Rationalizable Implementation in Finite Mechanisms*

Yi-Chun Chen† Takashi Kunimoto‡ Yifei Sun§ Siyang Xiong¶

December 24, 2020

Abstract

We prove that the Maskin monotonicity* condition (by Bergemann, Morris, and Tercieux (2011)) fully characterizes exact rationalizable implementation in an environment with lotteries and transfers. Different from previous papers, our approach possesses many appealing features simultaneously, e.g., finite mechanisms with no integer game or modulo game are used; no transfer is imposed on any rationalizable profile; the message space is small; the implementation is robust to information perturbations and continuous in the sense of Oury and Tercieux (2012).

JEL Classification: C72, D78, D82.

Keywords: Complete information, continuous implementation, implementation, information perturbations, Maskin monotonicity*, rationalizability, social choice function.

---

*Financial support from the Singapore Ministry of Education Academic Research Fund Tier 1 (R-122-000-296-115) (Chen) and the National Natural Science Foundation of China (NSFC71703020) (Sun) is gratefully acknowledged.

†Department of Economics and Risk Management Institute, National University of Singapore. Email: ecsycc@nus.edu.sg

‡School of Economics, Singapore Management University. Email: tkunimoto@smu.edu.sg

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

¶Department of Economics, University of California, Riverside. Email: siyang.xiong@ucr.edu.

1 Introduction

The design of institution to be interacted among strategic agents has been an important research agenda in economic theory. Suppose a society has decided on social choice rule – a recipe for choosing the socially optimal alternatives on the basis of individuals' preferences over alternatives. To tackle the problem of how to *implement* the rule, Maskin (1999), in his classic paper, (i) describes a decentralized decision making process as a *mechanism*, which specifies the possible actions available to members of a society, as well as the outcomes of these actions; and (ii) asks to what extent one can design a mechanism which makes its "all" Nash equilibrium outcomes socially desirable. This is called *Nash implementation*. Maskin proposes a well-known monotonicity condition, which we refer to as *Maskin monotonicity*, and shows it to be necessary and almost sufficient for Nash implementation.

A Nash equilibrium of the game is often captured as a strategy profile with the following two properties: (i) the players' strategies are best reply to their belief about other players' strategies; and (ii) their beliefs are correct. This paper is primarily concerned with redoing Nash implementation theory by using a more robust solution concept that drops (ii) and retains (i). This leads us to use the notion of *rationalizability*. An advantage of using rationalizability lies in its clean epistemic foundation, as it is the strategic consequence that comes solely from common certainty of rationality. This exhibits a contrast with a rather involved epistemic condition for Nash equilibrium (See Aumann and Brandenburger (1995)).

We study a finite environment where a finite implementing mechanism is anticipated. In such a finite environment, however, the recent contributions to rationalizable implementation by Bergemann, Morris, and Tercieux (2011)(henceforth, BMT), Jain (2019), Kunimoto and Serrano (2019), and Xiong (2019) all construct an infinite mechanism which makes use of the *integer game* for their sufficiency results. In the integer game, each agent announces some integer and the person who announces the highest integer gets to name his favorite outcome. Although the use of the integer game in the mechanism has been prevalent in the literature, it has also been considered a questionable feature (See Jackson (1992)). There is one notable exception by Abreu and Matsushima (1992), who construct a finite mechanism but consider *virtual* (as opposed to exact) implementation in rationalizable strategies. Virtual implementation means that the planner contents herself with implementing the socially desirable outcome with arbitrarily high probability. Thus, the main purpose of our paper is to characterize the class of *social choice functions* (henceforth, SCFs) that are *exactly* implementable in rationalizable strategies by a finite mechanism, which necessarily excludes the integer game constructions. Rationalizable strategies are defined as the set of strategies that survive the iterated elimination of never best responses. In finite mechanisms, as in this paper, rationalizable strategies are equivalent to the strategies that survive the iterated elimination of strictly dominated strategies.

In the environments where the designer can impose lottery allocations and transfers, our Theorem 1 shows that an SCF is implementable in rationalizable strategies by a finite mechanism if and only if it satisfies Maskin monotonicity*. Note how the *no-worst-alternative* condition (often abbreviated as NWA), used in their sufficiency results by BMT, Jain (2019), Kunimoto and Serrano (2019), and Xiong (2019), is automatically satisfied in our setup with transfers.[1] Maskin monotonicity* is proposed by BMT and stronger than Maskin monotonicity.[2] Theorem 1 handles the case of two agents as well as more than two agents; moreover, no transfer is imposed on any rationalizable profile and the message space of our implementing mechanism is small because our mechanism is slightly more complex than a direct mechanism in which each agent announces only a state and we need a mechanism more complex than a direct mechanism, which is discussed in Section 5.3. The sufficiency result of BMT, on the other hand, needs at least three agents and uses an infinite mechanism.[3]

We now highlight how this result provides new insights on classical as well as recent results in the literature. First, Oury and Tercieux (2012) advocate rationalizable implementation by finite mechanisms as a way of achieving continuous implementation. They consider the following situation: the planner wants not only that there is an equilibrium that implements the SCF but also that the same equilibrium continues to implement the SCF in all the models *close* to her initial model. Hence, the SCF is *continuously* implementable. Theorem 4 of Oury and Tercieux (2012) shows that an SCF is continuously implementable by a finite mechanism if it is exactly implementable in rationalizable strategies by a finite mechanism.[4]

[1] The no-worst-alternative condition requires that the social choice outcome never be the worst for any agent in any state.

[2] BMT show that Maskin monotonicity* is a necessary condition for rationalizable implementation using mechanisms satisfying what they call the best-response property (which include finite mechanisms).

[3] The mechanisms constructed to prove the positive results in Bergemann, Morris, and Tercieux (2011), Jain (2019), Kunimoto and Serrano (2019), and Xiong (2019) all invoke an infinite message space with integer games and state-contigent allocations. In contrast, the message space of our implementing mechanism is "small" in the sense that it only consists of few reports of payoff-relevant information such as types or states. Note that the message space in Abreu and Matsushima (1992) can also be made "small" in a similar sense if they allow for large transfers as we do.

[4] Oury and Tercieux (2012) also prove the "only if" part of the result under a further assumption that sending messages is slightly costly.

This leaves open a characterization of SCFs which are exact implementable in rationalizable strategies by a finite mechanism. Our Theorem 1 addresses this important open issue. In particular, it follows from Theorem 1 that any SCF which satisfies Maskin monotonicity* is continuously implementable.[5]

Second, we also discuss rationalizable implementation when the SCF is *responsive*. A responsive SCF assigns distinct outcomes to different states. BMT observe that when the SCF is responsive, Maskin monotonicity* reduces to Maskin monotonicity. We show that, for any SCF f, we can construct an SCF f^ε that is ε-close to f such that f^ε is responsive and satisfies Maskin monotonicity. This is summarized as our Corollary 3: "any" SCF is virtually implementable in rationalizable strategies by a finite mechanism, which is first proved by Abreu and Matsushima (1992) in the case with three or more agents. Finally, we construct an example in which some Maskin monotonic* SCF cannot be implemented in rationalizable strategies by any direct mechanism.

The rest of the paper is organized as follows. In Section 2, we present the basic setup and definitions. In Section 3, we adopt rationalizability and identify Maskin monotonicity* as a necessary and sufficient condition for rationalizable implementation by a finite mechanism. We also compare the result of this paper with Chen, Kunimoto, Sun, and Xiong (2020) who investigate mixed Nash implementation in the same class of environments. We extend our result to the case where only small transfers are allowed on and off rationalizable strategy profiles in Section 4. Section 5 discusses a number of extensions of our main result.

2 Preliminaries

2.1 Environment

Consider a finite set of agents $\mathcal{I} = \{1, 2, ..., I\}$ with $I \geq 2$; a finite set of possible states Θ; and a set of pure alternatives A. We consider an environment with lotteries and transfers. Specifically, we work with the space of allocations/outcomes $X \equiv \Delta(A) \times \mathbb{R}^I$ where $\Delta(A)$ denotes the set of lotteries on A that have a countable support, and $\mathbb{R}^I$ denotes the set of transfers to the agents.

Each state $\theta \in \Theta$ induces for each agent $i \in \mathcal{I}$ a type θ_i. Each type θ_i is associated with a bounded expected utility function $v_i(\cdot, \theta_i) : \Delta(A) \rightarrow \mathbb{R}$, and conversely, each bounded

[5] See Section 5.1 for more discussion as well as some caveats in connecting our result with Theorem 4 of Oury and Tercieux (2012).

expected utility function identifies at most one type. Let Θ_i denote the set of types/expected utility functions of agent i which can be induced from Θ. As in Abreu and Matsushima (1992), we will take for granted that distinct elements of Θ_i induce different preference orderings over $\Delta(A)$. Assume also that for any type θ_i, there are alternatives a and a' in A such that $v_i(a, \theta_i) \neq v_i(a', \theta_i)$. For each $x = \left(\ell, (t_i)_{i \in \mathcal{I}}\right) \in X$, we denote by $u_i(x, \theta_i) = v_i(\ell, \theta_i) + t_i$ the quasilinear utility function induced by θ_i.

We focus on a *complete information* environment in which the state θ is common knowledge among the agents but unknown to a mechanism designer. Thanks to the complete-information assumption, it is without loss of generality to assume that agents' values are private. The designer's objective is specified by a *social choice function* $f : \Theta \to \Delta(A)$, namely, if the state is θ, the designer would like to implement the social outcome $f(\theta)$ which is allowed to be a lottery. We can also allow an SCF to be defined as a mapping from Θ to X. In this case, our implementation requires that no additional transfers be imposed on any rationalizable message profile beyond the transfers prescribed by the SCF.

2.2 Mechanism and Solution

A mechanism $\mathcal{M}$ is a triplet $((M_i), g, (\tau_i))_{i \in \mathcal{I}}$ where M_i is the nonempty finite set of *messages* available to agent i, and we write $M \equiv \times_{i=1}^{I} M_i$; $g : M \to X$ is the *outcome function*; and $\tau_i : M \to \mathbb{R}$ is the *transfer rule* which specifies the transfer to agent i. The environment and the mechanism together constitute a *game with complete information* at each state $\theta \in \Theta$ which we denote by $\Gamma(\mathcal{M}, \theta)$.

In studying implementation in rationalizable strategies, we adopt the notion of *correlated rationalizability* defined in Brandenburger and Dekel (1987) as a solution concept. We define rationalizability for the finite game $\Gamma(\mathcal{M}, \theta)$ as follows. Let $S_i^0(\mathcal{M}, \theta) = M_i$, and we define $S_i^k(\mathcal{M}, \theta)$ inductively: for any $k > 0$, we set

$$S_i^k(\mathcal{M}, \theta) = \left\{ m_i \in M_i \,\middle|\, \begin{array}{l} \text{there exists } \lambda_i \in \Delta(M_{-i}) \text{ such that} \\ (1)\ \lambda_i(m_{-i}) > 0 \Rightarrow m_j \in S_j^{k-1}(\mathcal{M}, \theta) \text{ for each } j \neq i, \\ (2)\ m_i \in \arg\max_{m_i' \in M_i} \lambda_i(m_{-i})\, u_i(g(m_i', m_{-i}), \theta_i). \end{array} \right\}.$$

Then, $S_i^\infty(\mathcal{M}, \theta) = \bigcap_{k=0}^{\infty} S_i^k(\mathcal{M}, \theta)$ denotes the set of rationalizable messages of agent i, $S^\infty(\mathcal{M}, \theta) = \prod_{j \in \mathcal{I}} S_j^\infty(\mathcal{M}, \theta)$ the set of rationalizable message profiles, and $S_{-i}^\infty(\mathcal{M}, \theta) = \prod_{j \in \mathcal{I}} S_j^\infty(\mathcal{M}, \theta)$ the set of rationalizable message profiles of the opponents of agent i in $\Gamma(\mathcal{M}, \theta)$.

We abuse notation to identify $\ell \in \Delta(A)$ with $x^l = (\ell, 0, ..., 0) \in X$ and the range of social choice function, $f(\Theta)$, as a subset of X. While we allow the outcome function g to invoke transfers, the following implementation notion requires that for each rationalizable profile m at state θ, the outcome $g(m)$ is *exactly* the social outcome $f(\theta)$ at state θ with no transfer. In other words, we require *exact* implementation.

Definition 1 *An SCF f is* ***implementable in rationalizable strategies*** *if there exists a mechanism $\mathcal{M} = ((M_i), g, (\tau_i))_{i\in\mathcal{I}}$ such that for any $\theta \in \Theta$, (i) $S^{\infty}(\mathcal{M}, \theta) \neq \varnothing$; and (ii) for any $m \in S^{\infty}(\mathcal{M}, \theta)$, we have $g(m) = f(\theta)$ and $\tau_i(m) = 0$.*

Remark: Since we propose a finite implementing mechanism below, $S^{\infty}(\mathcal{M}, \theta)$ is always nonempty, namely, requirement (i) of rationalizable implementation is automatically satisfied.

2.2.1 Maskin Monotonicity*

In this section, we introduce a central condition to our rationalizable implementation result, which is called *Maskin monotonicity**. In our environment with penalties, *Maskin monotonicity** is equivalent to strict Maskin *monotonicity** proposed by BMT as a necessary condition for rationalizable implementation using "well behaved" (such as finite) mechanisms.

For $(\theta_i, x) \in \Theta_i \times X$, let

$$\mathcal{L}_i(x, \theta_i) = \{y \in X : u_i(x, \theta_i) \geq u_i(y, \theta_i)\}$$

denote the lower-contour set at allocation x for type θ_i of agent i. Let

$$\mathcal{U}_i(x, \theta_i) = \{y \in X : u_i(y, \theta_i) \geq u_i(x, \theta_i)\}$$

denote the upper-contour set at allocation x for type θ_i of agent i. Replacing the weak inequality with a strict one, we can define $\mathcal{SL}_i(x, \theta_i)$ and $\mathcal{SU}_i(x, \theta_i)$ as the strict lower and upper contour sets for type θ_i of agent i, respectively. For a given SCF f, we let $\mathcal{P}_f = \{\Theta_z\}_{z\in f(\Theta)}$ be the partition on Θ induced by f, i.e., $\Theta_z \equiv \{\theta \in \Theta | \ f(\theta) = z\}$. For each partition $\mathcal{P}$ on Θ, we denote by $\mathcal{P}(\theta)$ the atom in $\mathcal{P}$ which contains state θ and by $\mathcal{P}_i(\theta)$ the projection of the set of type profile $\mathcal{P}(\theta)$ on Θ_i. Moreover, for each $x \in X$, let

$$\mathcal{L}_i(x, \mathcal{P}(\theta)) \equiv \bigcap_{\tilde{\theta}\in\mathcal{P}(\theta)} \mathcal{L}_i(x, \tilde{\theta}_i).$$

The following definition is obtained by adapting Definition 5 of BMT to our setup that accommodates both lotteries and penalties.

Definition 2 *Say an SCF f satisfies **Maskin monotonicity*** if there exists a partition $\mathcal{P}$ of Θ such that (i) $\mathcal{P}$ is at least as fine as $\mathcal{P}_f$; (ii) for any $\tilde{\theta}, \theta \in \Theta$, whenever $\tilde{\theta} \notin \mathcal{P}(\theta)$, there exists $i \in \mathcal{I}$ for whom*

$$\mathcal{L}_i(f(\tilde{\theta}), \mathcal{P}(\tilde{\theta})) \cap \mathcal{SU}_i(f(\tilde{\theta}), \theta_i) \neq \varnothing. \tag{1}$$

Although Maskin monotonicity* implies Maskin monotonicity, it was not a priori clear whether the two conditions are different. Jain (2019) constructs an example showing that Maskin monotonicity* is strictly stronger than Maskin monotonicity. In Appendix A.1, we modify the example of Jain (2019) to make the same point in our setup, which accommodates the case with two agents, lotteries, and transfers. Accordingly, rationalizable implementation is strictly more restrictive than Nash implementation, when we focus on finite mechanisms and allow for lotteries and transfers.[6]

2.2.2 Challenge Scheme

Let $\mathcal{P}$ be the partition in the definition of Maskin monotonicity*. First, a *challenge scheme* for an SCF f is a set of allocations $\{x(\tilde{\theta}, \theta_i)\}$, one for each pair of state $\tilde{\theta}$ and type θ_i of agent i, such that

$$\begin{aligned} \text{if } \mathcal{L}_i(f(\tilde{\theta}), \mathcal{P}(\tilde{\theta})) \cap \mathcal{SU}_i(f(\tilde{\theta}), \theta_i) &\neq \varnothing, \text{ then } x(\tilde{\theta}, \theta_i) \in \mathcal{L}_i(f(\tilde{\theta}), \mathcal{P}(\tilde{\theta})) \cap \mathcal{SU}_i(f(\tilde{\theta}), \theta_i); \\ \text{if } \mathcal{L}_i(f(\tilde{\theta}), \mathcal{P}(\tilde{\theta})) \cap \mathcal{SU}_i(f(\tilde{\theta}), \theta_i) &= \varnothing, \text{ then } x(\tilde{\theta}, \theta_i) = f(\tilde{\theta}), \end{aligned}$$

where we omit the reference to $\mathcal{P}$ in $x(\cdot, \cdot)$ to simplify the notation. We call $x(\tilde{\theta}, \theta_i)$ a *test allocation* when $x(\tilde{\theta}, \theta_i) \in \mathcal{L}_i(f(\tilde{\theta}), \mathcal{P}(\tilde{\theta})) \cap \mathcal{SU}_i(f(\tilde{\theta}), \theta_i)$ and in such a case, agent i called a *whistle-blower* for $\tilde{\theta}$ at state θ.

The following lemma shows that there is a challenge scheme under which truth-telling induces the best allocation within $\{x(\tilde{\theta}, \theta_i)\}_{\tilde{\theta} \in \Theta, \theta_i \in \Theta_i}$.

Lemma 1 *There is a challenge scheme $\{x(\tilde{\theta}, \theta_i)\}$ for an SCF f such that for any state $\tilde{\theta}$ and type θ_i,*

$$u_i(x(\tilde{\theta}, \theta_i), \theta_i) \geq u_i(x(\tilde{\theta}, \theta_i'), \theta_i), \forall \theta_i' \in \Theta_i. \tag{2}$$

[6] In Chen, Kunimoto, Sun, and Xiong (2020), we prove that Maskin monotonicity fully characterizes mixed-strategy Nash implementation using finite mechanisms in the same setup. See Section 3.3 for the comparison between this paper and Chen, Kunimoto, Sun, and Xiong (2020) in terms of the differences in the implementing mechanism and underlying arguments.

Proof. Fix an arbitrary challenge scheme $\{\bar{x}(\tilde{\theta}, \theta_i)\}$ for an SCF f. For each state $\tilde{\theta}$ and each type θ_i, we can redefine $x(\tilde{\theta}, \theta_i)$ as the most preferred allocation of type θ_i in the (finite) menu of allocations $\{\bar{x}(\tilde{\theta}, \theta_i') : \theta_i' \in \Theta_i\}$. It is straightforward to show that $x(\tilde{\theta}, \theta_i)$ remains a challenge scheme. Thus, we satisfy the following two properties: for each $\theta_i \in \Theta_i$ and $\tilde{\theta} \in \Theta$,

$$\text{(i)}\ \ u_i(x(\tilde{\theta}, \theta_i), \theta_i) = \max_{\theta_i'' \in \Theta_i} u_i(\bar{x}(\tilde{\theta}, \theta_i''), \theta_i),$$

and (ii) for any $\theta_i^{'} \in \Theta_i$, there exists $\theta_i^{'''} \in \Theta_i$ such that

$$u_i(x(\tilde{\theta}, \theta_i'), \theta_i) = u_i(\bar{x}(\tilde{\theta}, \theta_i'''), \theta_i).$$

Then, combining the two equation above, we get

$$u_i(x(\tilde{\theta}, \theta_i), \theta_i) \geq u_i(x(\tilde{\theta}, \theta_i'), \theta_i),\ \ \forall \theta_i' \in \Theta_i.$$

This completes the proof. ■

In the following, we shall invoke a challenge scheme which satisfies (2) and we call it the *best challenge scheme*. The existence of the best challenge scheme proved in Lemma 1 demonstrates that the designer's twin goals of allowing for whistle-blowing (as in Maskin (1977, 1999)) and eliciting the truth (from the dictator lotteries as in Abreu and Matsushima (1992, 1994)) can be aligned with the test allocations pre-specified at the outset.

2.2.3 Dictator Lottery

Let $\tilde{X} \equiv A \cup \bigcup_{i \in \mathcal{I}, \theta_i \in \Theta_i, \tilde{\theta} \in \Theta} x(\tilde{\theta}, \theta_i)$ where each $a \in A$ is identified with $x^a = (a, 0, ..., 0) \in X$. Since $v_i(\cdot, \theta_i)$ is bounded and Θ is finite, we choose $\eta' > 0$ as an upper bound on the monetary value of a change in the selection of an alternative in $\tilde{X}$, that is,

$$\eta' > \sup_{i \in \mathcal{I}, \theta_i \in \Theta_i, x, x' \in \tilde{X}} \left| u_i(x, \theta_i) - u_i(x', \theta_i) \right|, \tag{3}$$

Now, we have the following lemma.

Lemma 2 *For each agent $i \in \mathcal{I}$, there exists a function $y_i : \Theta_i \cup \Theta \to X$ such that $y_i(\theta) = f(\theta)$ for each $\theta \in \Theta$ and for all types θ_i, θ_i' with $\theta_i \neq \theta_i$, we have*

$$u_i\left(y_i\left(\theta_i\right), \theta_i\right) > u_i\left(y_i\left(\theta_i'\right), \theta_i\right); \tag{4}$$

moreover, for each $j \in \mathcal{I}$ and θ_j', we also have that for every $x \in \tilde{X}$,

$$u_j(y_j(\theta_j'), \theta_j) < u_j(x, \theta_j). \tag{5}$$

From Abreu and Matsushima (1992) we can prove the existence of lotteries $\{y_i'(\cdot)\} \subset \Delta(A)$ which satisfy Condition (4). To satisfy Condition (5), we simply add a penalty of η' to each outcome of the lotteries $\{y_i'(\theta_i)\}_{\theta_i \in \Theta_i}$. More precisely, for each $\theta_i \in \Theta_i$, we set

$$y_i(\theta_i) = (y_i'(\theta_i), -\eta', \ldots, -\eta') \in X.$$

We call the resulting lotteries the *dictator lotteries* for agent i and denote them by $\{y_i(\cdot)\}$.

Condition (4) shows that under dictator lotteries, each agent has a strict incentive to reveal his true type(see the role in Step 1 of Section 3.2), whereas Condition (5) says that these dictator lotteries are strictly less preferred than any alternative a or test allocations in $\tilde{X}$(see the role in Step 3 of Section 3.2).

3 Rationalizable Implementation

We now state our main result on rationalizable implementation.

Theorem 1 *An SCF f is implementable in rationalizable strategies by a finite mechanism if and only if it satisfies Maskin monotonicity*.*

Since a finite mechanism satisfies the *best response property* defined in BMT (see Definition 6 of BMT), the "only if" part of Theorem 1 follows from Proposition 3 of BMT. In the following subsections, we will construct a mechanism to prove the "if" part of Theorem 1.

3.1 The Mechanism

3.1.1 Message Space:

A generic message of agent i is:

$$m_i = \left(m_i^1, m_i^2, m_i^3\right) \in M_i^1 \times M_i^2 \times M_i^3 = M_i = (\Theta_i \cup \Theta) \times \Theta \times \Theta.$$

That is, agent i is asked to make (1) an announcement of either his own type or the state (which we denote by m_i^1); and (2) another two announcements of the state (which we denote by m_i^2 and m_i^3).

3.1.2 Allocation Rule:

Let $\mathcal{P}$ be the partition in the definition of Maskin monotonicity*. Say two states θ and θ' are *equivalent* (denoted as $\theta \sim \theta'$) if they belong to the same atom of $\mathcal{P}$. Given a message profile m, we say that m is *consistent* if there exists $\tilde{\theta} \in \Theta$ such that

$$m_i^1 \sim \ m_i^2 \sim \ m_i^3 \sim \tilde{\theta} \text{ for every } i \in \mathcal{I}. \tag{6}$$

That is, consistency requires that every agent i announces three states m_i^1, m_i^2 and m_i^3 from the same atom of $\mathcal{P}$. Note that m is inconsistent whenever $m_i^1 \in \Theta_i$ for some agent i. We extend $x : \Theta \times \Theta_i \to X$ to $x : \Theta \times (\Theta_i \cup \Theta) \to X$ such that $x(\theta, \tilde{\theta}) = f(\theta)$ for any pair of states $(\theta, \tilde{\theta}) \in \Theta^2$. For every agent $i \in \mathcal{I}$, we say that agent i challenges his own report if $x\left(m_i^3, m_i^1\right) \neq f\left(m_i^3\right)$, and agent i does not challenge his own report if $x\left(m_i^3, m_i^1\right) = f\left(m_i^3\right)$.

For each message profile $m \in M$, the allocation is defined as follows:

$$g(m) = \frac{1}{I} \sum_{i \in \mathcal{I}} \left[e(m)\, y_i\left(m_i^1\right) \oplus (1 - e(m))\, x\left(m_i^3, m_i^1\right) \right]$$

where $y_k : \Theta_k \cup \Theta \to X$ is the dictator lottery for agent k defined in Lemma 2 and $\alpha x \oplus (1 - \alpha)x'$ denotes the outcome which corresponds to the compound lottery that with probability α, outcome x occurs, and with probability $1 - \alpha$, outcome x' occurs[7]; moreover, we define

$$e(m) = \begin{cases} 0, & \text{if } m \text{ is consistent;} \\ \varepsilon, & \text{otherwise.} \end{cases}$$

That is, the designer first chooses an agent, with equal probability, to be checked. In checking agent i, the designer will use agent i's first report to check i's third report in determining the allocation.

After the designer picked agent i to be checked, the outcome function distinguishes two cases: (1) if $e(m) = 0$, then we implement $f\left(m_i^3\right)$; (2) if $e(m) = \varepsilon$, we implement the compound lottery:

$$\varepsilon \times y_i\left(m_i^1\right) \oplus (1 - \varepsilon) \times x\left(m_i^3, m_i^1\right).$$

That is, with probability ε, we implement the lottery $y_i\left(m_i^1\right)$ and with probability $1 - \varepsilon$, we implement the lottery $x(m_i^3, m_i^1)$. We elaborate on how we choose ε together with other parameters in Section 3.1.4.

[7] More precisely, if $x = (\ell, (t_i)_{i \in \mathcal{I}})$ and $x' = (\ell', (t_i')_{i \in \mathcal{I}})$ are two outcomes in X, we identify $\alpha x \oplus (1 - \alpha)x'$ with the outcome $(\alpha\ell + (1 - \alpha)\ell', (\alpha t_i + (1 - \alpha)t_i')_{i \in \mathcal{I}})$.

3.1.3 Transfer Rule:

In order to define the transfer rule, we introduce a few pieces of notation. For each message profile m, let $\mathcal{I}^0(m^1) \equiv \{j \in \mathcal{I} : m_j^1 \in \Theta\}$ be the set of agents who report a state in their first announcement. Fixing an arbitrary $\theta''' \in \Theta$, we define for each message profile m,

$$\hat{\theta}(m^1) = \begin{cases} \theta', & \text{if } \mathcal{I}^0(m^1) = \varnothing \text{ and } m^1 = (\theta'_i)_{i \in \mathcal{I}} \text{ for some } \theta' \in \Theta; \\ \theta'', & \text{if } \mathcal{I}^0(m^1) \neq \varnothing \text{ and } m_j^1 = \theta'' \text{ for all } j \in \mathcal{I}^0(m^1); \\ \theta''', & \text{otherwise.} \end{cases}$$

We may interpret $\hat{\theta}(m^1)$ as a state "identified" by the first announcement profile $(m_i^1)_{i \in \mathcal{I}}$. In the first two cases, such an identification is clear: either every one report a type in their first announcement and the joint type profile can be induced from a single state θ' or some agent(s) announce a state in their first announcement and these agents reach an unanimous agreement in announcing a common state θ'' (in which case we ignore the agents who announce a type in their first announcement, in this identification). When there is no such a clear identification, we simply set $\hat{\theta}(m^1)$ equal to an arbitrarily pre-specified state θ'''.

Equipped with the definition of $\hat{\theta}(m^1)$, for any message profile m and any agent i, we specify the transfer to agent i as follows:

$$\tau_i(m) = \tau_i^1(m) + \tau_i^2(m) + \tau_i^3(m), \tag{7}$$

where

$$\tau_i^1(m) = \begin{cases} 0, & \text{if } m_i^2 \sim \hat{\theta}(m^1); \\ -\eta'', & \text{otherwise.} \end{cases}$$

$$\tau_i^2(m) = \begin{cases} 0 & \text{if } m_i^3 \sim m_{i+1}^2; \\ -\eta & \text{otherwise.} \end{cases}$$

$$\tau_i^3(m) = \begin{cases} 0 & \text{if } m_i^1 \in \Theta_i \text{ or } [m_i^1 \in \Theta,\ m_i^1 \sim\ m_j^2 \sim m_j^3, \text{ and } x(m_i^1, m_j^1) = f(m_i^1), \forall j \in \mathcal{I}]; \\ -\eta & \text{otherwise.} \end{cases}$$

In words, $\tau_i^1(m)$ and $\tau_i^2(m)$ each agent i will only want to announce states which are also equivalent to $\hat{\theta}(m^1)$ in reporting m_i^2 and m_i^3. Specifically, $\tau_i^1(m)$ requires that agent i pay η'' if his announcement m_i^2 is *not* equivalent to $\hat{\theta}(m^1)$; likewise, $\tau_i^2(m)$ requires that agent i pay η if his announcement m_i^3 is *not* equivalent to agent $(i+1)$'s announcement m_{i+1}^2 where $I+1 \equiv 1$. In addition, $\tau_i^3(m)$ requires that agent i pay η if he announces a state in m_i^1 which is not equivalent to his own or some other agents' second and third announcements of state;

or agent i's announced state m_i^1 is challenged by some other agent (i.e., $x(m_i^1, m_j^1) \neq f(m_i^1)$ for some $j \neq i$).

3.1.4 Choice of parameters

Since Θ is finite, we can find $d > 0$ such that for any $i \in \mathcal{I}$ and any pair of types $\theta_i, \theta_i' \in \Theta_i$ with $\theta_i \neq \theta_i'$, the dictator lotteries satisfy

$$u_i\left(y_i\left(\theta_i\right), \theta_i\right) > u_i\left(y_i\left(\theta_i'\right), \theta_i\right) + d. \tag{8}$$

By (3), we can choose $\varepsilon > 0$ and $\eta'' > 0$ sufficiently small and $\eta > 0$ sufficient large such that the following three conditions hold:

- The penalty scale η dominates any incentive from a change in allocations in outcomes $g(\cdot)$ together with the penalty η'' resulted from τ_i^1, i.e.,

$$\eta > \eta'' + \sup_{i \in \mathcal{I}, \theta_i \in \Theta_i, m, m' \in M} \left| u_i(g\left(m\right), \theta_i) - u_i(g\left(m'\right), \theta_i) \right|. \tag{9}$$

- The penalty scale η'' and ε do disturb the "effectiveness" of agent i's challenge. More precisely, whenever agent i is checked, if he has reported a false state in m_i^3 for which he is a whistle-blower at the true state, it is still strictly better for him to tell the truth in m_i^1 to challenge m_i^3. That is,

$$\begin{aligned} &x\left(m_i^3, m_i^1\right) \neq f(m_i^3) \Rightarrow \text{for any } \tilde{\theta} \in \mathcal{P}(m_i^3), \\ &\tfrac{1}{I}\left(1-\varepsilon\right)\left[u_i(x(\tilde{\theta}, m_i^1), m_i^1) - u_i(f(\tilde{\theta}), m_i^1)\right] - \varepsilon\eta > \eta''. \end{aligned} \tag{10}$$

- The penalty scale η'' does not disturb the truth-telling incentive from the dictator lotteries. That is,

$$\frac{\varepsilon}{I} d > \eta''. \tag{11}$$

3.2 Proof of Theorem 1

We start by outlining the proof of Theorem 1. In this proof, we show in Step 1 that if the SCF f satisfies Maskin monotonicity* and a message $m_i = (m_i^1, m_i^2, m_i^3)$ is rationalizable, then either m_i^1 is the true type of agent i or m_i^1 is a state which is equivalent to the true state. To wit, if an agent reports a type, then reporting his true type is strictly better than reporting another type due to (4) of Lemma 2. Moreover, we show that if an agent reports a state, then it must be equivalent to the true state. This results from the combined force of the transfer rule $\tau_i^3(\cdot)$ and Maskin monotonicity*, which is at the heart of our proof of Theorem 1. As a consequence of these two claims, $\hat{\theta}(m^1)$ must be equivalent to the true state.

In Step 2, the cross-checking penalties $\tau_i^1(m)$ and $\tau_i^2(m)$ ensure that m_i^2 and hence also m_i^3 are both equivalent to the true state. Finally, in Step 3 we conclude that if a message profile m_i is rationalizable, then m_i^1 must also be a state rather than a type and the state must be equivalent to the true state. This follows from the fact that a type announcement in m_i^1 triggers a worse outcome by (5) of Lemma 2 than a state announcement in m_i^1. In summary, if m is rationalizable, then m_i^1, m_i^2, and m_i^3 must be all equivalent to the true state. Hence, the social outcome designated to the true state is implemented and no transfers are incurred.

Recall that the agents commonly know the true state of the world which is unknown to the designer. Denote the true state by $\theta \in \Theta$. We now prove the "if" part of Theorem 1 in three steps.

Step 1: *For every* $m \in S^\infty(\mathcal{M}, \theta)$, if $m_i^1 \in \Theta_i$, then $m_i^1 = \theta_i$; if $m_i^1 \in \Theta$, then $m_i^1 \sim \theta$.

Fix agent $i \in \mathcal{I}$ and message $m_i \in S_i^\infty(\mathcal{M}, \theta)$. Then, there is $\lambda_i \in \Delta\left(S_{-i}^\infty(\mathcal{M}, \theta)\right)$ against which m_i is a best reply. We prove Step 1 in each of the following two substeps.

Step 1A. If $m_i^1 \in \Theta_i$, then $m_i^1 = \theta_i$.

Fix m_{-i} with $\lambda_i(m_{-i}) > 0$. We show that for every $m_i \in S_i^\infty(\mathcal{M}, \theta)$ with $m_i^1 \in \Theta_i$, we have $m_i^1 = \theta_i$. Suppose not. We construct $\tilde{m}_i = (\theta_i, m_i^2, m_i^3)$, which is the same as m_i except that $\tilde{m}_i^1 = \theta_i \neq m_i^1$. Note that for any $m_{-i} \in S_{-i}^\infty(\mathcal{M}, \theta)$ we have $e(m_i, m_{-i}) = e(\tilde{m}_i, m_{-i}) = \varepsilon$ since both m_i^1 and $\tilde{m}_i^1$ are in Θ_i so that (m_i, m_{-i}) and $(\tilde{m}_i, m_{-i})$ are not consistent. Thus, in terms of allocation, the gain from choosing $\tilde{m}_i$ rather than m_i is at least $\frac{\varepsilon}{I}d$ when agent i is chosen to be checked, and the potential loss in terms of transfers is bounded by η'' due to $\tau_i^1(\cdot)$. It follows from (11) that $\tilde{m}_i$ is a strictly better reply than m_i. This is a contradiction.

Step 1B. If $m_i^1 \in \Theta$, then $m_i^1 \sim \theta$.

Fix $m_i \in S_i^\infty(\mathcal{M}, \theta)$ with $m_i^1 \in \Theta$. Say $m_i^1 = \tilde{\theta}$. We first show that there exists some m_{-i} with $\lambda_i(m_{-i}) > 0$ such that $m_i^1 \sim m_j^2 \sim m_j^3$ and $x(m_i^1, m_j^1) = f(m_i^1)$ for every agent $j \in \mathcal{I}$. Suppose not. Then, by $\tau_i^3(\cdot)$, agent i is penalized by η. Consider $\tilde{m}_i = (\theta_i, m_i^2, m_i^3)$, which is identical to m_i except that $\tilde{m}_i^1 = \theta_i \neq \tilde{\theta} = m_i^1$. The potential loss in terms of transfers is bounded by η'' due to $\tau_i^1(\cdot)$. The potential loss in terms of allocation from choosing $\tilde{m}_i$ rather than m_i is bounded by $\sup_{i \in \mathcal{I}, \theta_i \in \Theta_i, m, m' \in M} |u_i(g(m), \theta_i) - u_i(g(m'), \theta_i)|$, while the gain due to $\tau_i^3(\cdot)$ from choosing $\tilde{m}_i$ rather than m_i is at least η. By (9), $\tilde{m}_i = (\theta_i, m_i^2, m_i^3)$ is a strictly better reply against λ_i than m_i. This contradicts to the hypothesis that $m_i \in S_i^{\infty}(\mathcal{M}, \theta)$.

Note that we have kept fixing $m_i \in S_i^{\infty}(\mathcal{M}, \theta)$ with $m_i^1 = \tilde{\theta}$. In addition to this, we consider $m_{-i} \in S_{-i}^{\infty}(\mathcal{M}, \theta)$ such that, $\forall k \in \mathcal{I}$,

$$\tilde{\theta} \sim m_k^2 \sim m_k^3,$$

and

$$x(\tilde{\theta}, \theta_k) = f(\tilde{\theta}).$$

Suppose on the contrary that $\tilde{\theta} \not\sim \theta$. Then, since the SCF f satisfies Maskin monotonicity*, there exists some agent $j \in \mathcal{I}$ for whom $x(\tilde{\theta}, \theta_j) \neq f(\tilde{\theta})$ and

$$u_j(x(\tilde{\theta}, \theta_j), \theta_j) > u_j(f(\tilde{\theta}), \theta_j). \tag{12}$$

Now we construct $\tilde{m}_j = \left(\theta_j, m_j^2, m_j^3\right)$, which is the same as m_j except that $\tilde{m}_j^1 = \theta_j$. In the following, we shall show that $\tilde{m}_j$ strictly dominates m_j, which contradicts the hypothesis that $m_j \in S_j^{\infty}(\mathcal{M}, \theta)$.

Fix $\tilde{m}_{-j} \in S_{-j}^{\infty}(\mathcal{M}, \theta)$ arbitrarily. Observe first that $e(\tilde{m}_j, \tilde{m}_{-j}) = \varepsilon$ since $\tilde{m}_j^1 \in \Theta_j$ so that $(\tilde{m}_j, \tilde{m}_{-j})$ is not consistent. Thus,

$$g(\tilde{m}_j, \tilde{m}_{-j}) = \frac{1}{I} \sum_{k \neq j} \left[\varepsilon y_k\left(\tilde{m}_k^1\right) \oplus (1-\varepsilon)\, x\left(\tilde{m}_k^3, \tilde{m}_k^1\right)\right] \oplus \frac{1}{I}\left[\varepsilon y_j(\tilde{m}_j^1) \oplus (1-\varepsilon)\, x(m_j^3, \tilde{m}_j^1)\right],$$

where $x(m_j^3, \tilde{m}_j^1) = x(\tilde{\theta}, \theta_j) \neq f(m_j^3)$. In contrast,

$$g(m_j, \tilde{m}_{-j}) = \frac{1}{I} \sum_{k \in \mathcal{I}} \left[e(m_j, \tilde{m}_{-j})\, y_k\left(\tilde{m}_k^1\right) \oplus (1 - e(m_j, \tilde{m}_{-j}))\, x\left(\tilde{m}_k^3, \tilde{m}_k^1\right)\right],$$

where $x(m_j^3, m_j^1) = f(m_j^3)$. We now show that $\tilde{m}_j$ strictly dominates m_j by considering the following two subcases of j's opponents' announcement $\tilde{m}_{-j}$:

Case 2.1. $e(m_j, \tilde{m}_{-j}) = \varepsilon$.

In this case, we have

$$g(m_j, \tilde{m}_{-j}) = \frac{1}{I} \sum_{k \neq j} \left[\varepsilon y_k\left(\tilde{m}_k^1\right) \oplus (1-\varepsilon)\, x\left(\tilde{m}_k^3, \tilde{m}_k^1\right)\right] \oplus \frac{1}{I}\left[\varepsilon y_j\left(m_j^1\right) \oplus (1-\varepsilon)\, f(m_j^3)\right].$$

In this case, $g(\tilde{m}_j, \tilde{m}_{-j})$ differs from $g(m_j, \tilde{m}_{-j})$ only when agent j is chosen to be checked. In terms of dictator lotteries, the loss of agent j from choosing m_j to choosing $\tilde{m}_j$ is bounded by $\frac{1}{I}\varepsilon d$; in terms of allocations from the best challenge scheme, the gain of agent j from choosing m_j to choosing $\tilde{m}_j$ is $(1-\varepsilon)\frac{1}{I}\left(u_j(x(\tilde{\theta}, \theta_j), \theta_j) - u_j(f(\tilde{\theta}), \theta_j)\right)$; and in terms of transfers, the loss of agent j from choosing m_j to choosing $\tilde{m}_j$ is bounded by η'' (due to τ_j^1). By (10), we conclude that agent j obtains a strictly higher expected utility under $\tilde{m}_j$ than m_j against $\tilde{m}_{-j}$.

Case 2.2. $e(m_j, \tilde{m}_{-j}) = 0$.

In this case, $(m_j, \tilde{m}_{-j})$ is consistent, and $x(\tilde{m}_k^2, \tilde{m}_k^1) = x(m_j^2, m_j^1) = f(\tilde{m}_k^2) = f(\tilde{\theta})$ for every $k \neq j$. Hence,

$$g(m_j, \tilde{m}_{-j}) = f(\tilde{\theta}).$$

In contrast,

$$\begin{aligned} g(\tilde{m}_j, \tilde{m}_{-j}) &= \frac{1}{I}\sum_{k\neq j}\left[\varepsilon y_k\left(\tilde{m}_k^1\right) \oplus (1-\varepsilon) f\left(\tilde{m}_k^2\right)\right] \oplus \frac{1}{I}\left[\varepsilon y_j(\tilde{m}_j^1) \oplus (1-\varepsilon) x(\tilde{\theta}, \theta_j)\right] \\ &= \varepsilon\left[\frac{1}{I}\sum_{k\in\mathcal{I}} y_k\left(\tilde{m}_k^1\right)\right] \oplus (1-\varepsilon)\left[\frac{1}{I}\sum_{k\neq j} f\left(\tilde{m}_k^2\right) \oplus \frac{1}{I}x(\tilde{\theta}, \theta_j)\right] \\ &= \varepsilon\left[\frac{1}{I}\sum_{k\in\mathcal{I}} y_k\left(\tilde{m}_k^1\right)\right] \oplus (1-\varepsilon)\left[\frac{I-1}{I}f(\tilde{\theta}) \oplus \frac{1}{I}x(\tilde{\theta}, \theta_j)\right]. \end{aligned}$$

Hence, it follows from (5) of Lemma 2 (or more precisely, the argument for how to construct $y_i(\theta_i)$ right after Lemma 2) that the net gain from choosing $\tilde{m}_j$ rather than m_j is at least

$$-\varepsilon\eta + (1-\varepsilon)\frac{1}{I}\left(u_j(x(\tilde{\theta}, \theta_j), \theta_j) - u_j(f(\tilde{\theta}), \theta_j)\right) > 0$$

where the strict inequality follows from (10). In addition, the loss of agent j from choosing m_j to choosing $\tilde{m}_j$ is bounded by η'' (due to τ_j^1). By (10), we know that agent j obtains a strictly higher expected utility under $\tilde{m}_j$ than m_j against $\tilde{m}_{-j}$. This is a contradiction.

Therefore, the consideration of the above two cases concludes that $\tilde{m}_j$ strictly dominates m_j. This implies that $m_i^1 \sim \theta$.

Step 2: *For every agent $i \in \mathcal{I}$ and every $m_i \in S_i^\infty(\mathcal{M}, \theta)$, we have $m_i^2 \sim m_i^3 \sim \theta$.*

We first show that $m_i^2 \sim \theta$. It follows from Step 1, for each $m_i \in S_i^\infty(\mathcal{M}, \theta)$, $m_i^1 = \theta_i$ or $m_i^1 \sim \theta$. If $m_i^1 = \theta_i$ for each agent $i \in \mathcal{I}$, then in m^1, we have a type profile induced by the true state. Hence, for each $m \in S_i^\infty(\mathcal{M}, \theta)$, we must have $\hat{\theta}(m^1) \sim \theta$. Suppose by way of contradiction that $m_i^2 = \theta' \not\sim \theta$. Note that by Step 1 for any $m_{-i} \in S_{-i}^\infty(\mathcal{M}, \theta)$, (m_i, m_{-i}) is inconsistent. Now, we construct $\tilde{m}_i = (m_i^1, \theta, m_i^3)$ which is identical to m_i except that $\tilde{m}_i^2 = \theta$. Thus, $(\tilde{m}_i, m_{-i})$ implements some allocation at least better than (m_i, m_{-i}) since m_i and $\tilde{m}_i$ only differ in their second report. In terms of transfers incurred, the gain from $(\tilde{m}_i, m_{-i})$ is η'' from $\tau_i^1(\cdot)$, while the loss from (m_i, m_{-i}) is η'' from $\tau_i^1(\cdot)$. Hence, $\tilde{m}_i$ is a better reply than m_i against any $m_{-i} \in S_{-i}^\infty(\mathcal{M}, \theta)$. This is the desired contradiction.

Then, we show that $m_i^3 \sim \theta$. By the previous argument, we know that in the second report, $m_{i+1}^2 \sim \theta$. Suppose by way of contradiction that $m_i^3 = \theta' \not\sim \theta$. Now, we construct

$\tilde{m}_i = (m_i^1, m_i^2, \theta)$ which is identical to m_i except that $\tilde{m}_i^3 = \theta$. Note that $(\tilde{m}_i, m_{-i})$ and (m_i, m_{-i}) may implement different allocations, and in terms of transfer, the gain is η from $(\tilde{m}_i, m_{-i})$ rather than (m_i, m_{-i}). Hence, we conclude that $\tilde{m}_i$ is a strictly better response than m_i against m_{-i}. This completes the proof of Step 2.

Step 3: *For any agent $i \in \mathcal{I}$ and any $m \in S^{\infty}(\mathcal{M}, \theta)$, we have $g(m) = f(\theta)$ and $\tau_i(m) = 0$.*

By Steps 1 and 2, for any $m \in S^{\infty}(\mathcal{M}, \theta)$, we have that $\hat{\theta}(m^1) \sim \theta$ and $m_i^2 \sim m_i^3 \sim \theta$ for every agent $i \in \mathcal{I}$. We next show that for *any* $m \in S^{\infty}(\mathcal{M}, \theta)$, $m_i^1 \sim \theta$ for every agent i. Suppose not. By Step 1, we know m_i is such that $m_i^1 = \theta_i$. Given an arbitrary $m_{-i} \in S_{-i}^{\infty}(\mathcal{M}, \theta)$, we first know that m is inconsistent. Thus,

$$g(m_i, m_{-i}) = \frac{1}{I}\varepsilon \sum_{k \neq i} y_k\left(m_k^1\right) \oplus \frac{1}{I}\varepsilon y_i(\theta_i) \oplus (1-\varepsilon) f(\theta). \tag{13}$$

Consider $\tilde{m}_i = (\theta, m_i^2, m_i^3)$ which is identical to m_i except that $m_i^1 = \theta$. When $e(\tilde{m}_i, m_{-i}) = \varepsilon$,

$$g(\tilde{m}_i, m_{-i}) = \frac{1}{I}\varepsilon \sum_{k \neq i} y_k\left(m_k^1\right) \oplus \frac{1}{I}\varepsilon f(\theta) \oplus (1-\varepsilon) f(\theta); \tag{14}$$

when $e(\tilde{m}_i, m_{-i}) = 0$,

$$g(\tilde{m}_i, m_{-i}) = f(\theta).$$

By choosing $\tilde{m}_i$ rather than m_i, there is positive gain from (14) rather than (13) by (5); while there is no loss since $m_j^2 \sim m_j^3 \sim \theta$ for every agent $j \in \mathcal{I}$. Hence, we have a contradiction.

We thus conclude that for every $m \in S^{\infty}(\mathcal{M}, \theta)$, we have $e(m) = 0$ so that no transfer is invoked and $f(\theta)$ is implemented. This completes the proof of Step 3.

3.3 Comparison with Chen, Kunimoto, Sun, and Xiong (2020)

The current paper was developed from our earlier paper which addresses the problem of mixed-strategy Nash implementation in finite mechanisms. Here we comment on the main differences in the implementing mechanisms and arguments adopted in these two papers.

The mechanism in Chen, Kunimoto, Sun, and Xiong (2020) (henceforth, M-AM) rests on the elicitation and cross-checking of type profiles to identify the allocation to be implemented. Specifically, the implementing mechanism of M-AM asks each agent to report a type (the first report) and a *type profile* (the second report). The mechanism is structured to achieve two main steps which we coined consistency (a common type profile being announced in their second report and the profile identifies a state) and no challenge (no agent has an incentive to "overturn" the allocation induced by the consistent second report with a test allocation). The Nash implementation result follows from these two steps and Maskin-monotonicity.

For rationalizable implementation, however, we need to work with the Maskin monotonicity* condition proposed by BMT instead of Maskin monotonicity. The Maskin monotonicity* condition is associated with a partition over the *states* rather than the type profiles. In order for a type profile to identify the atom which contains the true state, the partition must satisfy what we could call a *product structure*, by which we mean that each atom can be identified with a product set of type profiles. Unfortunately, this additional requirement entails a loss of generality, as it rules out some SCFs which satisfy Maskin-monotonicity*.[8]

The current mechanism resolves this difficulty by giving the agents a choice of reporting a type *or* a state in their first report, as opposed to only a type as in M-AM. Moreover, through structuring the incentive of inducing "truth-telling" in the first report, we make sure that any rationalizable message profile m identifies (via $\hat{\theta}(m^1)$) the atom which contains the true state. This is accomplished by introducing three new ideas: (i) we revise the notion of consistency which now refers only to states but not type profiles; moreover, the notion involves all the three reports m_i^1, m_i^2, and m_i^3 of each agent i; (ii) we include two state reports so that the first state report (m_i^2) does not affect the allocation and η'' can be made small; and (iii) we introduce the new transfer rule $\tau_i^3(\cdot)$. We elaborate this crucial idea in Step 1 of the proof of Theorem 1 which finds no counterpart in M-AM.

This new transfer rule $\tau_i^3(\cdot)$ plays a crucial role for our result. Recall that in the proof of Theorem 1, we first make sure that in the first report, the agents either report a true type or a state equivalent the true state. The former case is ensured by (4) of Lemma 2 as well as point (ii) above and the latter case is ensured by $\tau_i^3(\cdot)$. Then, via $\hat{\theta}(m^1)$, the cross-checking between the second report and the first report guarantees the state reported in the second report belongs to the true atom. As the notion of consistency differs from that of M-AM (point (i) above), we also need to rule out the possibility that some agent reports a type in his first report. The construction of dictator lotteries incentivizes the agent to choose a state report rather than any type report. In contrast, the first report in M-AM is a type and need not be the true type even in equilibrium.

[8] To see this, suppose that there are two agents and each of them has two types, i.e., $\mathcal{I} = \{1, 2\}$, and $\Theta_i = \{\theta_i, \theta_i'\}$ for each $i \in \mathcal{I}$. Let f be an SCF f such that $f(\theta_1, \theta_2) \neq f(\theta_1, \theta_2') = f(\theta_1', \theta_2) = f(\theta_1', \theta_2')$ and the associated partition $\mathcal{P}_f$ is given as $\mathcal{P}_f = \left\{\{(\theta_1, \theta_2)\}, \{(\theta_1', \theta_2), (\theta_1, \theta_2'), (\theta_1', \theta_2')\}\right\}$. This clearly violates product structure. Assume further that $v_i(f(\theta_1, \theta_2), \theta_i) > v_i(f(\theta_1', \theta_2'), \theta_i)$ and $v_i(f(\theta_1, \theta_2), \theta_i') < v_i(f(\theta_1', \theta_2'), \theta_i')$ for each agent $i \in \mathcal{I}$. Then, f satisfies Maskin-monotonicity* associated with $\mathcal{P}_f$.

4 Small Transfers

One issue with the mechanism which we propose for Theorem 1 is that the size of transfers may be large. However, since we allow lottery allocations, we can use the idea of Abreu and Matsushima (1994) to show that if the SCF satisfies *Maskin monotonicity* without transfer* (see Definition 5 below), then it is implementable in rationalizable strategies with arbitrarily small transfers. We first propose a notion of rationalizable implementation with bounded transfers.

Definition 3 *An SCF f is implementable in rationalizable strategies* ***with transfers bounded by*** $\bar{\tau}$ *if there exists a mechanism $\mathcal{M} = ((M_i, \tau_i)_{i\in\mathcal{I}}, g)$ such that for every state $\theta \in \Theta$, (i) $S^{\infty}(\mathcal{M}, \theta) \neq \varnothing$; (ii) for any $m \in S^{\infty}(\mathcal{M}, \theta)$, we have $g(m) = f(\theta)$ and $\tau_i(m) = 0$; and (iii) $|\tau_i(m)| \leq \bar{\tau}$ for every $m \in M$ and every agent $i \in \mathcal{I}$.*

In other words, Definition 3 strengthens Definition 1 in requiring that the transfer be bounded by $\bar{\tau}$ even for message profiles which are not rationalizable. Next, we propose a notion of Rationalizable implementability in which there are no transfers on any rationalizable strategy profile and only arbitrarily small transfers on every strategy profile.

Definition 4 *An SCF f is implementable in rationalizable strategies* ***with arbitrarily small transfers*** *if, for every $\bar{\tau} > 0$, the SCF f is implementable in rationalizable strategies with transfers bounded by $\bar{\tau}$.*

We say that an SCF f satisfies *Maskin monotonicity* without transfer* if $\tilde{\theta} \notin \mathcal{P}(\theta)$ implies that there are an agent i and a lottery $x(\tilde{\theta}, \theta_i)$ in $\Delta(A)$ such that $x(\tilde{\theta}, \theta_i)$ belongs to $\mathcal{L}_i(f(\tilde{\theta}), \mathcal{P}(\theta)) \cap \mathcal{SU}_i(f(\tilde{\theta}), \theta_i)$. Here, for $(\ell, \theta_i) \in \Delta(A) \times \Theta_i$, we use $\mathcal{L}_i(\ell, \theta_i)$ to denote the lower-contour set in $\Delta(A)$ at allocation ℓ for type θ_i, i.e.,

$$\mathcal{L}_i(\ell, \theta_i) = \{\ell' \in \Delta(A) : v_i(\ell, \theta_i) \geq v_i(\ell', \theta_i)\}.$$

Then, for each $\ell \in \Delta(A)$, let

$$\mathcal{L}_i(\ell, \mathcal{P}(\theta)) \equiv \bigcap_{\tilde{\theta} \in \mathcal{P}(\theta)} \mathcal{L}_i(\ell, \tilde{\theta}_i).$$

Similarly, we also define $\mathcal{SU}_i(\ell, \theta_i)$ without invoking transfers.

Definition 5 *Say an SCF f satisfies Maskin monotonicity** ***without transfer*** *if there exists a partition $\mathcal{P}$ of Θ such that (i) $\mathcal{P}$ is at least as fine as $\mathcal{P}_f$; (ii) for any $\tilde{\theta}, \theta \in \Theta$, whenever $\tilde{\theta} \notin \mathcal{P}(\theta)$, there exists $i \in \mathcal{I}$ for whom*

$$\mathcal{L}_i(f(\tilde{\theta}), \mathcal{P}(\tilde{\theta})) \cap \mathcal{SU}_i(f(\tilde{\theta}), \theta_i) \neq \varnothing.$$

Clearly, Maskin monotonicity* without transfer implies the Maskin monotonicity* condition defined in Definition 2, since the former requires that the test allocation be a lottery over alternative without transfers. To prove our implementation result with arbitrarily small transfers, we assume that there exists a "uniformly worse outcome" $w \in A$ such that $u_i(f(\tilde{\theta}), \theta_i) > u_i(w, \theta_i)$ for every $\tilde{\theta} \in \Theta$, and every agent i of every type θ_i. Note that this assumption is stronger than the no-worst-alternative (NWA) condition used by BMT. We now state the result formally:

Theorem 2 *Suppose that there exists an outcome $w \in A$ such that for every agent $i \in \mathcal{I}$, $u_i(f(\tilde{\theta}), \theta_i) > u_i(w, \theta_i)$ for every $\tilde{\theta} \in \Theta$ and every type $\theta_i \in \Theta_i$. Then, an SCF $f_A : \Theta \to \Delta(A)$ is implementable in rationalizable strategies with arbitrarily small transfers if f_A satisfies Maskin monotonicity* without transfer.*

We present the formal proof of Theorem 2 in Appendix A.2. In the proof, for each $\bar{\tau} > 0$, we construct an augmented mechanism which assigns a small weight/probability α to the baseline mechanism (so that its transfers rescaled by α do not exceed $\bar{\tau}$) and weight $1-\alpha$ assigned to part of the mechanism in Abreu and Matsushima (1994) (so that the outcome manipulation incentive can be disciplined with small transfers bounded by $\bar{\tau}$). Moreover, any rationalizable message in the augmented mechanism must involve a rationalizable message in the baseline component.[9]

Abreu and Matsushima (1992) achieves virtual implementation in rationalizable strategies with a domain restriction instead of transfers. However, to achieve exact implementation with the Maskin monotonicity* condition, we can only have the agents report (in their rationalizable messages) states which are equivalent to the true state. Since the agents' type/preference can still vary within these equivalent states, we cannot use the technique of Abreu and Matsushima (1992) with the domain restriction. More precisely, in Abreu

[9] We formalize the step as Claim 4 in Appendix A.2. Roughly speaking, the baseline mechanism plays the same role as the dictator lotteries in Abreu and Matsushima (1994), except that the baseline mechanism only elicits the atom in the partition which contains the true state instead of the true state. With Claim 4, Theorem 2 follows from a construction and argument similar to those of Abreu and Matsushima (1994).

and Matsushima (1992), with a true state in first report the designer can use the reported state (which is the truth under rationalizability by their construction) to pick a good/bad outcome to incentivize a particular agent. This is also the case when the partition in the Maskin monotonicity* condition is the finest, and there is no other state equivalent to the true state. In this case, we can invoke the domain restriction in Abreu and Matsushima (1992) without transfers to achieve our implementation result. In a similar vein, we can also dispense with the outcome w in this case.

5 Discussion

We conclude this paper by providing a few discussions that allow us to locate our contribution in the literature. First, we relate our result to continuous implementation of Oury and Tercieux (2012). Second, we discuss how responsiveness makes our result simpler and tightly connected to virtual implementation of Abreu and Matsushima (1992). Finally, we argue that it is generally impossible to simplify our implementing mechanism into a direct mechanism where every agent only announces a state.

5.1 Continuous Implementation

Oury and Tercieux (2012) consider the following situation: the planner wants not only that there is an equilibrium that implements the SCF but also that the same equilibrium continues to achieve implementation of the SCF in all the models *close* to his initial model. Hence, the SCF is *continuously* implementable. Oury and Tercieux (2012) obtain the following characterization of continuous implementation in their Theorem 4: an SCF is continuously implementable by a finite mechanism if it is exactly implementable in rationalizable strategies by a finite mechanism.[10] Under "local payoff uncertainty", Oury (2015) obtains the same characterization. Since these result say nothing about the class of SCFs that are exactly implementable in rationalizable strategies by finite mechanisms, we view this as an important open question in the literature. We establish the following continuous implementation result which is a direct consequence of our Theorem 1 and Theorem 4 of Oury and Tercieux (2012).

[10]In fact, assuming that sending messages is slightly costly, Oury and Tercieux also prove the converse: an SCF is continuously implementable by a finite mechanism only if it is rationalizably implementable by a finite mechanism.

Corollary 1 *If an SCF satisfies Maskin monotonicity*, it is continuously implementable by a finite mechanism.*

To the best of our knowledge, our Proposition 1 is the first result which continuously implements all Maskin monotonic* SCFs by a finite mechanism. The identified condition, Maskin monotonicity*, is strictly stronger than Maskin monotonicity, as we will show in Appendix A.1. However, two caveats remain in relating Corollary 1 to Theorem 4 of Oury and Tercieux (2012). The first caveat is that we focus on complete information environments, whereas Oury and Tercieux deal with incomplete information environments where the baseline model can be an arbitrary finite type space. The second caveat is that we specialize in environments with lottery and transfer, whereas Oury and Tercieux impose no condition on the environments.

In incomplete information environments with lottery and transfer, Chen, Kunimoto, and Sun (2019) made some progress in this direction. They show that any incentive compatible SCF is continuously implementable by a finite mechanism, provided that (i) we allow for arbitrarily small ex post penalties both on the equilibrium and off the equilibrium; (ii) each agent knows his own payoff type; and (iii) agents' beliefs satisfy a generic correlation condition. In other words, under the three assumptions above, incentive compatibility is the only constraint for continuous implementation.

5.2 Responsive SCFs

Here we draw a connection between our result and the virtual implementation result proved by Abreu and Matsushima (1992). To do so, consider the following condition on SCFs introduced by BMT:

Definition 6 *An SCF f is* ***responsive*** *if, for any pair of states $\theta, \theta' \in \Theta$, $f(\theta) = f(\theta') \Rightarrow \theta = \theta'$.*

Responsiveness requires that the SCF "respond" to a change in the state with a change in the social choice outcome. Observe that a responsive SCF that satisfies Maskin monotonicity must satisfy Maskin monotonicity*. Indeed, since $\mathcal{P}_f$ is the finest partition on Θ, for any two states θ and θ', $\theta' \in \mathcal{P}(\theta)$ is equivalent to $\theta' = \theta$.

In the case of responsive SCFs, Maskin monotonicity*, which is a necessary condition for rationalizable implementation, reduces to Maskin monotonicity. We formalize this result whose proof is omitted.

Lemma 3 *If an SCF f is responsive and satisfies Maskin monotonicity, it also satisfies Maskin monotonicity*.*

Theorem 1 and Lemma 3 together imply the following corollary for the case of responsive SCFs.

Corollary 2 *Any responsive SCF f is implementable in rationalizable strategies by a finite mechanism if and only if it satisfies Maskin monotonicity.*

Note that the "only if" part of Corollary 2 follows from BMT, who prove that under the no-worst alternative condition (See Definition 4 of BMT, p. 1259)[11], if there are at least three agents, f is responsive, and satisfies strict Maskin monotonicity, then it is implementable in rationalizable strategies by an infinite mechanism. In contrast, Corollary 2 covers the case of two agents.

An SCF f is said to be virtually implementable if, for any $\varepsilon \in (0,1)$, the SCF f is exactly implementable with probability $1-\varepsilon$. Abreu and Matsushima (1992) show that when there are at least three agents, any SCF is *virtually* implementable in rationalizable strategies by a finite mechanism. In Appendix A.3, we show that given any SCF f, there exists a responsive and Maskin-monotonic SCF which is "close" to f. Hence, the following corollary follows from Theorem 1 and Lemma 3.

Corollary 3 *Any SCF f is virtually implementable in rationalizable strategies by a finite mechanism.*

Recall that our mechanism is different from that of Abreu and Matsushima (1992), who do not use penalties but rather introduce a domain restriction in the lottery space. The domain restriction in Abreu and Matsushima (1992) requires that for every agent i and state θ, there exist a pair of lotteries which are strictly ranked for agent i and for which other agents have the (weakly) opposite ranking.

5.3 Direct Mechanisms

The message in our implementing mechanism is remarkably parsimonious. To recap, each agent is only asked to announce a type or state together with another two states. As in our

[11] No-worst alternative requires that any social outcome cannot be the worst outcome in any state. In our setup with transfers, no-worst alternative is automatically satisfied.

setup different types corresponds to different cardinal preferences over lottery allocations, we only ask the agents to announce payoff-relevant information, free of using integer/modulo games. With the feature in mind, we may still investigate to what extent we could simplify the mechanism further.

A prominent benchmark is to ask whether we could actually achieve rationalizable implementation for any Maskin monotonic* SCF via some direct mechanism. In our setup, a direct mechanism is a mechanism $((M_i), g, (\tau_i))_{i \in \mathcal{I}}$ in which (i) agents are asked to report the state (i.e., $M_i = \Theta$ for every agent i), and (ii) a unanimous report leads to the social outcome with no transfers (i.e., $g(\theta, ..., \theta) = f(\theta)$ and $\tau_i(\theta) = 0$ for every agent i and for each state θ). In Appendix A.4, we construct an SCF which satisfies Maskin monotonicity*; hence, by Theorem 1, it is implementable in rationalizable strategies. Moreover, we show that the SCF cannot be implemented in rationalizable strategies in a direct mechanism.

A Appendix

In this Appendix, we provide the proofs omitted from the main body of the paper.

A.1 Maskin Monotonicity and Maskin Monotonicity*

We first recap the definition of Maskin monotonicity.

Definition 7 *An SCF f satisfies **Maskin monotonicity** if, for any pair of states $\tilde{\theta}$ and θ with $f(\tilde{\theta}) \neq f(\theta)$, there is some agent $i \in \mathcal{I}$ such that*

$$\mathcal{L}_i(f(\tilde{\theta}), \tilde{\theta}_i) \cap \mathcal{SU}_i(f(\tilde{\theta}), \theta_i) \neq \varnothing. \tag{15}$$

The following example shows that Maskin monotonicity* is strictly stronger than Maskin monotonicity.

Example 1 *Let $A = \{a, b, c, d\}$, $\mathcal{I} = \{1, 2\}$, $X = \Delta(A) \times \mathbb{R}^2$, and $\Theta = \{\alpha, \beta, \gamma, \delta\}$. The agents' utility functions are given in the two tables below. Consider the following SCF $f(\alpha) = f(\beta) = f(\gamma) = (a, 0, 0) \in X$ and $f(\delta) = (b, 0, 0) \in X$. For simplicity of notation, we write $a \in A$ for $(a, 0, 0) \in X$ and $b \in A$ for $(b, 0, 0) \in X$, each of which is a degenerate* *allocation with no transfer to any agent.*

v_1	α	β	γ	δ
a	3	2	2	2
b	2	3	1	3
c	1	1	3	1
d	0	0	0	0

v_2	α	β	γ	δ
a	3	2	2	2
b	1	0	1	1
c	2	1	3	3
d	0	3	0	0

In the following three claims below, we show that the SCF is Maskin monotonic, but it does not satisfy Maskin monotonicity*.

Claim 1 *For every agent $i \in \mathcal{I}$ and $\theta \in \Theta$, $\mathcal{L}_i(a,\theta) \subset \mathcal{L}_i(a,\alpha)$.*

Proof. Observe that for any agent $i \in \mathcal{I}$, any $\tilde{a} \in A\backslash \{a\}$, and any $\theta \in \Theta$, the utility difference between a and $\tilde{a}$ is weakly larger at α than that at θ. That is,

$$v_i(a,\alpha) - v_i(\tilde{a},\alpha) \geq v_i(a,\theta) - v_i(\tilde{a},\theta).$$

Hence, for any $x \in X$, $i \in \mathcal{I}$, and $\theta \in \Theta$, we have $u_i(a,\theta) \geq u_i(x,\theta)$ whenever $u_i(a,\alpha) \geq u_i(x,\alpha)$. ■

Claim 2 *The SCF f violates Maskin monotonicity*.*

Proof. Consider an arbitrary partition $\mathcal{P}$ finer than $\mathcal{P}_f = \{\{\alpha,\beta,\gamma\},\{\delta\}\}$. Note that $\mathcal{P}(\delta) = \{\delta\}$ for any partition $\mathcal{P}$ finer than $\mathcal{P}_f$.

Case 1. $\alpha \in \mathcal{P}(\beta)$ and $\alpha \in \mathcal{P}(\gamma)$.

In this case, $\mathcal{P} = \mathcal{P}_f$ and hence $\mathcal{P}(\alpha) = \{\alpha,\beta,\gamma\}$. Since $\mathcal{L}_1(a,\beta) = \mathcal{L}_1(a,\delta)$ and $\mathcal{L}_2(a,\gamma) = \mathcal{L}_2(a,\delta)$. Thus, $\mathcal{L}_i(a,\mathcal{P}(\alpha)) \subset \mathcal{L}_i(a,\delta)$ for all $i \in \{1,2\}$ but $f(\alpha) \neq f(\delta)$. Hence, f violates Maskin monotonicity* for such $\mathcal{P}$.

Case 2. $\alpha \notin \mathcal{P}(\beta)$ or $\alpha \notin \mathcal{P}(\gamma)$.

We derive a contradiction for $\alpha \notin \mathcal{P}(\beta)$ and the argument for the case with $\alpha \notin \mathcal{P}(\gamma)$ is similar and so omitted. If $\alpha \notin \mathcal{P}(\beta)$, then by Claim 1, we have $\mathcal{L}_i(a,\mathcal{P}(\beta)) \subset \mathcal{L}_i(a,\alpha)$ for all $i \in \{1,2\}$. Then, f violates Maskin monotonicity* for $\mathcal{P}$ since $\mathcal{L}_i(a,\mathcal{P}(\beta)) \subset \mathcal{L}_i(a,\alpha)$ for all $i \in \{1,2\}$ and $\alpha \notin \mathcal{P}(\beta)$. ■

Claim 3 *The SCF f satisfies Maskin monotonicity.*

Proof. This can be confirmed by observing that $b \in \mathcal{L}_1(a,\alpha) \cap \mathcal{SU}_1(a,\delta)$, $c \in \mathcal{L}_2(a,\beta) \cap \mathcal{SU}_2(a,\delta)$, $b \in \mathcal{L}_1(a,\gamma) \cap \mathcal{SU}_1(a,\delta)$, $a \in \mathcal{L}_1(b,\delta) \cap \mathcal{SU}_1(b,\alpha)$, $d \in \mathcal{L}_2(b,\delta) \cap \mathcal{SU}_2(b,\beta)$, and $a \in \mathcal{L}_A(b,\delta) \cap \mathcal{SU}_1(b,\gamma)$. ■

A.2 Small Transfers

The following lemma is the counterpart of Lemma 2. We prove this by making use of the outcome w as opposed to transfers:

Lemma 4 *Suppose that there exists an outcome $w \in A$ such that for every agent i, $u_i(f(\tilde{\theta}), \theta_i) > u_i(w, \theta_i)$ for every $\tilde{\theta} \in \Theta$ and every type $\theta_i \in \Theta_i$. Then, for each agent $i \in \mathcal{I}$, there exists a function $y_i : \Theta_i \cup \Theta \to X$ such that $y_i(\theta) = f(\theta)$ for each $\theta \in \Theta$, and for all types $\theta_i, \theta_i' \in \Theta_i$ with $\theta_i' \neq \theta_i$, we have*

$$u_i(y_i(\theta_i), \theta_i) > u_i(y_i(\theta_i'), \theta_i); \tag{16}$$

moreover, for each $j \in \mathcal{I}$ and θ_j', we also have for every $\tilde{\theta} \in \Theta$,

$$u_j(y_j(\theta_j'), \theta_j) < u_j(f(\tilde{\theta}), \theta_j). \tag{17}$$

Proof. From Abreu and Matsushima (1992) we can prove the existence of lotteries $\{y_i'(\cdot)\} \subset \Delta(A)$ which satisfy Condition (16). To satisfy Condition (17), we simply choose $\delta \in (0,1)$ large enough and define $y_j(\theta_j) = \delta w \oplus (1-\delta) y_j(\theta_j)$ such that $u_j(y_j(\theta_j), \theta_j) < u_j\left((f(\tilde{\theta}), \theta_j\right)$ for each $j \in \mathcal{I}$ and each $\theta_j' \in \Theta_j$, and for every $\tilde{\theta} \in \Theta$. ■

Equipped with Lemma 4, we can reproduce a mechanism $\mathcal{M} = ((M_i), g, (\tau_i))_{i \in \mathcal{I}}$ in Section 3, η'' and η is determined according to (8) and (9) and the transfers all arise from τ_i. We call such $\mathcal{M}$ the baseline mechanism. Based on the baseline mechanism $\mathcal{M}$, we now construct an augmented mechanism $\tilde{\mathcal{M}} = ((\tilde{M}_i), \tilde{g}, (\tilde{\tau}_i))_{i \in \mathcal{I}}$ to prove Theorem 2.

A.2.1 The Mechanism

We augment the message space of the implementing mechanism in Section 3 by adding to it $K+1$ additional reports of states, where K will be chosen later. Formally, each message in the augmented mechanism is written as follows.

Player i's message space is

$$\tilde{M}_i = M_i \times \tilde{M}_i^0 \times \cdots \times \tilde{M}_i^K = M_i \times \underbrace{\Theta \times \cdots \times \Theta}_{K+1 \text{ terms}},$$

where each player i simultaneously makes an announcement in M_i (of the baseline mechanism $\mathcal{M}$) and $K+1$ announcements of the state. We write a generic message $\tilde{m}_i \in \tilde{M}_i$ of agent i as

$$\tilde{m}_i = \left(m_i, \tilde{m}_i^0, ..., \tilde{m}_i^K\right).$$

A.2.2 Outcome

The outcome is specified as follows. Define $\rho : \Theta^I \to \Delta(A)$ such that for every $k \geq 1$, we define

$$\rho\left(\tilde{m}^k\right) = \begin{cases} f(\tilde{\theta}), & \text{if there exists } \tilde{\theta} \in \Theta \text{ such that } \tilde{m}_i^k = \tilde{\theta} \in \Theta \text{ for all } i \in \mathcal{I}; \\ w & \text{otherwise.} \end{cases}$$

Let D be the bound of the utility difference across all different outcomes of $\rho(\cdot)$. That is,

$$D = \sup_{i \in \mathcal{I}, \theta_i \in \Theta_i, \theta', \theta'' \in \Theta} \left| u_i(\rho\left(\theta'\right), \theta_i) - u_i(\rho\left(\theta''\right), \theta_i) \right|.$$

For any message profile $\tilde{m} = \left(m, \tilde{m}^0, ..., \tilde{m}^K\right)$, the outcome is defined as follows:

$$\tilde{g}(\tilde{m}) = \alpha \times g(m) \oplus (1-\alpha) \times \frac{1}{K} \sum_{k=1}^{K} \rho(\tilde{m}^k)$$

where $\alpha > 0$ will be chosen later, and $g(\cdot)$ is the outcome function of the baseline mechanism in Section 3. That is, the outcome is a lottery combining the outcome of the baseline mechanism, and the equally weighted sum of all outcome functions over all rounds 1 through K.

A.2.3 Transfers

For every agent i, we specify the transfer to agent i as follows:

$$\tilde{\tau}_i(\tilde{m}) = \alpha \times \tau_i(m) + (1-\alpha) \times \left(\tau_i^4\left(m_{-i}, \tilde{m}_i^0\right) + \tau_i^5\left(\tilde{m}^0, ..., \tilde{m}^K\right)\right)$$

where $\tau_i(\cdot)$ is the transfer rule of the baseline mechanism defined in (7); moreover, we include two additional transfers $\tau_i^4(\cdot)$ and $\tau_i^5(\cdot)$:

$$\tau_i^4\left(m_{-i}, \tilde{m}_i^0\right) = \begin{cases} -\gamma & \text{if } \tilde{m}_i^0 \sim m_{i+1}^3; \\ 0 & \text{otherwise.} \end{cases}$$

That is, agent i receives a fine of γ if their 4th report is not equivalent to agent $i+1$'s 3rd report in the baseline mechanism.

$$\tau_i^5\left(\tilde{m}^0, ..., \tilde{m}^K\right) = \begin{cases} -\xi, & \begin{array}{l}\text{if there exists } k \in \{1, \ldots, K\} \text{ such that } \tilde{m}_i^k \not\sim \tilde{m}_i^0 \text{ and } \tilde{m}_j^{k'} = \tilde{m}_j^0, \\ \text{for all } k' \in \{1, \ldots, k-1\} \text{ and all } j \neq i;\end{array} \\ 0, & \text{otherwise.} \end{cases}$$

That is, agent i receives a fine of ξ if he is the first one who deviates from his own 0th report.

Finally, given any $\bar{\tau} > 0$, we choose positive numbers α, γ, ξ, and K such that

$$\begin{aligned} \bar{\tau} &> \alpha(\eta'' + 2\eta) + (1-\alpha)(\gamma + \xi) \\ \xi &> \frac{1}{K}D \\ \gamma &> \xi + \frac{1}{K}D. \end{aligned} \tag{18}$$

We choose ξ small and K large such that $\xi > \frac{1}{K}D$ and $\xi + \frac{1}{K}D < \frac{\bar{\tau}}{4}$. Then, we choose γ such that $\xi + \frac{1}{K}D < \gamma < \frac{\bar{\tau}}{4}$. Thus, $(1-\alpha)(\gamma + \xi) < \frac{\bar{\tau}}{2}$ for any $\alpha \in (0,1)$. Given η'' and η chosen according to (8) and (9), we choose α small enough such that $\alpha(\eta'' + 2\eta) < \frac{\bar{\tau}}{2}$.

A.2.4 Proof of Theorem 2

Note that all allocations and transfers used in the baseline mechanism are multiplied by α, which can be interpreted as "the baseline mechanism is played with probability α." By the construction of $\tilde{g}(\tilde{m})$, we know that m_i^1, m_i^2 and m_i^3 have no influence on the outcome designated by ρ nor agent i's transfers specified by τ_i^4 or τ_i^5. In particular, m_i^3 affects τ_{i-1}^4 but not τ_i^4. Thanks to the features, we can establish the following claim.

Claim 4 *For every agent $i \in \mathcal{I}$, every state $\theta \in \Theta$, and every $\tilde{m}_i = \left(m_i, \tilde{m}_i^0, ..., \tilde{m}_i^K\right) \in S_i^\infty(\tilde{\mathcal{M}}, \theta)$, we have $m_i \in S_i^\infty(\mathcal{M}, \theta)$.*

Proof. We prove $\tilde{m}_i = \left(m_i, \tilde{m}_i^0, ..., \tilde{m}_i^K\right) \in S_i^k(\tilde{\mathcal{M}}, \theta)$ implies that $m_i \in S_i^k(\mathcal{M}, \theta)$ by induction on k. The case with $k = 0$ is trivial. Suppose $\tilde{m}_i = \left(m_i, \tilde{m}_i^0, ..., \tilde{m}_i^K\right) \in S_i^k(\theta, \tilde{\mathcal{M}})$ implies that $m_i \in S_i^k(\mathcal{M}, \theta)$. We shall show that $\tilde{m}_i = \left(m_i, \tilde{m}_i^0, ..., \tilde{m}_i^K\right) \in S_i^{k+1}(\tilde{\mathcal{M}}, \theta)$ implies that $m_i \in S_i^{k+1}(\mathcal{M}, \theta)$. We prove the contrapositive. Suppose that $m_i \notin S_i^{k+1}(\mathcal{M}, \theta)$. Thus, in the game $\Gamma(\mathcal{M}, \theta)$, a standard duality argument implies that there exists some $\beta_i \in \Delta(M_i)$ such that β_i delivers a strictly better payoff than m_i for agent i against any $m_{-i} \in S_{-i}^k(\mathcal{M}, \theta)$. Now consider $\tilde{\beta}_i \in \Delta(\tilde{M}_i)$ such that $\tilde{\beta}_i$ takes the same distribution among M_i as β_i does, and assigns probability one on $(\tilde{m}_i^0, ..., \tilde{m}_i^K)$. Note that $\tilde{\beta}_i$ and $\tilde{m}_i$ each generate different payoffs for agent i only when the baseline mechanism $\mathcal{M}$ is chosen. Moreover, against any $\tilde{m}_{-i} = \left(m_{-i}, \tilde{m}_{-i}^0, ..., \tilde{m}_{-i}^K\right) \in S_{-i}^k(\tilde{\mathcal{M}}, \theta)$, it follows from the induction hypothesis that $m_{-i} \in S_{-i}^k(\mathcal{M}, \theta)$. Since $\alpha > 0$, we know that $\tilde{\beta}_i$ delivers a strictly better payoff than $\tilde{m}_i$ for agent i against any $\tilde{m}_{-i} = \left(m_{-i}, \tilde{m}_{-i}^0, ..., \tilde{m}_{-i}^K\right) \in S_{-i}^k(\tilde{\mathcal{M}}, \theta)$. Hence, $\tilde{m}_i \notin S_i^{k+1}(\tilde{\mathcal{M}}, \theta)$. ■

By Claim 4 and the proof of Theorem 3, we have $m_j^3 \sim \theta$. Then, we can follow verbatim the argument on p. 12 of Abreu and Matsushima (1994) to show that every agent j reports a state equivalent to θ in his k-th report for every $k = 0, ..., K$. Hence, for every agent $i \in \mathcal{I}$ and every $\tilde{m}_i \in S_i^{\infty}(\tilde{\mathcal{M}}, \theta)$, we have $\tilde{g}(\tilde{m}) = f(\theta)$ and $\tilde{\tau}_i(\tilde{m}) = 0$. By (18), we also have $|\tau_i(m)| \le \bar{\tau}$ for every $m \in M$.[12]

A.3 Responsive SCFs

We show that any SCF can be "perturbed" a little bit to satisfy responsiveness and Maskin monotonicity. Fix an SCF f and $\varepsilon \in (0,1)$. Define $f^{\varepsilon} : \Theta \to \Delta(A)$ as follows: for any $\theta \in \Theta$,

$$f^{\varepsilon}(\theta) = \varepsilon y_i(\theta_i) + (1-\varepsilon) f(\theta),$$

where $y_i(\theta_i)$ is the dictator lottery for type θ_i, as constructed in Lemma 2. Moreover, by adding small penalties to the dictator lotteries, we can make $y_i(\theta_i) \neq y_i(\theta_i')$ whenever $\theta \neq \theta'$, without affecting the conclusion of Lemma 2 (i.e., (19) below). Therefore, $\theta \neq \theta'$ implies $f^{\varepsilon}(\theta) \neq f^{\varepsilon}(\theta')$. In other words, we can make f^{ε} responsive. We now argue that f^{ε} is also Maskin monotonic. Fix two states θ and θ' with $\theta \neq \theta'$ (and hence $f^{\varepsilon}(\theta) \neq f^{\varepsilon}(\theta')$). Since $\theta \neq \theta'$ and due to the construction of dictator lotteries, there must exist agent i for whom

$$u_i(y_i(\theta_i), \theta_i) > u_i(y_i(\theta_i'), \theta_i) \text{ and } u_i(y_i(\theta_i'), \theta_i') > u_i(y_i(\theta_i), \theta_i'). \tag{19}$$

We construct the following lottery $x(\theta', \theta_i) \in X$:

$$x(\theta', \theta_i) \equiv \varepsilon y_i(\theta_i) + (1-\varepsilon) f(\theta').$$

That is, $x(\theta', \theta_i)$ is constructed by replacing $y_i(\theta_i')$ in $f^{\varepsilon}(\theta')$ with $y_i(\theta_i)$. By (19), we have

$$x(\theta', \theta_i) \in \mathcal{L}_i\left(f^{\varepsilon}\left(\theta'\right), \theta_i'\right) \cap \mathcal{SU}_i\left(f^{\varepsilon}\left(\theta'\right), \theta_i\right).$$

This shows that f^{ε} satisfies Maskin monotonicity.

A.4 Direct Mechanisms

Example 2 *Suppose that there are two agents: $\{1,2\}$; two states: $\{\alpha, \beta\}$; and four pure alternatives: $\{a, b, c, d\}$. Define an SCF f such that $f(\alpha) = (a, 0, 0)$ and $f(\beta) = (b, 0, 0)$.*

[12] Thanks to the existence of such outcome w, the construction of $\rho(\cdot)$ allows us to penalize any unilateral deviation from an unamimous announcement. Hence, we do not need the additional transfer used in Abreu and Matsushima (1994) (which they denote by η).

Agents' utilities across different states are described in the following table:

v_1	α	β
a	2	3
b	0	0
c	-4	0
d	1	-1

v_2	α	β
a	0	0
b	3	2
c	-1	1
d	0	-4

Since $d \in \mathcal{L}_1(f(\beta), \beta) \cap \mathcal{SU}_1(f(\beta), \alpha)$ and $c \in \mathcal{L}_2(f(\alpha), \alpha) \cap \mathcal{SU}_2(f(\alpha), \beta)$, it follows that f satisfies Maskin monotonicity*. Hence, by Theorem 1, f is implementable in rationalizable strategies by a finite (indirect) mechanism. A direct mechanism $\mathcal{M} = ((M_i)_{i\in\{1,2\}}, h)$ in this environment has message space $M_i = \{\alpha, \beta\}$ and we denote its outcome and transfer rule altogether by $h = (g(\cdot), \tau_1(\cdot), \tau_2(\cdot))$. To derive a contradiction, we hypothesize that f is implementable in rationalizable strategies by a direct mechanism. Without loss of generality, we assume that $h(\alpha, \alpha) = (a, 0, 0)$ and $h(\beta, \beta) = (b, 0, 0)$; moreover, $(\alpha, \alpha) \in S^\infty(\mathcal{M}, \alpha)$ and $(\beta, \beta) \in S^\infty(\mathcal{M}, \beta)$. Our argument is decomposed into the following three claims.

The first claim states how to proceed the first step of elimination of messages under the hypothesis.

Claim 5 *At state α, message α is strictly dominated by message β for agent 1; moreover, at state β, message α is strictly dominated by message β for agent 2.*

Proof. Note that $h(\beta, \beta) = (b, 0, 0)$. Since $(b, 0, 0)$ is the best outcome for agent 2 at state α, there is no better outcome for agent 2 at state α. Hence, it is a best response for agent 2 to report β given that agent 1 reports β. Under our hypothesis of rationalizable implementation by a direct mechanism, message β is strictly dominated by message α at state α. A similar argument proves the second part of the claim. ■

The next claim states that agent 1 is the only whistle-blower when (β, β) is misreported at state α, and agent 2 is the only whistle-blower when (α, α) is misreported at state β; moreover, the common test allocation is $h(\alpha, \beta)$.

Claim 6 *$h(\alpha, \beta) \in \mathcal{SL}_1(f(\beta), \beta) \cap \mathcal{SU}_1(f(\beta), \alpha)$ and $h(\alpha, \beta) \in \mathcal{SL}_2(f(\alpha), \alpha) \cap \mathcal{SU}_2(f(\alpha), \beta)$.*

Proof. By our hypothesis, (α, α) is a rationalizable message profile at state α and (β, β) is a rationalizable message profile at state β. Moreover, (α, α) cannot be a rationalizable message profile at state β and (β,β) cannot be a rationalizable message profile at state α. By Claim 5, it must be that at state α, agent 1 strictly prefers $h(\alpha,\beta)$ to $h(\beta,\beta)$ and agent 2 strictly prefer $h(\alpha,\alpha)$ to $h(\alpha,\beta)$. Similarly, at state β, agent 2 must strictly prefer $h(\alpha,\beta)$ to $h(\alpha,\alpha)$ and agent 1 strictly prefers $h(\beta,\beta)$ to $h(\alpha,\beta)$. Hence, we obtain Claim 6. ■

Claim 7 *There are no direct mechanisms that implement f in rationalizable strategies.*

Proof. Suppose we can achieve rationalizable implementation in a direct mechanism. Then, by Claim 6, we have such an allocation $h(\alpha,\beta)$, which in general can be a lottery over the four pure alternatives as well as penalties. Let p_a, p_b, p_c, and p_d be the probabilities assigned over the four alternatives induced by $h(\alpha,\beta)$. By Claim 6, for agent 1, we obtain the following inequalities:

$$h(\alpha,\beta) \in \mathcal{SL}_1(f(\beta),\beta) \Leftrightarrow 3p_a - p_d + \tau_1(\alpha,\beta) < 0;$$
$$h(\alpha,\beta) \in \mathcal{SU}_1(f(\beta),\alpha) \Leftrightarrow 2p_a - 4p_c + p_d + \tau_1(\alpha,\beta) > 0.$$

Hence, we have $2p_d - 4p_c > p_a \geq 0$, which further implies that $p_d > 2p_c$.

For agent 2, we have the following inequalities:

$$h(\alpha,\beta) \in \mathcal{SL}_2(f(\alpha),\alpha) \Leftrightarrow 3p_b - p_c + \tau_2(\alpha,\beta) < 0;$$
$$h(\alpha,\beta) \in \mathcal{SU}_2(f(\alpha),\beta) \Leftrightarrow 2p_b + p_c - 4p_d + \tau_2(\alpha,\beta) > 0.$$

Hence, we have $2p_c - 4p_d > p_b \geq 0$, which further implies that $p_c > 2p_d$. Therefore, we obtain $p_d > 2p_c > 4p_d$, which is a contradiction. ■

References

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

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

Aumann, R., and A. Brandenburger (1995): "Epistemic conditions for Nash equilibrium," *Econometrica: Journal of the Econometric Society*, pp. 1161–1180.

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

Brandenburger, A., and E. Dekel (1987): "Rationalizability and Correlated Equilibria," *Econometrica*, 55, 1391–1402.

Chen, Y.-C., T. Kunimoto, and Y. Sun (2019): "Continous Implementation with Small Transfers," *mimeo*.

Chen, Y.-C., T. Kunimoto, Y. Sun, and S. Xiong (2020): "Maskin Meets Abreu and Matsushima," *mimeo*.

Jackson, M. O. (1992): "Implementation in Undominated Strategies: A Look at Bounded Mechanisms," *The Review of Economic Studies*, 59(4), 757–775.

Jain, R. (2019): "Rationalizable Implementation of Social Choice Correspondences," *mimeo*.

Kunimoto, T., and R. Serrano (2019): "Rationalizable implementation of correspondences," *Mathematics of Operations Research*, 44, 1326–1344.

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

Oury, M. (2015): "Continuous Implementation with Local Payoff Uncertainty," *Journal of Economic Theory*, 159, 656–677.

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

Xiong, S. (2019): "Rationalizable Implementation of Social Choice Functions: Complete Characterization," *mimeo*.