m_{-i}), \hat{\theta}_i(t_i)) \geq u_i(g(\tilde{m}_i, m_{-i}), \hat{\theta}_i(t_i))$, $m_i^{-1} = t_i$, and $\tilde{m}_i^{-1} \neq t_i$, then

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

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

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

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 any message containing truth-telling in round -2 *strictly* dominates any other message containing non-truth telling in round -2. Hence, given partial honesty, every dominance becomes *strict* in our mechanism. This means that we can improve upon our previous results by replacing $S^\infty W$ 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.[10] By Claim 7, we know that this rationalizability correspondence is upper hemi-continuous. Hence, we obtain the following result:

Proposition 2 *Suppose that the environment $\mathcal{E}$ satisfies Assumptions 1 and 2. Assume $I \geq 2$. Assume further that all agents are partially honest. Then, 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.*

Proof. We simply combine all the arguments we made above for Theorems 2 and 5. This completes the proof. ■

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, both of these are critical for our rationalizable implementation result. For any SCF f, we denote by f^τ the augmentation of f by ex post transfers τ. We interpret f^τ as an

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

SCF that is very close to f. We show that when all players are partially honest and an SCF f is incentive compatible, then f^τ is implementable in rationalizable strategies by a finite mechanism. Kunimoto and Serrano (2014) show that if an SCF is implementable in rationalizable strategies by a finite mechanism, it satisfies interim rationalizable monotonicity. Combining these results, we conclude that when all agents are partially honest, for any incentive compatible SCF f, one can find a nearby SCF f^τ such that f^τ is implementable in rationalizable strategies by a finite mechanism if and only if it satisfies interim rationalizable monotonicity.

Since interim rationalizable monotonicity implies Bayesian monotonicity (see Kunimoto and Serrano (2014)), as long as all agents are partially honest and the planner can allow a tiny number of ex post transfers in designing the mechanism, Bayesian monotonicity or any version of monotonicity condition can be fully dispensed with for continuous implementation. However, this argument applies only to private-value environments. In the next subsection, we discuss to which extend we can extend our results to interdependent-value environments.

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. 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 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.

6.2 Private Values vs. Interdependent Values

We now deal with the case of interdependent-value environments in which each player i's utility function is defined as $u_i : A \times \Theta \to \mathbb{R}$. This section is organized as follows: we first provide a class of interdependent-value environments to which all our results in private-value environments can be extended. Such an environment is said to satisfy *Condition (S)*. Second, we elaborate on the implications of Condition (S). Finally, we show by example that our mechanism fails to work when Condition (S) is violated. We thus conclude that we need a completely different mechanism if we want to deal with more general interdependent-value environments.

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 $t_{-i} \in \bar{T}_{-i}$,

$$u_i(x_i(t_i), (\hat{\theta}_i(t_i), \hat{\theta}_{-i}(t_{-i}))) - u_i(x_i(t_i'), (\hat{\theta}_i(t_i), \hat{\theta}_{-i}(t_{-i}))) > \zeta. \tag{30}$$

Although we can extend all our results to interdependent-value environments satisfying Condition (S), we restrict our discussion here to the extension of Theorem 2.[11]

Proposition 3 *Suppose that the environment $\mathcal{E}$ satisfies Condition (S) and Assumption 2. Assume $I \geq 2$. An SCF f is implementable in $S^\infty W$ with arbitrarily small transfers where $S^\infty W(t|\mathcal{M}, \bar{\mathcal{T}})$ is a singleton for each $t \in \bar{T}$ if and only if it is incentive compatible.*

Proof. We only focus on the if-part of Theorem 2. From the proof of the Theorem 2, we observe that the proof of Claim 1 exploits the private-value assumption, while Claim 2, 3, and 5 hold even in interdependent-value environments. Therefore, it suffices to show that Claim 1 still holds here.

In this class of interdependent-value environments, $\left\{u_i(x_i(\bar{t}_i), \bar{\theta}_i) - u_i(x_i(m_i^{-2}), \bar{\theta}_i)\right\}$ in (14) is replaced by

$$u_i(x_i(t_i), (\hat{\theta}_i(t_i), \hat{\theta}_{-i}(t_{-i}))) - u_i(x_i(t_i'), (\hat{\theta}_i(t_i), \hat{\theta}_{-i}(t_{-i}))).$$

By inequality (30), the last inequality in (14) is strict whenever $e(m^{-1}, m^0) = \epsilon$ for some m_{-i}. This completes the proof. ■

To illustrate the strength of Condition (S), we use the concept of *type diversity*, which is introduced by Serrano and Vohra (2005). Type diversity is a natural counterpart of Assumption 1 in interdependent-value environments.

To define type diversity, I need to introduce some notation. Let A be a finite set of alternatives. For each $a \in A$ and $i \in I$, define $u_i^a(t_i)$ to be the interim utility of player i of type $t_i \in \bar{T}_i$ for a constant lottery which assigns a in each state, i.e.,

$$u_i^a(t_i) = \sum_\theta u_i(a, \theta) h_i^1(t_i)[\theta].$$

Let $u_i^A(t_i) = (u_i^a(t_i))_{a \in A}$

[11]This restriction is justified because one can easily see that all other results of our paper crucially rely on the validity of Theorem 2. Note also that Theorem 1 can be seen as a special case of Theorem 2.

Assumption 3 *The environment $\mathcal{E}$ satisfies **type diversity** if the following two properties hold*[12]*:*

1. *there does not exist $i \in I$, and $t_i, t_i' \in \bar{T}_i$ with $t_i \neq t_i'$ such that*

$$u_i^A(t_i) = \alpha u_i^A(t_i') + \beta$$

for some $\alpha > 0$ and $\beta \in \mathbb{R}$.

2. *for every $i \in I$ and $t_i \in \bar{T}_i$, there exist $a, a' \in A$ such that*

$$u_i^a(t_i) \neq u_i^{a'}(t_i).$$

Serrano and Vohra (2005) establish the following lemma, which can be considered an extension of Lemma 1 of the current paper.

Lemma 4 *(Serrano and Vohra (2005)) Suppose that the environment $\mathcal{E}$ satisfies Assumption 3. Then, for each $i \in I$, there exists a function $x_i : \bar{T}_i \to \Delta(A)$ such that for all $t_i, t_i' \in \bar{T}_i$ with $t_i \neq t_i'$,*

$$\sum_\theta u_i(x_i(t_i), \theta) h_i^1(t_i)[\theta] > \sum_\theta u_i(x_i(t_i'), \theta) h_i^1(t_i)[\theta], \tag{31}$$

where $h_i^1(t_i) \in \Delta(\Theta)$ denotes the first-order belief of type t_i.

Remark 8 *It is easy to see that Condition (S) implies type diversity.*

In Example 1 below, we will construct an interdependent-value environment satisfying type diversity but violating Condition (S) in which there exists a message profile in $S^\infty W$ but it induces an outcome different from the one specified by the social choice function. The main difficulty lies in eliciting each player's true type in round -2 announcement.

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 $\bar{T}_i \to \Delta(\bar{T}_{-i})$:*

$$\pi_i(t_i)[t_{-i}] = \begin{cases} 2/3 & \text{if } t_{i+1} = t_{i+2} = t_i; \\ 1/3 & \text{if } t_{i+1} = t_{i+2} \neq t_i. \end{cases}$$

That is, in player i's view, player $(i+1)$'s type and player $(i+2)$'s type are perfectly correlated but they are only partially correlated with player i's type.

[12] To be precise, the second property of our type diversity was not included in its original definition of Serrano and Vohra (2005). Thus, our version of type diversity is slightly stronger than theirs.

Each player i has the following preferences: for any $a \in A$ and $t \in \bar{T}$,

$$u_i(a,t) = (1-\delta) \times a \cdot t_i + \delta \times a \cdot t_{i+1},$$

where $\delta \in [0,1]$ and $a \cdot t_i$ denotes the dot (or, inner) product of the two vectors a and t_i. That is, player i's preferences depend on his own type and player $(i+1)$'s type, but not depend on player $(i+2)$'s type.

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

We parameterize the class of environments by the value of $\delta \in [0,1]$: when $\delta = 0$, the environment corresponds to a private-value one and also satisfies Assumptions 1 and 2 so that our mechanism can implement f^*; When $\delta \in (0,1/2)$, it corresponds to an interdependent-value environment which satisfies Condition (S) and Assumption 2 so that our mechanism can implement f^*; and when $\delta \in [1/2,1]$, it corresponds to an interdependent-value environment which satisfies Assumptions 2 and 3, but violates Condition (S).

Consider Example 1 with $\delta = 1$. By Lemma 4, we can find a set of lotteries $\{x_i(t_i)\}_{t_i \in \bar{T}_i, i \in I}$ satisfying inequality (31). Therefore, for any $\bar{\tau} > 0$, we can adopt the corresponding mechanism $(\mathcal{M}, \bar{\tau})$ defined in Section 3.1 with this set of lotteries. We claim that in the case of $\delta = 1$, 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^*. This shows some difficulty of extending our results to general interdependent-value environments. We formally state this claim as follows:

Claim 8 *Consider Example 1 with $\delta = 1$. Fix any set of lotteries $\{x_i(t_i)\}_{t_i \in \bar{T}_i, i \in I}$ satisfying inequality (31) and the corresponding mechanism $(\mathcal{M}, \bar{\tau})$ defined in Section 3.1. For any $i \in I$ and any $t_i \in \bar{T}_i$, we have that $(t'_i, \ldots, t'_i) \in S_i^\infty W_i(t_i | \mathcal{M}, \bar{\mathcal{T}})$ where $t'_i \neq t_i$.*

Proof. See Appendix A.2. ■

In their Theorem 4 Oury and Tercieux (2012) show that a social choice function f is continuously implementable by a finite mechanism if and only if it is implementable in rationalizable strategies by a finite mechanism. They do not need any ex post payment, but assume that sending messages in the mechanism is (slightly) costly. We assume that sending messages is costless, but allow for small transfers. We show that all of our results can be extended to the class of interdependent-value environments which satisfy Condition (S).

Bergemann and Morris (2009) show that their robust measurability, which is a necessary condition for robust virtual implementation, is closely connected to the degree of interdependence of preferences. They also show that robust measurability is equivalent to requiring that the notion of measurability originally suggested by Abreu and Matsushima (1992b)–henceforth, AM measurability–holds on the union of all type spaces. Following this idea, in our paper, AM measurability must be a necessary condition.

This example 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 here has nothing to do with the measurability condition. In other words, we must seek another explanation if we consider (full) exact implementation, not virtual one.[13]

6.3 Budget Balance

Assume $I \geq 3$. By constructing d_i^0 under a stronger (and yet still generic) version of Assumption 2, following d'Aspremont et al. (2003), 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).

6.4 Implementation with Arbitrarily Small Transfers vs. Virtual Implementation

Virtual implementation means that the planner contents himself with implementing the social choice rule with arbitrarily high probability. For example, under complete information, Abreu and Sen (1991), Abreu and Matsushima (1992a), and Matsushima (1988) all show that essentially any SCF is virtually implementable. 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. If we interpret a mechanism as a contract between the two parties, it is natural to worry about the possibility of renegotiation and seek to design renegotiation-proof mechanisms. This argument leads us to the conclusion that virtual implementation will not be renegotiation-proof, which potentially upsets its very permissive results. When we are

[13]For example, Artemov et al. (2013) show that robust measurability almost always becomes a vacuous constraint for robust virtual implementation. This seems to be consistent with our finding in this example: AM measurability has nothing to do with the problem of interdependent preferences, while Condition (S) indeed does.

satisfied with virtual implementation, we might simply overlook a big cost of designing a credible 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. This makes our mechanism less susceptible to renegotiation and therefore more credible.

A Appendix

There are two subsections in the appendix. In Section A.1, we show that our mechanism also works under iterative deletion of weakly dominated strategies, i.e., W^{∞} and moreover, the order of removal of strategies in W^{∞} is irrelevant in our mechanism. In Section A.2, we prove the claim we have made in the argument in Example 1 of Section 6.2.

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

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

for all $\sigma_{-i} : \bar{T}_{-i} \to M_{-i}$ such that $\sigma_{-i}(t_{-i}) \in H_{-i}(t_{-i})$ and a strict inequality holds for some σ_{-i}.[14]

Let $\{W^k\}_{k=0}^{\infty}$ be a sequence of profiles of correspondences such that (i) $W_i^0(\bar{t}_i | \mathcal{M}, \bar{\mathcal{T}}) = M_i$; (ii) any $m_i \in W_i^{k+1}(\bar{t}_i | \mathcal{M}, \bar{\mathcal{T}}) \backslash W_i^k(\bar{t}_i | \mathcal{M}, \bar{\mathcal{T}})$ is weakly dominated with respect to W^k

[14] 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., $\sigma_{-i} : \bar{T}_{-i} \to \Delta(M_{-i})$ such that $\sigma_{-i}(t_{-i})[H_{-i}(t_{-i})] = 1$.

for player i of type $\bar{t}_i$; (iii) any $m_i \in W_i^\infty\left(\bar{t}_i|\mathcal{M},\bar{\mathcal{T}}\right)$ is weakly undominated with respect to W^∞ for player i of type $\bar{t}_i$ where $W_i^\infty(\bar{t}_i|\mathcal{M},\bar{\mathcal{T}}) \equiv \bigcap_{l=1}^\infty W_i^l(\bar{t}_i|\mathcal{M},\bar{\mathcal{T}})$.

Let $W^\infty\left(\bar{t}|\mathcal{M},\bar{\mathcal{T}}\right) = \prod_{i\in I} W_i^\infty\left(\bar{t}|\mathcal{M},\bar{\mathcal{T}}\right)$ for any $\bar{t}\in\bar{T}$. 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 9 *Assume that the environment $\mathcal{E}$ satisfies Assumption 2. For $\gamma'>0$, there exist $\lambda>0$ and a proper scoring rule d_i^0 such that for any t_i', $t_i''\in\bar{T}_i$ with $t_i'\neq t_i''$ and any $\hat{\sigma}_{-i}^{-2}:\bar{T}_{-i}\to\bar{T}_{-i}$, we have that*

$$\lambda\left|\sum_{t_{-i}\in\bar{T}_{-i}}\left[d_i^0\left(\hat{\sigma}_{-i}^{-2}\left(t_{-i}\right),t_i'\right)-d_i^0\left(\hat{\sigma}_{-i}^{-2}\left(t_{-i}\right),t_i''\right)\right]\pi_i\left(t_i\right)\left[t_{-i}\right]\right|>\gamma'. \tag{32}$$

Proof. Fix any i. Let

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

D_i^0 is the set of proper scoring rules in $\mathbb{R}^{\bar{T}}$. By Lemma 2, D_i^0 is a nonempty open set. Let

$$I_i^0=\left\{d_i^0\in\mathbb{R}^{\bar{T}}:\sum_{t_{-i}\in\bar{T}_{-i}}\left[d_i^0\left(\hat{\sigma}_{-i}^{-2}\left(t_{-i}\right),t_i'\right)-d_i^0\left(\hat{\sigma}_{-i}^{-2}\left(t_{-i}\right),t_i''\right)\right]\bar{\pi}_i\left(t_i\right)\left[t_{-i}\right]\neq 0,\forall t_i\neq t_i',\forall\hat{\sigma}_{-i}^{-2}\right\}.$$

Since $\bar{T}$ is finite, the complement of I_i^0 has measure zero in $\mathbb{R}^{\bar{T}}$.

Therefore, $\bigcup_{i\in I}\left(D_i^0\cap I_i^0\right)$ has a positive measure in $\mathbb{R}^{\bar{T}}$. Thus we can find a proper scoring rule d_i^0 such that for any $\hat{\sigma}_{-i}^{-2}:\bar{T}_{-i}\to\bar{T}_{-i}$ and t_i', $t_i''\in\bar{T}_i$ with $t_i'\neq t_i''$,

$$\sum_{t_{-i}\in\bar{T}_{-i}}\left[d_i^0\left(\hat{\sigma}_{-i}^{-2}\left(t_{-i}\right),t_i'\right)-d_i^0\left(\hat{\sigma}_{-i}^{-2}\left(t_{-i}\right),t_i''\right)\right]\pi_i\left(t_i\right)\left[t_{-i}\right]\neq 0.$$

Finally, since $\bar{T}$ is finite, for any $\gamma'>0$, we can find some $\lambda>0$ such that for any $\hat{\sigma}_{-i}^{-2}:\bar{T}_{-i}\to\bar{T}_{-i}$ and t_i', $t_i''\in\bar{T}_i$ with $t_i'\neq t_i''$, inequality (32) holds. ■

Proposition 4 *Suppose that the environment $\mathcal{E}$ satisfies Assumptions 1 and 2. Assume $I\geq 2$. Given any incentive compatible SCF f, for all $\bar{\tau}>0$, 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 $g(m)=f(t)$.*

Fix $\bar{\tau} > 0$. Choose the mechanism $(\mathcal{M}, \bar{\tau})$ defined in Section 3.1, with the proper scoring rule d_i^0 given in Claim 8, and λ under $\gamma' = \gamma$ (which is defined in Section 3.1). To prove Proposition 4, it suffices to show that for any $i \in I$ and $\bar{t}_i \in \bar{T}_i$, 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 is because from here we can fill the gap of the argument by adapting the proof of Theorem 2. The rest of the proof builds upon the following three claims.

Claim 10 *Fix any player i of type $\bar{t}_i$. 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. Let σ_i be defined such that $\sigma_i(\bar{t}_i) = (\bar{t}_i, ..., \bar{t}_i)$ for player i of type $\bar{t}_i$. 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, and $\bar{t}_i$.

Fix $\bar{t} \in \bar{T}$. Note first that we trivially have $\sigma(\bar{t}) \in W^0(\bar{t}|\mathcal{M}, \bar{\mathcal{T}})$. For any $k \geq 0$, assume that $\sigma(\bar{t}) \in W^k(\bar{t}|\mathcal{M}, \bar{\mathcal{T}})$. Then, we shall show that $\sigma(\bar{t}) \in W^{k+1}(\bar{t}|\mathcal{M}, \bar{\mathcal{T}})$. This is equivalent to showing the following: for any $\tilde{m}_i \in M_i$, either $\sigma_i(\bar{t}_i)$ is always at least as good as $\tilde{m}_i$ or $\sigma_i(\bar{t}_i)$ is a strictly better reply to some strategies of the other players than $\tilde{m}_i$. 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 $k \geq -1$; (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 $\hat{\sigma}_{-i} : \bar{T}_{-i} \to M_{-i}$ by inequality (14). In Case (ii), against the conjecture σ_{-i}, $\sigma_i(\bar{t}_i)$ is a strictly better message than $\tilde{m}_i$ by the argument in Claims 2, 3 and 5. Therefore, no $\tilde{m}_i$ can weakly dominate $\sigma_i(\bar{t}_i)$. 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 $i \in I$ of type $\bar{t}_i$, 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 $(m_i^{-2}, \bar{t}_i, ..., \bar{t}_i) \in W_i^\infty\left(\bar{t}_i|\mathcal{M}, \bar{\mathcal{T}}\right)$ even when $m_i^{-2} \neq \bar{t}_i$. We shall show that no $\tilde{m}_i$ can weakly dominate $(m_i^{-2}, \bar{t}_i, ..., \bar{t}_i)$ 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, $(m_i^{-2}, \bar{t}_i, \ldots, \bar{t}_i)$ is at least as good as $\tilde{m}_i$ for any $\hat{\sigma}_{-i} : \bar{T}_{-i} \to M_{-i}$ 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 σ_{-i} by the argument in Case (ii) of Step 1. Thus, no $\tilde{m}_i$ can weakly dominate $(m_i^{-2}, \bar{t}_i, ..., \bar{t}_i)$. This completes the proof. ■

Claim 11 *Fix any player i and type $\bar{t}_i$. 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 10, 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: We show that 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 $\hat{\sigma}_{-i}$ such that $\hat{\sigma}_{-i}(\bar{t}_{-i}) \in W_{-i}^{\infty}\left(\bar{t}_{-i}|\mathcal{M}, \bar{\mathcal{T}}\right)$ 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 σ'_{-i} such that $\sigma'_{-i}(\bar{t}_{-i}) \in W_{-i}^{\infty}(\bar{t}_{-i}|\mathcal{M}, \bar{\mathcal{T}})$ for all $\bar{t}_{-i}$; and (2) m_i is strictly better than $\tilde{m}_i$ for player i of type $\bar{t}_i$ against some conjecture $\hat{\sigma}_{-i}$ such that $\hat{\sigma}_{-i}(\bar{t}_{-i}) \in W_{-i}^{\infty}(\bar{t}_{-i}|\mathcal{M}, \bar{\mathcal{T}})$ for all $\bar{t}_{-i} \in \bar{T}_{-i}$.

By Claim 9, 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 (32). 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: We show that for any $\tilde{m}_i$, if $\tilde{m}_i^{-1} \neq m_i^{-1}$ and $\tilde{m}_i^k = \bar{t}_i$ for all $k \neq -1$, $(\bar{t}_i, m_i^{-1}, \bar{t}_i, ..., \bar{t}_i)$ is strictly better than $\tilde{m}_i$ against some conjecture $\tilde{\sigma}_{-i}$ such that $\tilde{\sigma}_{-i}(\bar{t}_{-i}) \in W_{-i}^{\infty}(\bar{t}_{-i}|\mathcal{M}, \bar{\mathcal{T}})$ for all $\bar{t}_{-i} \in \bar{T}_{-i}$.

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 9, there exist a nonempty set of players $J \subset I\backslash\{i\}$ 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 $j \in J$ and $\bar{t}_j \in \bar{T}_j$. From Claim 10, 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 $\tilde{\sigma}_{-i}$. 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 10.

Thus, no $\tilde{m}_i$ can weakly dominate $(\bar{t}_i, m_i^{-1}, \bar{t}_i, ..., \bar{t}_i)$. This completes the proof. ■

Claim 12 *Fix any $i \in I$ and $\bar{t}_i \in \bar{T}_i$. 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 11, $(\bar{t}_i, m_i^{-1}, \bar{t}_i, ..., \bar{t}_i) \in W_i^{\infty}(\bar{t}_i|\mathcal{M}, \bar{\mathcal{T}})$. 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. ■

A.2 Proof of Claim 8

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 we set $\delta = 1$ in Claim 8. So, 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 $t_i \in \bar{T}_i$.

First, 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,*

$$\sum_{t_{-i} \in \bar{T}_{-i}} u_i(f^*(t_i', \alpha_{-i}(t_{-i})), t_{-i})\pi_i(t_i)[t_{-i}] \geq \sum_{t_{-i} \in \bar{T}_{-i}} u_i(f^*(t_i, \alpha_{-i}(t_{-i})), t_{-i})\pi_i(t_i)[t_{-i}]. \tag{33}$$

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. ■

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:

$$\sum_{t_{-i} \in \bar{T}_{-i}} u_i(a, t_{-i})\pi_i(t_i)[t_{-i}] = \frac{2}{3} a \cdot t_i + \frac{1}{3} a \cdot t_i',$$

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 $\{x_i(t_i)\}_{i \in I, t_i \in \bar{T}_i}$ satisfies inequality (31) and there are only two outcomes contained in A, it must be that $x_i(t_i)[a] > 1/2$ if and only if $t_i = a$. Since $u_i\left(a, t_{-i}\right) = a \cdot t_{i+1}$, $u_i(x_i(t_i'), t_{i+1}') - u_i(x_i(t_i), t_{i+1}') > 0$ if $t_i \neq t_i' = t_{i+1}'$. ■

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}', $\sigma_i'\left(t_i\right)$ 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 (33).

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 (33). 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 be a conjecture of type t_i such that $\nu(\bar{m}_{-i}|t_{-i}) = 1$ and $\nu(\tilde{m}_{-i}|t_{-i}') = 1$ 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 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. ■

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*, 997–1021.

Artemov, G., T. Kunimoto, and R. Serrano (2013): "Robust virtual implementation: Toward a reinterpretation of the Wilson doctrine," *Journal of Economic Theory*, 148, 424–447.

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

Bergemann, D. and S. Morris (2009): "Robust virtual implementation," *Theoretical Economics*, 4, 45–88.

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

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

Chen, Y.-C. and S. Xiong (2013): "Genericity and robustness of full surplus extraction," *Econometrica*, 81, 825–847.

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

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

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

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

Demski, J. S. and D. Sappington (1984): "Optimal incentive contracts with multiple agents," *Journal of Economic Theory*, 33, 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, 154–169.

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

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

Kunimoto, T. and R. Serrano (2014): "Implementation in Rationalizable Strategies," *mimeo.*

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

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

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

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

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

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

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

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

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

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

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

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