THREE ESSAYS ON IMPLEMENTATION THEORY

SUN YIFEI

(*B.A. University of International Business and Economics, China*)

A THESIS SUBMITTED

FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

DEPARTMENT OF ECONOMICS

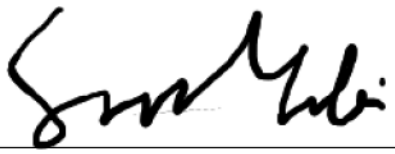
NATIONAL UNIVERSITY OF SINGAPORE

2015

Declaration

I hereby declare that this thesis is my original work and it has been written by me in its entirety. I have duly acknowledged all the sources of information which have been used in the thesis.

This thesis has also not been submitted for any degree in any university previously.

Signed: 

Date: **20 May, 2015**

Acknowledgements

In the past five years, I am lucky enough to all these important people around who have helped me and made this dissertation possible.

Firstly, it is with immense gratitude that I acknowledge the guidance and support of my supervisor, Professor Yi-Chun Chen. His enthusiasm, patience, knowledge and inspiration for research have encouraged me and helped me since the first day I decided to try myself as a researcher. It is an honor to be under his supervision.

Moreover, I would like to thank Professor Takashi Kunimoto, Professor Satoru Takahashi, Professor Yeneng Sun, Professor Xiao Luo, Professor Jingfeng Lu, Professor Songfa Zhong and Professor Parimal Bag, for their valuable comments and suggestions. I have benefited a lot from both of them.

I would also like to thank all my colleagues and friends for their support and suggestions along the way.

Finally, I would like to gratefully dedicate this dissertation to my parents and my love.

Contents

Summary

This thesis is on full implementation theory. In this literature, the mechanism is designed such that all its equilibria reveal players' true information and achieve a given social choice function. The fundamental question addressed in this literature is that which social choice functions are implementable and under what assumptions. Most of the first results is negative (e.g., Satterthwaite, 1975, and Gibbard, 1973, for implementation in dominant strategies). Starting with Maskin (1977), who gave necessary and sufficient conditions for Nash implementation, researchers have studied implementation problems under various solution concepts. Abreu and Matsushima (1992) made an important step in this direction. They showed that almost any social choice function is virtually implementable. We explicitly and fully exploit the power of monetary transfers and lotteries which are usually used in virtual implementation.

The first chapter shows that in a complete-information environment with two or more players and a finite type space, any truthfully implementable social choice function can be fully implemented in backwards induction via a finite perfect-information stochastic mechanism with arbitrarily small transfers. This provides an improvement from the virtual implementation result by Glazer and Perry (1996). With arbitrarily small transfers only off the equilibri-

um path, the mechanism we construct is much less susceptible to renegotiation problem.

the second chapter, we provides a dynamic mechanism which fully implements any social choice function under initial rationalizability in complete information environments. Accommodating any belief revision assumption, initial rationalizability is the weakest among all the rationalizability concepts in extensive form games. This mechanism is also robust to small amounts of incomplete information about the state of nature. That is, the mechanism not only fully implements any social choice function in complete information environments but also does so in all nearby environments where players' values are private. Although our mechanism allows for monetary transfers out of the solution path, we can make them arbitrarily small and even achieve its budget balance when there are more than two players.

In the third chapter, we further exploit the transfers in an incomplete information environments and show in private-value environments that any incentive compatible rule is implementable with small transfers. Our mechanism only needs small ex post transfers to make our implementation results completely free from the multiplicity of equilibrium problem. In addition, our mechanism possesses the unique equilibrium that is robust to higher-order belief perturbations. We also provide a sufficient condition for implementation

in interdependent-value environments and discuss the difficulty of extending our results to interdependent values environments in general.

Chapter 1

Full Implementation in Backward Induction

1.1 Introduction

In a complete-information environment with two or more players and a finite type space, we show that any truthfully implementable social choice function[1] can be fully implemented in backward induction using a finite perfect-information stochastic mechanism. Our result is achieved by invoking (1) a dynamic stochastic mechanism, (2) arbitrarily small transfers, and (3) the domain restriction which rules out identical preferences and preference orderings with complete indifference over all outcomes.

It is known that subgame-perfect implementation is more permissive than

[1]A social choice function is truthfully implementable if there exists a direct revelation mechanism where truth-telling (i) is a Nash equilibrium, and (ii) implements the social choice function. It is well known that any Nash implementable social choice function is truthfully implementable. In Section 3, we show that truthful implementability is also a necessary condition for our notion of implementation. When there are three or more players, any social choice function is truthfully implementable, that is, truthful implementability is trivially satisfied.

Nash implementation (Moore and Repullo (1988)). Our result can be contrasted with two existing perfect-information mechanisms which implement an arbitrary social choice function in subgame-perfect equilibrium.[2] The mechanism in Moore and Repullo (1988, Section 5.1) (henceforth, the MR mechanism) imposes large off-equilibrium transfers, while the mechanism in Glazer and Perry (1996) (henceforth, the GP mechanism) requires at least three players and that the implementation be virtual, i.e., the desirable social outcome is obtained only with large probability.[3] Both mechanisms have thus been criticized for their susceptibility to renegotiation (see Jackson (2001, p. 690)). In contrast, our mechanism is a finite stochastic game with perfect information, which ensures full implementation via backward induction through arbitrarily small transfers off the equilibrium path, and no transfers on the equilibrium path.

In a generic perfect-information game, the backward induction outcome is induced by several notions of extensive-form rationalizability.[4] Since we allow

[2]See Glazer and Perry (1996, p. 28) for a discussion of practical and theoretical reasons to favor sequential/perfect-information mechanisms. In particular, they argue that "sequential mechanisms, with backward induction as their solution concept, seem to be more intuitive and simpler to understand than their simultaneous counterparts." Nevertheless, since the length of our constructed game form will grow as the imposed transfers vanish, the simplicity of solving the game is subject to debate.

[3]See also Osborne and Rubinstein (1994, pp. 193-195) for an exposition of the result in Glazer and Perry (1996).

[4]These solution concepts include, for example, the subgame rationalizability in Bernheim (1984) and the extensive-form rationalizability in Pearce (1984). See also Battigalli and Siniscalchi (2002) for an epistemic characterization of extensive-form rationalizability.

for small transfers, our mechanism can be made generic to implement any truthfully implementable social choice function in these notions of extensive-form rationalizability. In contrast, Bergemann et al. (2011) show that a stronger version of the monotonicity condition due to Maskin (1999) is necessary for implementation in normal-form rationalizability.

Our result can also be contrasted with the static mechanism in Abreu and Matsushima (1994) which fully implements any social choice function in iterated deletion of weakly dominated strategies.[5] The GP mechanism is a dynamic counterpart of the mechanism in Abreu and Matsushima (1992a) which achieves virtual implementation for any social choice function in a static mechanism; in contrast, our result provides a dynamic counterpart of the mechanism in Abreu and Matsushima (1994) which fully implements an arbitrary social choice function in a static mechanism.[6] Abreu and Matsushima (1994) extend the result in Abreu and Matsushima (1992a) from virtual implementation to full implementation, but strengthen the solution concept from

[5]In Abreu and Matsushima (1994), implementation in iterated deletion of weakly dominated strategies is achieved by one round of removal of weakly dominated strategies followed by iterative removal of strictly dominated strategies. Since they study the implementation problem in the environment with more than two players, truthful implementability is automatically satisfied.

[6]Glazer and Perry (1996) make a simple modification of the normal form mechanism in Abreu and Matsushima (1992), where the GP mechanism is an extensive form game with the same outcome function. Nevertheless, the difficulty of modifying the normal form mechanism in Abreu and Matsushima (1994) is due to their adopting an indication in their outcome function, for which we know of no counterpart in an extensive form game except for using the MR mechanism.

iterated deletion of strictly dominated strategies in Abreu and Matsushima (1992a) to iterated deletion of weakly dominated strategies. In contrast, we achieve full implementation in the same solution concept as in Glazer and Perry (1996), i.e., backward induction.

Glazer and Rubinstein (1996) argue that an extensive-form game provides a "guide" for solving a normal-form game and thereby reduces the computational burden on the players. They define a solution concept called *guided iteratively undominated strategies* and prove that a social choice function is implementable in guided iteratively undominated strategies if and only if it is implementable in subgame-perfect equilibrium in a perfect-information mechanism. It follows that our mechanism also implements any truthfully implementable social choice function in guided iteratively undominated strategies.

The paper is organized as follows. Section 2 describes the environment. Section 3 presents the main result and the mechanism. Section 4 provides the proof, and Section 5 concludes.

1.2 Environment

Let $N = \{1, 2, ..., n\}$ denote the set of players. 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 $l \in \Delta(A)$ denotes a lottery on A.

For each player $i \in N$, let Θ_i denote a finite set of types of player i. The utility index of player i over the set A is denoted by $v_i : A \times \Theta_i \to \mathbb{R}$, where $v_i(a, \theta_i)$ specifies the bounded utility of player i from the social alternative a, when he is of type θ_i. Player i's expected utility from a lottery $l \in \Delta(A)$ under type θ_i is $u_i(l, \theta_i) = \sum_{a \in A} l(a) v_i(a, \theta_i)$, which is well defined since $v_i(a, \theta_i)$ is bounded.

Following Abreu and Matsushima (1992a) and Glazer and Perry (1996), we assume that (i) for each $\theta_i \in \Theta_i$, $v_i(\cdot, \theta_i)$ is not a constant function on A; and (ii) for any two distinct types θ_i and θ_i', $v_i(\cdot, \theta_i)$ is not a positive affine transformation of $v_i(\cdot, \theta_i')$. This restriction guarantees the reversal property which is used to elicit players' true type (see (1.3)).

A planner aims to implement a social choice function that is a mapping $f : \Theta \to \Delta(A)$, where $\Theta = \Theta_1 \times \Theta_2 \times \cdots \times \Theta_n$.[7] We assume that the true type profile $\psi \in \Theta$ is commonly known to the players but unknown to the planner.

We assume that the planner can fine or reward a player $i \in N$, and we denote by $t_i \in \mathbb{R}$ the transfer from player i to the planner. We also assume that player i's utility is quasilinear in transfers, and is denoted by $u_i(l, \theta_i) + t_i$. A finite sequential stochastic mechanism is a finite perfect-information game

[7] Here we follow Abreu and Matsushima (1994) and Glazer and Perry (1996) in assuming that the space of type profiles is a product space.

tree Γ together with an outcome function ζ, including an allocation function g which specifies for each terminal history a lottery $l \in \Delta(A)$ and a transfer rule $t = (t_1, t_2, ..., t_n)$. A sequential mechanism (Γ, ζ) has fines and rewards bounded by $\bar{t}$ if $|t_i| \leq \bar{t}$ for every $i \in N$ and every terminal history.

1.3 Mechanism

In this section, we provide a full characterization of social choice functions which are fully implemented in backward induction with arbitrarily small transfers. It is well known that if f is implementable, then it must be truthfully implementable. That is, there must exist a "direct revelation mechanism" $\tilde{f} : \Theta^n \to \Delta(A)$, such that for any $\theta \in \Theta$, the following hold:

- $P1 : \tilde{f}(\theta^n) = f(\theta)$, i.e., if all individuals announce θ, the outcome is $f(\theta)$.

- $P2$: the unanimous announcement of θ is a Nash equilibrium at state θ.

That is, truth-telling is a Nash equilibrium. Observe that any social choice function f can then be truthfully implemented when $n \geq 3$. This can be achieved by constructing a direct revelation mechanism with the following property: if at least $n-1$ individuals announce θ, then the outcome is $f(\theta)$. No individual can change the outcome by deviating from a unanimous an-

nouncement, so that truth-telling is clearly a Nash equilibrium. The restriction $n \geq 3$ is crucial because it allows the planner to identify a deviant from a truth-telling strategy combination. If instead $n = 2$ and player 1 announces θ and player 2, ϕ, then there is no way for the planner to ascertain whether state θ has occurred and 2 is lying, or state ϕ has occurred and 1 is lying. Clearly, if truth telling is to be sustained as an equilibrium, there must exist an outcome which is simultaneously no better than $f(\theta)$ for 2 in state θ and no better than $f(\phi)$ for 1 in state ϕ. That is, not every social choice function is truthfully implementable when $n = 2$.[8]

Definition 1.1. *A social choice function f is truthfully implementable if there exists a direct revelation mechanism $\tilde{f}$ which satisfies P1 and P2.*

It is well known result that any Nash-implementable social choice function (even if only partially implementable) must be truthfully implementable (see Dasgupta et al. (1979)). Proposition 1.1 states that truthful implementability is a necessary condition for our notion of implementation which allows arbitrarily small transfers off equilibrium path.

Proposition 1.1. *Assume A is finite. Suppose that for any $\bar{t} > 0$, there exists a finite sequential stochastic mechanism with fines and rewards bounded*

[8]Dutta and Sen (1991) have a detailed discussion in which they provide a full characterization of the class of two-person social choice correspondences which are Nash-implementable.

by $\bar{t}$, such that for each type profile ψ, $f(\psi)$ with no transfer is the unique subgame-perfect equilibrium outcome. Then, f is truthfully implementable.

Proof. For convenience, let $\bar{t} = \frac{1}{q}$ where $q \in \mathbb{N}$. Suppose $f : \Theta \to \Delta(A)$ is implementable in SPE by a mechanism (Γ, ζ) with fines and rewards bounded by $\frac{1}{q}$. Let g^q be the function which specifies the lottery associated with the terminal node and let t^q be the transfer rule.

Let $\tilde{f}_{\bar{t}}$ be a direct revelation mechanism such that

$$\tilde{f}_{\bar{t}}\left(\left(\theta^i\right)_{i\in N}\right) = \left(g^q\left(\left(m_i^{\theta^i}\right)_{i\in N}\right), t_i^q\left(\left(m_i^{\theta^i}\right)_{i\in N}\right)\right),$$

where θ^i denotes that player i announce θ for any $\theta \in \Theta$.

Suppose ψ is the true state. Let ψ^{-i} denotes that all the players other than i announce ψ. We have

$$u_i\left(g^q\left(m_i^{\psi}, m_{-i}^{\psi}\right), \psi_i\right) \geq u_i\left(g^q\left(m_i^{\phi}, m_{-i}^{\psi}\right), \psi_i\right) + t_i^q\left(m_i^{\phi}, m_{-i}^{\psi}\right).$$

Note that this inequality holds for any q and $t_i^q(m) < \frac{1}{q}$. Since A is finite, $\Delta(A)$ is compact. There exists some $g^0\left(m_i^{\phi}, m_{-i}^{\psi}\right) \in \Delta(A)$ such that

$$u_i\left(g^q\left(m_i^{\phi}, m_{-i}^{\psi}\right), \psi_i\right) \to u_i\left(g^0\left(m_i^{\phi}, m_{-i}^{\psi}\right), \psi_i\right) \text{ as } q \to \infty.$$

That is, we have some $\tilde{f}_0$ with no transfer such that,

$$u_i\left(\tilde{f}_0\left(\psi^n\right), \psi_i\right) \geq u_i\left(\tilde{f}_0\left(\phi^i, \psi^{-i}\right), \psi_i\right).$$

This completes the proof. □

Remark 1.1. *The compactness of the set of alternatives is to guarantee the existence of the limit of the bad outcomes as the bound of transfers approaches zero. If A is compact, our result holds with two technical assumption: (1) $\Delta(A)$ is the set of all probability measure over A; (2) $v_i(\cdot, \theta_i)$ is continuous.*

Theorem 1.1. *For any $n \geq 2$, any truthfully implementable social choice function f, and any $\bar{t} > 0$, there exists a finite sequential stochastic mechanism with fines and rewards bounded by $\bar{t}$ such that for each type profile ψ, the outcome $f(\psi)$ with no transfer is the unique subgame-perfect equilibrium outcome.*

1.3.1 The preliminaries

Given a social choice function f, since Θ_i is finite for any i, we let

$$\xi = \max_{\theta_i \in \Theta_i, \theta, \theta' \in \Theta, i \in N} \left| u_i(f(\theta), \theta_i) - u_i(f(\theta'), \theta_i) \right|. \tag{1.1}$$

That is, ξ is the maximal difference in payoffs of all implementable outcomes for all players of all types. Choose an integer K and $\varepsilon > 0$ such that

$$\xi / K < \varepsilon < \bar{t}/6. \tag{1.2}$$

Hence, K is large when $\bar{t}$ is small. For any distinct types θ_i and θ_i', let $x_{\theta_i,\theta_i'}$ and x_{θ_i',θ_i} be two lotteries such that

$$u_i(x_{\theta_i,\theta_i'},\theta_i) > u_i(x_{\theta_i',\theta_i},\theta_i);$$

$$u_i(x_{\theta_i,\theta_i'},\theta_i') < u_i(x_{\theta_i',\theta_i},\theta_i'). \tag{1.3}$$

The existence of $x_{\theta_i,\theta_i'}$ and x_{θ_i',θ_i} is guaranteed by the assumption on the preferences. Let $L \equiv \left\{x_{\theta_i,\theta_i'}, x_{\theta_i',\theta_i}\right\}_{\theta_i \neq \theta_i', i \in N}$. Observe that L is a finite set since Θ_i and N are both finite.

1.3.2 The Mechanism

The mechanism has $K+2$ rounds. In each round $k \leq K+1$, the players move sequentially. Player 1 moves first, player 2 moves second, and so on. In round $k \leq K$, each player i announces a type profile $m_i^k \in \Theta$.

In round $K+1$, each player i announces his own type, $m_i^{K+1} \in \Theta_i$. Let $m^{K+1} = \left(m_1^{K+1}, ..., m_n^{K+1}\right)$.

Let

$$l = \sum_{k=1}^{K} \frac{1}{K} \tilde{f}\left(m^k\right),$$

where $\tilde{f}$ satisfies P1 and P2.

Then, by the finiteness of L and Θ_i, choose $p_l \in (0,1)$ such that for any $l' \in L$, any $i \in N$, and any $\theta_i \in \Theta_i$,

$$|u_i(l,\theta_i) - u_i((1-p_l)l + p_l l', \theta_i)| < \varepsilon/2. \tag{1.4}$$

Let

$$x_{l,\theta_i,\theta_i'} = (1-p_l)l + p_l x_{\theta_i,\theta_i'};$$

$$x_{l,\theta_i',\theta_i} = (1-p_l)l + p_l x_{\theta_i',\theta_i}.$$

Consequently, we have

$$u_i(x_{l,\theta_i,\theta_i'},\theta_i) > u_i(x_{l,\theta_i',\theta_i},\theta_i); \tag{1.5}$$

$$u_i(x_{l,\theta_i,\theta_i'},\theta_i') < u_i(x_{l,\theta_i',\theta_i},\theta_i').$$

Remark 1.2. *The conditions in (1.5) will guarantee that truth telling is strictly better when players face the constructed lotteries (see the proof of Claim 3.1 in Section 4 below).*

In round $K+2$, in the order of player $n+1(\equiv 1)$, n,..., 2, player i has an opportunity to announce his predecessor's preference $m_i^{K+2} \in \Theta_{i-1}$ if and only if $m_j^{K+2} = m_{j-1}^{K+1}$ for every $j > i$.[9]

- If $m_i^{K+2} \neq m_{i-1}^{K+1}$, then player $i-1$ chooses $x_{l,m_{i-1}^{K+1},m_i^{K+2}}$ or $x_{l,m_i^{K+2},m_{i-1}^{K+1}}$ and the game ends;

- If $m_i^{K+2} = m_{i-1}^{K+1}$, then the game continues and player $i-1$ gets the opportunity to announce his predecessor's preference $m_{i-1}^{K+2} \in \Theta_{i-2}$.

[9]Note that player 1 always has the opportunity to announce player n's type.

If $m_i^{K+2} = m_{i-1}^{K+1}$ for all i, then the social alternative is determined by the lottery l and the game ends.

The transfers are specified as follows:

$$t_i = \eta_i + \tau_i + \delta_i.$$

$$\eta_i = \begin{cases} -3\varepsilon, & \text{if } m_i^{K+2} \neq m_{i-1}^{K+1}, \text{ and } i-1 \text{ chooses } x_{l,m_{i-1}^{K+1},m_i^{K+2}}; \\ \varepsilon, & \text{if } m_i^{K+2} \neq m_{i-1}^{K+1}, \text{ and } i-1 \text{ chooses } x_{l,m_i^{K+2},m_{i-1}^{K+1}}; \\ 0, & \text{otherwise.} \end{cases}$$

$$\tau_i = \begin{cases} -2\varepsilon, & \text{if } m_{i+1}^{K+2} \neq m_i^{K+1}; \\ 0, & \text{otherwise.} \end{cases}$$

$$\delta_i = \begin{cases} -\varepsilon, & \text{if } i \text{ is the last person who chooses } m_i^k \neq m^{K+1} \text{ for some } k \leq K; \\ 0, & \text{otherwise.} \end{cases}$$

Note first that along any history, a player is fined at most 6ε and is rewarded at most ε, which are bounded by $\bar{t}$ (by (1.2)). Second, when $m_i^{K+2} \neq m_{i-1}^{K+1}$, player $i-1$ will be fined 2ε regardless of her choice between $x_{l,m_{i-1}^{K+1},m_i^{K+2}}$ and $x_{l,m_i^{K+2},m_{i-1}^{K+1}}$; on the other hand, whether i will get ε or -3ε depends on player $i-1$'s choice. We draw the game tree for rounds $K+1$ and $K+2$ in Figure 1 and highlight the equilibrium path in boldface.

Remark 1.3. *The "direct revelation mechanism" $\tilde{f}$ works in the same way as ρ (a majority rule), used in the GP mechanism.*[10] *With this construction, we generalize the implementation result in Glazer and Perry (1996) to a two-*

[10] We restate the majority rule from Glazer and Perry (1996, p. 30) as follows:
For each stage k, $k = 1, ..., K$, a probability of $(1-\varepsilon)/K$ is assigned to $f(\psi)$ if $m_i^k = \psi$, for at least $n-1$ players; otherwise, a probability of $(1-\varepsilon)/K$ is assigned to some arbitrarily chosen alternative b.

person setting. Note that truthful implementability is trivially satisfied by the majority rule when there are three or more players. The following corollary holds immediately if we replace the majority rule in the GP mechanism with $\tilde{f}$.

Corollary 1.1. *For any* $n \geq 2$*, any truthfully implementable social choice function* f*,* $\varepsilon > 0$*, and* $\bar{t} > 0$*, there exists a finite sequential stochastic mechanism with fines and rewards bounded by* $\bar{t}$ *for which the unique subgame-perfect equilibrium outcome is such that for each type profile* ψ*, the outcome* $f(\psi)$ *is chosen with probability of at least* $1 - \varepsilon$.

Remark 1.4. *The main difference between our mechanism and the GP mechanism is that we adopt a modified MR mechanism to elicit the players' true types in round* $K + 1$ *and round* $K + 2$*. The modified MR mechanism further differs from the MR mechanism in an essential way: by using randomization, we can (by (1.4)) make the lottery assigned to each terminal history arbitrarily close to lottery* l*, which is determined by the announcements from round* 1 *to round* K*. Consequently, relative to the transfers, the announcement made in either round* $K + 1$ *or round* $K + 2$ *has a negligible effect on the lotteries associated to terminal histories. We can therefore elicit each player's true type in round* $K + 1$ *without the large transfers required in the MR mechanism.*

If we keep the first K rounds identical to the setting in the GP mechanism,

we have the following corollary.

Corollary 1.2. *For any $n \geq 3$, social choice function f, and $\bar{t} > 0$, there exists a finite sequential stochastic mechanism with fines and rewards bounded by $\bar{t}$ such that for each type profile ψ, the outcome $f(\psi)$ with no transfer is the unique subgame-perfect equilibrium outcome.*

Remark 1.5. *Moore and Repullo (1988) provide a necessary condition for subgame-perfect implementation for general preferences. The necessary condition is actually indispensable in quasilinear environment which our paper studies. In their section 5, they construct a simple finite mechanism with perfect information in quasilinear environment. With sufficiently large transfers, this simple mechanism can implement any social choice function (see the detailed discussion on pp. 1214–1215 in Moore and Repullo (1988)). That is, with large enough transfers, the necessary condition they identify in their Theorem 1 is automatically satisfied. Our mechanism breaks up the large transfers into a small scale by adopting a large horizon and making full use of lotteries. See the detailed discussion in Appendix.*

1.4 Implementation

Denote the true type profile by ψ.

Claim 1.1. *In any subgame-perfect equilibrium where player i moves in round $K+2$, player i will announce $m_i^{K+2} = \psi_{i-1}$ if $m_{i-1}^{K+1} = \psi_{i-1}$ and will announce $m_i^{K+2} \neq m_i^{K+1}$ otherwise.*

Proof. First, consider player 2's choice in round $K+2$. This is the last move in the game tree. There are two cases:

Case 1. $m_1^{K+1} = \psi_1$: If player 2 announces $m_2^{K+2} = \psi_1$, then l is implemented and $\eta_2 = 0$. If, instead, player 2 announces $m_2^{K+2} \neq \psi_1$, then by (1.5) player 1 will choose $x_{l,m_1^{K+1},m_2^{K+2}}$, while player 2 will be fined $\eta_2 = -3\varepsilon$. By (1.4), player 2 will announce ψ_1.

Case 2. $m_1^{K+1} \neq \psi_1$: If player 2 announces $m_2^{K+2} = m_1^{K+1}$, then l is implemented and $\eta_2 = 0$. If, instead, player 2 announces $m_2^{K+2} = \psi_1$, then by (1.5) player 1 will choose $x_{l,m_2^{K+2},m_1^{K+1}}$, while player 2 will be rewarded with $\eta_2 = \varepsilon$. By (1.4), player 2 will announce some $m_2^{K+2} \neq m_1^{K+1}$.

Similarly, since the payoff difference between any two lotteries in the set $\{l\} \cup L$ is at most ε, each player i (where $2 \leq i \leq n$) will confirm his predecessor's announcement in $K+1$ (i.e., $m_i^{K+2} = m_{i-1}^{K+1}$) if $m_{i-1}^{K+1} = \psi_{i-1}$; while player i will challenge his predecessor's announcement in $K+1$ (i.e., $m_i^{K+2} \neq m_{i-1}^{K+1}$) if $m_{i-1}^{K+1} \neq \psi_{i-1}$.

Now consider player 1 (i.e., player $n+1$)'s choice in round $K+2$. Again, there are two cases:

Case 1. $m_n^{K+1} = \psi_n$: If player 1 announces $m_1^{K+2} = \psi_n$, then one outcome from $\{l\} \cup L$ is implemented, $\eta_1 = 0$, and player 1 will be fined $\tau_1 = -2\varepsilon$ if he is challenged by player 2 later. In total, the potential loss from announcing $m_1^{K+2} = \psi_n$ is less than 3ε. If, instead, player 1 announces $m_1^{K+2} \neq \psi_n$, then by (1.5) player n will choose $x_{l,m_n^{K+1},m_1^{K+2}}$, while player 1 will be fined $\eta_1 = -3\varepsilon$. Therefore, player 1 will announce ψ_n.

Case 2. $m_n^{K+1} \neq \psi_n$: If player 1 announces $m_1^{K+2} = m_n^{K+1}$, then one outcome from $\{l\} \cup L$ is implemented, $\eta_1 = 0$. In total, the potential gain from announcing $m_1^{K+2} = m_n^{K+1}$ is less than ε. If, instead, player 1 announces $m_1^{K+2} = \psi_n$, then by (1.5) player n will choose $x_{l,m_1^{K+2},m_n^{K+1}}$, while player 1 will be rewarded with $\eta_1 = \varepsilon$. Therefore, player 1 will announce some $m_1^{K+2} \neq m_n^{K+1}$. □

Claim 1.2. *In any subgame-perfect equilibrium, every player truthfully announces his own type in round $K+1$, i.e., $m_i^{K+1} = \psi_i$ for all $i \in N$.*

Proof. Consider player n first. Suppose that player n announces $m_n^{K+1} \neq \psi_n$. Since player 1 moves first in round $K+2$, then by Claim 3.1, this announcement will be challenged by player 1 and result in a penalty $\tau_n = -2\varepsilon$. It follows from (1.4) that by announcing $m_n^{K+1} \neq \psi_n$, player n's utility from the induced lottery is affected by an amount less than ε. In addition, player n potentially reduces the penalty $\delta_n = -\varepsilon$. Therefore, player n will announce $m_n^{K+1} = \psi_n$. Thus, by Claim 3.1, player n will have an opportunity move in round $K+2$,

and by a similar argument, $m_{n-1}^{K+1} = \psi_{n-1}$. We can inductively argue that $m_i^{K+1} = \psi_i$ for all $i \in N$. □

Claim 1.3. *In any subgame-perfect equilibrium, if player i is not the last one to announce a type profile that is different from m^{K+1} along a history up to round $k \leq K$, then $m_i^k = \psi$.*

Proof. Note that by Claim 3.3 $m^{K+1} = \psi$ in any subgame-perfect equilibrium. Consider player n's decision in round K. Suppose that player n is not the last one who lies along a given history. Then, player n will be fined $\delta_n = -\varepsilon$ if he lies by announcing $m_n^K \neq \psi$, but will not be fined if he announces $m_n^K = \psi$. The maximal gain from the change in lottery chosen by lying is ξ/K. By (1.2), he strictly prefers to tell the truth. Inductively we can show that any player $i \leq n-1$ strictly prefers to tell the truth in round K if player i is not the last one who lies along a given history.

Suppose that for any player i, he strictly prefers to tell the truth in round k' if player i is not the last one who lies along a given history for any $k \leq k' \leq K$. We show that player i strictly prefers to tell the truth in round $k-1$ if player i is not the last one who lies along a given history for any player i.

If player i lies, then by the induction hypothesis, all the players will tell the truth in the following histories. Thus, player i will be fined $\delta_1 = -\varepsilon$. The maximal gain from the change in lottery chosen by lying is bounded by ξ/K

in round k. From P2 of $\tilde{f}$, the maximal gain from the change in lottery chosen by lying is 0 in round $k'' \geq k$. If he tells the truth, instead of player 1, player i' will be fined $\delta_{i'} = -\varepsilon$. In total, the potential gain is less than the loss. It follows that truth-telling is strictly better for player i in round $k+1$. □

This completes the proof.

Claim 1.4. *In any subgame-perfect equilibrium, $m_i^k = \psi$, for all $i \in N$, and for all $1 \leq k \leq K$.*

Proof. No player has lied in round $k = 1$. It then follows from Claim 1.3 that $m_i^1 = \psi$ for all i. Inductively, $m_i^k = \psi$ for all $i \in N$ and for all $1 \leq k \leq K$. □

1.5 Concluding Remarks

Our result is proved by observing the complementarity between Moore and Repullo (1988) and Glazer and Perry (1996). Specifically, we modify the MR mechanism by allowing randomization on the pure outcomes. We can strengthen the result of Glazer and Perry (1996) to full implementation from virtual implementation, if we adopt the MR mechanism in the last two rounds, round $K+1$ and round $K+2$. In addition, the result of Moore and Repullo (1988) (which holds with large payments) can be proved with arbitrarily small transfers, if we adopt the idea of Glazer and Perry (1996) (which is due to Abreu and Matsushima (1992a)) in breaking the large fine into K small pieces.

If there are three or more players, our argument is essentially unaltered if the fines (resp. rewards) imposed on some player are to be paid to (resp. paid by) some other player instead of the planner. In other words, with three or more players, we can achieve budget balance (i.e., the transfers add up to zero) both on and off the equilibrium path.[11]

Our result crucially relies on the assumption of complete information and is therefore subject to the criticism by Aghion et al. (2012), namely, that our mechanism still admits undesirable sequential equilibria when some in-

[11] When there are only two players, as in Moore and Repullo (1988), there may be an additional surplus generated off the equilibrium path.

formation perturbation (as defined in Aghion et al. (2012)) is introduced to the complete-information environment. An extension of our analysis to an incomplete-information environment is left for future research.[12]

The finiteness of the mechanism relies crucially on the assumption that the state space is finite. We cannot hope for a finite mechanism to fully implement any social choice function when the state space is infinite. In addition, the finiteness assumption guarantees the existence of lotteries to elicit the true preference of each player. This is crucial for our result as well as for the results in Abreu and Matsushima (1992a), Abreu and Matsushima (1994), and Glazer and Perry (1996).

[12]Instead of using dynamic mechanisms, Chen et al. (2014) use a finite static mechanism to show that, in incomplete information environments, any truthfully implementable social choice function is implementable in one round deletion of weakly dominated strategies followed by iterative removal of strictly dominated strategies.

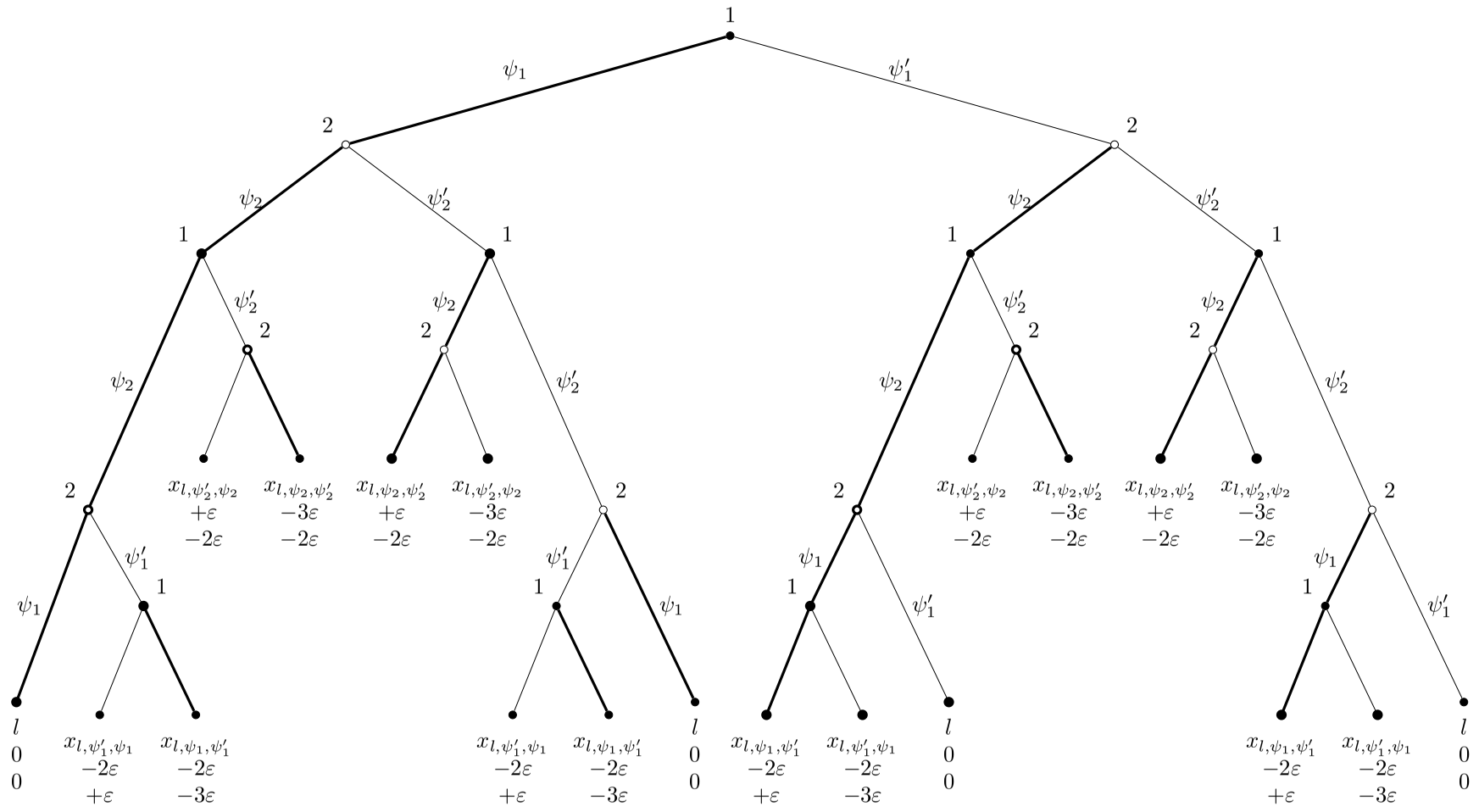


Figure 1.1: The game tree in round $K+1$ and round $K+2$ when (i) there are two players, and (ii) each player has two types, ψ_i and ψ_i', where $\psi = (\psi_1, \psi_2)$ is the true type profile. For each payoff vector associated with the specific terminal node, the first coordinate is the lottery implemented, the second is the fine or reward imposed on player 1, and the third is the fine or reward imposed on player 2. The equilibrium path is indicated in boldface.

1.6 Appendix

In this section, we restate the necessary condition, i.e., Condition C, in Theorem 1 of Moore and Repullo (1988) and show that Condition C is trivially satisfied in qusilinear environment. We incorporate their setting into our environment. In this section, f is a social choice correspondence from Θ to $\Delta(A)$.

Condition C For each pair of profiles θ and ϕ in Θ, and for each $a \in f(\theta)$ but $a \notin f(\phi)$, there exists a finite sequence

$$a(\theta, \phi; a) \equiv \{a_0 = a, a_1, ..., a_k, ..., a_h = x, a_{h+1} = y\} \subset A,$$

with $h = h(\theta, \phi; a) \geq 1$, such that:

(1) for each $k = 0, ..., h-1$, there is some particular agent $j(k) = j(k|\theta, \phi; a)$, say, for whom

$$u_{j(k)}(a_k, \theta) \geq u_{j(k)}(a_{k+1}, \theta); \text{ and}$$

(2) there is some particular agent $j(h) = j(h|\theta, \phi; a)$, say, for whom

$$u_{j(h)}(x, \theta) \geq u_{j(h)}(y, \theta) \text{ and } u_{j(h)}(y, \phi) > u_{j(h)}(x, \phi).$$

Further, $h(\theta, \phi; a)$ is uniformly bounded by some $\bar{h} < \infty$.

We first show that with sufficiently large transfers, Condition C is automatically satisfied in qusilinear environment.

To see Condition C is trivially satisfied when large enough transfers are allowed, we consider a pair of states $\{(\theta_i, \theta_{-i}), (\theta_i', \theta_{-i})\}$ and $a \in f(\theta_i, \theta_{-i})$ but $a \notin f(\theta_i', \theta_{-i})$.

Since the state space is finite, there exist a pair of outcomes $x, y \in \Delta(A)$ and a pair of transfers $t_x, t_y \in \mathbb{R}$, such that

$$\begin{array}{l} u_i(x, \theta_i) - t_x > u_i(y, \theta_i) - t_y, \\ u_i(x, \theta_i') - t_x < u_i(y, \theta_i') - t_y. \end{array} \tag{1.6}$$

Furthermore, $u_i(a, \theta_i) > u_i(a', \theta_i) - t$, for all $\theta_i \in \Theta_i$, all $a' \in \Delta(A)$ and for any $t \in \{t_x, t_y\}$.

Now, let the finite sequence be

$$a(\theta, \phi; a) \equiv \{a_0 = a, a_1 = \{x, t_x\}, a_2 = \{y, t_y\}\}.$$

Let $j(0) = j(1) = i$. We have

$$u_i(a, \theta_i) > u_i(x, \theta_i) - t_x > u_i(y, \theta_i) - t_y$$

that is, (1) in Condition C holds; morever, (2) follows from (1.6).

We show that we can make use of lotteries to decrease the large payments into an arbitrarily small scale.

Recall that for any distinct types θ_i and θ_i', there exists a pair of lotteries $\left\{x_{\theta_i, \theta_i'}, x_{\theta_i', \theta_i}\right\}$ such that

$$u_i(x_{\theta_i, \theta_i'}, \theta_i) > u_i(x_{\theta_i', \theta_i}, \theta_i);$$

$$u_i(x_{\theta_i, \theta_i'}, \theta_i') < u_i(x_{\theta_i', \theta_i}, \theta_i').$$

For any $\bar{t} > 0$, we can find some small enough $p_a > 0$, such that there exists $t < \bar{t}$,

$$u_i\left((1-p_a)a + p_a x_{\theta_i,\theta_i'}, \theta_i\right) - t > u_i((1-p_a)a + p_a x_{\theta_i',\theta_i}, \theta_i) - t;$$

$$u_i((1-p_a)a + p_a x_{\theta_i,\theta_i'}, \theta_i') - t < u_i\left((1-p_a)a + p_a x_{\theta_i',\theta_i}, \theta_i'\right) - t.$$

In our mechanism, the finite sequence is

$$a\left(\theta, \phi; a\right) \equiv \left\{a_0 = a, a_1 = \left\{(1-p_a)a + p_a x_{\theta_i,\theta_i'}, -t\right\}, a_2 = \left\{(1-p_a)a + p_a x_{\theta_i',\theta_i}, -t\right\}\right\}.$$

Chapter 2

Robust Dynamic Implementation

2.1 Introduction

Consider a society consisting of a group of individuals. Assume that this society agrees upon some *social choice rule* (or welfare criterion) as a mapping from states to outcomes where each state can be interpreted as the relevant information needed to pin down desirable outcomes at that state. Then, the theory of *implementation* and *mechanism design* poses the following institutional design question: what class of social choice rules can be realized by mechanisms (institutions)? The answer to this question precisely relies on how we hypothesize about the following two ingredients: (1) what class of mechanisms are we allowed to use? (2) how does each agent behave in the mechanism? It is already well known in the literature that one can obtain very permissive implementation results by using *dynamic (or sequential)* mechanisms and exploiting the

assumption of *complete information*. In complete information environments, Moore and Repullo (1988) construct a dynamic mechanism (henceforth, the MR mechanism) that implements "any" social choice rule as the unique subgame perfect equilibrium.

Subgame perfect implementation is particularly successful because it shows that most desirable outcomes are in fact uniquely implementable as subgame perfect equilibria. Nevertheless, there remain several criticisms: (1) It relies excessively on the agents' rationality. For deviations are always considered to be "one-shot deviations from rationality" that do not shatter the faith players have in the subsequent rationality of their opponents; (2) The punishment of all agents is often needed out of the equilibrium in the mechanism and this is clearly not in their collective interest: what if the agents decided to abandon the original mechanism after a Pareto inefficient outcome is realized as an out-of-equilibrium outcome and they renegotiate this into a new Pareto efficient outcome? (3) The introduction of even small information perturbations greatly reduces the power of subgame perfect implementation. Aghion, Fudenberg, Holden, Kunimoto, and Tercieux (2012, henceforth, AFHKT) show that under arbitrarily small information perturbations the MR mechanism does not yield (even approximately) truthful revelation and that in addition the mechanism has sequential equilibria with undesirable outcomes.

The main objective of this paper is to provide very permissive robust implementation results via dynamic mechanisms. More specifically, this paper proposes a two-stage mechanism which (1) has a unique truth-telling sequential equilibrium in pure strategies that is robust to any "private-value perturbation"; (2) is dominance-solvable in the weakest notion of "sequential rationalizability"; (3) is immune to renegotiation. Before getting into the details, from the outset, we want to be clear about the domain of problems to which our results apply. First, we consider environments where monetary transfers among the players are available and all players have quasilinear utilities in money. 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 and Morgenstern expected utility. Third, we focus on private values environments. That is, each player's utility depends only upon his own payoff type as well as the lottery chosen and his monetary payment.

In a dynamic mechanism, agents could have multiple beliefs, one at each information set. These beliefs are updated via Bayes' rule whenever possible; however, if an agent is surprised by a zero-probability event, Bayesian updating does not apply and the agent needs to revise her belief in another

fashion. The assumption on how this belief revision proceeds is precisely what distinguishes different existing solution concepts for dynamic games. Subgame perfection equilibrium entails *backwards induction*, which requires that there be rationality and common belief in rationality at "every" information set. This means that under backwards induction, each agent always attributes any out-of-equilibrium behavior of the opponents to mere mistakes and maintains her initial hypothesis of rationality and common belief in rationality in the subsequent stages of the game. Following Ben-Porath (1997), Dekel and Siniscalchi (2013) introduce the concept of *initial rationalizability*, which we take as this paper's solution concept in extensive form games. Initial rationalizability is like rationalizability in normal-form games in that it iteratively deletes strategies that are not best replies. Unlike backwards induction, initial rationalizaiblity only requires that there be rationality and common belief in rationality "at the beginning of the game." Accommodating any belief revision assumption at any subsequent stages of the game after a zero-probability event occurs, we acknowledge that initial rationalizability is the weakest rationalaizability concept among all in extensive-form games. Hence, implementation under initial rationalizability is the most robust concept of implementation among the existing concepts for implementation in dynamic mechanisms.

Our first result shows that one can construct a two-stage mechanism which

implements any social choice function under initial rationalizability. The requirement of initial rationalizable implementation can be decomposed into the following two parts: (1) there always exists an initial rationalizable strategy profile whose outcome coincides with the given rule; (2) there are no initial rationalizable strategy profile whose outcomes differ from those of the rule.

Since complete information entails common knowledge of states, which is very demanding and at best taken to be a simplifying assumption, it is a sensible exercise to ask for the robustness of the implementation results to small amounts of incomplete information. To pursue this line of research, we are motivated by the approach of Chung and Ely (2003), who consider the following scenario: if a planner is concerned that 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 closed graph. Specifically, our second result shows that it is possible to construct a finite two-stage mechanism which not only fully implements any social choice function under complete information but also does so in all the nearby environments. Therefore our result generates the following important corollary: any social choice function is implementable for all types in the model under study and it continues to be implementable for all types "close" to this initial model. Therefore, any social choice function

is *continuously* implementable in dynamic mechanism where the concept of continuity here is the same as the one proposed by Oury and Tercieux (2012). This robustness result still holds if we instead adopt other solution concepts such as subgame perfect equilibrium, subgame rationalizability (Bernheim (1984)), and extensive form rationalizability (Pearce (1984)) because these are simply the refinements of initial rationalizability.

Our results narrow several open questions in the literature. First, we contribute to the literature of rationalizable implementation. Bergemann, Morris, and Tercieux (2011) investigate the implications of rationalizable implementation by employing *infinite, static, stochastic* mechanisms. They show that strict *Maskin monotonicity* is a necessary condition. Note that Maskin monotonicity is known to be a necessary condition for Nash implementation.[1] Moore (1992) proposes a simple sequential mechanism where every player moves only once. His result does not rely excessively on the agents' rationality, since even when some player is surprised by his opponent's behavior, it does not matter whether he believes the one who surprised him is rational or not. However, there is a cost associated with it: his simple sequential mechanism needs large size of monetary penalties and this mechanism works only under a stringent condition on the environment. Moore (1992) argues that the most natural

[1]See Maskin (1999) for this.

examples where his simple mechanism works are either only one of the two players has a state dependent preference, or both of their preferences are perfectly correlated. Clearly, the applicability of his result is very limited. On the other hand, we obtain a very permissive implementation result in much more general environments: any social choice function is fully implementable under initial rationalizability by a finite dynamic mechanism.

Second, we contribute to the literature of the robustness of the implementation results to almost complete information. For instance, Chung and Ely (2003) investigate the robustness of undominated Nash implementation and AFHKT (2012) investigate the robustness of subgame perfect implementation. Exploiting "interdependent" values perturbations, they both conclude that Maskin monotonicity is a necessary condition for their robust implementation. We investigate the robustness of implementation under initial rationalizaibility. Our result shows that any social choice function is robustly implementable under "private" values perturbations. As shown by Qin and Yang (2013), the perturbations in Chung and Ely (2003) and AFHKT (2012) are both consider as order two perturbation; in contrast, our positive result extends to any high order perturbation in universal type space.

Third, we contribute to the literature of implementation with renegotiation. We sometimes interpret a mechanism as a *contract* between the agents. In

this case, they will presumably choose a mechanism that will deliver a Pareto efficient outcome in equilibrium. Suppose, for whatever the reason may be, that play of the mechanism results in an out-of-equilibrium outcome and this outcome is not Pareto efficient. Then, it is very likely that the agents tear up the contract and negotiate a new Pareto efficient outcome. To prevent this type of ex post renegotiation, Maskin and Tirole (1999) consider the buyer-seller bilateral trading model and assume that the agents sign a contract that uses a stochastic transfer from the seller to the buyer when out-of-equilibrium outcome is realized. If the buyer is risk-averse, then this fine can be designed so that it hurts both the seller and the buyer. However, this construction does not work for risk-neutral parties. If parties are close to risk-neutral, the stochastic fine that is required needs to have a very large variance, which is not very credible as it will violate the wealth constraints. Thus the applicability of their result is doubtful in this case. Our mechanism adopts the idea of Abreu and Matsushima (1992) to transform the required large payments into arbitrarily small scale. This makes our mechanism a lot more reasonable than that of Maskin and Tirole (1999).

Fourth, we contribute to the literature of the hold-up problem. It is often the case that, when two parties engage in a relationship, they are uncertain about the values of some parameter which will affect their future payoffs. This

uncertainty is represented by a set of parameters that take several values. Although they will both learn the value of the parameter in the future, they cannot write ex ante contracts contingent on the state of nature because this state of nature is not verifiable by a third party. When two parties sign an ex ante contract based on some parameter which will be realized ex post but not verifiable by a third party, it might entail *transaction cost* (Williamson (1975)). However, the mechanism we develop here can be used to ensure that truthful revelation occurs in equilibrium. Therefore the unverifiability alone does not create any transaction cost.

Our paper is also related to the literature motivated by King Solomon's dilemma. Qin and Yang (2009) provide a two-stage dynamic mechanism to implement the social desired allocation in one round deletion of weakly dominated strategy followed by iterative deletion of strictly dominated strategies. They allow the information is incomplete among players and use an infinite mechanism (the second stage they adopt second price auction to elicit players' true type). When we focus the complete information environment, we can adopt a much weaker solution to achieve the social desired allocation. The common feature is that both their mechanism and ours are robust to private value perturbations.

The robust dynamic implementation literature is also closely related to

our work. Müller (2013b) studies robust virtual implementation using dynamic mechanism under common strong belief in rationality. Müller (2013a) adopts the same solution as ours to study robust dynamic implementation. The difference between the robustness notion and ours is that instead of pursuing a mechanism to work in any type space, we focus on the bench mark type space and consider the class of type space around it.

The rest of the paper is organized as follows: Section 2 uses a simple buyer-seller example to introduce the MR mechanism and the general criticism on it. Then within the same example, we construct a two-stage mechanism which is immune to many of the criticisms. In Section 3, we introduce the preliminary notation and definitions. Section 4 provides our main results. More specifically, we establish Theorem 1 for implementation under initial rationalizability (Section 4.1); Theorem 2 for robust implementation to small perturbations around the benchmark model (Section 4.2); Corollary 1 for the robust implementation to small perturbations around complete information; and Theorem 3 for implementation with arbitrarily small transfers (Section 4.3). In Section 5 we discuss several issues. First, we show that it is possible to provide a perfect information mechanism based on the MR mechanism that not only implements any social choice function under complete information but also does so in all the nearby environments. However, the implementation

under information perturbations is successful only if the players adopt mixed strategies in the unique sequential equilibrium. This casts doubt on how the MR mechanism being played by the real people because it is not cognitively simple at all for a player to play mixed strategies. Finally, we propose a way of making the transfer rule satisfying budget balance when there are at least three individuals.

2.2 Illustration

To illustrate the main idea of this paper, we consider the following simple example adapted from Hart and Moore (2003). There are two parties, a B(uyer) and a S(eller) of a single unit of an indivisible good. If trade occurs then B's payoff is

$$V_B = \theta - p,$$

where p is the price and θ is the good's quality. S's payoff is

$$V_S = p,$$

thus we normalize the cost of producing the good to zero.

The good can be of either high or low quality. If it is high quality then B values it at $\theta_H = 14$, and if it is low quality then B values it at $\theta_L = 10$. We seek to implement the social choice function f^* whereby the good is always

traded ex post, and where the buyer always pays the true value of θ to the seller.

2.2.1 Moore-Repullo Mechanism

Suppose first that the quality θ is observable and common knowledge to both parties. The implementation of f^* can be achieved through the following Moore-Repullo (MR) mechanism:

(1) B announces either a "high" or "low" quality. If B announces "high" then B pays S a price equal to 14 in exchange of the good and the game stops.

(2) If B announces "low" and S does not "challenge" B's announcement, then B pays a price equal to 10 and the game stops.

(3) If S challenges B's announcement then:

- **(a)** B pays a fine $F = 9$ to T (a third party)
- **(b)** B is offered the good for 6
- **(c)** If B accepts the good then S receives F from T (and also a payment of 6 from B) and the game stops.
- **(d)** If B rejects at $3b$ then S pays F to T.
- **(e)** B and S each get the item with probability 1/2 and the game stops.

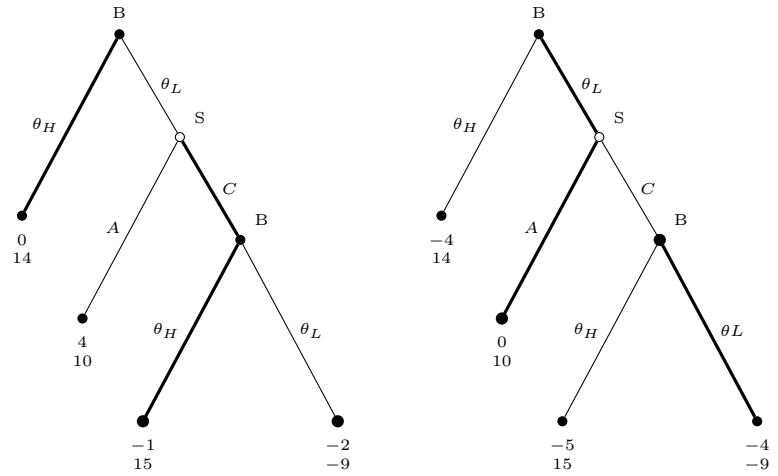


Figure 2.1: The left is under θ_H and the right is under θ_L. The equilibrium path is in boldface.

The game specified by the MR mechanism under different values are shown in Figure 1. The MR mechanism is extremely successful because it shows that most desirable outcomes are in fact implementable as a unique subgame-perfect equilibrium. However, the way MR mechanism delivers such a good performance is subject to several criticisms. First, the solution concept of backwards induction relies excessively on the assumption of common belief of rationality. For deviations are always considered to be "one-shot deviations from rationality" that do not shatter the faith players have in the subsequent rationality of their opponents (see Reny (1992), Ben-Porah (1997) and Battigali and Siniscalchi (1999) for more details on the criticisms on backwards induction). Second, the punishment of all agents is often needed out of the equilibrium in the mechanism and this is clearly not in their collective inter-

est. This feature is particularly problematic if the agents can decide ex post to abandon the original mechanism after a Pareto inefficient outcome is realized as an out-of-equilibrium outcome so that they can renegotiate for a new Pareto efficient outcome (see Laffont and Martimort (2002) and Bolton and Dewatripont (2005) for the detailed discussion on renegotiation). Third, the introduction of even small information perturbations greatly reduces the power of subgame perfect implementation. In particular, AFHKT (2012) show that under arbitrarily small information perturbations, the MR mechanism does not yield (even approximately) truthful revelation and that in addition, the mechanism has sequential equilibria with undesirable outcomes.

The first two criticisms are well known and we see no strong reason to illustrate them. However, we would like to illustrate the issues that come the last criticism. To do so, we first make a brief review of AFHKT (2012). Players have imperfect information about θ, which is generated from a common prior μ with $\mu(\theta_H) = 1-\alpha$ and $\mu(\theta_L) = \alpha$ for some $\alpha \in (0,1)$. Each player receives a draw from a signal structure with two possible signals s^h or s^l, where s^h is a high signal that is associated with θ_H, and s^l is a low signal associated with θ_L. We use the notation $s_B = s_B^h$ (resp. $s_B = s_B^l$) to refer to the event in which B receives the high signal s^h (resp. the low signal s^l) and similar notation applies to S. The following table shows the joint probability distribution μ :

μ	s_B^h, s_S^h	s_B^h, s_S^l	s_B^l, s_S^h	s_B^l, s_S^l
θ_H	$1-\alpha$	0	0	0
θ_L	0	0	0	α

Let ν^ε denote a perturbed information structure:

ν^ε	s_B^h, s_S^h	s_B^h, s_S^l	s_B^l, s_S^h	s_B^l, s_S^l
θ_H	$(1-\alpha)(1-\varepsilon-\varepsilon^2)$	$(1-\alpha)\varepsilon$	$\frac{(1-\alpha)\varepsilon^2}{2}$	$\frac{(1-\alpha)\varepsilon^2}{2}$
θ_L	$\frac{\alpha\varepsilon^2}{2}$	$\frac{\alpha\varepsilon^2}{2}$	$\alpha\varepsilon$	$\alpha(1-\varepsilon-\varepsilon^2)$

Note that as ε converges to 0, the marginal probability distribution of ν^ε on θ coincides with μ. That is, each player's signal is almost correct under ν^ε. The second feature of ν^ε is that when the agents receive different signals, B's signal becomes infinitely more accurate than S's. This implies that when S and B were informed of the signal, and the signals disagree, they will conclude that with high probability the true state corresponds to B's signal. This matters a lot when S decides whether to challenge B.

AFHKT (2012) first show that truth telling cannot be (even approximately) an equilibrium in pure strategies. This is easy to see in the previous example: if S does not challenge when observing low signal, B would like to announce "low" regardless of the signal he received. They also show that even allowing for mixed strategies, the probability of truthful announcement never goes to 1 as ε goes to 0 (see Proposition 1 in AFHKT (2012) for details). Furthermore,

under this information structure, there exists a persistently bad sequential equilibrium. Suppose that B always announces "high" regardless of the signal received. S always challenges when observing "low" regardless of her signal too. In the last stage, B accepts the offer when his signal is high, and rejects it otherwise. B holds his posterior belief given his private information and the initial prior. We specify the following belief system at the last stage of the game: S believes with probability 1 that B received high signal. Sequential rationality is easy to check with this belief system, which is also consistent indeed.

2.2.2 Two-Stage Mechanism

We will provide a sequential mechanism that implements the social choice function f^* under complete information. We also show that f^* is implementable under all the nearby environments. We define the mechanism as follows.

(1) Both B and S announce "high" or "low" simultaneously. If both of them announce "high" then B pays S a price equal to 14 in exchange of the good and the game stops; if both of them announce "low" then B pays S a price equal to 10 in exchange of the good and the game stops.

(2) If B announces differently from S's announcement then:

(a) B pays a fine $F = 9$ to T (a third party).

(b) B is offered the good for the price of 6.

(c) If B accepts the good, then S receives F from T (and also a payment of 6 from B) and the game stops.

(d) If B rejects the offer made at $2\,(b)$, then S pays F to T.

(e) B and S each get the item with probability $1/2$ and the game stops.

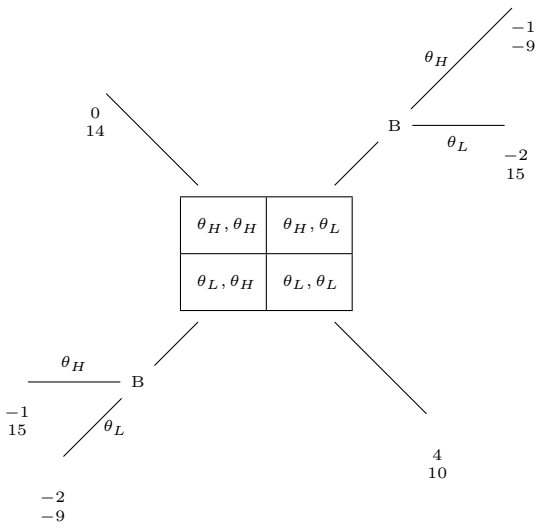


Figure 2.2: The payoff is specified by the mechanism under θ_H.

First, we focus on complete information environments. The game specified by two-stage mechanism under θ_H is shown in Figure 2. Since B is the sole player in the second stage, knowing the state is high, he will accept the offer. Therefore, the two stage game collapses into the following normal form game:

	s_S^h	s_S^l
s_B^h	$0, 14$	$-1, -9$
s_S^l	$-1, 15$	$0, 10$

Apparently, announcing high is a strictly dominant strategy for S. Knowing this, B will announce high too. Similarly, when the state is low, S and B will coordinate on low.

There is only one active player, B, in the second stage and B's choice only depends on her own type. This structure delivers a lot of advantages over the existing mechanisms. First, we discuss the agents' rationality in this mechanism. Although the active player might be surprised by other players' moves in the previous history, he will play in a rational way. Then, we can show that when we adopt the solution which requires only rationality and initial common belief of rationality, the outcome still coincides with the one induced by subgame perfect equilibrium.

Second, we discuss the role of information perturbations. As long as the active player almost knows his own payoff type (recall also that the buyer has infinitely more accurate information than the seller) then it will not change the outcome from subgame perfect equilibrium as well. Finally, all the payments needed out of the equilibrium in this mechanism can be reduced to arbitrarily small scale by adopting the idea from Abreu and Matsushima (1992), the details of which will be discussed in Section 2.6.

The key insights of this two-stage mechanism is as follows. First, we merge the first two stages in the MR mechanism into one. This prevents the information leakage from the buyer to the seller, which is the very reason that the MR is not robust to even small information perturbations (see AFHKT (2012)). Since S will not be influenced by whatever B announces, she will make decisions based on his own posterior belief, which is almost accurate when information is almost complete. Second, there is only one active player in the last stage. This construction makes the mechanism work with the least requirement of the active player. B chooses his type based on his own rationality no matter how he is surprised by S's previous choice. When the information structure is only slightly perturbed, since we consider the perturbation where B almost knows his own type regardless of what signal received by S, B will behave in the same way regardless of whether he knows S's signal or not.

The mechanism can be generalized to the environments where there are I players and each player has J types. Our mechanism is still a two-stage mechanism, while the MR mechanism needs I phases, which has $3I$ stages in total. To avoid some technical details involved in more general mechanisms, we will postpone the formal result until Sections 2.4, 2.5 and 2.6.

2.3 Preliminaries

2.3.1 The Environment

Let I denote a finite set of players and with abuse of notation, we denote by I the cardinality of I. Assume also that $I \geq 2$. The set of simple lotteries over an arbitrary set of outcomes is denoted by A. We assume that players' values are private. That is, the utility index of player i over the set A is denoted by a bounded utility function $u_i : A \times \Theta_i \to \mathbb{R}$, where Θ_i is the finite set of payoff types and $u_i(a, \theta_i)$ specifies the utility of player i from the social alternative $a \in A$ under $\theta_i \in \Theta_i$. We assume that any two distinct types θ_i and θ_i' induce different preference orders over A and there is no total indifference over the outcomes under any θ_i. 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 $\tau_i \in \mathbb{R}$.

Lemma 2.1. *(Abreu and Matsushima (1992)) For each $i \in I$, there exists a function $x_i : \Theta_i \to A$ such that for any $\theta_i, \theta_i' \in \Theta_i$ with $\theta_i \neq \theta_i'$,*

$$u_i\left(x_i\left(\theta_i\right), \theta_i\right) > u_i\left(x_i\left(\theta_i'\right), \theta_i\right).$$

Let $\overline{u} = \sup_{i,a,\theta_i} u_i(a, \theta_i)$ be a uniform upper bound of all players' utility functions. Similarly, let $\underline{u}$ be a uniform lower bound of all players' utility functions. We can choose a large enough money $D \in \mathbb{R}_+$ such that, $D > \overline{u} - \underline{u}$.

Until the end of Section 2.5, we assume that the true type profile $\theta^* \in \Theta$ is commonly known to the players but unknown to the planner. This is what we mean by *complete information* environments. We consider a *planner* who aims to implement a *social choice function* $f : \Theta \to A$. We assume that the planner can fine or reward a player $i \in I$ and denote by τ_i the transfer from the planner to player i. Throughout the paper, we define a *dynamic mechanism* as a multistage with observed actions, which means that at each history h, all players know the entire history of the play, and if more than one player moves at h, they do so simultaneously. The class of mechanisms we consider in the present paper is exactly the same as the one AFHKT (2012) allowed. A dynamic mechanism is then an extensive game form $\Gamma = (\mathcal{H}, M, \mathcal{Z}, g)$ where (1) $\mathcal{H}$ is the set of all histories; (2) $M = \times_{i \in I} M_i$, $M_i = \times_{h \in \mathcal{H}} M_i(h)$ for all $i \in I$ where $M_i(h)$ denotes the set of available messages for i at history h; (3) $\mathcal{Z}$ describes the history that immediately follows history h given that the strategy profile m has been played; and (4) g is the outcome function that maps the set of terminal histories into the set of lotteries $\Delta(A)$ with a transfer profile $\tau = (\tau_1, \tau_2, ..., \tau_I)$.

Let $\Gamma(\theta)$ denote an extensive form game associated with dynamic mechanism Γ at state θ. Let $\sigma_i : \Theta_i \to M_i$ be a strategy of player i. Let Σ_i denote the set of strategies of player i and $\Sigma = \times_{i \in I} \Sigma_i$ denote the set of strategy

profiles. A *solution concept* is a correspondence $S : \Theta \rightrightarrows \Sigma$ as a mapping from states to a subset of strategies. The outcome correspondence associated with a solution S is a mapping O_S from Θ to $A \times \mathbb{R}^I$ with the following property: $O_S(\theta)) = \left\{(a, \tau) \in A \times \mathbb{R}^I | \exists\ m \in S(\Gamma(\theta)) \text{ s.t. } g(m) = (a, \tau)\right\}$ for each $\theta \in \Theta$.

We say that a mechanism Γ implements a social choice function f via a solution concept S, if $O_S(\theta)) = f(\theta)$ for all $\theta \in \Theta$. Then, f is said to be implementable via the solution S if there exists a mechanism Γ which implements it via the solution S.

2.3.2 Mechanism

We shall construct a two stage finite dynamic mechanism and call it Γ^*.

The outcome

Let $1 = I + 1$.

First Stage: Each player i announces a pair of types, his own and player $i-1$'s, that is

$$m_i = \left(m_i^0, m_i^1\right),$$

where $m_i^0 \in M_i^0 = \Theta_i$ and $m_i^1 \in M_i^1 = \Theta_{i-1}$. We write $m^1 = \left(m_i^1\right)_{i \in I}$.

If $m_i^0 = m_{i+1}^1$, for all $i \in I$, then $f\left(m^1\right)$ is implemented. STOP. Otherwise, we proceed to the Second Stage.

Second Stage: Let $i^* = \min_{1 \le i \le I} \left\{ i \in I | m_i^0 \neq m_{i+1}^1 \right\}$. Player i^* announces one of his types, that is,

$$m_{i^*}^2 \in M_{i^*}^2 = \Theta_{i^*},$$

and $x_{i^*}\left(m_{i^*}^2\right)$ is implemented. STOP. Recall that $x_{i^*} : \Theta_{i^*} \to A$ is constructed as in Lemma 1.

The transfer rule

The transfers are specified as follows:

- Player i^* pays a penalty $(I + 1 - i^*) \times D$.
- If $m_{i^*+1}^1 = m_{i^*}^2$ then player $i^* + 1$ gets a reward $(I + 1 - i^*) \times D$;
- if $m_{i^*+1}^1 \neq m_{i^*}^2$ then player $i^* + 1$ pays a penalty $(I + 1 - i^*) \times D$.

This two-stage mechanism is quite simple. In the first stage, each player i announces a pair of types in the first stage, his own type and his predecessor's type ($i - 1$'s). For player i's type, if i's own announcement about his type is the same as his successor $(i+1)$'s announcement about i (i.e., $m_i^0 = m_{i+1}^1$), we say this player's announcement is *consistent*. If every player's announcement is consistent, then we implement $f(m^1)$.

Otherwise, we have a nonempty set of players whose announcements are not consistent. We pick the smallest index of this set of players, denoted by

i^* who is the sole active player at the second stage and makes an additional choice over the set of lotteries $\{x_{i^*}(\theta_i)\}_{\theta_i \in \Theta_i}$. Then the lottery based on his choice is implemented.

The transfers in this mechanism is specified in a straightforward way. First, player i^* is penalized by $(I+1-i^*)D$ because he is the smallest index which exhibits an inconsistent announcement. Second, whether player i^*+1 is penalized or rewarded depends upon his announcement about i^* (i.e., $m^1_{i^*+1}$) and player i^*'s second stage announcement (i.e., $m^2_{i^*}$): if player i^*+1 made the same announcement for i^* as i^*'s second stage announcement (i.e., $m^1_{i^*+1} = m^2_{i^*}$), player i^*+1 will be rewarded by $(I+1-i^*)D$; if player i^*+1 made a different announcement from i^*'s second stage announcement (i.e., $m^1_{i^*+1} \neq m^2_{i^*}$), player i^*+1 will be penalized by $(I+1-i^*)D$.

The size of the transfer is designed in a decreasing way with respect to the index, while the priority of being player i^* is given to the smaller index. This construction will prevent players from triggering or not triggering the second stage with the intention that he will be involved in the pair at the later stage.

2.4 Complete information

2.4.1 Solution and implementation

A mechanism Γ together with a type profile θ defines a two-stage game denoted by $\Gamma(\theta)$. The game proceeds as follows. At the initial history $\emptyset$, each player chooses a message m_i from his message space $M_i(\emptyset) = \Theta_i \times \Theta_{i-1}$, and we write m for the message profile obtained at the first stage. Given any m, there are two possibilities: (1) the game ends; (2) or the game proceeds to the second stage, where there is a unique player "i^*", who makes a choice out of his message space, $M_{i^*}(m) = \Theta_{i^*}$. Let $M[i]$ denote the set of histories after which player i is picked as the unique player "i^*". We write $m[i] \in M[i]$, and $M_i(m[i]) = \Theta_i$ for all $m[i] \in M[i]$.

Formally, each player's strategy is a function

$$\sigma_i : \{\emptyset\} \cup M[i] \to M_i(\emptyset) \cup \bigcup_{m_i[i] \in M[i]} M_i(m[i])$$

where $\sigma_i(\emptyset) = \left\{\sigma_i^i(\emptyset), \sigma_i^{i-1}(\emptyset)\right\} \in M_i(\emptyset) = \Theta_i \times \Theta_{i-1}$ and $\sigma_i(m[i]) \in M_i(m[i]) = \Theta_i$. Given $\Gamma(\theta)$, conditional on history $h \in \mathcal{H}$, player i's payoff from a strategy profile σ is given by

$$v_i(\sigma, \theta_i | h) = u_i(g(\sigma(\theta); h), \theta_i) + \tau_i(\sigma(\theta)).$$

In particular, conditional on $m[i]$ (a history where i's the player i^*), player i's payoff from a strategy σ_i is given by

$$v_i\left(\sigma_i, \theta_i | m\left[i\right]\right) = v_i\left(\sigma, \theta_i | m\left[i\right]\right) = u_i\left(x_i\left(\sigma_i^2(m\left[i\right])\right), \theta_i\right) + \tau_i\left(m\left[i\right]\right).$$

In order to analyze players' reasoning at each point in the game, it is necessary to adopt a model of conditional beliefs. Following Ben-Porath (1997) (see also Battigalli and Siniscalchi, 1999), we adopt the following notion, originally proposed by Renyi (1955).

Definition 2.1. *Fix a measurable space $(\Omega, \mathcal{X})$ and a countable collection $\mathcal{B} \subset \mathcal{X}$. A conditional probability system, or CPS, is a map $\mu : \mathcal{X} \times \mathcal{B} \to [0,1]$ such that:*

1. *For each $B \in \mathcal{B}$, $\mu\left(\cdot|B\right) \in \Delta\left(\Omega\right)$ and $\mu\left(B|B\right) = 1$.*

2. *If $A \in \Sigma$ and $B, C \in \mathcal{B}$ with $B \subset C$, then $\mu\left(A|C\right) = \mu\left(A|B\right) \cdot \mu\left(B|C\right)$.*

The set of CPSs on (Ω, Σ) with conditioning events $\mathcal{B}$ is denoted $\Delta^{\mathcal{B}}\left(\Omega\right)$.

Let $\mathcal{B}_{-i}$ a collection of Σ_{-i} and $\Omega = \Sigma_{-i}$. Due to the simplicity of the two stage game, this is enough to characterize any conditional belief system.

Note that the unique active player after the first stage makes his choice purely based on his own payoff type. Therefore, what kind of initial belief he holds has nothing to do with his choice, as long as he knows his own payoff type. This point is straightforward in the following definition. This will be clear when we introduce the details of our mechanism.

Definition 2.2. *(Sequential Rationality) Fix a player $i \in I$, a CPS $\mu \in \Delta^{\mathcal{B}_{-i}}(\Sigma_{-i})$ and a strategy $\sigma_i \in \Sigma_i$. Say that σ_i is a sequential best response to μ iff, for all $\sigma_i' \in \Sigma_i$, for all $h \in \mathcal{H}$,*

$$\sum_{\sigma_{-i}} v_i(\sigma, \theta_i|h)\, \mu[\sigma_{-i}|\Sigma_{-i}] \geq \sum_{\sigma_{-i}} v_i((\sigma_i', \sigma_{-i}), \theta_i|h)\, \mu[\sigma_{-i}|\Sigma_{-i}].$$

We represent the definition of initial rationalizability given by Dekel and Siniscalchi (2013). The epistemic foundation is rationality and initial common belief of rationality, which is provided by Ben-Porath (1997) to study perfect information games. The solution can be characterized via an iterative deletion algorithm in Battigalli and Siniscalchi (1999) which deals with general multistage games.

Definition 2.3. *(Initial Rationalizability) Fix a multistage game $\Gamma(\theta)$. For every player $i \in I$, let $R_{i,0}^{\Gamma(\theta)} = \Sigma_i$. Inductively, for every integer $k > 0$, let $R_{i,k}^{\Gamma(\theta)}$ be the set of strategies $\sigma_i \in \Sigma_i$ that are sequential best replies to a CPS $\mu \in \Delta^{\mathcal{B}_{-i}}(\Sigma_{-i})$ such that $\mu\left(R_{-i,k-1}^{\Gamma(\theta)}|\Sigma_{-i}\right) = 1$. Finally, the set of initially rationalizable strategies for i is $R_i^{\Gamma(\theta)} = \bigcap_{k=1}^{\infty} R_{i,k}^{\Gamma(\theta)}$.*

Definition 2.4. *A social choice function f is implementable in initial rationalizable strategies if there exists a mechanism Γ such that, for all θ and $m \in M$, $R^{\Gamma(\theta)} \neq \varnothing$ and $m \in R^{\Gamma(\theta)} \Rightarrow g(m) = f(\theta)$.*

2.4.2 Main result

Theorem 2.1. *If $I \geq 2$, any social choice function f is implementable in initial rationalizable strategies.*

We use the following claims to prove Theorem 2.1.

Claim 2.1. *If $\sigma_i \in R_{i,1}^{\Gamma(\theta)}$, then $\sigma_i(m[i]) = \theta_i$.*

Proof. From Lemma 2.1,

$$u_i(x_i(\theta_i), \theta_i) + \tau_i(m[i]) > u_i(x_i(\theta_i'), \theta_i) + \tau_i(m[i]),$$

for any $\theta_i' \neq \theta_i$. □

As the game proceeds to the second stage, the outcome of the game purely depends on player i^*'s choice. From the construction of the mechanism, the transfers to (or from) player i^* is regardless of his choice and the choice of player i^* is purely over his own payoff types. The Lemma 2.1 guarantees that every player i will truthfully reveals his own payoff type whenever $i = i^*$.

Claim 2.2. *If $\sigma_2 \in R_{2,2}^{\Gamma(\theta)}$, then $\sigma_2^1(\emptyset) = \theta_1$.*

Proof. If $\sigma_2 \in R_{2,2}^{\Gamma(\theta)}$, then $\mu_2\left(\left(R_{j,1}^{\Gamma(\theta)}\right)_{j\neq 2} | \Sigma_{-2}\right) = 1$, particularly,

$$\mu_2^1(\sigma_1(m[1]) = \theta_i | \Sigma_{-2}) = 1$$

.(We write μ_i^j for i's belief over Σ_j, μ_i for i's belief over Σ_{-i}.)

Consider the following two cases:

Case 2.1 $(\sigma_1^1(\emptyset) = \theta_1)$**.** *If* $\sigma_2^1(\emptyset) \neq \theta_1$*, then the game proceeds to the second stage and player* 1 *is the player* i^**. From Claim 2.1, player* 1 *will announce* θ_1 *in the second stage. Since* $\sigma_2^1(\emptyset) \neq \sigma_1(m[1])$*, player* 2 *gets punished by* nD *and the outcome is* $x_1(\theta_1)$*. If* $\sigma_2^1(\emptyset) = \theta_1$*, the game will proceed in the following possible ways: (1) Player* 2 *is the player* i^**. Player* 2 *gets punished by* $(I-1)D$ *and the outcome is* $x_2(\theta_2)$*. (2) Player* 2 *is not the player* i^* *and player* 2 *gets neither reward nor penalty and the outcome is in* A*.*

In case (1), for player 2*, consider* $\{ID, x_1(\theta_1)\}$ *and* $\{(I-1)D, x_2(\theta_2)\}$*, the potential gain from the different outcomes is bounded by the loss from the different penalties by the construction of* D*. That is, player* 2 *gets strictly better since the penalty is less. In case (2), it is straightforward that player* 2 *gets strictly better since he avoids the penalty.*

Case 2.2 $(\sigma_1^1(\emptyset) \neq \theta_1)$**.** *If* $\sigma_2^1(\emptyset) = \theta_1$*, the game proceeds to the second stage and from Claim 2.1 player* 2 *gets rewarded by* nD *and the outcome is* $x_1(\theta_1)$*. This is uniquely the best player* 2 *can expect in this game by the construction of* D*. Obviously, any announcement of player* 1*'s type rather than* θ_1 *delivers a strictly worse payoff to player* 2*.*

Therefore, for player 2*, it is a strictly dominated strategy to announce player* 1*'s type,* θ_1*.*

This completes the proof of Claim 2. □

Claim 2.3. *If* $\sigma_1 \in R_{1,3}^{\Gamma(\theta)}$, *then* $\sigma_1^1(\emptyset) = \theta_1$.

Proof. If $\sigma_1 \in R_{1,3}^{\Gamma(\theta)}$, then $\mu_1^2(\sigma_2^1(\emptyset) = \theta_1 | \Sigma_{-1}) = 1$. If $\sigma_1^1(\emptyset) \neq \theta_1$, then the game proceeds to the second stage and player 1 is the player i^*. Player 1 gets punished by nD and the outcome is $x_1(\theta_1)$. If $\sigma_1^1(\emptyset) = \theta_1$, the game will proceed in the following possible ways: (1) Player I is the player i^*. The worst situation for player 1 is that player 1 gets punished by D the outcome is $x_I(\theta_I)$. This happens when I is the player i^* and player 1 announces $\sigma_1^I(\emptyset) \neq \theta_I$; (2) Player I is not the player i^*. Thus player 1 gets neither reward nor penalty. Clearly, in either case, player 1 gets strictly better if he announces θ_1 rather than any other type. □

Claim 2.4. *If* $\sigma_{i+1} \in R_{i+1,2(i-1)}^{\Gamma(\theta)}$, *then* $\sigma_{i+1}^i(\emptyset) = \theta_i$; *if* $\sigma_i \in R_{i,2(i-1)+1}^{\Gamma(\theta)}$, *then* $\sigma_i^i(\emptyset) = \theta_i$.

Proof. We have established Claims 2.2 and 2.3. By induction, it suffices to show if "If $\sigma_{i+1} \in R_{i+1,2(i-1)}^{\Gamma(\theta)}$, then $\sigma_{i+1}^i(\emptyset) = \theta_i$; if $\sigma_i \in R_{i,2(i-1)+1}^{\Gamma(\theta)}$, then $\sigma_i^i(\emptyset) = \theta_i$." is true for all $i \leq j$, then "If $\sigma_{j+2} \in R_{j+2,2j}^{\Gamma(\theta)}$, then $\sigma_{j+2}^{j+1}(\emptyset) = \theta_{j+1}$; if $\sigma_{j+1} \in R_{j,2j+1}^{\Gamma(\theta)}$, then $\sigma_{j+1}^{j+1}(\emptyset) = \theta_{j+1}$." is true.

First we show that if $\sigma_{j+2} \in R_{j+2,2j}^{\Gamma(\theta)}$, then $\sigma_{j+2}^{j+1}(\emptyset) = \theta_{j+1}$. If $\sigma_{j+2} \in R_{j+2,2j}^{\Gamma(\theta)}$, by the induction hypothesis, $\mu_{j+2}\left(\sigma_{i+1}^i(\emptyset) = \sigma_i^i(\emptyset) = \theta_i, \cdot | \Sigma_{-(j+2)}\right) = 1$ for all $i \leq j$.

Consider the following two case:

Case 2.3 $(\sigma_{j+1}^{j+1}(\emptyset) = \theta_{j+1})$**.** *If* $\sigma_{j+2}^{j+1}(\emptyset) \neq \theta_{j+1}$*, then the game proceeds to the second stage and player* $j+1$ *is the player* i^**. Since* $\sigma_{j+2}^{j+1}(\emptyset) \neq \sigma_{j+1}(m[j+1])$*, player* $j+2$ *gets punished by* $(I-j)D$ *and the outcome is* $x_{j+1}(\theta_{j+1})$*. If* $\sigma_{j+2}^{j+1}(\emptyset) = \theta_{j+1}$*, the game will proceed in the following possible ways: (1) Player* $j+2$ *is the player* i^**. Player* $j+2$ *gets punished by* $(I-j-1)D$ *and the outcome is* $x_{j+2}(\theta_{j+2})$*. (2) Player* $j+2$ *is not the player* i^* *and player* $j+2$ *gets neither reward nor penalty and the outcome is in* A*. In either case, player* $j+2$ *gets strictly better.*

Case 2.4 $(\sigma_{j+1}^{j+1}(\emptyset) \neq \theta_{j+1})$**.** *If* $\sigma_{j+2}^{1}(\emptyset) = \theta_{j+1}$*, the game proceeds to the second stage and from Claim 2.1 player* $j+2$ *gets rewarded by* $(j+1)D$ *and the outcome is* $x_{j+1}(\theta_{j+1})$*. This is the best player* $j+2$ *can expect in this game by the construction of* D*.*

Second we show if $\sigma_{j+1} \in R_{j,2j+1}^{\Gamma(\theta)}$*, then* $\sigma_{j+1}^{j+1}(\emptyset) = \theta_{j+1}$*.*

If $\sigma_{j+1} \in R_{j,2j+1}^{\Gamma(\theta)}$*, then* $\mu_{j+1}\left(\sigma_{i+1}^{i}(\emptyset) = \sigma_{i}^{i}(\emptyset) = \theta_i, \sigma_{j+2}^{j+1}(\emptyset) = \theta_{j+1}, \cdot|\Sigma_{-(j+1)}\right) = 1$*. If* $\sigma_{j+1}^{j+1}(\emptyset) \neq \theta_{j+1}$*, then the game proceeds to the second stage and player* $j+1$ *is the player* i^**. Player* $j+1$ *gets punished by* $(I-j)D$ *and the outcome is* $x_{j+1}(\theta_{j+1})$*. If* $\sigma_{j+1}^{j+1}(\emptyset) = \theta_{j+1}$*, the game will proceed in a ways such that player* $j+1$ *will be neither* i^* *nor* i^*+1*. In any possible outcome of* A*, player* $j+1$ *will get neither reward nor penalty. Therefore, player* 1 *gets strictly better.*

This completes the proof of Claim 4. □

Claim 2.5. *If* $\sigma \in R^{\Gamma(\theta)}$, *then* $\sigma_i^i(\emptyset) = \sigma_{i+1}^i(\emptyset) = \theta_i$.

Proof. This follows directly from Claim 2.4. □

As discussed in Section 2.2, one general criticism about subgame perfect implementation is that many results rely on the heavy use of the power of backwards inductions. Indeed, the mechanism employed here is immune to this criticism. Although at this point, the size of transfers needed can be large, we will show the transfers can be made arbitrarily small.

2.5 Almost complete information

2.5.1 Solution and implementation

Now we consider a situation where the designer (1) is willing to fully implement in initial rationalizable strategies, and (2) wants to implement in a continuous manner. More specifically, we require that, in any model that embeds the initial model, initial rationalizable strategy exists and any initial rationalizable strategy profile yields the desired outcome, not only at all types of the initial model but also at all types "close" to initial types. We follow Oury and Tercieux (2012) to define the notion of closeness in types, which formally described by the product topology in the universal type space, captures the restrictions on the modeler's ability to observe the players' (high order) beliefs.

A *model* $\mathcal{T}$ is a pair (T, κ), where $T = T_1 \times T_2 \times \cdots \times T_I$ is a countable type space, and $\kappa(t_i) \in \Delta(\Theta \times T_{-i})$ denotes the associated belief for each $t_i \in T_i$. Let $\kappa(t_i)[E]$ denote the probability of any measurable set $E \subset \Theta \times T_{-i}$ given by $\kappa(t_i)$. Let $\kappa_\Theta(t_i) = \text{marg}_\Theta \kappa(t_i)$, $\kappa_{T_j}(t_i) = \text{marg}_{T_j} \kappa(t_i)$, and $\kappa_{T_{-i}}(t_i) = \text{marg}_{T_{-i}} \kappa(t_i)$.

For two models $\mathcal{T} = (T, \kappa)$ and $\mathcal{T}' = (T', \kappa')$, we will write $\mathcal{T} \supset \mathcal{T}'$ if $T \supset T'$ and for every $t_i \in T_i'$, we have $\kappa(t_i)[E] = \kappa_i'(t_i)\left[T_{-i}' \cap E\right]$ for any measurable $E \subset T_{-i}$.

A planner aims to implement a social choice function that is a mapping $f : T \to \Delta(A)$, where $T = T_1 \times T_2 \times \cdots \times T_I$.

Given a model (T, κ) and any type t_i in type space T_i, the first-order belief of t_i on Θ is computed as

$$h_i^1(t_i) = \text{marg}_\Theta \kappa(t_i).$$

Second-order belief of t_i is his belief about $(\theta, h_1^1(t_1), ..., h_I^1(t_I))$, set as

$$h_i^2(t_i)[F] = \kappa(t_i)\left[\left\{(\theta, t_{-i}) \mid \left(\theta, h_1^1(t_1), ..., h_I^1(t_I)\right) \in F\right\}\right],$$

where $F \subset \Theta \times \Delta(\Theta)^I$ is a measurable set. An entire hierarchy of beliefs can be computed similarly. A type of a player i induces an infinite hierarchy of beliefs $\left(h_i^1(t_i), h_i^2(t_i), ..., h_i^l(t_i), ...\right)$. 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 $\{t_i[k]\}_{k=0}^{\infty}$ converges to a type t_i, if, for every $\ell \in \mathbb{N}$, $h_i^{\ell}(t_i)[k] \to h_i^{\ell}(t_i)$ as $k \to \infty$. We write $t_i[k] \to_p t_i$ for this class of convergent sequences.

A model $\mathcal{T} = (T, \kappa)$ is finite if each T_i is a finite set and $\text{supp}\kappa(t_i)$ is finite for each $t_i \in T_i$. In Section 2.4, we focused on the complete information environment. In this situation, given a finite set of states of nature Θ, whenever the true state is θ, it is assumed to be common belief among agents. To incorporate this in our setting, we define the complete information finite model $\bar{\mathcal{T}} = (\bar{T}_\theta, \bar{\kappa})$ such that for each player i, for any $\bar{t}_\theta \in \bar{T}_\theta, \bar{\kappa}(\bar{t}_{i,\theta})[(\theta, \bar{t}_{-i,\theta})] = 1$.

Now we are ready to define almost complete information formally. As players' values are private, when we consider incomplete information, we sometimes model the uncertainty from other players' values while each player still knows his own value. In the following lines, we consider a slightly more general incomplete information environments around complete information. Specifically, each player holds a small uncertainty about his own payoff types, that is, each player almost knows his own payoff type. In addition, whether or not some player knows other players' type will not change his conjecture over his own payoff type. We write $\kappa(t_i)[\theta_i] = \text{marg}_{\Theta_i}\kappa(t_i)[\theta_i]$ for the belief on his own payoff type for player i with t_i and $\kappa(t_i)[\theta_i|t_{-i}] = \left(marg_{\Theta_i \times T_{-i}}\kappa(t_i)\right)[\theta_i|t_{-i}]$ for the belief conditional on some t_{-i}. Formally, it is captured by the following

definition.

Definition 2.5 (convergence in private values)**.** *Fix a model $\mathcal{T}$. We say a sequence of types $\{t_i[k]\}_{k=0}^{\infty}$ converges to a type t_i in private values where $t_i[k] \in T_i$ and $t_i \in T_i$ if, for any $t_{-i} \in T_{-i}$, such that $\kappa(t_i[k])[t_{-i}] > 0$,*

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

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

Now let us take a close look at the comparison between the perturbed informations structure we defined and the one used in Theorems 1 and 2 from AFHKT (2012). They define a small perturbation of the information structure of the following form: each player $i = 1, 2$ receives a signal $s_i^{k,l}$ where k and l are both integers in $\{1, ..., n\}$; the set of signals of player i is denoted by S_i. We assume the prior joint probability distribution ν^{ε} over the product of signal pairs and state of nature is such that, for each (k, l) :

$$\nu^{\varepsilon}\left(s_1^{k,l}, s_2^{k,l}, \theta_1^k, \theta_2^l\right) = \mu\left(\theta_1^k, \theta_2^l\right)\left[1 - \varepsilon - \varepsilon^2\right]$$

$$\nu^{\varepsilon}\left(s_1^{k,l_1}, s_2^{k_2,l}, \theta_1^k, \theta_2^l\right) = \mu\left(\theta_1^k, \theta_2^l\right)\frac{\varepsilon}{n^2 - 1}\text{for } (k_2, l_1) \neq (k, l)$$

$$\nu^{\varepsilon}\left(s_1^{k_1,l_1}, s_2^{k_2,l_2}, \theta_1^k, \theta_2^l\right) = \mu\left(\theta_1^k, \theta_2^l\right)\frac{\varepsilon^2}{n^4 - n^2}\text{for } k_1 \neq k \text{ or } l_2 \neq l,$$

where μ is a complete information prior over states of nature and signal pairs (i.e., a prior satisfying $\mu(s_1^{k_1,l_1}, s_2^{k_2,l_2}, \theta_1^k, \theta_2^l) = 0$ whenever $(k_i, l_i) \neq (k, l)$ for

some player i). In these expressions, we abuse notation and write $\mu\left(\theta_1^k, \theta_2^l\right)$ for the $\text{marg}_\Theta \mu\left(\theta_1^k, \theta_2^l\right)$. This corresponds to an information perturbation with the property that each player i's signal is much more informative about his own preferences than about the preferences of other player.

Let $\mathcal{P}$ denote the set of priors over $\Theta \times S$ with the following metric $d : \mathcal{P} \times \mathcal{P} \to \mathbb{R}_+$:for any $\mu, \mu' \in \mathcal{P}$,

$$d(\mu, \mu') = \max_{(\theta,s)\in\Theta\times S} |\mu(\theta, s) - \mu'(\theta, s)|.$$

Obviously the perturbation $\nu^\varepsilon \to \mu$, as $d(\nu^\varepsilon, \mu) \to 0$ as $\varepsilon \to 0$.

Aghion et al. (2012) model the incomplete information using a standard type space. That is, there is an ex ante stage during which each player observes a private signal about the payoffs, and the joint distribution of signals and payoffs is commonly known. Instead, we focus on the alternative class of situations, genuine situations of incomplete information. There is no ex ante stage; each player begins with some first order beliefs, some second-order beliefs and so on. This method is introduced by Harsanyi (1967) and developed in Mertens and Zamir (1985). We follow the interim approach due to Weinstein and Yildiz (2007) and define the "nearby" types. This notion formally described by the product topology in the universal type space.

The relation between our model and the structure in AFHKT (2012) is summarized as follows. First, instead of assuming the joint distribution of

signals and payoffs are common knowledge, we start with interim beliefs of each player and also capture the restrictions on the modeler's ability to observe the players' beliefs. Second, AFHKT (2012) fix a finite type space, where the signal is a one-to-one mapping into the payoff types; while we model the nearby types with belief hierarchy which allows for infinite types close to the benchmark types. Third, both AFHKT (2012) and our model explore the private values environment and naturally assume that players' signals are much more informative over their own payoff types than others' payoff types.

We define initial rationalizability in a general model as follows. We model player i's uncertain over the states (payoff type profiles), other players' types and other players' strategies, denoted by $\Omega = \Theta \times T_{-i} \times \Sigma_{-i}$. Let $\mathcal{B}_{-i}$ a collection of subsets of Σ_{-i} conditional on which player i forms a belief over Ω. We say that μ is *consistent* if $\sum_{\sigma_{-i}} \mu(t_i)[\theta, t_{-i}, \sigma_{-i}|\Sigma_{-i}] = \kappa(t_i)[\theta, t_{-i}]$ for any θ and t. Note that throughout this paper, players' values are private, that is $\Theta = \times_i \Theta_i$ and each player i's utility function is given as $u_i : \Delta(A) \times \Theta_i \to \mathbb{R}$. We write $\mu[\theta_i] = \text{marg}_{\Theta_i}\mu[\theta_i]$ and in case player i knows other players types t_{-i}, we write $\mu[\theta_i|t_{-i}]$ for the conditional belief. The following definition is specifically given under the game $U(\Gamma, \mathcal{T})$.

When a strategy σ_i is used, player i's type is t_i, player i's payoff type is θ_i, player i holds CPS μ, and history h is realized, the expected payoff of player

i is given as follows:

$$V_i(\sigma_i, t_i, \mu|h) = \sum_{\theta, t_{-i}} \sum_{\sigma_{-i}} \{u_i(g((\sigma_i(t_i), \sigma_{-i}(t_{-i})); h), \theta_i) + \tau_i(\sigma(t))\} \mu\left[(\theta, t_{-i}, \sigma_{-i})|\Sigma_{-i}(h)\right].$$

Definition 2.6 (Sequential Rationality)**.** *Fix a player $i \in I$, $t_i \in T_i$, a CPS $\mu \in \Delta^{\mathcal{B}_{-i}}(\Omega)$ and a strategy $\sigma_i \in \Sigma_i$. Say that σ_i is a sequential best response to μ iff, for all $\sigma_i' \in \Sigma_i$, for all $h \in \mathcal{H}$,*

$$V_i(\sigma_i, t_i, \mu|h) \geq V_i(\sigma_i', t_i, \mu|h).$$

We know that for any active player in the second stage, his payoff is only based on his own strategy and his own payoff type. Therefore, the requirement is regardless of other players' payoff types, types or strategies when player i knows his payoff type. We can decompose the definition of sequential rationality into two parts.

Definition 2.7. *Define*

$$\Sigma_i^*(t_i) = \left\{\sigma_i \in argmax_{\sigma_i' \in \Sigma_i} V_i(\sigma_i', t_i, \mu' |\ m[i]),\ \textit{for any } m[i],\ \textit{for any } \mu' \in \Delta^{\mathcal{B}_{-i}}(\Omega)\right\}.$$

*Fix a player $i \in I$, $t_i \in T_i$, a CPS $\mu \in \Delta\left(\Theta \times T_{-i} \times \Sigma_{-i}^*(T_{-i})\right)$ and a strategy $\sigma_i \in \Sigma_i^*(t_i)$. Say that σ_i is a **sequential best response** to μ iff, for all $\sigma_i' \in \Sigma_i$*

$$V_i(\sigma_i, t_i, \mu|\emptyset) \geq V_i(\sigma_i', t_i, \mu|\emptyset).$$

Therefore, the sequential best reply property coincides with the best reply property in static games. The initial rationalizability collapses to interim correlated rationalizability after we refine the strategy profiles according to the sequential rationality in the second stage.

Definition 2.8 (Initial Rationalizability)**.** *Fix a multistage game form* Γ*. For every player* $i \in I$*, let* $R_i^0(t_i|\Gamma,\mathcal{T}) = \Sigma_i^*(t_i)$*. Inductively, for every integer* $k > 0$*, let*

$$R_i^k(t_i|\Gamma,\mathcal{T}) = \left\{ \sigma_i \in \Sigma_i : \begin{array}{l} \text{there exists } \mu \in \Delta\left(\Theta \times T_{-i} \times \Sigma_{-i}^*(T_{-i})\right) \text{ such that} \\ (1)\ \mu[(\theta, t_{-i}, \sigma_{-i})] > 0 \Rightarrow \sigma_{-i} \in R_{-i}^{k-1}(t_{-i}|\Gamma,\mathcal{T}) \\ (2)\ \sigma_i \in \arg\max V_i(\sigma_i', t_i, \mu|\emptyset) \\ (3)\ \sum_{\sigma_{-i}} \mu(t_i)[\theta, t_{-i}, \sigma_{-i}|\Sigma_{-i}] = \kappa(t_i)[\theta, t_{-i}] \end{array} \right\}$$

and $R_i(t_i|\Gamma,\mathcal{T}) = \bigcap_{k=1}^{\infty} R_i^k(t_i|\Gamma,\mathcal{T})$.

Definition 2.9. *A social choice function* f *is* ***implementable in initial rationalizable strategies*** *if there exists a mechanism* Γ *such that, for all* $t \in T$ *and* $m \in M$*,* $R(t|\Gamma,\mathcal{T}) \neq \varnothing$ *and* $m \in R(t|\Gamma,\mathcal{T}) \Rightarrow g(m) = f(t)$.

Note that in complete information, the conjecture μ of player i of type $t_{i,\theta}$ is degenerate with respect to (θ, t_{-i}). The definitions above are the same as defined in Section 2.4. Now we give the formal definition of robust implementation.

Definition 2.10. *A social choice function* f *is* ***robustly implementable*** *if there exists a finite mechanism* $\Gamma = (M, g)$ *such that (i) for all* t*,* $R(t|\Gamma,\mathcal{T}) \neq$

$\varnothing$; *(ii) for any* $\bar{t} \in \bar{T}$ *and any sequence* $t[k] \to_{pp} \bar{t}$, *whenever* $t[k] \in T$ *for each* k, *we have* $g\left(m^k\right) \to f(\bar{t})$, *for any* $m^k \in R\left(t\left[k\right]|\Gamma, \mathcal{T}\right)$.

As discussed above, the perturbation in AFHKT (2012) is a special case of nearby environment defined by the universal type space. In contrast to the negative result AFHKT (2012) got using the MR mechanism, our mechanism achieves robust implementation under the same perturbation. This is because we take advantage of the simultaneous move in the two-stage game and make full use of stochastic mechanisms.

2.5.2 Main result

Theorem 2.2. *Suppose* $I \geq 2$, *then any social choice function is robustly implementable.*

We use the following claim to prove Theorem 2.2.

Claim 2.6. *For any* $\bar{t} \in \bar{T}$ *and any sequence* $t[k] \to_{pp} \bar{t}$, *whenever* $t[k] \in T$ *for* k *large enough, we have* $\Sigma_i^*\left(t_i\left(k\right)\right) = \Sigma_i^*\left(\bar{t}_i\right)$.

Proof. By convergence in private values, we know that, for any $\bar{t}_{i,\theta} \in \bar{T}_\theta$ and any sequence $t[k] \to_{pp} \bar{t}_\theta$,

$$\kappa\left(t_i\left[k\right]\right)\left[\theta_i|t_{-i}\right] \to \kappa\left(\bar{t}_{i,\theta}\right)\left[\theta_i\right] \text{ as } k \to \infty \text{ for any } t_{-i}.$$

Recall the definitions of $\Sigma^*(t_i)$ and $V_i(\sigma_i^{'}, t_i, \mu|m[i])$:

$$\begin{aligned}
\Sigma_i^*(t_i) &= \left\{\sigma_i \in \text{argmax} V_i(\sigma_i^{'}, t_i, \mu|m[i]) \quad \text{for any } m[i]\right\} \\
V_i(\sigma_i, t_i, \mu|m[i]) &= \sum_{\theta, t_{-i}} \sum_{\sigma_{-i}} \{u_i(g((\sigma_i(t_i), \sigma_{-i}(t_{-i})); m[i]), \theta_i) + \tau_i(m[i])\} \mu\left[(\theta, t_{-i}, \sigma_{-i})|\Sigma_{-i}(m[i])\right]
\end{aligned}$$

$\Sigma_i^*(t_i)$ is the best response set of player i of type t_i, which only depends upon what player i believes as his own payoff type. Hence, for each $\sigma_i \in \Sigma_i^*$, $V_i(\sigma_i, t_i, \mu|m[i])$ can be rearranged as follows:

$$V_i(\sigma_i, t_i, \mu|m[i]) = \sum_{t_{-i}} \sum_{\theta_i} u_i\left(x_i(\sigma_i^2(t_i)), \theta_i\right) \mu\left[\theta_i|t_{-i}\right] \sum_{\sigma_{-i}} \mu\left[t_{-i}|\sigma_{-i}\right] \mu\left[\sigma_{-i}|\Sigma_{-i}\left(m\left[i\right]\right)\right] + \tau_i(m[i])$$

for any $m[i]$.

Note that no matter what other players' types t_{-i} are, we obtain

$$\kappa\left(t_i\left[k\right]\right)\left[\theta_i|t_{-i}\right] \to \kappa\left(t_{i,\theta}\right)\left[\theta_i\right] = 1 \text{ as } k \to \infty.$$

This implies that for any t_{-i},

$$\mu\left(t_i\left[k\right]\right)\left[\theta_i|t_{-i}\right] \to 1 \text{ as } k \to \infty.$$

From Lemma 2.1, we have that for all $\theta_i, \theta_i^{'} \in \Theta_i$ with $\theta_i \neq \theta_i^{'}$,

$$u_i\left(x_i\left(\theta_i\right), \theta_i\right) > u_i\left(x_i\left(\theta_i^{'}\right), \theta_i\right).$$

Fix any such $\theta_i, \theta_i^{'}$. Then, there exists some $\bar{k}$ such that for any $k \geq \bar{k}$,

$$\begin{aligned}
&\sum_{\theta, t_{-i}} \sum_{\sigma_{-i}} u_i\left(x_i\left(\theta_i\right), \theta_i\right) \mu\left(t_i\left[k\right]\right) \mu[t_{-i}|\sigma_{-i}] \mu[\sigma_{-i}|\Sigma_{-i}(m[i])] \left[\theta_i|t_{-i}\right] \\
> &\sum_{\theta, t_{-i}} \sum_{\sigma_{-i}} u_i\left(x_i\left(\theta_i^{'}\right), \theta_i\right) \mu\left(t_i\left[k\right]\right) \left[\theta_i|t_{-i}\right] \mu[t_{-i}|\sigma_{-i}] \mu[\sigma_{-i}|\Sigma_{-i}(m[i])].
\end{aligned}$$

That is, $\Sigma_i^*(t_i(k)) = \Sigma_i^*(\bar{t}_i)$. □

We then recall the following well known lemma.

Lemma 2.2. *(Dekel, Fudenberg, and Morris (2006)) Fix any model $\mathcal{T} = (T, \theta, \pi)$ such that $\bar{\mathcal{T}} \subset \mathcal{T}$, and any finite mechanism Γ. (i) For any $\bar{t} \in \bar{T}$ and any sequence $\{t[k]\}_{k=0}^{\infty}$ in T, if $t[n] \to_p \bar{t}$, then, for k large enough, we have $R(t[k]|\Gamma, \mathcal{T}) \subset R(\bar{t}[k]|\Gamma, \mathcal{T})$. (ii) For any type $t \in T$, $R(t|\Gamma, \mathcal{T})$ is nonempty.*

This lemma completes the proof with Claim 2.6.

2.6 Application

There are two players. We follow Maskin and Tirole (1999) to assume players are risk averse, and follow their assumptions:

(a) for all $\theta \in \Theta$ functions take the form

$$u_i^{\theta}(a, t_i) = U_i(u_i(a, \theta) + t_i) \text{ for } i = 1, 2,$$

where $U_i : \mathbb{R} \to \mathbb{R}$ is increasing and strictly concave;

(b) individual players' transfers are denoted by

$$T = \{(t_1, t_2) | t_1 + t_2 = 0\}.$$

Contrast to Maskin and Tirole (1999), we drop the assumption that there is no bound on the magnitude of the transfers. Instead, we will show that our result

extends to the environment with restricted transfers $T = \{(t_1, t_2) \,|t_1 + t_2 = 0, \text{and } \, |t_i| < \bar{\tau}\}$ for any $\bar{\tau} > 0$. We use $(\Gamma, \bar{\tau})$ to denote a mechanism with transfers bounded by $\bar{\tau}$.

The renegotiation process can be expressed as a function: $h : A \times \Theta \to \tilde{A}$, where $\tilde{A} = A \times \mathbb{R}^2$ is the set of outcomes (alternatives A and transfers $\mathbb{R}^2$). We write $h(a, \theta)$ for the equilibrium renegotiated outcome, starting from the mechanism-prescribed outcome (a, t_1, t_2) in state θ. That is, we adopt the assumption in Maskin and Tirole (1999) that renegotiation is independent of (t_1, t_2) for expositional convenience. Given any (a, t_1, t_2), any θ,

$$h(a, \theta) = \left(a^\theta, t_1 + \Delta t_1^\theta(a), t_2 + \Delta t_2^\theta(a)\right),$$

where $\Delta t_i^\theta(a)$ is the renegotiation-transfer and $\Delta t_1^\theta(a) + \Delta t_2^\theta(a) = 0$. Let

$$u_i^\theta(h(a, \theta)) = U_i\left(u_i\left(a^\theta, \theta\right) + t_1 + \Delta t_1^\theta(a)\right).$$

Remark 2.1. *Note that when we say the transfer is arbitrarily small in our mechanism, the transfer is specified by the mechanism. It is natural that the renegotiation-transfer can be arbitrary subject to players' wealth constraint. In addition, if* Γ *implements* f *subject to renegotiation, the renegotiation-transfer can be large. For example, if* $f(\theta)$ *is inefficient in* ϕ *and* $h(f(\theta), \phi) = \left(f(\theta)^\phi, \Delta t_1^\phi(f(\theta)), \Delta t_2^\phi(f(\theta))\right)$, *then it is possible that the maginitude of* $\Delta t_i^\phi(f(\theta))$ *is large for some player* i.

We follow Maskin and Moore (1999) to restate the assumptions about $h(\cdot,\cdot)$ as follows.

Assumption A1 (Renegotiation is predictable). $h(\cdot,\cdot)$ is a function that is common knowledge to the individuals.

Assumption A2 (Renegotiation is efficient). $h(a,\theta)$ is Pareto optimal for all $(a,t_1,t_2)\in\tilde{A}$ and $\theta\in\Theta$ (that is, there does not exist $(a',t')\in\tilde{A}$ such that $u_i^\theta(a',t_i')\geq u_i^\theta(h(a,\theta))$ for all i, with strict preference for some i).

Assumption A3 (Renegotiation is individually rational). For all $(a,t_1,t_2)\in\tilde{A}$ and $\theta\in\Theta$, and all i, $u_i^\theta(h(a,\theta))\geq u_i^\theta(a,t_i)$.

Given a social choice function f and a renegotiation function h, we say that f is implementable in SPE with renegotiation function h if there exists a mechanism Γ such that $f(\theta)=h\circ g\left(m^\theta\right)$, where m^θ is a subgame perfect equilibrium in Γ^θ subject to renegotiation function h for any $\theta\in\Theta$.

For simplicity, we assume states are describle and show how our mechanism works subject to renegotiation. By Maskin and Tirole (1999), indescribility does not constrains the set of implementable social choice rules. It follows immediately that if f is implementable subject to renegotiation, $f(\theta)$ must be Pareto efficient in state θ for any θ.

We adopt a modified version of individual players' preference assumption in the renegotiation environment.

For any state θ, player 1 has a preference ordering over the set of outcomes $\{h(\tilde{a}, \theta)\}_{\tilde{a} \in \tilde{A}}$. We assume that under any two distinct state θ and θ', player i has two different preference orderings over the outcome set after renegotiation and there is no total indifference over the outcomes under any θ. Formally, we have the following assumption.

Assumption

(i) For any $\theta, \theta' \in \Theta$, $u_i^{\theta}(h(\cdot, \theta))$ is not a positive affine transformation of $u_i^{\theta'}(h(\cdot, \theta'))$;

(ii) For any θ, $u_i^{\theta}(h(\cdot, \theta))$ is not a constant function on A.

We abuse notation to use $h(x, \theta)$ to denote the lottery after renegotiation, that is, with probability $x[a]$ the outcome $h(a, \theta)$ is the one after renegotiation. Thus, $u_i^{\theta}(h(x, \theta))$ denotes player i's expected utility from x subject to renegotiation function h. Now we obtain an important lemma in the environment allowing renegotiation. We consider Lemma 2.3 in renegotiation environment as a counter part of Lemma 2.1.

Lemma 2.3. *For any state θ, we construct a lottery $x^{\theta} \in \Delta(A)$, such that*

$$u_1^{\theta}\left(h\left(x^{\theta}, \theta\right)\right) > u_1^{\theta}\left(h\left(x^{\theta'}, \theta\right)\right),$$

for any $\theta' \neq \theta$.

Remark 2.2. *Maskin and Tirole (1999) assume that the efficient outcome in any state is unique and adopt an implicit assumption that for any distinct pair* $\{\theta, \theta'\}$ *there exist a pair of outcomes* $\{a, a'\} \subset A$ *such that* $\Delta t_1^{\theta}(a) > \Delta t_1^{\theta}(a')$ *and* $\Delta t_1^{\theta'}(a) < \Delta t_1^{\theta'}(a')$.

Theorem 2.3. *Assume that utility functions take the form* $\tilde{u}_i^{\theta}(a, t_i) = U_i\left(u_i^{\theta}(a) + t_i\right)$ *for* $i = 1, 2$ *with* U_i *increasing and strictly concave, and that* f *is Pareto-optimal. Then for any* $\bar{\tau} > 0$, f *is implementable in subgame perfect equilibrium by a mechanism* $(\Gamma, \bar{\tau})$ *subject to renegotiation.*

2.6.1 Mechanism

The allocation

First Stage : Each player i announces $K+1$ times the possible state, $m_i = \left(m_i^0, m_i^1 ..., m_i^K\right)$, where $m_i^k \in \Theta$, for all $k \in \{0, 1, ..., K\}$. If $m_1^0 = m_2^0$, then

$$l = \frac{1}{K} \sum_{k=1}^{K} \tilde{f}\left(m^k\right)$$

(before renegotiation) is implemented. Otherwise, we proceed to the Second Stage.

Second Stage : Player 1 announces $m_1^{K+1} \in \Theta$, and

$$l\left(\varepsilon, m_1^{K+1}\right) = \frac{1-\varepsilon}{K} \sum_{k=1}^{K} \tilde{f}\left(m^k\right) + \varepsilon x_1^{m_1^{K+1}}$$

(before renegotiation) is implemented.

2.6.2 The transfer

Let γ, ξ and η be positive numbers.

Second Stage

(i) Player 1 pays γ to player 2;

(ii) If $m_1^{K+1} = m_2^0$, then there is no extra transfer;

If $m_1^{K+1} \neq m_2^0$ (let $\hat{\theta} \equiv m_1^0$ for simplicity of notation), then $Q > 0 > L$ are chosen so that for any θ', θ

$$\begin{aligned} & |\frac{1}{2}U_1\left(h\left(x_1^{\theta'},\theta\right)-\gamma+Q\right)+\frac{1}{2}U_1\left(h\left(x_1^{\theta'},\theta\right)-\gamma+L\right) \\ & -U_1\left(h\left(x_1^{\theta'},\theta\right)-\gamma\right)| \\ < \ & \frac{\delta}{2}; \end{aligned}$$

and

$$\begin{aligned} & \frac{1}{2}U_2\left(h\left(l\left(\varepsilon,\hat{\theta}\right),\hat{\theta}\right)-Q+\gamma\right)+\frac{1}{2}U_2\left(h\left(l\left(\varepsilon,\hat{\theta}\right),\hat{\theta}\right)-L+\gamma\right) \\ < \ & U_2\left(h\left(l,\hat{\theta}\right)-\gamma\right); \end{aligned} \tag{2.1}$$

$$\max\{Q,|L|\} < \tau.$$

Remark 2.3. *The "closeness" between l and $l\left(\varepsilon, m_1^{K+1}\right)$ guarantees that we can choose Q and L to get the second inequality.*

First Stage

Player i is to pay player $j \neq i$:

1. • 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).

$$d_i\left(m^0, ..., m^K\right) = \begin{cases} \xi & \text{if there exists } k \in \{1, ..., K\} \text{ s.t. } m_i^k \neq m_i^0, \\ & \text{and } m_j^{k'} = m_j^0 \text{ for all } k' \in \{1, ..., k-1\} \text{ for all } j; \\ 0 & \text{otherwise.} \end{cases} \tag{2.2}$$

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

$$d_i^k\left(m_i^0, m_i^k\right) = \begin{cases} \eta & \text{if } m_i^k \neq m_i^0; \\ 0 & \text{otherwise.} \end{cases} \tag{2.3}$$

Let

$$B = \max_{\tilde{a}, \tilde{a}' \in A^\Gamma, \theta \in \Theta, i \in I} \left| U_i\left(h\left(\tilde{a}, \theta\right)\right) - U_i\left(h\left(\tilde{a}', \theta\right)\right) \right|,$$

where $\tilde{A}^\Gamma$ is the set of outcomes specified by the mechanism. For any $\tau > 0$, we can choose γ, ξ, and η such that

$$
\begin{aligned}
\tau &> \gamma + \xi + K\eta \\
\left(1 - \frac{1}{K}\right)\left\{U_i\left(h\left(\tilde{a}, \theta\right)\right) - U_i\left(h\left(\tilde{a}, \theta\right) - \xi\right)\right\} &> \frac{1}{K}B \\
\eta &> 0 \\
\gamma - \xi - K\eta &> 0 \qquad (2.4) \\
(1-\varepsilon)\left\{U_i\left(h\left(\tilde{a}, \theta\right) + \gamma - \xi - K\eta\right) - U_i\left(h\left(\tilde{a}, \theta\right)\right)\right\} &> \varepsilon B. \qquad (2.5)
\end{aligned}
$$

Suppose the true state is θ.

Claim 2.7. *At second stage, it is sequential rational for player* 1 *to choose* $m_1^{K+1} = \theta$.

Proof. Suppose $m_1^{K+1} = \theta' \neq \theta$. Let $\bar{m}_1$ be such that $\bar{m}_1^q = m_1^q$ for all $q \neq K+1$, and $\bar{m}_1^{K+1} = \theta$. For any $h \neq \emptyset$,

$$
\begin{aligned}
& U_1\left(g\left(\bar{m}_1, m_2\right)|h\right) - U_1\left(g\left(m_1, m_2\right)|h\right) \\
> \; & \varepsilon \left\{ \begin{array}{l} \left\{U_1\left(h\left(x_1^{\theta}, \theta\right) - \gamma\right) - \frac{\delta}{2}\right\} - \\ \left\{U_1\left(h\left(x_1^{\theta'}, \theta\right) - \gamma\right) + \frac{\delta}{2}\right\} \end{array} \right\} \\
> \; & 0
\end{aligned}
$$

Since $\bar{m}_1$ and m_1 differs only at the second stage, the utility difference is from the "ε'' lottery and whether player 1 gets paid by the (Q, L) lottery. We can only focus on this difference. The first inequality is from the fact that $U_1\left(h\left(x_1^{\theta}, \theta\right) - \gamma\right) - \frac{\delta}{2}$ is the minimum payoff from playing $\bar{m}_1$, while

$U_1\left(h\left(x_1^\theta,\theta\right)-\gamma\right)+\frac{\delta}{2}$ is maximum payoff from playing m_1; the last inequality follows from Lemma 2.3. Therefore, it is sequential rational for player 1 to choose $m_1^{K+1}=\theta$. □

Claim 2.8. *If m is a SPE of Γ^θ, then $m_2^0=\theta$.*

Proof. We prove Claim 2.8 in the following two cases: (i) $m_1^0=\theta$; (ii) $m_1^0\neq\theta$.

In case (i), suppose $m_2^0\neq\theta$, then the game proceeds to the second stage. From Claim 2.7, player 1 will announce θ in the second stage. Since $m_2^0\neq\theta$, player 2 will pay the lottery (Q,L) to player 1. In terms of the transfers in the first stage, the possible gain from choosing m_2^0 rather than θ is bounded above by $\xi+K\eta$. In total, the possible payoff for player 2 from m_2^0 is bounded above by

$$U_2\left(h\left(l,\theta\right)-\gamma+\xi+K\eta\right)$$

by (2.1), where $l=\frac{1}{K}\sum_{k=1}^{K}\tilde{f}\left(m^k\right)$. Let $\bar{m}_2$ be such that $\bar{m}_2^q=m_2^q$ for all $q\neq 0$, and $\bar{m}_2^0=\theta$. The payoff from choosing $\bar{m}_2$ is

$$U_2\left(h\left(l,\theta\right)\right).$$

By (2.4), player 2 is worse off from choosing $m_2^0\neq\theta$ rather than $\bar{m}_2$.

In case (ii), if $m_2^0=\theta$, then the game proceeds to the second stage. From Claim 2.7, player 1 will announce θ in the second stage. From the transfer rule, player 2 gets paid γ. The possible loss is from "ε'' lottery by triggering

the second stage. The possible loss from choosing $m_2^0 = \theta$ is bounded above by $\xi + K\eta$. Let $\bar{m}_2$ be such that $\bar{m}_2^q = m_2^q$ for all $q \neq 0$, and $\bar{m}_2^0 = m_1^0$. In total, the least possible payoff difference for player 2 from m_2^0 rather than choosing $\bar{m}_2^0$ is bounded below by

$$\begin{aligned}
& U_2\left(h\left(l\left(\varepsilon,\theta\right),\theta\right)+\gamma-\xi-K\eta\right)-U_2\left(h\left(l,\theta\right)\right) \\
\geq\ & \left(1-\varepsilon\right)\left\{U_2\left(h\left(l,\theta\right)\right)+\gamma-\xi-K\eta\right)-U_2\left(h\left(l,\theta\right)\right)\right\} \\
& +\varepsilon\left\{U_2\left(h\left(x_1^\theta,\theta\right)+\gamma-\xi-K\eta\right)-U_2\left(h\left(l,\theta\right),\theta\right)\right\} \\
>\ & \left(1-\varepsilon\right)\left\{U_2\left(h\left(l,\theta\right)\right)+\gamma-\xi-K\eta\right)-U_2\left(h\left(l,\theta\right)\right)\right\} \\
>\ & 0
\end{aligned}$$

Therefore, player 2 is better off from choosing $m_2^0 = \theta$ rather than telling a lie. It is easy to check player 2 will get worse off if he chooses $m_2^0 \neq m_1^0$, where $m_2^0 \neq \theta$. □

Claim 2.9. *If m is a SPE of Γ^θ, then $m_1^0 = \theta$.*

Proof. From Claim 2.7 and Claim 2.8, if $m_1^0 \neq \theta$, then the second stage is triggered and player 1 pays γ to player 2. The minimum loss is γ, while the possible gain is from "ε" lottery and the transfers in the first stage. Let $\bar{m}_1$ be such that $\bar{m}_1^q = m_1^q$ for all $q \neq 0$, and $\bar{m}_1^0 = \theta$. In total, the least possible payoff difference for player 1 from $m_1^0 \neq \theta$ rather than choosing θ is bounded

above by

$$\begin{aligned} & U_1\left(h\left(l\left(\varepsilon,\theta\right),\theta\right)-\gamma+\xi+K\eta\right)-U_1\left(h\left(l,\theta\right)\right) \\ \leq\ & \left(1-\varepsilon\right)\left\{U_1\left(h\left(l,\theta\right)-\gamma+\xi+K\eta\right)-U_1\left(h\left(l,\theta\right)\right)\right\} \\ & +\varepsilon\left\{U_1\left(h\left(x_1^{\theta},\theta\right)-\gamma+\xi+K\eta\right)-U_1\left(h\left(l,\theta\right)\right)\right\} \\ <\ & 0 \end{aligned}$$

Therefore, player 1 is worse off from choosing $m_2^0 \neq \theta$ rather than telling the truth. □

Claim 2.10. *If m is a SPE of Γ^{θ}, then $m_i^k = \theta$, for any i, any $k \geq 0$.*

Proof. We prove Claim 2.10 inductively. We have established that if m is a SPE of Γ^{θ}, then $m_i^0 = \theta$. Suppose $m_i^0 = ... = m_i^{k-1} = \theta$ for all i. We show that $m_i^k = \theta$ for all i. Suppose not, let $\bar{m}_i$ be the message such that $\bar{m}_i^q = m_i^q$ for all $q \neq k$, and $\bar{m}_i^k = \theta$. Suppose $m_j^k \neq \theta$ for $j \neq i$. Then minimum loss from playing m_i is $\xi+\eta$ by the transfer rule. In terms of allocation, the possible gain from playing m_i is $\frac{1}{K}B$. By (2.4), player i is worse off from playing m_i rather than $\bar{m}_i$. Suppose $m_j^k = \theta$ for $j \neq i$. Then the minimum loss from playing m_i is η. In terms of allocation, player i cannot get better by the truthful implementability of f. Therefore, player i is worse off from playing m_i rather than $\bar{m}_i$. This completes the proof. □

The existing literature is concerned about the renegotiation problem. Maskin and Tirole (1999) introduce a stochastic transfer from the seller to the buyer. If the buyer is risk-averse, then this fine can be designed so that it hurts both the seller and the buyer. However, if parties are close to risk-neutral, the stochastic fine that is required needs to have a very large variance, which is not very credible as it will violate the wealth constraints. Thus the applicability of the irrelevance theorem is doubtful in this case. Our mechanism adopts the idea of Abreu and Matsushima (1992) to break up the large payments into arbitrarily small scale. Therefore, this permissive mechanism is immune to renegotiation with arbitrarily small cost.

2.7 Discussion

We first provide a way to achieve budget balance when there are at least three players. We conclude with a comparison between dynamic mechanisms and static ones.

2.7.1 Budget balance

When there are at least three players, the transfers specified at the last stage can be made between the pair of players (i^*, i^*+1) and the other players. This mild modification does not change the incentive of any players. Moreover, all the arguments above still hold. Therefore we can achieve budget balance

everywhere (both on and off the solution outcome).

2.7.2 Dynamic vs static mechanisms

Robustness of the implementation problem is studied by researchers recently. The pioneering work of Chung and Ely (2003) shows that when players' values are interdependent if we adopt undominated Nash equilibrium as the solution concept, then Maskin monontonicity is a necessary condition for robust implementation. When players' values are private, Chen et al. (2014) show that any incentive compatible social choice function is robustly implementable if we use the solution $S^{\infty}W$, which is obtained by deletion of weakly dominated strategies followed by iterative deletion of strictly dominated strategies. Consider dynamic mechanisms. Aghion et al. (2012) show that when players' values are interdependent if we adopt subgame-perfect equilibrium as the solution concept, then Maskin monontonicity is also a necessary condition for robust implementation. This paper shows that if players' values are private, any social choice function is robustly implementable.

Chapter 3

Implementation with Transfers

3.1 Introduction

The theory of *implementation* and *mechanism design* is mainly concerned with the following question: what is the set of outcomes that can be achieved by institutions (or mechanisms)? This institutional design problem is particularly relevant when a group of individuals with conflicting interests has to make a collective decision. The key question then becomes: when can individuals, acting in their own self-interest, arrive at the outcomes consistent with a given welfare criterion (or social choice rule)? To characterize the set of Pareto efficient allocations, for instance, we must know the preferences of those individuals, which is dispersed among the individuals involved. If Pareto efficiency is 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 exam-

ple, 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 con-

tinuous 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 envi-

ronments. 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.

3.2 Preliminaries

3.2.1 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, then we say $(T_i, \hat{\theta}_i, \pi_i)_{i \in I}$ is a *finite* model.

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

Let $\pi_i(t_i)[E]$ denote the probability that $\pi_i(t_i)$ assigns to any set $E \subset T_{-i}$.

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

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

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

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

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

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

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

Throughout the paper, we consider a fixed environment $\mathcal{E}$ which is a triplet $\left(A, (u_i)_{i\in I}, \bar{\mathcal{T}}\right)$ with a finite model $\bar{\mathcal{T}} = \left(\bar{T}_i, \bar{\theta}_i, \bar{\pi}_i\right)_{i\in I}$ and a *planner* who aims to implement a *social choice function* (henceforth, SCF) $f : \bar{T} \to \Delta(A)$.[2]

3.2.2 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* from player $i \in I$ to the designer. 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 $m \in M$. Note that there is a class of such mechanisms given $\bar{\tau}$. We denote one of the mechanisms by $(\mathcal{M}, \bar{\tau})$.

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

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

$m_i' \in M_i$ such that

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

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

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

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 $m_{-i} \in S_{-i}^l\left(t_{-i}|\mathcal{M}, \mathcal{T}\right) = \prod_{j \neq i} S_j^l\left(t_j|\mathcal{M}, \mathcal{T}\right)$. 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.,

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

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

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 3.1. *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 implementability in $S^\infty W$. First, we impose no conditions on the magnitude of transfers and propose the concept of implementation with transfers.

Definition 3.2 (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 3.3 (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 3.4 (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 3.1. *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.*

3.2.3 Assumptions

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

Assumption 3.1. *An environment* $\mathcal{E} = (A, (u_i)_{i \in I}, \bar{T})$ *satisfies Assumption 3.1 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 3.1 guarantees the existence of a function that can elicit each player's type.

Lemma 3.1. *(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 $t_i \neq t_i'$,*

$$u_i(x_i(t_i), \hat{\theta}_i(t_i)) > u_i(x_i(t_i'), \hat{\theta}_i(t_i)) \tag{3.1}$$

We next introduce the following assumption.

Assumption 3.2. *An environment $\mathcal{E}$ satisfies Assumption 3.2 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 3.2. *Since $\bar{T}$ is finite, if $|T_i| = 1$ or $|T_{-i}| \geq 2$, Assumption 3.2 generically holds in the space of the probability distributions over $\bar{T}$. Note, however, that Assumption 3.2 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 3.2. *Suppose that an environment $\mathcal{E}$ satisfies Assumption 3.2. For all $i \in I$ and $(t_i, t_{-i}) \in \bar{T}$, define*

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

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

$$\sum_{t_{-i} \in T_{-i}} \left[d_i^0(t_i, t_{-i}) - d_i^0(t_i', t_{-i}) \right] \bar{\pi}_i(t_i)[t_{-i}] > 0. \tag{3.2}$$

Remark 3.3. *Lemma 3.2 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 3.6)*

Proof. The construction of $d_i^0(t_i, t_{-i})$ makes itself a proper scoring rule. By Assumption 3.2, the strict inequality of (3.2) always holds. □

3.3 The Mechanism and its Basic Properties

3.3.1 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 $m \in M$,

$$g(m) = e(m^{-1}, m^0) \frac{1}{I} \sum_{i \in I} x_i\left(m_i^{-2}\right) + \left\{1 - e(m^{-1}, m^0)\right\} \frac{1}{K} \sum_{k=1}^{K} f\left(m^k\right), \tag{3.3}$$

The outcome function contains a "random dictator" component (recall the function x_i defined in (3.1)) 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).

$$d_i\left(m^0, ..., m^K\right) = \begin{cases} \xi & \text{if there exists } k \in \{1, ..., K\} \text{ s.t. } m_i^k \neq m_i^0, \\ & \text{and } m_j^{k'} = m_j^0 \text{ for all } k' \in \{1, ..., k-1\} \text{ for all } j; \\ 0 & \text{otherwise.} \end{cases} \tag{3.4}$$

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

$$d_i^k\left(m_i^0, m_i^k\right) = \begin{cases} \eta & \text{if } m_i^k \neq m_i^0; \\ 0 & \text{otherwise.} \end{cases} \tag{3.5}$$

In total,

$$\tau_i(m) = -\lambda d_i^0\left(m_{-i}^{-2}, m_i^{-1}\right) - \lambda d_i^0(m_{-i}^{-1}, m_i^0) + d_i\left(m^0, ..., m^K\right) + \sum_{k=1}^{K} d_i^k\left(m_i^0, m_i^k\right). \tag{3.6}$$

[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.

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

$$E = \max_{m_i^{-2} \in M_i^{-2}, m^k \in M^k, \bar{\theta}_i \in \bar{\Theta}_i, i \in I} \left| \frac{1}{I} \sum_{j \in I} u_i \left(x_j(m_j^{-2}), \bar{\theta}_i \right) - u_i \left(f \left(m^k \right), \bar{\theta}_i \right) \right|; \tag{3.7}$$

$$D = \max_{\bar{m}_i^k \in M_i^k, m^k \in M^k, \bar{\theta}_i \in \bar{\Theta}_i, i \in I} \left\{ u_i \left(f \left(m^k \right), \bar{\theta}_i \right) - u_i(f(m_{-i}^k, \bar{m}_i^k), \bar{\theta}_i) \right\}, \tag{3.8}$$

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

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

$$\bar{\tau} > 2\lambda \bar{d}_i^0 + \xi + K\eta; \tag{3.9}$$

$$\sum_{t_{-i} \in \bar{T}_{-i}} \left[\lambda d_i^0 \left(t_{-i}, t_i' \right) - \lambda d_i^0 \left(t_{-i}, t_i \right) \right] \bar{\pi}_i \left(t_i \right) \left[t_{-i} \right] > \gamma; \tag{3.10}$$

$$\eta > \epsilon E; \tag{3.11}$$

$$\xi > \frac{1}{K} D; \tag{3.12}$$

$$\gamma > \epsilon E + \xi + K\eta, \tag{3.13}$$

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

3.3.2 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 3.1. *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 $\nu : \bar{T}_{-i} \to \Delta(M_{-i})$.

The difference of the expected values between m_i' and m_i for player i of

[4]Given any $\bar{\tau} > 0$ exogenously, we first choose λ small enough so that $\lambda \bar{d}_i^0 < \frac{1}{4}\bar{\tau}$. Second, by (3.2), we can choose γ small enough so that (3.10) 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 (3.9), (3.11), (3.12) and (3.13) hold.

type $\bar{t}_i$ is shown as follows:

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

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 (3.3) and the definition of τ in (3.6)); by (3.1) the last inequality is strict whenever $e\left(m^{-1},m^0\right)=\epsilon$ for some m_{-i} with $\sum_{t_{-i}}\nu(m_{-i}|t_{-i})\pi_i(\bar{t}_i)[t_{-i}]>0$. □

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 3.2. *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:

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

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 (3.3)) lies in function $e(\cdot)$ and the difference in terms of transfer is summarized in functions d_i^0 (see the transfer rule in (3.6)).

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

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

From (3.7), 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 3.1, 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 (3.10), 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 (3.13).

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

Claim 3.3. *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 3.1, 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 (3.3)) 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 (3.6)).

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

3.4 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.

3.4.1 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 3.1. *Suppose that the environment $\mathcal{E}$ satisfies Assumptions 3.1 and 3.2. Assume $I \geq 2$. Any SCF is implementable in $S^{\infty}W$ with transfers.*

We use the following claim to prove Theorem 3.1.

Claim 3.4. *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 3.3, 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

$$\sum_{t_{-i}, m_{-i}} \left\{\tau_i\left(\tilde{m}_i, m_{-i}\right) - \tau_i\left(m\right)\right\} \nu(m_{-i}|t_{-i})\pi_i(\bar{t}_i)[t_{-i}] = \eta + \xi, \tag{3.15}$$

where player i of type $\bar{t}_i$ will get punished by η according to rule d_i^1 (by (3.5)) and ξ according to rule d_i (by (3.4)).

Note that $e\left(m^{-1}, m^0\right) = 0$. In terms of outcome function $g(\cdot)$ (defined in (3.3)): the possible gain from playing m_i rather than $\tilde{m}_i$ is

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

From (3.8), we also have the following inequality on the expected gain of type t_i when playing m_i rather than $\tilde{m}_i$:

$$\sum_{t_{-i},m_{-i}} \left\{u_i(f(m^1),\bar{\theta}_i) - u_i(f(\bar{t}_i,m^1_{-i}),\bar{\theta}_i)\right\} \nu(m_{-i}|t_{-i})\pi_i(\bar{t}_i)[t_{-i}] \leq D. \quad (3.16)$$

When $K = 1$, we know from Section 3.1 that $\xi > D$ (see (3.12)).[5] So, we obtain

$$\eta + \xi > D. \quad (3.17)$$

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 (3.3) and the transfer rule (3.6); the second inequality follows from (3.16); the last inequality follows from (3.17). Therefore, player i of type $\bar{t}_i$ will report $\bar{t}_i$ rather than m^1_i. □

[5]When $K = 1$, we can appropriately choose λ, γ, ϵ, ξ, and η to satisfy those conditions on transfers and utilities in Section 3.3.1. 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 (3.13)). Hence, ξ can be chosen large enough to satisfy $\xi > D$ (inequality (3.12)).

3.4.2 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 $\left(\bar{T}, f\right)$ for player i of type t_i when he announces t_i' and the other players all make truthful announcements.

Definition 3.5. *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 3.2. *Suppose that the environment $\mathcal{E}$ satisfies Assumptions 3.1 and 3.2. 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 3.4. *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 3.2.

Claim 3.5. *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 3.3. 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

$$\begin{aligned} & \sum_{t_{-i},m_{-i}} \left\{u_i(g(\tilde{m}_i, m_{-i}), \bar{\theta}_i) + \tau_i(\tilde{m}_i, m_{-i})\right\} \nu(m_{-i}|t_{-i})\pi_i(\bar{t}_i)[t_{-i}] \\ & - \sum_{t_{-i},m_{-i}} \left\{u_i(g(m_i, m_{-i}), \bar{\theta}_i) + \tau_i(m_i, m_{-i})\right\} \nu(m_{-i}|t_{-i})\pi_i(\bar{t}_i)[t_{-i}] \quad (3.18) \\ > \ & 0. \end{aligned}$$

Note the left hand side of inequality is equal to

$$\begin{aligned} & \sum_{t_{-i},m_{-i} \notin M_{-i}^*} \left\{ \begin{array}{c} \left\{u_i(g(\tilde{m}_i, m_{-i}), \bar{\theta}_i) + \tau_i(\tilde{m}_i, m_{-i})\right\} - \\ \left\{u_i(g(m_i, m_{-i}), \bar{\theta}_i) + \tau_i(m_i, m_{-i})\right\} \end{array} \right\} \nu(m_{-i}|t_{-i})\pi_i(\bar{t}_i)[t_{-i}] \quad (3.19) \\ + & \sum_{t_{-i},m_{-i} \in M_{-i}^*} \left\{ \begin{array}{c} \left\{u_i(g(\tilde{m}_i, m_{-i}), \bar{\theta}_i) + \tau_i(\tilde{m}_i, m_{-i})\right\} - \\ \left\{u_i(g(m_i, m_{-i}), \bar{\theta}_i) + \tau_i(m_i, m_{-i})\right\} \end{array} \right\} \nu(m_{-i}|t_{-i})\pi_i(\bar{t}_i)[t_{-i}]. \end{aligned}$$

Step 1:

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

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 (3.5) in Section 3.1) and ξ according to rule d_i (see (3.4)), 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 (3.3)), we have

$$\sum_{t_{-i},m_{-i}\notin M^*_{-i}} \frac{1}{K}\left\{u_i(f(m^{k-1}),\bar{\theta}_i)-u_i(f(\tilde{m}_i^{k-1},m_{-i}^{k-1}),\bar{\theta}_i)\right\}\nu(m_{-i}|t_{-i})\pi_i(\bar{t}_i)[t_{-i}] \leq \frac{1}{K}D. \tag{3.20}$$

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 (3.12) in Section 3.1), we have

$$\eta+\xi > \frac{1}{K}D. \tag{3.21}$$

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 (3.5) 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 (3.6) 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:

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

In terms of $g(\cdot)$ (see the outcome function in (3.3)), 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 3.2 is proved as follows.

Proof. Fix $\bar{\tau}>0$ arbitrarily small. Given $f:\bar{T}\to\Delta(A)$ 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 such that

$$\begin{aligned} f(t_i, t_{-i}) &= g(m_i(t_i), m_{-i}(t_{-i})), \\ \tau_i(t_i, t_{-i}) &= \tau_i(m_i(t_i), m_{-i}(t_{-i})), \end{aligned}$$

where g and τ is specified in $(\mathcal{M}, \bar{\tau})$. Then truth telling must be a Bayesian Nash equilibrium. That is, for any $t_i, t_i' \in \bar{T}_i$,

$$\begin{aligned} & \sum_{t'_{-i}} \pi_i(t_i)[t'_{-i}] \left\{ u_i(f(t_i, t'_{-i}), \hat{\theta}_i(t_i)) + \tau_i(t_i, t'_{-i}) \right\} \\ \geq & \sum_{t'_{-i}} \pi_i(t_i)[t'_{-i}] \left\{ u_i(f(t_i', t'_{-i}), \hat{\theta}_i(t_i)) + \tau_i(t_i', t'_{-i}) \right\} \end{aligned} \tag{3.22}$$

Note that (3.22) holds for any $\bar{\tau}$ since from "if" part, given any $\bar{\tau}$ we can constructed a desirable $(\mathcal{M}, \bar{\tau})$. Since $\bar{\tau}$ can be arbitrarily close to 0, we must have

$$\sum_{t'_{-i}} \pi_i(t_i)[t'_{-i}] u_i(f(t_i, t'_{-i}), \hat{\theta}_i(t_i)) \geq \sum_{t'_{-i}} \pi_i(t_i)[t'_{-i}] u_i(f(t_i', t'_{-i}), \hat{\theta}_i(t_i)) \tag{3.23}$$

That is, f is incentive compatible. □

3.4.3 Implementation with No Transfer

In Theorem 3.2, 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.

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

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

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 3.3. *Suppose that the environment $\mathcal{E}$ satisfies Assumptions 3.1 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 3.3.1 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 3.2. □

Strict Incentive Compatibility and Separability

Following Sjöström (1994), we introduce the following class of environments. We assume the outcome space $A = A_1 \times A_2 \times ... \times A_I$, and $u_i : A_i \times \Theta_i \to \mathbb{R}$. For each SCF f and type $t \in \bar{T}$, we denote $f(t) = (f_1(t), \ldots, f_I(t))$ where $f_i(t)$ denotes the marginal distribution of $f(t)$ on A_i where $A = A_1 \times A_2 \times ... \times A_I$. 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 3.7. *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 3.4. *Suppose that a separable environment* $\mathcal{E}$ *satisfies Assumptions 3.1. 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.

$$\tau_i\left(m_i^0, m_i^1\right) = \begin{cases} \eta & \text{if } m_i^1 \neq m_i^0; \\ 0 & \text{otherwise.} \end{cases} \tag{3.24}$$

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

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

$$\bar{\tau}_i > \eta; \tag{3.25}$$

$$\tilde{\lambda}_q \sum_{t_{-i} \in \bar{T}_{-i}} \left[u_i(f_i(t_i, t_{-i}), \hat{\theta}_i(t_i)) - u_i(f_i(t_i', t_{-i}), \hat{\theta}_i(t_i)) \right] \bar{\pi}_i(t_i)\,[t_{-i}] > \gamma, \text{ for } q \in \{1, 2\}; \tag{3.26}$$

$$\eta > \epsilon E + (1 - \tilde{\lambda}_1 - \tilde{\lambda}_2) D; \tag{3.27}$$

and

$$\gamma > \epsilon E + (1 - \tilde{\lambda}_1 - \tilde{\lambda}_2) D + \eta. \tag{3.28}$$

Since f is strictly incentive compatible, the existence of γ is guaranteed in (3.26).

Remark 3.5. *In a separable environment, a proper adjustment of the weight between the* 0*th round report and the* 1*st 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 3.6. *We omit the proof of Theorem 3.4 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 $-1th$ and $0th$ rounds, the constructed SCF $\tilde{f}$ works in a similar way as the scoring rule (d_i^0) did in the proofs of Claims 2 and 3. Specifically, the function $\tilde{f}$ is constructed such that each player i's payoff from $\tilde{f}$ is affected only by his own $-1th$ (resp. $0th$) round report and the other players' $-2th$ (resp. $-1th$) round report. By the strict incentive compatibility, each player will announce truthfully in the $-1th$ (resp. $0th$) round(given the truth telling in the $-2th$ (resp. $-1th$) reports for everyone). When all players tell the truth in every round, the constructed function $\tilde{f}$ coincides with the SCF f. This enables the mechanism to implement f without any ex post transfers. Finally, the last round of elimination of strictly dominated strategies works in a way that is parallel to the proof of Claim 4.

3.5 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.

3.5.1 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

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

Definition 3.8. *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 $t_i \in T_i$,*

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

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 $\pi_i(t_i)[E] = \bar{\pi}_i(t_i)\left[\bar{T}_{-i} \cap E\right]$ for any measurable $E \subset T_{-i}$.

Definition 3.9. *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 $U(\mathcal{M}, \bar{\tau}, \mathcal{T})$ **(strictly) continuously***

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

We introduce two variants of continuous implementation:

Definition 3.10. *An SCF $f : \bar{T} \to \Delta(A)$ is continuously implementable with* ***transfers*** *if there exists a mechanism $(\mathcal{M}, \bar{\tau})$ such that for each model $\mathcal{T}$ with $\bar{\mathcal{T}} \subset \mathcal{T}$, there is a Bayes Nash equilibrium σ in $U(\mathcal{M}, \bar{\tau}, \mathcal{T})$ that continuously implements f.*

Definition 3.11. *An SCF $f : \bar{T} \to \Delta(A)$ is continuously implementable with* ***arbitrarily small transfers*** *if for any $\bar{\tau} > 0$, there exists a mechanism $(\mathcal{M}, \bar{\tau})$ such that for each model $\mathcal{T}$ with $\bar{\mathcal{T}} \subset \mathcal{T}$, there is a Bayes Nash equilibrium σ in $U(\mathcal{M}, \bar{\tau}, \mathcal{T})$ that continuously implements f.*

First, we establish the following important lemma.

Lemma 3.3. *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 $t[n] \to_p \bar{t}$, then, for each n large enough, we have $S^{\infty}W(t[n]|\mathcal{M}, \mathcal{T}) \subset S^{\infty}W(\bar{t}|\mathcal{M}, \mathcal{T})$.*

Let $\mathcal{M}$ be any one of the mechanisms used in Section 3.4. The proof of Lemma 3.3 builds upon the following claims.

Claim 3.6. *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 $t[n] \to_p \bar{t}$, there exists $N_1 \in \mathbb{N}$ such that for any $n \geq N_1$, we have if $m_i \in W_i^1(t_i[n]|\mathcal{M}, \mathcal{T})$, then $m_i^{-2} = \bar{t}_i$.*

Proof. Fix $\bar{t} \in \bar{T}$. Let $\{t[n]\}_{n=0}^{\infty}$ be such that $t[n] \to_p \bar{t}$. There exists a natural number $N_1 \in \mathbb{N}$ such that for each $n > N_1$, we have $\hat{\theta}_i(t_i[n]) = \hat{\theta}_i(\bar{t}_i) = \bar{\theta}_i$ 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 3.1 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 3.2.2 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 such that $t[n] \to_p \bar{t}$, 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 3.6.

Claim 3.7. *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 *such that* $t[n] \to_p \bar{t}$, *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 Claims 2, 3, and 5 in Section 3.3, 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 follows from showing 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 3.6, 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

$$\begin{aligned} & \sum_{t_{-i},m_{-i}} \left[u_i(g(m_i,m_{-i}),\bar{\theta}_i) + \tau_i(m_i,m_{-i})\right] \nu^{[n]}(m_{-i}|t_{-i})\pi_i(t_i[n])[t_{-i}] \\ \geq & \sum_{t_{-i},m_{-i}} \left[u_i(g(m_i',m_{-i}),\bar{\theta}_i) + \tau_i(m_i',m_{-i})\right] \nu^{[n]}(m_{-i}|t_{-i})\pi_i(t_i[n])[t_{-i}], \end{aligned} \tag{3.29}$$

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

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

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

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

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

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

where $(\bar{t}_{-i})^{\varepsilon_n}$ denotes an open ball consisting of the set of types t_{-i} whose $(k-1)$-order beliefs are ε_n-close to those of types $\bar{t}_{-i}$.[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:

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

[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).

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

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

Therefore,

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

By construction, $\nu(m_{-i}|\bar{t}_{-i})\pi_i(\bar{t}_i)[\bar{t}_{-i}] > 0$ implies that $m_{-i} \in S_{-i}^k(t_{-i}[n]|\mathcal{M}, \mathcal{T})$. By our induction hypothesis, $S_{-i}^k(t_{-i}[n]|\mathcal{M}, \mathcal{T}) \subset S_{-i}^k(\bar{t}_{-i}|\mathcal{M}, \mathcal{T})$. 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 3.5. *Suppose that the environment $\mathcal{E}$ satisfies Assumptions 3.1 and 3.2. 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 $\bar{t} \in \bar{T}$, $m \in S^{\infty}W\left(\bar{t}|\mathcal{M}, \bar{\mathcal{T}}\right) \Rightarrow g(m) = f(\bar{t})$. Note that $S^{\infty}W\left(\bar{t}|\mathcal{M}, \bar{\mathcal{T}}\right) = \{(\bar{t}, ..., \bar{t})\}$. We write σ^* such that $\sigma_i^*(\bar{t}_i) = (\bar{t}_i, ..., \bar{t}_i)$ for all $\bar{t}_i \in \bar{T}_i$. Now pick any $\mathcal{T}$ such that $\bar{\mathcal{T}} \subset \mathcal{T}$. It is well known that a trembling hand perfect equilibrium[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 $\sigma_i(\bar{t}_i) = \sigma_i^*(\bar{t}_i)$ for each $\bar{t}_i \in \bar{T}_i$. Because M is finite and T is countable, standard arguments (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 $\{t[n]\}_{n=0}^{\infty}$ such that $t[n] \to_p \bar{t}$. It is clear that, for each n : $\text{Supp}(\sigma(t[n])) \subset S^{\infty}W(t[n]|\mathcal{M}, \mathcal{T})$. In addition, for n large enough, we know by Lemma 3.3 that $S^{\infty}W(t[n]|\mathcal{M}, \mathcal{T}) \subset S^{\infty}W\left(\bar{t}|\mathcal{M}, \bar{\mathcal{T}}\right)$ and so, $(g \circ \sigma)(t[n]) = f(\bar{t})$ as claimed. □

[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 $U(\mathcal{M}, \bar{\tau}, \mathcal{T})$ 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, $m_i \in \text{supp}(\sigma_i(t_i)) \Rightarrow m_i \in \text{argmax}_{m_i' \in M_i} V_i\left(\left(m_i', \sigma_{-i}^k\right), t_i\right)$, for every k.

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.

Theorem 3.6. *Suppose that the environment $\mathcal{E}$ satisfies Assumptions 3.1 and 3.2. 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 Theorem 3.5.

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 3.2, we conclude that f is incentive compatible. □

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

Proposition 3.1 (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).

3.5.2 $\overline{UNE}$ 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 3.12. *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*

[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.

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 3.5, we record the following result:

Corollary 3.1. *Suppose that the environment $\mathcal{E}$ satisfies Assumption 3.1 and $\bar{\mathcal{T}}$ is a complete-information model. Assume $I \geq 2$. **Any** SCF f is $\overline{UNE}$-implementable with transfers.*

More importantly, we obtain the following permissive result:

Corollary 3.2. *Suppose that the environment $\mathcal{E}$ satisfies Assumption 3.1 and $\bar{\mathcal{T}}$ is a complete-information model. Assume $I \geq 2$. Any incentive-compatible SCF f is $\overline{UNE}$-implementable with no transfers.*

Remark 3.7. *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 deviation-*

s. This new SCF is equivalent to the original SCF under the hypothesis of complete information so that we can make any SCF be incentive-compatible. So, when $I \geq 3$, we can drop incentive compatibility completely from Corollary 3.2. 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 3.2. The rest of the proof is completed by Theorem 3.6. □

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.

3.5.3 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 equilib-

ria. 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 3.2, we obtain the following:

Corollary 3.3. *Suppose that the environment $\mathcal{E}$ satisfies Assumptions 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

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

use weakly dominated actions.

3.6 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.

3.6.1 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 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

[10]In finite games, it is well known that an action is strictly dominated if and only if it is

rationalizability correspondence is upper hemi-continuous. Hence, we obtain the following result:

Proposition 3.2. *Suppose that the environment $\mathcal{E}$ satisfies Assumptions 3.1 and 3.2. 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, without either of these we know of no rationalizable implementation result with finite mechanism. For any SCF f, we denote by f^τ the augmentation of f by ex post transfers τ. We interpret f^τ as an 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

a never best response.

(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.

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

$$u_i(x_i(t_i), (\hat{\theta}_i(t_i), \hat{\theta}_{-i}(t_{-i}))) - u_i(x_i(t_i'), (\hat{\theta}_i(t_i), \hat{\theta}_{-i}(t_{-i}))) > \zeta. \qquad (3.30)$$

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 3.3. *Suppose that the environment $\mathcal{E}$ satisfies Condition (S) and Assumption 3.2. 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 Claims 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 (3.14) is replaced by

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

[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.

By inequality (3.30), the last inequality in (3.14) 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 $u_i^A(t_i) = (u_i^a(t_i))_{a \in A}$

Assumption 3.3. *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*

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

[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.

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 3.4. *(Serrano and Vohra (2005)) Suppose that the environment* $\mathcal{E}$ *satisfies Assumption 3.3. 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* $t_i \neq t_i'$,

$$\sum_{\theta} u_i(x_i(t_i), \theta) h_i^1(t_i)[\theta] > \sum_{\theta} u_i(x_i(t_i'), \theta) h_i^1(t_i)[\theta], \tag{3.31}$$

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

Remark 3.8. *It is easy to see that Condition* (S) *implies type diversity.*

In Example 3.1 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 3.1. *$A = \{a_1, a_2\}$; $I = \{1, 2, 3\}$; $\bar{T}_i = \{t_i^1, t_i^2\}$ 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.

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 3.4, we can find a set of lotteries $\{x_i(t_i)\}_{t_i \in \bar{T}_i, i \in I}$ satisfying inequality (3.31). Therefore, for any $\bar{\tau} > 0$, we can adopt the corresponding mechanism $(\mathcal{M}, \bar{\tau})$ defined in Section 3.3.1 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 3.8. *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 (3.31) and the corresponding mechanism $(\mathcal{M}, \bar{\tau})$ defined in Section 3.3.1. 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 for (full) exact rationalizable implementation in interdependent value environment.

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]

[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.

3.6.3 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).

3.6.4 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 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.

.1 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.

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

[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 $\sigma_{-i}(t_{-i})[H_{-i}(t_{-i})] = 1$.

$W_i^\infty\left(\bar{t}_i|\mathcal{M},\bar{\mathcal{T}}\right)$ 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 (B.1) holds. □

Proposition .4. *Suppose that the environment $\mathcal{E}$ satisfies Assumptions 3.1 and 3.2. 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 3.3.1, with the proper scoring rule d_i^0 given in Claim 8, and λ under $\gamma'=\gamma$ (which is defined in Section 3.3.1). To prove Proposition B.1, 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 (3.14). 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 3.5. 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 (3.14). 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 B.2, 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 B.1, 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 (B.1). Next, the utility loss comes from the random dictator component of the outcome function, which is bounded above from ϵE. By inequality (3.13), 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 B.1, 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 B.2, 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 B.2.

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 B.3, $(\bar{t}_i, m_i^{-1}, \bar{t}_i, ..., \bar{t}_i) \in W_i^\infty(\bar{t}_i|\mathcal{M}, \bar{\mathcal{T}})$. Since the indicator function $e(\cdot)$ has a positive weight in this case, by inequality (3.14), 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 3.2, 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. □

.1.2 Proof of Claim 3.8

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 3.8. 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 B.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 (3.31) 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 $u_i(a, t_{-i}) = a \cdot t_{i+1}$, $u_i(x_i(t_i'), t_{i+1}') - u_i(x_i(t_i), t_{i+1}') > 0$ if $t_i \neq t_i' = t_{i+1}'$. □

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 B.7 in the following three steps.

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

Fix any $\tilde{m}_i$. First, consider the case that $\tilde{m}_i^k \neq t_i'$ for some $k \in \{-1, 0\}$.

The utility gain in payment rule λd_i^0 from using $\sigma_i'(t_i)$ rather than $\tilde{m}_i$ is

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

where $t_{i+1} = t_{i+2} = t_i \neq t'_i = t'_{i+1} = t'_{i+2}$ and the first equality follows from that $\pi_i(t_i)[t_{-i}] = \pi_i(t'_i)[t'_{-i}]$ in this example; the last inequality follows from inequality (3.10). 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 (B.2).

By inequality (3.13), 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 (B.2). 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 B.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. □

Bibliography

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.

Aghion, P., D. Fudenberg, R. Holden, T. Kunimoto, and O. Tercieux (2012): "Subgame-Perfect Implementation Under Information Perturbations," *The Quarterly Journal of Economics*, 1843–1881.

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.

Battigalli, P. and M. Siniscalchi (1999): "Hierarchies of conditional beliefs and interactive epistemology in dynamic games," *Journal of Economic Theory*, 88, 188–230.

——— (2002): "Strong Belief and Forward Induction Reasoning," *Journal of Economic Theory*, 106, 356–391.

Ben-Porath, E. (1997): "Rationality, Nash equilibrium and backwards induction in perfect-information games," *The Review of Economic Studies*, 64, 23–46.

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

Bergemann, D., S. Morris, and O. Tercieux (2011): "Rationalizable Implementation," *Journal of Economic Theory*, 146, 1253–1274.

Bernheim, B. D. (1984): "Rationalizable Strategic Behavior," *Econometrica*, 52, 1007–1028.

Bolton, P. and M. Dewatripont (2005): *Contract theory*, MIT press.

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., T. Kunimoto, and Y. Sun (2014): "Implementation with transfers," *working paper*.

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.

Dasgupta, P., P. Hammond, and E. Maskin (1979): "The Implementation of Social Choice Rules: Some General Results on Incentive Compatibility." *Review of Economic Studies*, 46.

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.

Dekel, E. and M. Siniscalchi (2013): "Epistemic game theory," Tech. rep., Mimeo.

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 (1991): "A necessary and sufficient condition for two-person Nash implementation," *The Review of Economic Studies*, 58, 121–128.

——— (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*.

Fudenberg, D. and J. Tirole (1991): *Game theory, 1991*, Cambridge, Massachusetts.

Glazer, J. and M. Perry (1996): "Virtual Implementation in Backwards Induction," *Games and Economic Behavior*, 15, 27–32.

Glazer, J. and A. Rubinstein (1996): "An extensive game as a guide for solving a normal game," *Journal of Economic Theory*, 70, 32–42.

Hart, O. and J. Moore (2003): "Some (Crude) Foundations of incomplete contracts," *Mimeo*.

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

——— (2001): "A crash course in implementation theory," *Social Choice and Welfare*, 18, 655–708.

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

Laffont, J.-J. and D. Martimort (2001): *The Theory of Incentives: The Principal-Agent Model*, Princeton University Press.

Maskin, E. (1999): "Nash Equilibrium and Welfare Optimality," *Review of Economic Studies*, 66, 23–38.

Maskin, E. and J. Moore (1999): "Implementation and Renegotiation," *Review of Economic Studies*, 66, 39–56.

Maskin, E. and J. Tirole (1999): "Unforeseen contingencies and incomplete contracts," *The Review of Economic Studies*, 66, 83–114.

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.

Moore, J. (1992): "Implementation, contracts, and renegotiation in environments with complete information," in *Advances in Economic Theory (Proceedings of the Sixth World Congress of the Econometric Society)*, ed. by J.-J. Laffont, Cambridge, England: Cambridge University Press.

Moore, J. and R. Repullo (1988): "Subgame Perfect Implementation," *Econometrica*, 56, 1191–1220.

Müller, C. (2013a): "Robust Implementation in Weakly Rationalizable Strategies," .

——— (2013b): "Robust Virtual Implementation under Common Strong Belief in Rationality," .

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.

Pearce, D. G. (1984): "Rationalizable Strategic Behavior and the Problem of Perfection," *Econometrica*, 52, 1029–1050.

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

Qin, C.-Z. and C.-L. Yang (2009): "Make a guess: a robust mechanism for King Solomons dilemma," *Economic Theory*, 39, 259–268.

——— (2013): "Finite-order type spaces and applications," *Journal of Economic Theory*, 148, 689–719.

Reny, P. J. (1992): "Backward induction, normal form perfection and explicable equilibria," *Econometrica: Journal of the Econometric Society*, 627–649.

Rényi, A. (1955): "On a new axiomatic theory of probability," *Acta Mathematica Hungarica*, 6, 285–335.

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.

Weinstein, J. and M. Yildiz (2007): "A structure theorem for rationalizability with application to robust predictions of refinements," *Econometrica*, 75, 365–400.

Williamson, O. E. (1975): *Markets and hierarchies: antitrust analysis and implications*, The Free Press New York.

Appendix A

Proofs of Chapter One

Revisit to the necessary condition in Moore and Repullo (1988)

In this section, we restate the necessary condition, i.e., Condition C, in Theorem 1 of Moore and Repullo (1988) and show that Condition C is trivially satisfied in qusilinear environment.

Condition C For each pair of profiles θ and ϕ in Θ, and for each $a \in f(\theta)$ but $a \notin f(\phi)$, there exists a finite sequence

$$a(\theta, \phi; a) \equiv \{a_0 = a, a_1, ..., a_k, ..., a_h = x, a_{h+1} = y\} \subset A,$$

with $h = h(\theta, \phi; a) \geq 1$, such that:

(1) for each $k = 0, ..., h-1$, there is some particular agent $j(k) = j(k|\theta, \phi; a)$, say, for whom

$$a_k R^{j(k)}(\theta) a_{k+1}; \text{ and}$$

(2) there is some particular agent $j(h) = j(h|\theta, \phi; a)$, say, for whom

$$[x =] a_h R^{j(h)}(\theta) a_{h+1} [= y] \text{ and } [y =] a_{h+1} P^{j(h)} a_h [= x].$$

Further, $h(\theta, \phi; a)$ is uniformly bounded by some $\bar{h} < \infty$.

We first show that with sufficiently large transfers, Conditon C is trivially satisfied in qusilinear environment.

To see Condition C is trivially satisfied when large enough transfers are allowed, we consider a pair of states $\{(\theta_i, \theta_{-i}), (\theta_i', \theta_{-i})\}$ and $a = f(\theta_i, \theta_{-i}) \neq f(\theta_i', \theta_{-i})$.

Since the state space is finite, there exists a large enough bound $\bar{T} \in \mathbb{R}_+$, and $t_x, t_y \leq \bar{T}$, $x, y \in A$, such that $\{x, t_x\}$ and $\{y, t_y\}$ is a pair of outcomes, satisfying

$$u_i(x, \theta_i) - t_x > u_i(y, \theta_i) - t_y,$$

$$u_i(x, \theta_i') - t_x < u_i(y, \theta_i') - t_y.$$

Further, $u_i(a, \theta_i) > u_i(a', \theta_i) - t$, for all $\theta_i \in \Theta_i$, for any $t \in \{t_x, t_y\}$.

Now, let the finite sequence be

$$a(\theta, \phi; a) \equiv \{a_0 = a, a_1 = \{x, t_x\}, a_2 = \{y, t_y\}\}.$$

Let $j(0) = j(1) = i$. We have

$$u_i(a, \theta_i) > u_i(x, \theta_i) - t_x > u_i(y, \theta_i) - t_y$$

that is, (1) in Condition C holds;

and

$$u_i\left(x,\theta_i\right)-t_x>u_i\left(y,\theta_i\right)-t_y,$$

$$u_i\left(x,\theta_i'\right)-t_x<u_i\left(y,\theta_i'\right)-t_y$$

that is, (2) in Condition C holds.

We show that with full use of lotteries, the large payments can be decreased into arbitrarily small scale.

Recall that for any distinct types θ_i and θ_i', there exists a pair of lotteries $\left\{x_{\theta_i,\theta_i'},x_{\theta_i',\theta_i}\right\}$ such that

$$u_i(x_{\theta_i,\theta_i'},\theta_i)>u_i(x_{\theta_i',\theta_i},\theta_i);$$

$$u_i(x_{\theta_i,\theta_i'},\theta_i')<u_i(x_{\theta_i',\theta_i},\theta_i').$$

For any $\bar{t}>0$, we can find some small enough $p_a>0$, such that there exists $t<\bar{t}$,

$$u_i\left((1-p_a)a+p_ax_{\theta_i,\theta_i'},\theta_i\right)-t>u_i((1-p_a)a+p_ax_{\theta_i',\theta_i},\theta_i)-t;$$

$$u_i((1-p_a)a+p_ax_{\theta_i,\theta_i'},\theta_i')-t<u_i\left((1-p_a)a+p_ax_{\theta_i',\theta_i},\theta_i'\right)-t.$$

In our mechanism, the finite sequence is

$$a\left(\theta,\phi;a\right)\equiv\left\{a_0=a,a_1=\left\{(1-p_a)a+p_ax_{\theta_i,\theta_i'},-t\right\},a_2=\left\{(1-p_a)a+p_ax_{\theta_i',\theta_i},-t\right\}\right\}.$$

Appendix B

Proofs of Chapter Three

Order Independence

In this Appendix, 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.

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}.[1]

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

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. Note that $W^{\infty}\left(\bar{t}|\mathcal{M}, \bar{\mathcal{T}}\right)$ is dependent 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.

[1] 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 $\sigma_{-i}(t_{-i})[H_{-i}(t_{-i})] = 1$.

Claim B.1. *Assume that the environment $\mathcal{E}$ satisfies Assumption 2. Given $\gamma' > 0$. 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*

$$\lambda \left| \sum_{t_{-i} \in \bar{T}_{-i}} \left[d_i^0 \left(\hat{\sigma}_{-i}^{-2} \left(t_{-i} \right), t_i' \right) - d_i^0 \left(\hat{\sigma}_{-i}^{-2} \left(t_{-i} \right), t_i'' \right) \right] \pi_i \left(t_i \right) \left[t_{-i} \right] \right| > \gamma'. \tag{B.1}$$

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 (B.1) holds. $\square$

Proposition B.1. *Suppose that the environment $\mathcal{E}$ satisfies Assumptions 3.1 and 3.2. 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 3.3.1, with the proper scoring rule d_i^0 given in Claim 8, and λ under $\gamma' = \gamma$ (which is defined in Section 3.3.1). To prove Proposition B.1, it suffices to show 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 $m_i^{-1} = \bar{t}_i$. The rest of the proof is identical to the proof of Theorem 3.2. We prove this result in the following two claims.

Claim B.2. *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. Define σ_i such that $\sigma_i\left(\bar{t}_i\right) = \left(\bar{t}_i, ..., \bar{t}_i\right)$ for player i of type $\bar{t}_i$. We prove this claim in two steps.

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

Note that $\sigma\left(\bar{t}\right) \in W^0\left(\bar{t}|\mathcal{M}, \bar{\mathcal{T}}\right)$. Suppose $\sigma\left(\bar{t}\right) \in W^k\left(\bar{t}|\mathcal{M}, \bar{\mathcal{T}}\right)$, for some $k \geq 0$, we show that $\sigma\left(\bar{t}\right) \in W^{k+1}\left(\bar{t}|\mathcal{M}, \bar{\mathcal{T}}\right)$. For any $\tilde{m}_i \in M_i$, we show that $\tilde{m}_i$ cannot weakly dominate $\sigma_i\left(\bar{t}_i\right)$ in two cases: (i) $\tilde{m}_i^{-2} \neq \sigma_i^{-2}\left(\bar{t}_i\right)$ and $\tilde{m}_i^k = \sigma_i^k\left(\bar{t}_i\right)$ for all $k \geq -1$; (ii) $\tilde{m}_i^k \neq \sigma_i^k\left(\bar{t}_i\right)$ for some $k \geq -1$. In Case (i), $\sigma_i\left(\bar{t}_i\right)$ is weakly better than $\tilde{m}_i$ for any $\hat{\sigma}_{-i} : \bar{T}_{-i} \to M_{-i}$ by inequality

(3.14). Therefore, $\tilde{m}_i$ cannot weakly dominate $\sigma_i(\bar{t}_i)$. 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 3.5. Therefore, $\tilde{m}_i$ cannot weakly dominate $\sigma_i(\bar{t}_i)$. Thus, $\sigma(\bar{t}) \in W^{k+1}\left(\bar{t}|\mathcal{M}, \bar{\mathcal{T}}\right)$. 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 $\left(m_i^{-2}, \bar{t}_i, ..., \bar{t}_i\right) \in W_i^{\infty}\left(\bar{t}_i|\mathcal{M}, \bar{\mathcal{T}}\right)$ for $m_i^{-2} \neq \bar{t}_i$. For any $\tilde{m}_i \in M_i$, we show that $\tilde{m}_i$ cannot weakly dominate $\left(m_i^{-2}, \bar{t}_i, ..., \bar{t}_i\right)$ in two cases: (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), since $m_i^{-2} \neq \bar{t}_i$, then we must have that $e(\bar{m}^0, \bar{m}^1) = 0$ for any $\bar{m} \in W^{\infty}\left(\bar{t}\mathcal{M}, \bar{\mathcal{T}}\right)$, for any $\bar{t}$. (Note that m_i is weakly dominated whenever $e(\bar{m}^0, \bar{m}^1) \neq 0$ for some $\bar{m} \in W^{\infty}\left(\bar{t}\mathcal{M}, \bar{\mathcal{T}}\right)$. See inequality (3.14)). Therefore, player i of type $\bar{t}_i$ is indifferent between $\tilde{m}_i$ and $\left(m_i^{-2}, \bar{t}_i, ..., \bar{t}_i\right)$. 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, $\tilde{m}_i$ cannot weakly dominate $\left(m_i^{-2}, \bar{t}_i, ..., \bar{t}_i\right)$. This completes the proof. □

Claim B.3. *Fix any player i and type $\bar{t}_i$. If $m_i \in W_i^{\infty}\left(\bar{t}_i|\mathcal{M}, \bar{\mathcal{T}}\right)$, then $\left(\bar{t}_i, m_i^{-1}, \bar{t}_i, ..., \bar{t}_i\right) \in W_i^{\infty}\left(\bar{t}_i|\mathcal{M}, \bar{\mathcal{T}}\right)$.*

Proof. By Step 1 in the proof of Claim B.2, it suffices to consider the case that $m_i^{-1} \neq \bar{t}_i$. For any $\tilde{m}_i \in M_i$, we show that $\tilde{m}_i$ cannot weakly dominate

$\left(\bar{t}_i, m_i^{-1}, \bar{t}_i, ..., \bar{t}_i\right)$ in two cases: (i) $\tilde{m}_i^{-1} \neq m_i^{-1}$ and $\tilde{m}_i^k = \sigma_i^k\left(\bar{t}_i\right)$ 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 $\tilde{m}_i$ such that $\tilde{m}_i^{-1} \neq m_i^{-1}$ 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 $\hat{\sigma}_{-i}\left(\bar{t}_{-i}\right) \in W_{-i}^{\infty}\left(\bar{t}_{-i} | \mathcal{M}, \bar{\mathcal{T}}\right)$ for all $\bar{t}_{-i}$.

Since $m_i \in W_i^{\infty}\left(\bar{t}_i | \mathcal{M}, \bar{\mathcal{T}}\right)$, 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 $\sigma'_{-i}\left(\bar{t}_{-i}\right) \in W_{-i}^{\infty}\left(\bar{t}_{-i} | \mathcal{M}, \bar{\mathcal{T}}\right)$ *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 $\hat{\sigma}_{-i}\left(\bar{t}_{-i}\right) \in W_{-i}^{\infty}\left(\bar{t}_{-i} | \mathcal{M}, \bar{\mathcal{T}}\right)$ for all $\bar{t}_{-i}$.

By Claim B.1, Case (1) is impossible. Thus, we must have Case (2). Since m_i and $\tilde{m}_i$ only differs in round -1, the utility difference 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 (larger than γ by inequality (B.1) together with a potential utility loss through e function (bounded above by ϵE), which is at least larger than $\gamma - \epsilon E$. By inequality (3.13), $\gamma - \epsilon E > 0$.

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

Since $m_i^{-1} \neq \bar{t}_i$ and $m \in W_i^\infty\left(\bar{t}_i|\mathcal{M},\bar{\mathcal{T}}\right)$, by Claim 2, there exists a nonempty set of players $J \subset I\backslash\{i\}$ such that $\hat{\sigma}_j^{-2}\left(\bar{t}_j\right) \neq \bar{t}_j$ for all $j \in J$, of type $\bar{t}_j$. From Claim B.2, we know that $\left(\hat{\sigma}_j^{-2}\left(\bar{t}_j\right), \bar{t}_j, ..., \bar{t}_j\right) \in W_j^\infty\left(\bar{t}_j|\mathcal{M},\bar{\mathcal{T}}\right)$ for all $j \in J$. Define $\tilde{\sigma}_{-i}$ such that $\tilde{\sigma}_{-i}^{-2}\left(\bar{t}_{-i}\right) = \hat{\sigma}_{-i}^{-2}(\bar{t}_{-i})$ and $\tilde{\sigma}_{-i}^{k}\left(\bar{t}_{-i}\right) = \sigma_{-i}\left(\bar{t}_{-i}\right)$ for all $\bar{t}_{-i}$ and $k \geq -1$. Thus, $\tilde{\sigma}_{-i}\left(\bar{t}_{-i}\right) \in W_{-i}^\infty\left(\bar{t}_{-i}|\mathcal{M},\bar{\mathcal{T}}\right)$ for all $\bar{t}_{-i}$.

Fix conjecture $\tilde{\sigma}_{-i}$. Since $\left(\bar{t}_i, m_i^{-1}, \bar{t}_i, ..., \bar{t}_i\right)$ and $\tilde{m}_i$ only differs in round -1, the utility difference for player i of type $\bar{t}_i$ by using $\left(\bar{t}_i, m_i^{-1}, \bar{t}_i, ..., \bar{t}_i\right)$ rather than $\tilde{m}_i$ is concentrated in the payment rule λd_i^0 together with a potential utility loss through e function, which is larger than $\gamma - \epsilon E$ by the proof of Step 1. Therefore, $\tilde{m}_i$ cannot weakly dominate $\left(\bar{t}_i, m_i^{-1}, \bar{t}_i, ..., \bar{t}_i\right)$.

In Case (ii), $\tilde{m}_i$ cannot weakly dominate $\left(\bar{t}_i, m_i^{-1}, \bar{t}_i, ..., \bar{t}_i\right)$, as we can make an argument parallel to Step 2 in the proof of Claim B.2.

Thus, $\tilde{m}_i$ cannot weakly dominate $\left(\bar{t}_i, m_i^{-1}, \bar{t}_i, ..., \bar{t}_i\right)$. This completes the proof. □

Claim B.4. *Fix 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 $m_i^{-1} = \bar{t}_i$.*

Proof. Suppose not, that is, there exists some $m_i \in W_i^\infty\left(\bar{t}_i|\mathcal{M},\bar{\mathcal{T}}\right)$ such that $m_i^{-1} \neq \bar{t}_i$. Then by Claim B.3, $\left(\bar{t}_i, m_i^{-1}, \bar{t}_i, ..., \bar{t}_i\right) \in W_i^\infty\left(\bar{t}_i|\mathcal{M},\bar{\mathcal{T}}\right)$. By inequality (3.14), we conclude that for any $j \in I\backslash\{i\}$ and $\bar{t}_j \in \bar{T}_j$, if $m_j \in W_j^\infty\left(\bar{t}_j|\mathcal{M},\bar{\mathcal{T}}\right)$, then $m_j^{-2} = \bar{t}_j$. Suppose $m_i \in W_i^\infty\left(\bar{t}_i|\mathcal{M},\bar{\mathcal{T}}\right)$. Then, by Claim 3.2, we have $m_i^{-1} = \bar{t}_i$. This is a contradiction. □

Proof of Claim in Example 1

Since now player i's preferences only depends on player $i+1$'s type, for simplicity of notation, we write player i's preference as follows, $u_i(a,t) \equiv u_i(a,t_{i+1}) = a \cdot t_{i+1}$, for any a and any t.

For any $\bar{\tau} > 0$, for Example 1, we adopt a mechanism $(\mathcal{M}, \bar{\tau})$ defined in Section 3.3.1. Let σ' be a strategy profile such that $\sigma'_i(t_i) = (t'_i, ..., t'_i)$ such that $t_i \neq t'_i$ for all player $i \in I$ and all $t_i \in \bar{T}_i$. We will show that $\sigma'_i(t_i) \in S_i^{\infty} W_i\left(t_i | \mathcal{M}, \bar{\mathcal{T}}\right)$, for all i and t_i. We prove this in the following claims. Throughout this section, we write $t_i = t_j \neq t'_j = t'_i$ for all $i, j \in I$. Therefore, $t'_{-i} \neq t_{-i}$ if and only if $t'_j \neq t_i$ for all $j \neq i$.

Claim B.5. *For any player i of type t_i,*

$$\sum_{t_{-i} \in \bar{T}_{-i}} u_i\left(f\left(t'_i, \sigma'_i(t_{-i})\right), t_{i+1}\right) \pi_i(t_i)[t_{-i}] \geq \sum_{t_{-i} \in \bar{T}_{-i}} u_i\left(f\left(t_i, \sigma'_i(t_{-i})\right), t_{i+1}\right) \pi_i(t_i)[t_{-i}]. \tag{B.2}$$

Proof. For any $t'_{-i} \neq t_{-i}$, $\pi_i(t_i)[t_{-i}] = \pi_i(t'_i)\left[t'_{-i}\right]$ in this example. Therefore, by the construction of f, f does not depend on player i's type, from player i's perspective. □

Claim B.6. *Fix any set of lotteries $\{x_i(t_i)\}_{i \in I, t_i \in T_i}$ such that satisfying inequality (3.31). For any player i of type t_i, $x_i(t_i)[a] > \frac{1}{2}$ if and only if $t_i = a$.*

Proof. Consider any outcome a. Player i of type t_i's interim utility is as follows:

$$\sum_{t_{i+1}} u_i\left(a, t_{i+1}\right) = \frac{2}{3} a \cdot t_i + \frac{1}{3} a \cdot t_i'.$$

Therefore, we can see player i of type t_i strictly prefer a to the other outcome whenever $a = t_i$. Since $\left\{x_i\left(t_i\right)\right\}_{i \in I, t_i \in T_i}$ is such that inequality (3.31) holds, and there are only two outcome in A, we must have $x_i\left(t_i\right)[a] > \frac{1}{2}$ if and only if $t_i = a$. □

Claim B.7. *In the game* $U\left(\mathcal{M}, \bar{\mathcal{T}}\right)$*, for every* $i \in I$*,* $t_i \in \bar{T}_i$*,* $\sigma_i'\left(t_i\right) \in S_i^{\infty} W_i\left(t_i | \mathcal{M}, \bar{\mathcal{T}}\right)$.

Note that $\sigma'(t) \in W^0\left(t | \mathcal{M}, \bar{\mathcal{T}}\right)$. Suppose $\sigma'(t) \in S^{\tilde{k}}\left(t | \mathcal{M}, \bar{\mathcal{T}}\right)$, for some $\tilde{k} \geq 0$, we show that $\sigma'(\bar{t}) \in S^{\tilde{k}+1}\left(\bar{t} | \mathcal{M}, \bar{\mathcal{T}}\right)$. Consider player i of type t_i. For any $\tilde{m}_i \in M_i$, we show that $\tilde{m}_i$ cannot weakly dominate $\sigma_i'\left(t_i\right)$ in the following two cases.

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 such that $\bar{m}_j^{-1} = \bar{m}_j^0$ for all $j \neq i$, therefore $e(\left(m_i^{-1}, \bar{m}_{-i}^{-1}\right), \left(m_i^0, \bar{m}_{-i}^0\right)) = 0$ when $m_i^{-1} = m_i^0$. Let $\tilde{m}_{-i} \in M_{-i}$ be such that $\tilde{m}_j^{-1} \neq \tilde{m}_j^0$ for some $j \neq i$, therefore $e(\left(m_i^{-1}, \tilde{m}_{-i}^{-1}\right), \left(m_i^0, \tilde{m}_{-i}^0\right)) = \epsilon$ for all m_i. Let ν be a conjecture of type t_i such that $\nu[\bar{m}_{-i} | t_{i+1}, t_{i+2}] = 1$ and $\nu[\tilde{m}_{-i} | t_{i+1}', t_{i+2}'] = 1$. The expected

payoff gain for player i of type t_i from choosing $\sigma_i'(t_i)$ rather than $\tilde{m}_i$ is

$$\begin{aligned}
& \left\{0 \times u_i(x_i(t_i'), t_{i+1})\pi_i(t_i)[t_{-i}] + \epsilon \times u_i(x_i(t_i'), t_{i+1}')\pi_i(t_i)\left[t_{-i}'\right]\right\} \\
- & \left\{0 \times u_i(x_i(t_i), t_{i+1})\pi_i(t_i)[t_{-i}] + \epsilon \times u_i\left(x_i(t_i), t_{i+1}'\right)\pi_i(t_i)\left[t_{-i}'\right]\right\} \\
= & \epsilon\left\{u_i(x_i(t_i'), t_{i+1}') - u_i\left(x_i(t_i), t_{i+1}'\right)\right\}\pi_i(t_i)\left[t_{-i}'\right] \\
> & 0.
\end{aligned}$$

The last inequality follows from Claim B.6. Therefore, $\tilde{m}_i$ cannot weakly dominate $\sigma_i'(t_i)$.

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

We show that against conjecture σ_{-i}', $\sigma_i'(t_i)$ is a strictly better message than $\tilde{m}_i$. First, consider $\tilde{m}_i^k \neq t_i'$ where $k = 0$ or 1. In terms of outcome dependent on kth message profile where $k \geq 1$, if $\tilde{m}_i^k \neq t_i'$, $\sigma_i'(t_i)$ is better message than $\tilde{m}_i$ by (B.2). Therefore, the utility difference for player i of type $\bar{t}_i$ by using $\sigma_i'(t_i)$ rather than $\tilde{m}_i$ in the payment rule λd_i^0 together with a potential utility loss bounded above by ϵE. From the construction of d_i^0, we have

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

where the first equality follows because for any $t'_{-i} \neq t_{-i}$, $\pi_i(t_i)[t_{-i}] = \pi_i(t'_i)\left[t'_{-i}\right]$ in this example; the last inequality follows from inequality (3.10). By inequality (3.13), $\gamma > \epsilon E$. Therefore, $\tilde{m}_i$ cannot weakly dominate $\sigma'_i(t_i)$.

Finally, consider $\tilde{m}_i^k \neq t'_i$ for some $k \geq 1$.In terms of outcome dependent on kth message profile where $k \geq 1$, if $\tilde{m}_i^k \neq t'_i$, $\sigma'_i(t_i)$ is better message than $\tilde{m}_i$ by (B.2). In terms of payments, $\sigma'_i(t_i)$ is a strictly better message than $\tilde{m}_i$ by the construction of $\sigma'_i(t_i)$.