Continuous Implementation with Payoff Knowledge*

Yi-Chun Chen† Takashi Kunimoto‡ Yifei Sun§

November 24, 2022

Abstract

The robust mechanism design literature assumes players' knowledge about a fixed payoff environment and investigates the global robustness of optimal mechanisms to large changes in the information structure. Acknowledging the global robustness as a demanding requirement, we propose continuous implementation as a local robustness of optimal mechanisms to small changes in the information structure. Keeping the assumption of the payoff knowledge, we say that a social choice function is continuously implementable if there exists a mechanism which yields the outcome close to the desired one for all types close to the planner's initial model. We show that when a generic correlation condition is imposed on the class of interdependent-value environments, any (interim) incentive compatible social choice function is continuously implementable with arbitrarily small transfers imposed on and off the equilibrium. This exhibits a stark contrast with Bergemann and Morris (2005) who show that their global robustness amounts to ex post incentive compatibility as well as Oury and Tercieux (2012) who show that continuous implementation without payoff knowledge generates a substantial restriction, tightly connected to full implementation in rationalizable strategies.

*This paper is formerly titled as "Continuous Implementation with Small Transfers." We thank the editor and two anonymous referees for stimulating comments that substantially improve the paper. We also thank audiences at various seminar/conference presentations for helpful comments. Part of this paper was written while the three authors were visiting Academia Sinica, and we would like to thank the institution for its hospitality and support. All remaining errors are our own.

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

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

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

1 Introduction

While Bayesian mechanism design has been successful in generating many applications, it is rightly criticized for its sensitivity to the precise information that the agents and the planner have about the environment. To properly describe an incomplete information environment, an agent's private information is summarized by the notion of *type*. For an agent, a type specifies (i) his private information about his own preferences (*payoff type*); (ii) his belief about the payoff types of others (*first-order belief*); (iii) his belief about others' first-order beliefs (*second-order belief*), and so on, leading to an infinite hierarchy of beliefs. The set of all coherent belief hierarchies described above is called the *universal type space*.[1] Bayesian mechanism design theory typically works with a type space which is smaller than the universal type space and incorporates certain common knowledge among the agents, e.g., their beliefs are derived from an independent common prior. While the common knowledge assumptions often enhance the tractability of the model, it is at best an idealization of the reality.

The literature of *robust mechanism design* examines the restrictiveness of this kind of common knowledge assumption by fixing a payoff environment. The payoff environment specifies a set of outcomes, a set of payoff types, as well as utility functions for each agent, but makes "no" assumptions about all possible type spaces (including belief types as well as payoff types) which are constructed from the fixed payoff environment. We call this approach *global robustness*, which is pursued by Bergemann and Morris (2005). Bergemann and Morris (2005) define the planner's objective as a *social choice correspondence* (henceforth, SCC) that maps payoff type profiles into a nonempty subset of outcomes and say that an SCC is *interim (partially) implementable* on a type space if there exist a mechanism and *one* Bayes Nash equilibrium of that mechanism which yields the outcome in the set specified by the SCC for every payoff type profile.[2] Thus, this paper treats the problem of mechanism design and that

[1] See Section 3.1 for a formal definition and Mertens and Zamir (1985) or Brandenburger and Dekel (1993) for a formal construction of the universal type space.

[2] Following Bergemann and Morris (2005), we adopt partial implementation as the notion of implementation. Partial implementation requires that there be one equilibrium in the implementing mechanism that achieves the desirable outcome, whereas *full implementation* requires that all equilibria in the mechanism do so. Hereafter, we call partial implementation simply implementation and add the qualifier *full* only when we talk about full implementation.

of implementation interchangeably. In what they call separable environments, Bergemann and Morris (2005) show that an SCC is interim implementable over all type spaces if and only if it is *ex post* implementable, which is, by the revelation principle, equivalent to the existence of ex post incentive compatible social choice function (henceforth, SCF) which is contained in the SCC.[3] However, this equivalence result carries negative news for robust mechanism design because Jehiel et al. (2006) show that only *constant* SCFs are ex post incentive compatible when payoff types are multi-dimensional and interdependent value functions are generic.[4]

To seek positive results, we propose a notion of *local* robustness which weakens the notion of global robustness. Formally, we fix a benchmark type space associated with a given payoff environment and consider an SCF that maps (payoff and belief) type profiles into outcomes. Our notion of locally robust implementation adapts the notion of *continuous* implementation of Oury and Tercieux (2012) (henceforth, OT) to the setup in which the players always know their own payoff type as in Bergemann and Morris (2005). We say that an SCF is continuously implementable by a mechanism if there exists a (possibly mixed-strategy) equilibrium of the mechanism which yields the outcome close to the desired one for all types "close to" the planner's benchmark model. Following OT, we consider the closeness of types in terms of the product topology of weak convergence of infinite belief hierarchies in the universal type space. We also verify that any ex post incentive compatible SCF is indeed continuously implementable; hence, locally robust implementation in our sense is weaker than globally robust implementation in the sense of Bergemann and Morris (2005).

To establish our main result, we further assume that the agents' utility functions are quasilinear with respect to monetary transfers. We say that an SCF is continuously implementable *with small transfers* if it is continuously implementable by a mechanism in which arbitrarily small transfers are added to both on and off the equilibrium. Our main result (Theorem 1) shows that when a generic correlation condition, which we call Assumption 1, is imposed on the class of interdependent values environments, an SCF is continuously implementable with small transfers if and only if it is (interim) incentive compatible on the

[3]The reader is referred to Section 4.1 of Bergemann and Morris (2005) for the definition of separable environments in which they consider SCCs whose multi-valuedness is only permitted over the "private components," which correspond to the transfer component in our setup. Therefore, the result of Bergemann and Morris (2005) holds for such a class of SCCs.

[4]The reader is referred to Jehiel et al. (2006) for all the qualifications needed for their result.

benchmark type space. Since interim incentive compatibility is a necessary condition for interim implementation, our continuous implementation result is as permissive as it can be.

To achieve continuous implementation, we establish instead full implementation of any incentive compatible SCF under a permissive solution concept denoted as $S^{\infty}\hat{W}^{\infty}$, which is the set of message profiles surviving the iterative elimination of weakly dominated messages followed by the iterative elimination of interim strictly dominated messages.

We expand on the implication of Assumption 1. Assumption 1 is stronger than the BDP property proposed by Neeman (2004). A type space satisfies the BDP property if different types of any player have different beliefs. In contrast, Assumption 1 requires that different types of any player have different beliefs over *strategically distinguishable* payoff types in the sense of (Bergemann and Morris, 2009b, Proposition 2). In private-value environments where different types of a player have different preferences, Assumption 1 is equivalent to imposing the BDP property on the benchmark model. Since we are aiming for a permissive result to locally robust implement *any* interim incentive compatible SCF, it shall come at no surprise that our results leverage conditions on the benchmark type space such as Assumption 1 or the BDP property. This differentiates our exercise from the global robust implementation exercise due to Bergemann and Morris (2005) which aims to relax all common knowledge assumptions including the BDP property.[5]

The rest of the paper is organized as follows. Section 2 positions our contribution in a broader context of the literature and relegates the detailed comparisons with the related papers to Section 4.4. In Section 3, we introduce (i) the general setup for the paper; (ii) the notion of continuous implementation with small transfers; (iii) the notions of strategic distinguishability and the maximally revealing mechanism; and (iv) the generic correlation condition used in this paper (Assumption 1). In Section 4.1, we state the main result of this paper (Theorem 1) and discuss two special cases: the case with a complete-information benchmark model and the case with private-value environments. Section 4.2 describes how our main result is proved in a heuristic manner. In Sections 4.3, 4.3.1, and 4.3.2, we prepare all the machineries needed for the proof of our main result. Section 6 concludes the paper.

[5] Heifetz and Neeman (2006) show that, in auction setups, the BDP property is a necessary condition for full surplus extraction whose genericity is decisive on validity of the current mechanism design paradigm; see McAfee and Reny (1992). When all common knowledge assumptions are relaxed, Heifetz and Neeman (2006) establish the geometric as well as the measure-theoretic non-genericity of the BDP property, whereas Chen and Xiong (2011) establish the topological genericity of the BDP property.

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

2 A Broader Perspective

There have been several attempts in the literature to propose notions of locally robust implementation. Closest to our exercise are OT , Oury (2015), Jehiel et al. (2012) (henceforth, JMM, 2012), and Chen, Muller-Frank, Pai (2022). In contrast to the globally robust implementation exercise due to Bergemann and Morris (2005), these papers only consider some class of type spaces "nearby" a benchmark type space which the designer is reasonably confident of. We defer a detailed comparison with these papers and other related papers to Section 5 and provide here a broader perspective to situate our exercise within these existing contributions on locally robust implementation.

The main distinguishing feature that divides the aforementioned papers is the payoff knowledge assumption that the players know their own payoff type and their utility functions, as mappings from the product set of social alternatives and payoff type profiles, are common knowledge. The payoff knowledge assumption is imposed by Bergemann and Morris (2005) in their analysis of globally robust implementation and constitutes the basis of other ex post/globally robust implementation exercises in the literature. JMM and our paper impose the payoff knowledge assumption, whereas OT, Oury (2015), and Chen, Muller-Frank, Pai (2022) do not impose this assumption.

Without imposing the payoff knowledge assumption, OT and Oury (2015) allow for perturbations of the belief hierarchies as well as the payoff knowledge. In contrast, our paper (as well as JMM) maintain the payoff knowledge assumption and only perturb the belief hierarchies. As a result, OT and Oury's notions of local robustness are more demanding than ours. In particular, both OT's Theorem 4 and Oury (2015) prove that continuous implementation without payoff knowledge requires full interim rationalizable implementation.[6] Hence, an ex post implementable SCF need not be continuously implementable in the sense of OT or Oury (2015); see Appendix A.6, even though it must be continuously implementable in our

[6]More precisely, both equivalence results are proved for finite implementing mechanisms. Moreover, OT obtain their Theorem 4 with an additional assumption of costly messages. In contrast, Oury (2015) dispenses with the cost of sending messages but instead introduces "local payoff uncertainty," which means that the planner has some doubts on the payoffs of the outcomes and wants his prediction to be robust when these payoffs are close but not exactly equal to those in the initial model.

sense/with payoff knowledge. (See also Observation 1). Moreover, with small transfers and Assumption 1, we show that continuous implementation with payoff knowledge is as permissive as interim incentive compatibility which is substantially weaker than full implementation in interim rationalizable strategies.[7]

JMM also impose the payoff knowledge assumption and their notion of locally robust implementation is also weaker than ex post implementability. JMM prove that no "regular" allocation function is locally robust implementable in generic settings with quasi-linear utility. Among other differences, JMM prove their impossibility result for truthful equilibrium in direct mechanisms (defined on a perturbed type space) whereas we prove our result by constructing an indirect mechanism and invoking a mixed-strategy equilibrium. While our model setup and implementation notion are not directly comparable to those of JMM, the contrast points to an essential trade-off between restricting attention to direct mechanisms and using complex indirect mechanisms to locally robustly implement more (or all) incentive compatible SCFs.

The literature has also considered intermediate notions of robust implementation between global robustness and local robustness. In their study of full implementation, Artemov et al. (2013), Ollár and Penta (2017), and Ollár and Penta (2019) assume that the planner has partial knowledge about the agents' first-order beliefs over the payoff type space and the partial knowledge is always respected across all type spaces.[8] In the spirit of how these papers introduce the planner's partial knowledge to complement a belief-free approach, our paper introduces the payoff knowledge to the study of continuous implementation.

3 Preliminaries

In this section, we introduce the setup and concepts used throughout the paper. Section 3.1 introduces the setup for the paper. In Section 3.2, we introduce our notion of continuous implementation as a notion of locally robust implementation. Section 3.3 elaborates on the

[7] Bergemann and Morris (2008) show that interim rationalizable monotonicity is a necessary condition for full interim rationalizable implementation by finite mechanisms; moreover, OT show that interim rationalizable monotonicity implies (semi-strict) interim incentive compatibility and Bayesian monotonicity; see also Kunimoto et al. (2020).

[8] Ollár and Penta (2017), and Ollár and Penta (2019) insist on full implementation by *direct* mechanisms, while Artemov et al. (2013) study *virtual (or approximate)* full implementation by *indirect* mechanisms.

notion of strategic distinguishability and the maximally revealing mechanism, both of which are proposed by Bergemann and Morris (2009b).

3.1 The Environment

Let I denote a finite set of players and with abuse of notation, we also denote by I the cardinality of the set 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 $u_i : A \times \Theta \to \mathbb{R}$, where $\Theta = \Theta_1 \times \cdots \times \Theta_I$ is the finite set of payoff type profiles. We therefore assume that Θ has a product structure. We allow for interdependent values and $u_i(a, \theta)$ specifies the (bounded) utility of player i from the social alternative a under type profile $\theta \in \Theta$. We also write $\Theta_{-i} = \Theta_1 \times \cdots \times \Theta_{i-1} \times \Theta_{i+1} \times \cdots \times \Theta_I$.[9] We abuse notation to also denote by $u_i(x, \theta)$ player i's expected utility from a lottery allocation $x \in \Delta(A)$ under θ. Assume that player i's utility is quasilinear in transfers, denoted by $u_i(x, \theta) + \tau_i$ where $\tau_i \in \mathbb{R}$.

We follow the same setup as Bergemann and Morris (2005) and Bergemann and Morris (2011). Specifically, 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 interim belief for each $t_i \in T_i$. We assume that the model is common knowledge among all players. We also assume that each player knows his own type t_i (and hence his payoff type $(\hat{\theta}_i(t_i))$.[10] For each type profile $t = (t_i)_{i \in I}$, let $\hat{\theta}(t)$ denote the payoff type profile at t, i.e., $\hat{\theta}(t) \equiv (\hat{\theta}_i(t_i))_{i \in I}$. If T_i is a finite set for every player i, then we say that $(T_i, \hat{\theta}_i, \pi_i)_{i \in I}$ is a *finite* model. Let $\pi_i(t_i)[E]$ denote the probability that $\pi_i(t_i)$ assigns to any set $E \subset T_{-i}$.

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

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

The *second-order belief* of t_i is his belief about the set of payoff types and first-order beliefs

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

[10] As Oury (2015) argue in footnote 8 (p.659), except a special case of private values environments, OT's argument (in proving their Theorems 1-3) cannot be applied to a setup in which each player knows his payoff type and that the state space can be written as the product space of payoff types. See footnote 18 for further elaboration.

of player i's opponents. Formally, for any measurable set $F \subset \Theta \times \Delta(\Theta)^{I-1}$, we set

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

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.

The set of all belief hierarchies with "common certainty" that their beliefs are coherent (i.e., each player's beliefs at different orders are consistent with each other and this is commonly believed) is the *universal type space*; see Mertens and Zamir (1985) and Brandenburger and Dekel (1993). 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 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 $l \in \mathbb{N}$, $h_i^l(t_i^n) \to h_i^l(t_i)$ as $n \to \infty$. We write $t^n \to_p t$ if $t_i^n \to_p t_i$ for all $i \in I$.

This notion of convergence of a sequence of types also builds upon our crucial assumption that each agent "knows" his own payoff type and this knowledge is maintained as long as type t_i^n and t_i are close enough to each other. More precisely, by (1), $h_i^1(t_i)$ assigns probability one to $\hat{\theta}_i(t_i)$, and likewise, $h_i^1(t_i^n)$ assigns probability one to $\hat{\theta}_i(t_i^n)$. Hence, $h_i^1(t_i^n) \to h_i^1(t_i)$ only if $\hat{\theta}_i(t_i^n) = \hat{\theta}_i(t_i)$ for every n sufficiently large. As we discussed in the Introduction, this feature distinguishes our notion of continuous implementation from the notion of OT and is responsible for our permissive result of continuous implementation.[11] We will come back to this point in Sections 5.1 and 5.2.

Throughout the paper, we consider a fixed environment $\mathcal{E}$ which is a triplet $(A, \Theta, (u_i)_{i \in I})$ with a finite benchmark model $\bar{\mathcal{T}} = \left(\bar{T}_i, \bar{\theta}_i, \bar{\pi}_i\right)_{i \in I}$. We also consider a *planner* who aims to implement a *social choice function* (henceforth, SCF) $f : \bar{T} \to \Delta(A)$. Note that unlike the robust mechanism design literature, we consider a more general class of SCFs whose

[11] The distinction perhaps manifests itself best in a private-value setup where each agent's payoff type only determines his preference but not others'. In this case and with the terminology of Fudenberg et al. (1988), OT consider *elaborations with general types* where an agent may be uncertain about his own utility function, whereas we consider *elaborations with personal types* where each agent is certain of his own utility function. Fudenberg et al. (1988) wrote the following on p. 376 of their paper: "General elaborations may seem to the reader to be too large a class of perturbations, because they allow one player to know more about a second player's payoff than second player knows herself. By constraining ourselves to elaborations with the "type" structure, we suppose that each player has all the information about his own payoffs that any other player has." Our motivation to study continuous implementation with payoff knowledge is aligned with these remarks.

domain include agents' belief types as well payoff types. The following definition of incentive compatibility is standard.

Definition 1 *An SCF $f : \bar{T} \to \Delta(A)$ is (interim)* ***incentive compatible*** *if, for all $i \in I$ and all $t_i, t_i' \in \bar{T}_i$,*

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

3.2 Mechanisms and Continuous Implementation

We assume that the planner can penalize or reward any player by collecting or making *side payments*. A *mechanism* $\mathcal{M}$ is a triplet $((M_i), g, (\tau_i))_{i \in I}$ where M_i is the nonempty *finite message space* for player i; $g : M \to \Delta(A)$ is an *outcome function*; and $\tau_i : M \to \mathbb{R}$ is a *transfer rule* which specifies the payment from player i to the planner. For any $\alpha_i \in \Delta(M_i)$ and $\alpha_{-i} \in \Delta(M_{-i})$, we abuse the notation to denote by $g(\alpha_i, \alpha_{-i})$ the induced lottery in $\Delta(A)$ and by $\tau_i(\alpha_i, \alpha_{-i})$ the induced expected transfer. In the mechanism $\mathcal{M} = ((M_i), g, (\tau_i))_{i \in I}$, we define $\hat{\tau} = \max_{i \in I} \max_{m \in M} |\tau_i(m)|$ as the bound of transfer rule $(\tau_i)_{i \in I}$. We denote by $\mathcal{M}^{\hat{\tau}}$ a mechanism whose transfer rule is bounded by $\hat{\tau}$.

Given a mechanism $\mathcal{M}$ and a model $\mathcal{T}$, we write $U(\mathcal{M}, \mathcal{T})$ for the induced incomplete information game. In the game $U(\mathcal{M}, \mathcal{T})$, a (behavior) strategy of a player i is $\sigma_i : T_i \to \Delta(M_i)$. We follow Oury and Tercieux (2012) to write down the following definitions. A function $\nu_{-i} : T_{-i} \to \Delta(M_{-i})$ is called a *conjecture* of player i. We define the interim payoff for player i of type t_i when he chooses (mixed) message α_i against conjecture ν_{-i} as:

$$V_i((\alpha_i, \nu_{-i}), t_i) = \sum_{t_{-i}} \pi_i(t_i)[t_{-i}] \left[u_i(g(\alpha_i, \nu_{-i}(t_{-i})), \hat{\theta}(t)) + \tau_i(\alpha_i, \nu_{-i}(t_{-i})) \right].$$

Definition 2 *A profile of strategies $\sigma = (\sigma_1, ..., \sigma_I)$ is a* ***Bayes Nash equilibrium*** *in $U(\mathcal{M}, \mathcal{T})$ if, for each player $i \in I$ and each type $t_i \in T_i$,*

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

We say that a strategy profile σ is a *strict* Bayes Nash equilibrium if, for every $i \in I$ and $t_i \in T_i$, $\sigma_i(t_i)$ is the unique solution to $\max_{m_i' \in M_i} V_i((m_i', \sigma_{-i}), t_i)$. It is easy to see that a strict Bayes Nash equilibrium must be a pure-strategy equilibrium.

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

Definition 3 *Fix a mechanism $\mathcal{M}$ and a model $\mathcal{T}$ such that $\bar{\mathcal{T}} \subset \mathcal{T}$. We say that a Bayes Nash equilibrium σ in $U(\mathcal{M}, \mathcal{T})$* ***(strictly) continuously implements*** *the SCF $f : \bar{T} \to \Delta(A)$ if the following two conditions hold: (i) $\sigma_{|\bar{T}}$ is a (strict) Bayes Nash equilibrium in $U\left(\mathcal{M}, \bar{\mathcal{T}}\right)$; (ii) for any $t \in \bar{T}$ and any sequence $t^n \to_p t$, whenever $t^n \in T$ for each n, we have $(g \circ \sigma)(t^n) \to f(t)$.*

Remark: In their definition of continuous implementation, Oury and Tercieux (2012) require in addition that $\sigma_{|\bar{T}}$ be a *pure strategy* Bayes Nash equilibrium. As they mainly focus on strict continuous implementation in the paper, this restriction is inconsequential. Here we focus on continuous implementation rather than strict continuous implementation and do not impose the requirement that $\sigma_{|\bar{T}}$ be a pure strategy Bayes Nash equilibrium.

We introduce the notion of continuous implementation with arbitrarily small transfers:

Definition 4 *An SCF $f : \bar{T} \to \Delta(A)$ is continuously implementable with* ***arbitrarily small transfers*** *if, for any $\hat{\tau} > 0$, there exists a mechanism $\mathcal{M}^{\hat{\tau}}$ such that for each model $\mathcal{T}$ with $\bar{\mathcal{T}} \subset \mathcal{T}$, there is a Bayes Nash equilibrium σ in $U(\mathcal{M}, \mathcal{T})$ that continuously implements the SCF f.*

Recall that Proposition 2 of Bergemann and Morris (2005) proves that an SCF is (globally) robustly implementable if and only if it is ex post implementable. In our setup, we say that $f : \bar{T} \to \Delta(A)$ is ex post implementable if and only if there exists $f^* : \Theta \to \Delta(A)$ such that (1) $f(t) = f^*(\hat{\theta}(t))$ for every $t \in \bar{T}$; and (2) for all $i \in I$ and $\theta \in \Theta$,

$$u_i(f^*(\theta), \theta) \geq u_i(f^*(\theta_i', \theta_{-i}), \theta). \tag{2}$$

The following observation verifies that if f is ex post implementable (i.e., globally robustly implementation), then it is continuously implementable (i.e., locally robustly implementable in our sense). The proof follows from the proof of Proposition 1 in Bergemann and Morris (2005) and relies on the payoff type knowledge in constructing the equilibrium σ.

Observation 1 If f is ex post implementable, then it is continuously implementable.

Proof. Since f is ex post implementable, there is $f^* : \Theta \to \Delta(A)$ such that (1) and (2) above hold. Consider an arbitrary type space $\mathcal{T} = (T_i, \hat{\theta}_i, \pi_i)_{i \in I}$ such that $\bar{\mathcal{T}} \subset \mathcal{T}$. We claim that the strategy profile σ with $\sigma_i(t_i) = \hat{\theta}_i(t_i)$ is a Bayes Nash equilibrium σ in $U(f^*, \mathcal{T})$ and σ continuously implements f. Note that in writing $U(f^*, \mathcal{T})$, we identify f^* with a mechanism with $M_i = \Theta_i$. It follows from (2) that, for any $i \in I$ and $t_i \in T_i$,

$$\hat{\theta}_\iota(t_i) \in \arg\max_{\theta_i \in \Theta_i} u_i(f^*(\theta_i, \theta_{-i}), (\theta_i, \theta_{-i})),$$

Thus, σ is a Bayes Nash equilibrium in $U(f^*, \mathcal{T})$. Moreover, for any $t \in \bar{T}$ and any sequence $t^n \to_p t$, we know that $\hat{\theta}_i(t_i^n) = \hat{\theta}_i(t_i)$; hence, we have $(f^* \circ \sigma)(t^n) \to f^*(\hat{\theta}(t)) = f(t)$. ■

3.3 Maximally Revealing Mechanism

Given a mechanism $\mathcal{M} = (M, g)$, we first define the process of iterative elimination of strictly dominated messages, which makes no assumptions on each player's belief about the other players' payoff types. We set $\hat{S}_i^0(\theta_i|\mathcal{M}) = M_i$ and for each $l \geq 0$, we inductively define

$$\begin{aligned} &\hat{S}_i^{l+1}(\theta_i|\mathcal{M}) \\ &= \left\{ m_i \in \hat{S}_i^l(\theta_i|\mathcal{M}) \;\middle|\; \begin{array}{l} \exists \alpha_i \in \Delta(M_i) \text{ s.t. } u_i(g(\alpha_i, m_{-i}), (\theta_i, \theta_{-i})) + \tau_i(\alpha_i, m_{-i}) \\ > u_i(g(m_i, m_{-i}), (\theta_i, \theta_{-i})) + \tau_i(m_i, m_{-i}) \\ \text{for any } m_{-i} \in \hat{S}_{-i}^l(\theta_{-i}|\mathcal{M}) \text{ and any } \theta_{-i} \in \Theta_{-i}. \end{array} \right\}. \end{aligned}$$

Finally, we let $\hat{S}_i^\infty(\theta_i|\mathcal{M}) = \bigcap_{l \geq 0} \hat{S}_i^l(\theta_i|\mathcal{M})$ and call it the set of message profiles for payoff type θ_i which survive the iterative elimination of strictly dominated messages. Following Bergemann and Morris (2009b), we say that payoff types θ_i and θ_i' are *strategically indistinguishable* (we denote it by $\theta_i \sim \theta_i'$) if $\hat{S}_i^\infty(\theta_i|\mathcal{M}) \cap \hat{S}_i^\infty(\theta_i'|\mathcal{M}) \neq \varnothing$ for every mechanism $\mathcal{M}$. The following definition is also proposed by (Bergemann and Morris, 2009b, Proposition 2).

Definition 5 *We say that a mechanism $\mathcal{M}^* = ((M_i^*), g^*, (\tau_i^*))_{i \in I}$ is a maximally revealing mechanism if $\theta_i \not\sim \theta_i'$, then $\hat{S}_i^\infty(\theta_i|\mathcal{M}^*) \cap \hat{S}_i^\infty(\theta_i'|\mathcal{M}^*) = \varnothing$.*

That is, a maximally revealing mechanism is a mechanism where every pair of strategically distinguishable payoff types can be distinguished according to their messages which survive iterative strict dominance. (Bergemann and Morris, 2009b, Proposition 2) construct a maximally revealing mechanism which will be a building block of our implementing mechanism in proving the main result (see Section 4 for details).

Let $\sim^*$ be the transitive closure of the binary relation $\sim$. For each player i of payoff type θ_i, we define $P_i(\theta_i) = \{\theta_i' \in \Theta_i | \theta_i' \sim^* \theta_i\}$. Since $\sim^*$ is transitive, it follows that $\{P_i(\theta_i)\}_{\theta_i \in \Theta_i}$ forms a partition over Θ_i, which we denote by $\mathcal{P}_i$. For any $m_i \in \hat{S}_i^\infty(\theta_i | \mathcal{M}^*)$, we are able to identify the unique $P_i(\theta_i) \in \mathcal{P}_i$.

To formulate our assumption, observe first that $\mathcal{P}_i$ induces a partition Ψ_i^0 over $\bar{T}_i$, i.e., $\Psi_i^0 = \{\psi_i^0(t_i)\}_{t_i \in \bar{T}_i}$ such that, for any types t_i and t_i' in $\bar{T}_i$, $t_i' \in \psi_i^0(t_i)$ if and only if $\hat{\theta}_i(t_i') \in \mathcal{P}_i(\hat{\theta}_i(t_i))$. Let $\chi_i^0(t_i)$ denote the belief over Ψ_{-i}^0 for player i of type t_i, that is,

$$\chi_i^0(t_i)\left[\psi_{-i}^0\right] = \sum_{t_{-i} \in \psi_{-i}^0} \pi_i(t_i)[t_{-i}],$$

for any $\psi_{-i}^0 \in \Psi_{-i}^0$. Moreover, $\chi_i^0(\cdot)$ and Ψ_i^0 jointly induce another partition Ψ_i^1 over $\bar{T}_i$, i.e., $\Psi_i^1 = \{\psi_i^1(t_i)\}_{t_i \in \bar{T}_i}$ in which for any types t_i and t_i' in $\bar{T}_i$, we have $t_i' \in \psi_i^1(t_i)$ if and only if $\chi_i^0(t_i) = \chi_i^0(t_i')$ and t_i' belongs to $\psi_i^0(t_i)$. Let $\chi_i^1(t_i)$ denote the belief of type t_i over Ψ_{-i}^1. We are now ready to state our key assumption.

Assumption 1 *For any player $i \in I$, any pair of types t_i and t_i' in $\bar{T}_i$ with $t_i \neq t_i'$, we have $\chi_i^1(t_i) \neq \chi_i^1(t_i')$.*

Assumption 1 says that each player's type can fully be identified with their belief over $\prod_{j \neq i}\left(\Psi_j^0 \times \Delta(\Psi_{-j}^0)\right)$, i.e., their belief over the partition Ψ^0 (induced by strategically distinguishable payoff types of their opponents) and over their opponents' beliefs over Ψ^0. Assumption 1 holds if each player's distinct types hold different beliefs over Ψ_{-i}^0. Hence, provided that at least two players have nontrivial partition under $\mathcal{P}$, Assumption 1 generically holds over the space of probability distributions over $\bar{T}$. However, Assumption 1 does not hold if the players' types are independently distributed according to a common prior.

To elaborate on Assumption 1 further, we consider a complete-information model, i.e., a model $\mathcal{T}^{CI} = (T_i^{CI}, \hat{\theta}_i, \pi_i)_{i \in I}$ where for each $i \in I$, $T_i^{CI} = \bigcup_{\theta \in \Theta}\{t_{i,\theta}\}$ and for each $\theta = (\theta)_{i \in I} \in \Theta$, we have $\hat{\theta}_i(t_{i,\theta}) = \theta_i$ and $\pi_i(t_{i,\theta})[t_{-i,\theta}] = 1$. In other words, at any payoff type profile θ, it is common knowledge among all the players that payoff type profile is θ. In this case, Assumption 1 holds if $\mathcal{P}_i$ is the finest partition $\{\{\theta_i\} | \theta_i \in \Theta_i\}$. Then, it follows that $\psi_i^0(t_{i,\theta}) = \psi_i^0(t_{i,\theta'})$ only if $\theta_i = \theta_i'$; moreover, $\chi_i^0(t_{i,\theta}) = \chi_i^0(t_{i,\theta'})$ only if $\theta_{-i} = \theta_{-i}'$. Hence, $\psi_j^1(t_{j,\theta}) = \{t_{j,\theta}\}$ for each $\theta \in \Theta$ and each j. It follows that $\chi_i^1(t_{i,\theta}) = \chi_i^1(t_{i,\theta'})$ only if $t_{i,\theta} = t_{i,\theta'}$.

We name two prominent situations where $\mathcal{P}_i$ is the finest partition. First, if the players' values are private (i.e., $u_i : \Delta(A) \times \Theta_i \to \mathbb{R}$), then $\mathcal{P}_i$ is the finest partition if different payoff

types induce different preferences over the lottery allocations. This is the assumption made in Abreu and Matsushima (1994). Second, if the players' values are interdependent, Bergemann and Morris (2009a) show that $\mathcal{P}_i$ is the finest possible partition when the following three conditions are all satisfied: (1) there is a strictly ex post incentive compatible SCF; (2) players have single-crossing preferences; (3) the players' preferences satisfy a condition called *the contraction property*, which demands that value interdependence be not too large. Based upon the two prominent situations, we will derive Corollaries 1 and 2 from Theorem 1 in the next section.

It is also straightforward to see how we can generalize Assumption 1. To do so, we define partition Ψ_i^k for any $k \geq 2$. That is, Ψ_i^k is the partition over $\bar{T}_i$, which is induced by $\chi_i^{k-1}(\cdot)$ and Ψ_i^{k-1}. Since $\{\Psi_i^k\}_{k=1}^{\infty}$ is a sequence of increasingly finer partitions over $\bar{T}_i$ which is a finite set, Ψ_i^k becomes a fixed partition Ψ_i for any k sufficiently large. Then, we can prove our continuous implementation result by weakening Assumption 1 to the requirement that the SCF be measurable with respect to Ψ. Here we impose the stronger assumption for simplicity, as our goal is to include the special case with a complete-information benchmark model and P_i being the finest partition $\{\{\theta_i\} | \theta_i \in \Theta_i\}$ so that Corollaries 1 and 2 in the next section follow from Theorem 1.

In private-value environments where different types of any player have different preferences, it follows that P_i is the finest partition; hence, Assumption 1 is equivalent to the beliefs-determine-preferences (BDP) property proposed by Neeman (2004) in such environments. The BDP property says that distinct types must hold distinct beliefs over the opponents' types. The generalized version of Assumption 1 discussed in the preceding paragraph is then translated into a strengthening of the BDP property: any pair of distinct types holds distinct (higher-order) beliefs over strategically distinguishable payoff types. This is what we need for our continuous implementation result.

4 Main Result

In this section, we discuss our main result. In Section 4.1, we first state our main result formally and in Section 4.2, we next illustrate the logic of the proof in a heuristic manner. Section 4.3 introduces the solution concept of $S^{\infty}\hat{W}^{\infty}$, i.e., the set of message profiles which survive the iterative elimination of weakly dominated messages followed by the iterative

elimination of interim strictly dominated messages. In Sections 4.3.1 and 4.3.2, we explain our key augmentation step which "combines" a generic version of the maximally revealing mechanism and a mechanism akin to the one used in Abreu and Matsushima (1994) into a single implementing mechanism. Section 4.4 provides the proof of Theorem 1.

4.1 The Theorem

We now state the main result of this paper:

Theorem 1 *Suppose that Assumption 1 holds. Then, an SCF $f : \bar{T} \to \Delta(A)$ is continuously implementable with arbitrarily small transfers if and only if it is incentive compatible.*

We relegate the proof of this theorem to Section 4.4 and only outline the steps of the proof in the rest of the section. Observe that the equilibrium which continuously implements f also implements f in $\bar{\mathcal{T}}$. Then, a limiting argument taking the transfer bound to zero shows that incentive compatibility is a necessary condition for continuous implementation with arbitrarily small transfers. The main task is therefore to prove the "if" part of Theorem 1, which is the focus of our discussion below.

Let f be an SCF which is incentive compatible. We structure the main argument in two steps: first, we show that under Assumption 1, we can implement the SCF f under a solution concept denoted by $S^\infty\hat{W}^\infty$ (to be defined in Section 4.3) with arbitrarily small transfers. Second, we show that if the SCF f is implementable in $S^\infty\hat{W}^\infty$ with arbitrarily small transfers, then it must be continuously implementable with arbitrarily small transfers.

To grasp the basic idea, consider a benchmark model with complete information. First, assume, as in Abreu and Matsushima (1994), that values are private (i.e., $u_i : \Delta(A) \times \Theta_i \to \mathbb{R}$) *and* different payoff types induce different preferences over lottery allocations. Under this assumption, Abreu and Matsushima (1994) show that any social choice function can be implemented in one round deletion of interim weakly dominated messages followed by the iterative deletion of interim strictly dominated messages (i.e., $S^\infty W$) by a finite mechanism. Thanks to the private-value assumption, interim weak dominance is equivalent to iterative weak dominance, i.e., $S^\infty\hat{W}^\infty = S^\infty W$. Hence, we obtain the following corollary. The corollary can be proved by simply invoking the mechanism constructed in Abreu and Matsushima (1994) and observing that for any model $\mathcal{T}$, there is a trembling-hand perfect equilibrium σ which survives $S^\infty W$.

Corollary 1 *Consider a complete information model* $\mathcal{T}^{CI} = (T_i^{CI}, \hat{\theta}_i, \pi_i)_{i \in I}$ *in which the agents' values are private and different payoff types* θ_i *and* θ_i' *induce different preferences over lottery allocations* $\Delta(A)$*. Then, an SCF* $f : \bar{T} \to \Delta(A)$ *is continuously implementable with arbitrarily small transfers if and only if it is incentive compatible.*

When we consider a complete information model with interdependent values, however, the approach of Abreu and Matsushima (1994) applies only when different type *profiles* induce different preferences over lottery allocations.[12] We do not need to make this stronger assumption. Indeed, as we remark in Section 3.3, to make Assumption 1 satisfied under a complete-information benchmark, we only need to require that the partition $\mathcal{P}_i$ be the finest possible one. A leading example of such an interdependent-value environment has been studied by Bergemann and Morris (2009a) which we briefly recap at the end of the previous section. We document this more permissive special case as another corollary.

Corollary 2 *Consider a complete information model* $\mathcal{T}^{CI} = (T_i^{CI}, \hat{\theta}_i, \pi_i)_{i \in I}$ *in which* $\mathcal{P}_i$ *is the finest possible partition* $\{\{\theta_i\} | \theta_i \in \Theta_i\}$*. Then, an SCF* $f : \bar{T} \to \Delta(A)$ *is continuously implementable with arbitrarily small transfers if and only if it is incentive compatible.*

Note that in this second case, our result is not reduced to that of Abreu and Matsushima (1994) even though we consider a complete-information benchmark $\bar{\mathcal{T}}$. In other words, regardless of whether we deal with complete information or incomplete information, the mileage of our Proposition 1 over the existing literature lies in our handling the case with interdependent values.

4.2 Roadmap

In this section, we outline the proof of Theorem 1 according to Figure 1. By "$A \to B$" in the diagram, we mean that A is used for proving B. There are three propositions used for proving our Theorem 1. Among them, Proposition 1 is the key step which we will explain separately in Sections 4.3.1 and 4.3.2.

Proposition 1 shows that under Assumption 1, if an SCF is incentive compatible, then it is fully implementable with arbitrarily small transfers in $S^{\infty}\hat{W}^{\infty}$, which is the set of

[12] More precisely, when we translate the interdependent-value model to a private-value model (by means of the complete-information assumption and in order to apply Abreu and Matsushima (1994)), a payoff type in the latter corresponds to a payoff type *profile* in the former.

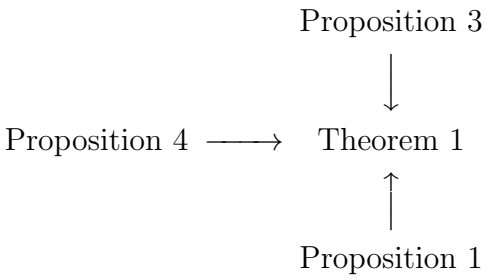

Figure 1: The Diagram of the Proof of Theorem 1

message profiles surviving the iterative elimination of weakly dominated messages followed by the iterative elimination of interim strictly dominated messages.[13] Proposition 3 shows that the solution correspondence $S^\infty\hat{W}^\infty$ in a finite mechanism is upper hemicontinuous. This result is considered an extension of the well known upper hemicontinuity of the interim correlated rationalizability correspondence (see Dekel et al. (2007)) to the case with payoff knowledge. Therefore, Proposition 3 establishes the continuity property of the implementing mechanism which is constructed in proving Proposition 1. Finally, Proposition 4 shows that any Bayesian game with a finite action/message space and a countable type space possesses a Bayes Nash equilibrium which survives $S^\infty\hat{W}^\infty$. This, together with Proposition 3, establishes the existence of an equilibrium which exhibits the desirable robustness property for continuous implementation.[14]

4.3 The Solution Concept of $S^\infty\hat{W}^\infty$

In proving Theorem 1, our major step is to show that any incentive compatible SCF is implementable in $S^\infty\hat{W}^\infty$ with arbitrarily small transfers. To formalize the step, we first define the solution concept of $S^\infty\hat{W}^\infty$. Given a mechanism $\mathcal{M}$, we first define the process of iterative elimination of weakly dominated messages. As the process of iterative elimination of strictly dominated messages, the iterative elimination of weakly dominated messages makes

[13] A message m_i is weakly dominated by m_i' if against any message profile and payoff type profile of the other agents, m_i yields at least as much payoff for agent i as m_i'; moreover, for some message profile and some payoff type profile of other agents, m_i yields strictly higher payoff than m_i'. The solution concept is proposed by Chen et al. (2015) but they do not consider the case with $\Theta = \times_{i\in I}\Theta_i$ in which players know their own payoff types (and the knowledge is never perturbed).

[14] The idea is similar to the result in Kohlberg and Mertens (1986), which shows that each stable set contains a stable set in the truncated game obtained by eliminating a weakly dominated strategy.

no assumption on each player's belief about other players' payoff types.

We set $\hat{W}_i^0(\theta_i|\mathcal{M}) = M_i$ and for each integer $l \geq 0$, we inductively define

$$\hat{W}_i^{l+1}(\theta_i|\mathcal{M})$$
$$= \left\{ m_i \in \hat{W}_i^l(\theta_i|\mathcal{M}) \;\middle|\; \begin{array}{l} \exists \alpha_i \in \Delta(M_i) \text{ s.t. } u_i(g(\alpha_i, m_{-i}), (\theta_i, \theta_{-i})) + \tau_i(\alpha_i, m_{-i}) \\ \geq u_i(g(m_i, m_{-i}), (\theta_i, \theta_{-i})) + \tau_i(m_i, m_{-i}) \\ \text{for any } m_{-i} \in \hat{W}_{-i}^l(\theta_{-i}|\mathcal{M}) \text{ and any } \theta_{-i} \in \Theta_{-i} \text{ and a strict inequality} \\ \text{holds for some } m_{-i} \in \hat{W}_{-i}^l(\theta_{-i}|\mathcal{M}) \text{ and some } \theta_{-i} \in \Theta_{-i} \end{array} \right\}.$$

Finally, we say that $\hat{W}_i^\infty(\theta_i|\mathcal{M}) \equiv \bigcap_{l\geq 0} \hat{W}_i^l(\theta_i|\mathcal{M})$ is the set of messages surviving the iterative deletion of *weakly* dominated messages for payoff type θ_i.

We define a solution concept $S^\infty\hat{W}^\infty$ as follows. We set $S_i^0\hat{W}^\infty(t_i|\mathcal{M},\mathcal{T}) = \hat{W}_i^\infty(\hat{\theta}_i(t_i)|\mathcal{M})$ and for each integer $l \geq 1$, we inductively define $m_i \in S_i^{l+1}\hat{W}^\infty(t_i|\mathcal{M},\mathcal{T})$ if and only if there does not exist $\alpha_i \in \Delta(M_i)$ such that

$$V_i((\alpha_i, \nu_{-i}), t_i) > V_i((m_i, \nu_{-i}), t_i)$$

for all conjecture $\nu_{-i} : T_{-i} \to M_{-i}$ and all $t_{-i} \in T_{-i}$ such that $\nu_{-i}(t_{-i}) \in S_{-i}^l\hat{W}^\infty(t_{-i}|\mathcal{M},\mathcal{T})$ for each $t_{-i} \in T_{-i}$ where $S_{-i}^l\hat{W}^\infty(t_{-i}|\mathcal{M},\mathcal{T}) \equiv \prod_{j\neq i} S_j^l\hat{W}^\infty(t_j|\mathcal{M},\mathcal{T})$. See Section 3.2 for the notation $V_i((\alpha_i, \nu_{-i}), t_i)$. Let $S^\infty\hat{W}^\infty$ denote the set of message profiles which survive the iterative deletion of weakly dominated messages followed by the iterative removal of interim strictly dominated messages, i.e.,

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

Finally, we define

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

We do not intend to justify the plausibility of the solution concept $S^\infty\hat{W}^\infty$. The solution concept $S^\infty\hat{W}^\infty$ is entirely instrumental in our proof. That is, implementation in $S^\infty\hat{W}^\infty$ is only an intermediate step toward achieving our result of continuous implementation. We now formally define the notion of implementation in $S^\infty\hat{W}^\infty$ with arbitrarily small transfers.

Definition 6 *An SCF $f : \bar{T} \to \Delta(A)$ is (fully) implementable in $S^\infty\hat{W}^\infty$ with **arbitrarily small transfers** if for any $\hat{\tau} > 0$, there exists a mechanism $\mathcal{M}^{\hat{\tau}}$ such that $g(m) = f(t)$ for every $m \in S^\infty\hat{W}^\infty\left(t|\mathcal{M}^{\hat{\tau}}, \bar{\mathcal{T}}\right)$ and every $t \in \bar{T}$.*

We can now formally state our first step of the proof of Theorem 1 as follows.

Proposition 1 *Suppose that Assumption 1 holds. If an SCF f is incentive compatible, then it is implementable in $S^{\infty}\hat{W}^{\infty}$ with arbitrarily small transfers.*

Proof. See Appendix A.3. ■

While we relegate its formal proof to the Appendix, we rather elaborate on its main idea and key step which we call *augmentation* in the next two subsections.

4.3.1 The Mechanism

To prove Proposition 1, we construct a mechanism $\bar{\mathcal{M}}^*$ which "connects" two mechanisms, $\mathcal{M}^*$ and $\bar{\mathcal{M}}$. First, we adopt the maximally revealing mechanism from Bergemann and Morris (2009b) and modify it into a "generic" maximally revealing mechanism $\mathcal{M}^*$. This is established in Lemma 3 in the Appendix. Second, the other building block $\bar{\mathcal{M}}$ is what we call an *extended direct* mechanism:

Definition 7 *We say that $\bar{\mathcal{M}} = \left((\bar{M}_i), \bar{g}, (\bar{\tau}_i)\right)_{i\in I}$ is an **extended direct mechanism** if for each $i \in I$, $\bar{M}_i = \bar{T}_i \times \cdots \times \bar{T}_i$ consists of finitely many copies of $\bar{T}_i$ and $\bar{g}(t, ..., t) = f(t)$ for every $t \in \bar{T}$.*

That is, an extended direct mechanism is a mechanism where each player announces his own type for finitely many times and truth-telling of everyone delivers the socially desirable outcome.

Third, we construct what we call an *augmented mechanism* $\bar{\mathcal{M}}^* = ((M_i), g, (\tau_i))_{i\in I}$ which builds upon and "combines" a maximally revealing mechanism $\mathcal{M}^*$ and an extended direct mechanism $\bar{\mathcal{M}} = \left((\bar{M}_i), \bar{g}, (\bar{\tau}_i)\right)_{i\in I}$.[15] In describing the construction of $\bar{\mathcal{M}}^*$, we fix $\bar{l}$ to be a positive number of iterations, which terminates the iterative deletion of strictly dominated messages in $\mathcal{M}^*$.[16] The augmented mechanism has the following components.

1. **The message space**:

[15] We construct the maximally revealing mechanism in Lemma 3 and the extended direct mechanism in the proof of Proposition 1.

[16] That is, for any $i \in I$, and $\theta_i \in \Theta_i$, we have $\hat{S}_i^l(\theta_i|\mathcal{M}^*) = \hat{S}_i^{\infty}(\theta_i|\mathcal{M}^*), \forall l \geq \bar{l}$.

Player i's message space is

$$M_i = M_i^0 \times M_i^1 \times \cdots \times M_i^{\bar{l}+3} \times M_i^{\bar{l}+4} \times M_i^{\bar{l}+5} = M_i^* \times \underbrace{\bar{T}_i \times \cdots \times \bar{T}_i}_{\bar{l}+3 \text{ copies of } \bar{T}_i} \times \bar{M}_i,$$

where $\bar{M}_i = M_i^{\bar{l}+4} \times M_i^{\bar{l}+5}$; $M_i^{\bar{l}+4} = \bar{T}_i$; and $M_i^{\bar{l}+5}$ consists of K copies of $\bar{T}_i$. That is, each player i simultaneously makes an announcement in M_i^*, $\bar{l}+3$ announcements of his own type, and finally an announcement in $\bar{M}_i$.

2. **The outcome function**:

Let $\epsilon \in (0,1)$ be a small positive number. Define $e : M \to \mathbb{R}$ by

$$e(m) = \begin{cases} \epsilon, & \text{if } m_i^l \neq m_i^2 \text{ for some } i \in I \text{ and some } l \in \{3, \ldots, \bar{l}+3\}, \\ 0, & \text{otherwise..} \end{cases} \tag{3}$$

Based on the outcome function g^* in the maximally revealing mechanism $\mathcal{M}^*$ and the outcome function $\bar{g}$ of the mechanism $\bar{\mathcal{M}}$, the outcome function of the augmented mechanism $g : M \to \Delta(A)$ is defined as follows: for each $m \in M$,

$$g(m) = e(m) \times g^*(m^0) + (1 - e(m)) \times \bar{g}(m^{\bar{l}+4}, m^{\bar{l}+5}). \tag{4}$$

3. **The transfer rule**:

In addition to τ_i^* (i.e., the transfer rule in $\mathcal{M}^*$) and $\bar{\tau}_i$ (i.e., the transfer rule in $\bar{\mathcal{M}}$), player i makes $\bar{l}+5$ number of payments of the three different sorts which we denote by $\tau_i^0(m_i^1, m_{-i}^0)$, $\tau_i^1(m_i^2, m_{-i}^1, m_{-i}^0)$, and $\tau_i^2(m_i^l, m_{-i}^{l-1}))$ for any $l = 2, \ldots, \bar{l}+4$, respectively, where τ_i^0, τ_i^1 and τ_i^2 will be defined in Section A.2.1. They are essentially the proper scoring rules (eliciting the players' true type) which satisfy a generic condition with a total bound denoted by τ. Hence, under a message profile m, player i pays a total equal to:

$$\tau_i(m) = \tau_i^*\left(m^0\right) + \tau_i^0(m_i^1, m_{-i}^0) + \tau_i^1(m_i^2, m_{-i}^1, m_{-i}^0) + \sum_{l=3}^{\bar{l}+4} \tau_i^2(m_i^l, m_{-i}^{l-1}) + \bar{\tau}_i(m_i^{\bar{l}+4}, m_i^{\bar{l}+5}), \tag{5}$$

The precise specification of the transfer rule and the choice of parameters of the mechanism (including the size of transfers) can be found in Appendix A.2.1 in proving Proposition 2. We will summarize the idea behind the construction of the mechanism $\bar{\mathcal{M}}^*$ as well as the role played by Proposition 2 in proving Proposition 1 in the next section.

4.3.2 Augmentation

We now discuss the augmentation step which is at the heart of our technical contribution. Formally, given any maximally revealing mechanism $\mathcal{M}^*$, Proposition 2 shows that we can choose the transfer rule and the parameters in the augmented mechanism $\bar{\mathcal{M}}^*$ such that the transfer rule is bounded by three times of the transfer size of the maximally revealing mechanism; moreover, each player reports his true type in the "bridge" component up until the first announcement of the extended direct mechanism $\bar{\mathcal{M}}$. More precisely, for each player i of type t_i we have $m_i^l = t_i$ for each $l = 2, ..., \bar{l}+4$ as long as m_i belongs to $S_i^\infty \hat{W}^\infty(t_i|\bar{\mathcal{M}}^*, \bar{\mathcal{T}})$.

Proposition 2 *Suppose that Assumption 1 holds. Let $\mathcal{M}^*$ be a maximally revealing mechanism with transfer rule bounded by $\hat{\tau}/3$ and $\bar{M}$ be an extended direct mechanism. Then, there exists an augmented mechanism $\bar{\mathcal{M}}^* = ((M_i), g, (\tau_i))_{i\in I}$ such that (a) the transfer rule $\tau_i(\cdot)$ is bounded by $\hat{\tau}$; (b) if the transfer size in $\bar{\mathcal{M}}$ is sufficiently small, then for each $i \in I$, each $t_i \in \bar{T}_i$, and each $m_i \in S_i^\infty \hat{W}^\infty(t_i|\bar{\mathcal{M}}^*, \bar{\mathcal{T}})$, we have $m_i^l = t_i$ for $l = 2, ..., \bar{l}+4$.*

Proof. See Appendix A.2. ■

The proof of Proposition 2 boils down to Lemma 1, which allows us to translate the agents' choice under $\hat{S}^\infty$ in the maximally revealing mechanism $\mathcal{M}^*$ into their choice under $\hat{W}^\infty$ in the augmented mechanism $\bar{\mathcal{M}}^*$. The main difficulty of this translation lies in showing that a strictly dominated message in $\mathcal{M}^*$ corresponds to a *weakly* dominated message in $\bar{\mathcal{M}}^*$ which we will elaborate further below.

Lemma 1 *Suppose that Assumption 1 holds. For any player i of type $t_i \in \bar{T}_i$ and any $l = 0, 1, ..., \bar{l}$, the following two statements, denoted by $P^1(l)$ and $P^2(l)$, hold:*

- *$P^1(l)$: for any $\hat{m}_i \in M_i$, $\hat{m}_i \in \hat{W}_i^l(\hat{\theta}_i(t_i)|\bar{\mathcal{M}}^*)$ implies $\hat{m}_i^0 \in \hat{S}_i^l(\hat{\theta}_i(t_i)|M^*)$;*
- *$P^2(l)$: there is some $\left(m_i^0, ..., m_i^l, m_i^{l+1}\right) \in \times_{k=0}^{l+1} M^k$ such that for every $t_i' \in \bar{T}_i$,*

$$\left(m_i^0, ..., m_i^l, m_i^{l+1}, t_i', t_i, ..., t_i\right) \in \hat{W}_i^l(\hat{\theta}_i(t_i)|\bar{\mathcal{M}}^*).$$

Proof. See Appendix A.2.2. ■

In words, $P^1(l)$ says that in each deletion step of weakly dominated messages, announcing $\hat{m}_i^0$ which is strictly dominated in the maximally revealing mechanism $\mathcal{M}^*$ must result in a weakly dominated message in the augmented mechanism $\bar{\mathcal{M}}^*$. $P^2(l)$ ensures that at least

two inconsistent announcements exist (i.e., there exists some $k > 2$ such that $m_i^k \neq m_i^2$ in message m_i) in a message surviving the previous round of deletion. It follows from $P^1(l)$ of Lemma 1 that $m_i \in \hat{W}_i^{\infty}(\theta_i|\bar{\mathcal{M}}^*)$ implies $\hat{m}_i^0 \in \hat{S}_i^{\infty}(\theta_i|\mathcal{M}^*)$. In fact, $P^1(l)$ will be proved via induction and $P^2(l)$ ensures that the induction argument goes through.

As Lemma 1 establishes the translation of each strictly dominated messages in $\mathcal{M}^*$ into a weakly dominated message in $\bar{\mathcal{M}}^*$, agent i with two strategically distinguishable payoff types must be associated with different set of $\hat{S}_i^{\infty}(\theta_i|\mathcal{M})$ and thereby report distinct messages in m_i^0. Then, Assumption 1 allows us to make use of proper scoring rules to incentivize each player to report his type truthfully in m_i^1 and similarly the additional $\bar{l}+3$ announcements in the "bridge" (i.e., $(m_i^1, \ldots, m_i^{\bar{l}+3}) = (t_i, \ldots, t_i)$) until the 1st announcement in $\bar{\mathcal{M}}$ (i.e., $m_i^{\bar{l}+4} = t_i$) in proving Proposition 2. Like the maximally revealing mechanism $\mathcal{M}^*$ constructed in Lemma 3, we also require that these proper scoring rules be "generic" and construct them in the proof of Lemma 4.[17]

Finally, we explain how we obtain Proposition 1 based on Proposition 2. In particular, we follow the idea of Abreu and Matsushima (1992) and Abreu and Matsushima (1994) in constructing the extended direct mechanism to prove Proposition 1; see Appendix A.3. This extended direct mechanism possesses two important properties. First, like the mechanisms constructed in Abreu and Matsushima (1992) and Abreu and Matsushima (1994), as long as the players are truthful in their first announcement in $\bar{\mathcal{M}}$ (which we guarantee by Proposition 2), the iterative deletion of interim strictly dominated messages implies that they will also truthfully announce their own types "all the way" in each of the subsequent announcements. As a result, we obtain the desirable social outcome in $\bar{\mathcal{M}}^*$. Second, each of the announcements will only get to determine the social alternative with probability $1/K$ where K is the number of announcements which an agent is asked to make in $\bar{\mathcal{M}}$. When K is large, the construction serves to piecemeal the players' incentive to misreport their type and thereby a small transfer suffices to incentivize truth-telling. By Proposition 2, the transfer size in $\bar{\mathcal{M}}^*$ can therefore be made arbitrarily small, as long as we can decrease both the transfer size of $\mathcal{M}^*$ and $\bar{\mathcal{M}}$ arbitrarily. We achieve the latter property with Lemma 3 and the construction of $\bar{\mathcal{M}}$ in the proof of Proposition 1 in the Appendix. Figure 2 summarizes how we structure the proof of Proposition 1.

[17] The generic property ensures that each player has a strict best response against any pure strategy profile of his opponents. This property plays a crucial role in our proof that the message in $P^2(l)$ survives $\hat{W}_i^l$.

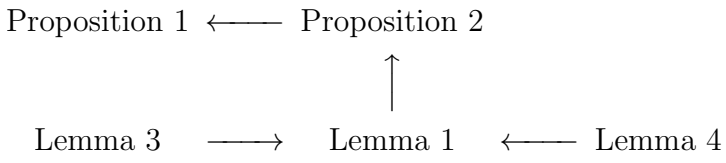

Figure 2: The Diagram of the Proof of Proposition 1

We conclude the section by briefly commenting on the difference between the augmentation established via Proposition 2 and the augmentation in Abreu and Matsushima (1992) and Bergemann and Morris (2009b). First, Abreu and Matsushima (1992) adopt an interim solution concept in both the maximally revealing mechanism and the augmented mechanism, whereas Bergemann and Morris (2009b) adopt an ex post/belief-free solution concept in both the maximally revelation mechanism and the augmented mechanism. Here our augmented mechanism starts by exploiting a "belief-free" solution concept $\hat{W}^\infty$ but switches to an interim perspective (S) once we succeed in eliciting the payoff type information from $\hat{W}^\infty$. Second, both Abreu and Matsushima (1992) and Bergemann and Morris (2009b) work with the solution concept of iterated strict dominance in establishing their virtual implementation results. In contrast, $P^1(l)$ in Lemma 1 shows that each weakly undominated message in the augmented mechanism $\bar{\mathcal{M}}^*$ must announce a strictly undominated message in the maximally revealing mechanism $\mathcal{M}^*$. As Abreu and Matsushima (1994), it is crucial for us to adopt iterative weak dominance because we would like to achieve exact implementation as opposed to virtual implementation in the social alternatives. Due to the two differences, the proofs of Lemma 1 and Proposition 2 are also substantially different from the proof of augmentation in Abreu and Matsushima (1992) or Bergemann and Morris (2009b).

4.4 Proof of Theorem 1

Based on Proposition 1, the proof of Theorem 1 is completed following two further steps. First, Proposition 3 establishes the upper hemicontinuity of the correspondence $S^\infty\hat{W}^\infty$ which is similar to the well known upper hemicontinuity of the interim correlated rationalizable strategies S^∞(see Dekel et al. (2007)).

Proposition 3 *Fix any model $\mathcal{T}$ such that $\bar{\mathcal{T}} \subset \mathcal{T}$ and any mechanism $\mathcal{M}$. Then, for any $t \in \bar{T}$ and any sequence $\{t^n\}_{n=0}^\infty$ in $\mathcal{T}$ such that $t^n \to_p t$, we have $S^\infty\hat{W}^\infty(t^n|\mathcal{M},\mathcal{T}) \subset S^\infty\hat{W}^\infty(t|\mathcal{M},\mathcal{T})$ for any n large enough.*

Proof. See Appendix A.4. ■

Second, Proposition 4 states that there exists a Bayes Nash equilibrium of the game $U\left(\bar{\mathcal{M}}^*, \mathcal{T}\right)$ which survives the iterative deletion of weakly dominated messages.

Proposition 4 *Fix any model $\mathcal{T}$ such that $\bar{\mathcal{T}} \subset \mathcal{T}$ and any mechanism $\mathcal{M}$. Then, there exists an equilibrium σ in the game $U(\mathcal{M}, \mathcal{T})$ such that for any player i of type t_i, we have $\sigma_i(t_i) \in \hat{W}_i^{\infty}(\hat{\theta}_i(t_i)|\mathcal{M})$.*

Proof. See Appendix A.5. ■

Now we are ready to prove Theorem 1 which we restate here for the ease of reference:

Theorem 1. *Suppose that Assumption 1 holds. Then, an SCF $f : \bar{T} \to \Delta(A)$ is continuously implementable with arbitrarily small transfers if and only if it is incentive compatible.*

Proof. We first prove the "if" part. For any $\hat{\tau} > 0$, by Proposition 1, for any $t \in \bar{T}$, there is some mechanism $\mathcal{M}^{\hat{\tau}}$ such that $m \in S^{\infty}\hat{W}^{\infty}\left(t|\mathcal{M}^{\hat{\tau}}, \bar{\mathcal{T}}\right)$ implies that $g(m) = f(t)$.

Now pick any model $\mathcal{T} \supset \bar{\mathcal{T}}$. We show that there exists an equilibrium which continuously implements f on $\bar{\mathcal{T}}$. By Proposition 4, there is an equilibrium σ in the game $U(\mathcal{M}^{\hat{\tau}}, \mathcal{T})$ such that $\sigma_i(t_i) \in \hat{W}_i^{\infty}(\hat{\theta}_i(t_i)|\mathcal{M}^{\hat{\tau}})$ for every type t_i of every player i. Since σ is an equilibrium in $U(\mathcal{M}^{\hat{\tau}}, \mathcal{T})$, $\sigma_{|\bar{T}}$ is an equilibrium in $U\left(\mathcal{M}^{\hat{\tau}}, \bar{\mathcal{T}}\right)$. Now, pick any sequence $\{t^n\}_{n=0}^{\infty}$ such that $t^n \to_p t$. By Proposition 3, $S^{\infty}\hat{W}^{\infty}\left(t^n|\mathcal{M}^{\hat{\tau}}, \mathcal{T}\right) \subset S^{\infty}\hat{W}^{\infty}\left(t|\mathcal{M}^{\hat{\tau}}, \bar{\mathcal{T}}\right)$ for any n large enough. Moreover, $\sigma(t^n) \in \hat{W}^{\infty}(\hat{\theta}(t^n)|\mathcal{M}^{\hat{\tau}})$. Since Θ is finite, for any n large enough, we have $\hat{\theta}(t^n) = \hat{\theta}(t)$, it follows that $\sigma(t) \in S^{\infty}\hat{W}^{\infty}\left(t^n|\mathcal{M}^{\hat{\tau}}, \bar{\mathcal{T}}\right)$. Thus, by Proposition 1, we have $(g \circ \sigma)(t^n) = f(t)$ for any n large enough.

The "only-if" part is proved as follows: Assume that the SCF f is continuously implementable with arbitrarily small transfers. Then, for any $\hat{\tau} > 0$, there is a mechanism $\mathcal{M}^{\hat{\tau}}$ and a Bayes Nash equilibrium σ in $U(\mathcal{M}^{\hat{\tau}}, \bar{\mathcal{T}})$ such that, for any $t \in \bar{T}$,

$$\begin{aligned} (g \circ \sigma)(t) &= f(t); \\ \tau(\sigma(t)) &< \hat{\tau}. \end{aligned} \tag{6}$$

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

$$V_i((\sigma_i, \sigma_{-i}), t_i) \geq V_i((m_i', \sigma_{-i}), t_i). \tag{7}$$

See Section 3.2 for the notation $V_i((\sigma_i, \sigma_{-i}), t_i)$. Then, by (6) and (7), the truth-telling is a Bayes Nash equilibrium in the incomplete information game induced by the direct mechanism

$(\bar{T}, f)$. That is, for any $t_i, t_i' \in \bar{T}_i$,

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

Since $\tau(\cdot)$ is bounded by $\hat{\tau}$ and $\hat{\tau}$ can be arbitrarily small, we have

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

That is, the SCF f is incentive compatible. ■

5 Discussion

We first discuss the implication for one crucial assumption that each player always knows his own payoff type while we perturb the model slightly (Section 5.1). We next discuss other works in the literature which also propose different notions of locally robust implementation as well as the different implications obtained in those related papers (Sections 5.2, 5.3, and 5.4).

5.1 No Knowledge about Payoff Types

We motivate our exercise as studying locally robust implementation which parallels the study of globally robust implementation in Bergemann and Morris (2005). In this vein, we view our assumption on payoff knowledge as an instrumental one. More precisely, our goal is to understand what the notion of continuous implementation entails as a locally robust implementation notion in comparison with its global counterpart, keeping other aspects of the two notions equal. We also verify the connection through establishing Observation 1 which shows that any ex post implementable SCF is continuously implementable.

In this section, we discuss a contrasting situation in which the agents know *nothing* about their own payoff types. Recall that the solution concept $S^\infty \hat{W}^\infty$ in Section 4.3 (and specifically how the definition of $\hat{W}_i^\infty(\hat{\theta}_i(t_i)|\mathcal{M})$) is based upon the assumption that each agent knows his own payoff type. If an agent does not know his own payoff type, then he needs to consider his every possible payoff type in deleting weakly dominated messages. For

each integer $l \geq 0$, we inductively define:

$$\tilde{W}_i^{l+1}(\mathcal{M}) = \left\{ m_i \in \tilde{W}_i^l(\mathcal{M}) \;\middle|\; \begin{array}{l} \exists \alpha_i \in \Delta(M_i) \text{ s.t. } u_i(g(\alpha_i, m_{-i}), \theta) + \tau_i(\alpha_i, m_{-i}) \\ \geq u_i(g(m_i, m_{-i}), \theta) + \tau_i(m_i, m_{-i}) \\ \text{for any } \theta \in \Theta, \text{ any } m_{-i} \in \tilde{W}_{-i}^l(\mathcal{M}) \text{ and a strict inequality} \\ \text{holds for some } m_{-i} \in \tilde{W}_{-i}^l(\mathcal{M}) \text{ and some } \theta \in \Theta \end{array} \right\}.$$

Finally, we say that $\tilde{W}_i^\infty(\mathcal{M}) \equiv \bigcap_{l \geq 0}^l \tilde{W}_i(\mathcal{M})$ is the set of messages surviving the iterative deletion of *weakly* dominated messages for agent i. $S_i^0 \tilde{W}^\infty(t_i|\mathcal{M}, \mathcal{T}) = \tilde{W}_i^\infty(\mathcal{M})$ and for each integer $l \geq 1$, we inductively define $m_i \in S_i^{l+1} \tilde{W}^\infty(t_i|\mathcal{M}, \mathcal{T})$ if and only if there does not exist $\alpha_i \in \Delta(M_i)$ such that

$$V_i((\alpha_i, \nu_{-i}), t_i) > V_i((m_i, \nu_{-i}), t_i)$$

for all conjecture $\nu_{-i} : T_{-i} \to M_{-i}$ and all $t_{-i} \in T_{-i}$ such that $\nu_{-i}(t_{-i}) \in S_{-i}^l \tilde{W}^\infty(t_{-i}|\mathcal{M}, \mathcal{T})$ for each t_{-i} where $S_{-i}^l \tilde{W}^\infty(t_{-i}|\mathcal{M}, \mathcal{T}) \equiv \prod_{j \neq i} S_j^l \tilde{W}^\infty(t_j|\mathcal{M}, \mathcal{T})$. Let $S^\infty \tilde{W}^\infty$ denote the set of message profiles which survive the iterative deletion of weakly dominated messages followed by the iterative removal of interim strictly dominated messages, i.e.,

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

Finally, we define $S^\infty \tilde{W}^\infty(t|\mathcal{M}, \mathcal{T}) = \prod_{i \in I} S_i^\infty \tilde{W}^\infty(t_i|\mathcal{M}, \mathcal{T})$. We propose the following definition of implementation in $S^\infty \tilde{W}^\infty$. In contrast to Definition 6, the following definition allows for transfers of any size to be used on and off the solution concept $S^\infty \tilde{W}^\infty$.

Definition 8 *An SCF $f : \bar{T} \to \Delta(A)$ is fully implementable in $S^\infty \tilde{W}^\infty$ with transfers if there exists a mechanism $\mathcal{M}$ such that $g(m) = f(t)$ for every $m \in S^\infty \tilde{W}^\infty(t|\mathcal{M}, \bar{\mathcal{T}})$ and every $t \in \bar{T}$.*

The following definition is adapted from Bergemann et al. (2011) by allowing for additional transfers to be made in rationalizable implementation.

Definition 9 *An SCF $f : \bar{T} \to \Delta(A)$ is fully implementable in rationalizable strategies with transfers if there exists a mechanism $\mathcal{M}$ such that $\tilde{W}_i^\infty(\mathcal{M}) = M_i$ for every $i \in \mathcal{I}$ and f is fully implementable in $S^\infty \tilde{W}^\infty$ by the mechanism $\mathcal{M}$.*

We introduce Maskin monotonicity as a condition for SCFs:

Definition 10 *An SCF f satisfies **Maskin monotonicity** if, for every pair of states t and t' with $f(t) \neq f(t')$, there exist some agent $i \in \mathcal{I}$ and some lottery $\alpha \in \Delta(A)$ such that*

$$u_i(f(t), \hat{\theta}(t)) \geq u_i(\alpha, \hat{\theta}(t);$$

and

$$u_i(\alpha, \hat{\theta}(t')) > u_i(f(t), \hat{\theta}(t').$$

The following result shows that implementation in $S^\infty\tilde{W}^\infty$ becomes a demanding requirement, even if we allow for additional transfers of any size to be imposed on and off the solution $S^\infty\tilde{W}^\infty$.

Proposition 5 *If an SCF f is fully implementable in $S^\infty\tilde{W}^\infty$ with transfers, it satisfies Maskin-monotonicity.*

Proof. We prove this claim by contradiction. Suppose that some mechanism $\mathcal{M}$ implements f in $S^\infty\tilde{W}^\infty$ with transfers. Define $\mathcal{M}'$ to be a restricted mechanism such that for all $i \in I$, $M_i' \equiv \tilde{W}_i^\infty(\mathcal{M})$ and g and τ_i are all restricted to $M' = \times_{i\in I} M_i'$. Then, that $\mathcal{M}$ implements f in $S^\infty\tilde{W}^\infty$ implies that $\mathcal{M}'$ implements f in rationalizable strategies with transfers. By Bergemann et al. (2011), f satisfies Maskin-monotonicity. ■

It follows from Proposition 5 and Corollary 1 that we can construct an SCF which is implementable in $S^\infty\hat{W}^\infty$ with arbitrarily small transfers but not implementable in $S^\infty\tilde{W}^\infty$ with transfers. This demonstrates that the situation where players know their own payoff types (as formulated in $S^\infty\hat{W}^\infty$) drastically differs from the one where the players do not know their payoff types (as formulated in $S^\infty\tilde{W}^\infty$). We construct such an example in Appendix A.6 to demonstrate this point formally. In particular, the example presents a complete-information environment with an SCF which is incentive compatible but not Maskin-monotonic even if we allow for transfers of any size.

In addition, virtual implementation by Abreu and Matsushima (1992) is achieved in the solution concept of rationalizability. Hence, the "nearby" SCF which Abreu and Matsushima (1992) implements must be Maskin-monotonic. In contrast, we construct an SCF in Appendix A.6 which is (1) not Maskin-monotonic even if we add transfers (large or small) to the SCF outcomes; and yet (2) it is implementable in $S^\infty\hat{W}^\infty$ with arbitrarily small transfers by our result. Thus, the "nearby" SCF which we implement with small transfers need not be Maskin-monotonic.

5.2 Other papers on continuous implementation

Oury and Tercieux (2012) is the first to propose the notion of continuous implementation as a strengthening of *partial* implementation. Theorem 3 of OT (which also implies Theorems 1 and 2 of OT) shows that strict *interim rationalizable monotonicity* is a necessary condition for strict continuous implementation. As strict interim rationalizable monotonicity implies strict Bayesian monotonicity, which is a well-known necessary and "almost sufficient" condition for full implementation in Bayes Nash equilibrium, a central message of OT is that while continuous implementation sounds weaker than full implementation, it is strong enough to obtain full implementation.

There are four differences between our paper and OT. First, as we mentioned in Section 2, our paper maintains the payoff knowledge assumption, whereas OT do not. Specifically, footnotes 8 and 17 in Oury (2015) remarked that for the main results of OT and Oury (2015) to hold, it is crucial that the players entertain some doubt about their own payoff type in the perturbed model nearby the benchmark.[18] Here we follow Bergemann and Morris (2005) in assuming that each player knows his payoff type and this knowledge as well as the common knowledge of utility functions is maintained even when we perturb their belief hierarchies. In contrast to OT, we obtain permissive continuous implementation results with the payoff knowledge assumption and our results cover SCFs which are not (Bayesian) monotonic.

Second, OT focus on "strict" continuous implementation rather than continuous implementation in their Theorems 1-3.[19] They show that "strict" continuous implementation generates its necessary conditions, which are tightly connected to that of full implementation. To dispense with "strictness" in their result, OT obtain the result of (non-strict) continuous implementation with an additional assumption of costly messages. Without the payoff

[18] In proving their Theorem 1, OT construct a sequence of types nearby the complete-information benchmark with the following property: the sequence of types assigns increasingly more probability on the payoff type θ' and vanishing probability on payoff type θ. The belief of these types (of agent i) are constructed so that they all believe that the opponents have a fixed type profile t_{-i} which has complete information about state θ, regardless of these types' belief about Θ. This is to ensure that the opponents of the type profile t_{-i} play the equilibrium message profile under state θ to start the contagion. When each player knows his own payoff type, however, it is not possible that player i assigns probability one that t_{-i} has complete information about θ yet player i also assigns increasingly more probability on θ' in which his opponents know $\theta'_{-i} \neq \theta_{-i}$.

[19] A related difference is that a strict Bayes Nash equilibrium constitutes, by definition, a pure strategy profile in the benchmark model, whereas we do not impose this requirement in part (i) of OT's Definition 2.

knowledge assumption, we know of no necessary condition for continuous implementation, with or without transfer. Moreover, we only know of one sufficient condition for continuous implementation, which is achieved by rationalizable implementation in a finite mechanism (if-part of OT's Theorem 4). In contrast to our result, rationalizable implementation in a finite mechanism (like implementation in $S^\infty\tilde{W}^\infty$) requires that the SCF satisfy Maskin-monotonicity, and in fact, a stronger condition called Maskin-monotonicity* (Bergemann et al. (2011)). To our knowledge, it was only proved recently in Chen et al. (2021) that every Maskin-monotonic* SCF can be implemented in rationalizable strategies in a finite mechanism in a complete information environment with lotteries and (off-path) transfers.

Third, OT's result holds whether the planner can use transfers or not, although their Theorem 4 builds on the assumption that sending messages incurs a small cost to the players and their preferences are quasilinear in the cost. In contrast, we follow the classical mechanism design literature in assuming that the messages are cheap talk and the planner can make use of arbitrarily small transfers, on and off the equilibrium. In other words, transfers are part of the planner's instrument in our setup whereas the cost of sending messages is part of the environmental constraint in OT.

Finally, we impose Assumption 1 on the benchmark model $\bar{\mathcal{T}}$, whereas OT consider an arbitrary (finite) benchmark model. The notion of continuous implementation builds upon the planner's uncertainty about the higher-order beliefs of the players. In this vein, the planner might also be concerned about having an alternative benchmark model where the players' hierarchies of beliefs lie in the neighborhood of the benchmark which she postulates. Indeed, provided that each player has strategically distinguishable payoff types (e.g., in Bergemann and Morris (2009a)), Assumption 1 holds for generic beliefs. As a result, for every benchmark model, there is a "nearby" benchmark model in which Assumption 1 holds and our result applies. We therefore can always choose a benchmark model satisfying Assumption 1 if the planner cannot distinguish these two nearby benchmark models due to the lack of full knowledge about the players' higher-order beliefs.

Oury (2015) also obtain a full characterization of continuous implementation in finite mechanisms in terms of rationalizable implementation in finite mechanisms. Instead of assuming that sending message is slightly costly as in Theorem 4 of OT, Oury (2015) assumes that the planner has some doubt on the payoffs of the outcomes and wants his prediction to be robust when these payoffs are close but not exactly equal to those in the initial model.

In our paper, on the contrary, we obtain a permissive result for continuous implementation in the setting of Bergemann and Morris (2005) where agents know their payoff type, the utility functions are common knowledge, and the planner maintains this payoff knowledge throughout.

Chen, Muller-Frank, Pai (2022, hereafter CMP) characterize when a social choice function is *truthfully continuously implementable*, i.e., continuously implementable using game forms corresponding to *direct revelation mechanisms* for the benchmark model. CMP show that whether the restriction to direct revelation mechanisms entails a loss of generality hinges on the formalization of the notion of "nearby types". In particular, when "nearby types" can have different higher-order beliefs (as in the case with the product topology), truthful continuous implementation is equivalent to requiring that the social choice function is implementable in unique rationalizable strategies in the initial model; moreover, some SCF is continuously implementable only with indirect mechanisms. Unlike OT and Oury (2015), CMP assume that the players' utility functions are common knowledge; however, unlike our paper, CMP do not assume that each player knows his own payoff type.

5.3 Comparisons with JMM

Jehiel et al. (2012) (hereafter, JMM) also define the notion of locally robust implementation which captures the idea that the planner may know the players' beliefs well, though not perfectly. Both JMM and our paper weaken the notion of global robust implementation due to Bergemann and Morris (2005) to a notion of locally robust implementation and both papers maintain the payoff knowledge assumption. JMM also allow transfers to be used in equilibrium and their impossibility result holds regardless of the size of the equilibrium transfers.

To focus on the most essential differences, we adapt JMM's notion of locally robust implementation to our setting. The benchmark model in JMM is a tuple $(\Theta_i, \pi_i^*)_{i\in I}$ where $\pi_i^* : \Theta_i \to \Delta(\Theta_{-i})$. In our terminology, their benchmark model is a model $(\bar{T}_i, \bar{\theta}_i, \bar{\pi}_i)_{i\in I}$ where $\bar{\theta}_i$ is the identity mapping from $\bar{T}_i$ to Θ_i, and $\bar{\pi}_i = \pi_i^*$. Let $B_\varepsilon(\pi_i^*(\theta_i))$ denote the open ε-balls in $\Delta(\Theta_{-i})$ which is endowed with the Euclidean topology.[20] To model the local uncertainty, they consider a larger (uncountably infinite) model $\mathcal{T}^\varepsilon = (T_i^\varepsilon, \hat{\theta}_i^\varepsilon, \pi_i^\varepsilon)$ which *includes all ε-perturbed beliefs*, i.e., $T_i^\varepsilon \subset \Theta_i \times \Delta(\Theta_{-i})$ and $\theta_i \times B_\varepsilon(\pi_i^*(\theta_i)) \subset T_i^\varepsilon$ for every θ_i.

[20] JMM allow for an infinite set of types, $\Theta_i = [0,1]^{d_i}$ and endow $\Delta(\Theta_{-i})$ with the total variation norm.

Moreover, for $t_i = (\theta_i, p_i)$ with $p_i \in \Delta(\Theta_{-i})$, $\pi_i(t_i) \in \Delta\left(T^{\varepsilon}_{-i}\right)$ is the unique measure with the marginal distribution on Θ_{-i} equal to p_i and $\pi_i(t_i)(\{(\theta_{-i}, \pi^*_i(\theta_{-i})) : \theta_{-i} \in \Theta_{-i}\}) = 1$. That is, agent i could have different beliefs about $-i$'s payoff types, but i believes with probability one that $-i$'s beliefs are specified by π^*_{-i}. Clearly, $\mathcal{T}^{\varepsilon} \supset \bar{\mathcal{T}}$.

JMM consider a problem where there are only two agents and two social alternatives (i.e., we can set $\Delta(A) = [0,1]$). The planner wants to implement an allocation function $q : \Theta \to [0,1]$. Let $v_i : \Theta \times \{0,1\} \to \mathbb{R}$ denote agent i's smooth interdependent value function. An allocation function q is said to be *locally robust implementable* if there exist $\varepsilon > 0$, a model $\mathcal{T}^{\varepsilon}$ which includes all ε-perturbed beliefs, and a payment function $p : T^{\varepsilon} \to \mathbb{R}^I$, such that the direct revelation mechanism (q,p) is incentive compatible on T^{ε}, i.e.

$$E_{\pi_i}[v_i(\theta)q(\theta) - p_i(t)] \geq E_{\pi_i}[v_i(\theta)q(\theta') - p_i(t')]$$

for all $t_i = (\theta_i, \pi_i)$, $t'_i = (\theta'_i, \pi'_i) \in T^{\varepsilon}_i$, where $\theta = (\theta_i, \theta_{-i})$, $\theta' = (\theta'_i, \theta_{-i}), t = (\theta_i, \pi_i, \theta_{-i}, \pi_{-i})$, and $t' = (\theta'_i, \pi'_i, \theta_{-i}, \pi_{-i})$. JMM show that no "regular" allocation function is locally robust implementable in generic settings with quasi-linear utility, interdependent and bilinear values, and multi-dimensional payoff types.[21]

There are three basic differences between JMM's notion of locally robust implementation and our notion of continuous implementation with payoff knowledge. First, the notion of continuous implementation allows us to make use of one particular (possibly mixed-strategy) equilibrium in an indirect implementing mechanism, while JMM's notion of locally robust implementation invokes the truthful equilibrium in a direct revelation mechanism (q,p). Second, in order to construct a finite implementing mechanism, we work with a finite benchmark model and invoke Proposition 4 to obtain a desirable equilibrium which continuously implements the SCF. Proposition 4 requires (1) the existence of a Bayesian Nash equilibrium for which we, like OT, work with a countable nearby model; and (2) a finite payoff type space so that the process $\hat{W}^{\infty}$ terminates in finitely many steps. In contrast, JMM consider a particular uncountably infinite benchmark model $\mathcal{T}^{\varepsilon}$ which includes all ε-perturbed beliefs. It is unclear whether our argument can be extended to this case. Third, JMM's notion of locally robust implementation is stated with respect to the direct revelation mechanism, which is defined with respect to the specific model $\mathcal{T}^{\varepsilon}$. In contrast, the notion of continuous implementation requires that our indirect implementing mechanism possess a good equilibrium for

[21] We refer the readers to JMM's paper for the formal definitions of regularity of allocation functions and bi-linearity of value functions.

every nearby model $\mathcal{T} \supset \bar{\mathcal{T}}$. For instance, while a countable nearby model $\mathcal{T}$ cannot include all ε-perturbed beliefs, it can include a countable dense subset of ε-perturbed beliefs. In any such nearby model $\mathcal{T}$, our implementing mechanism possesses a good equilibrium which continuously implement the SCF.

5.4 Other Related Papers

Meyer-ter-Vehn and Morris (2011) (hereafter, MM) also propose a notion of locally robust implementation but focus on robust full (as opposed to partial) implementation. MM prove that a mechanism that robustly implements optimal outcomes in a one-dimensional supermodular environment continues to robustly implement ε-optimal outcomes in all close-by environments. They adopt the notion of robust full implementation due to Bergemann and Morris (2009a) which necessitates ex post incentive compatibility as a requirement. It follows from Observation 1 that their notion is strictly stronger than our notion of continuous implementation with payoff knowledge. Moreover, although our Assumption 1 can be (generically) satisfied as long as the maximally revealing mechanism induces a non-trivial partition over the payoff type space, MM's Theorem 1 builds upon on the three assumptions which, according to Theorem 1 of Bergemann and Morris (2009a), imply that the partition induced by the maximally revealing mechanism is the finest one.

The current paper is developed from our earlier unpublished paper, Chen et al. (2016), although they differ in two important ways. First, Chen et al. (2016) focused on full implementation in the iterated deletion of *interim* weakly dominated messages. The current paper studies continuous implementation. Second, Chen et al. (2016) also prove a result on continuous implementation. However, this earlier result focuses on private-value environments, whereas our current result covers interdependent-value environments. In interdependent-value environments, we adopt the novel solution concept $S^{\infty}\hat{W}^{\infty}$ and our argument is built crucially on the maximally revealing mechanism due to Bergemann and Morris (2009a) and the augmentation step in Section 4.3.2. In contrast, the continuous implementation result in Chen et al. (2016) focuses on private-value environments and require none of these components. Indeed, when the benchmark model is the one with complete information, the main result of Chen et al. (2016) amounts to Corollary 1 in the current paper and can readily be proved by invoking the mechanism constructed in Abreu and Matsushima (1994). For a more delicate case with complete information which we document in Corollary 2, we can

only establish continuous implementation using our current implementing mechanism.

6 Conclusion

We show that continuous implementation with payoff knowledge is as permissive as it can be when small transfers are allowed and Assumption 1 is satisfied. In such situations, all we need is (interim) incentive compatibility, which is, by the revelation principle, a necessary condition for interim implementation. This exhibits a stark contrast with Bergemann and Morris (2005) who show that their global robustness amounts to ex post implementability as well as Oury and Tercieux (2012) who show that (strict) continuous implementation (without payoff knowledge) is tightly connected to full implementation in rationalizable strategies. We also compare our result with other existing results which are based on different notions of locally robust implementation. The contrasts exemplify the substantive difference between local robustness exercises which also involve payoff perturbations (Oury and Tercieux (2012) and Oury (2015)), and those which do not perturb common knowledge of the utility functions (this paper).

Our permissive result is based on constructing a complex indirect mechanism which leverages on the insight of Bergemann and Morris (2009a) and Abreu and Matsushima (1992). We view the result as a step toward understanding the subtleties of continuous implementation as a notion of locally robust implementation. In this regard, we focus on studying the scope of implementability and proving our result in a general quasilinear social choice environment. Can we find "simpler/more practical" mechanisms which continuously implement specific incentive compatible SCFs in specific settings?[22] Or, should we be concerned, if the formal framework with which we work necessitates the use of a complicated/unrealistic mechanism to prove the result? Last but not least, is it possible to obtain permissive results under an intermediate robustness notion, between the belief-free and local robustness

[22]Indeed, it is known that ex post implementability can be achieved with "simpler mechanisms" in specific setups such as the efficient allocation rule in auction settings; see Dasgupta and Maskin (2000), Bergemann and Morris (2009a), and Chung and Ely (2019). Here, what we mean by simpler mechanisms are direct mechanisms. More specifically, Chung and Ely (2019) consider the generalized VCG mechanism and Dasgupta and Maskin (2000) consider a version of the VCG mechanism in which each bidder announces a bidding "function," which does depend on the payoff type profile of other players. On the contrary, Bergemann and Morris (2009a) consider general direct mechanisms.

approach of this paper, which does not rely on properties like Assumption 1? These are important questions for future research.

A Appendix

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

A.1 Maximally Revealing Mechanism and Scoring Rules

In this section, we construct a generic maximally revealing mechanism and generic scoring rules which will be the building blocks of our augmented mechanism. We first prove a lemma which will be used to prove Lemmas 3 and 4.

A.1.1 A Preliminary Lemma

Let $\bar{r} > 0$ and $\mathcal{M} = ((M_i), g)_{i \in I}$ denote a mechanism with zero transfer (i.e., $\tau_i(m) = 0$ for every $m \in M$ and $i \in I$). Fix a player i. For any $t_i \in \bar{T}_i$, any $\sigma_{-i} : T_{-i} \to M_{-i}$, and any messages m_i and m_i' in M_i with $m_i \neq m_i'$, we define the set

$$C_{t_i,\sigma_{-i}}^{\mathcal{M},\bar{r}}(m_i, m_i') \equiv \left\{ \tau_i \in [-\bar{r}, \bar{r}]^{|M|} : V_i((m_i, \sigma_{-i}), t_i) \neq V_i((m_i', \sigma_{-i}), t_i) \right\}$$

where we recall that

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

In words, $C_{t_i,\sigma_{-i}}^{\mathcal{M},\bar{r}}(m_i, m_i')$ is the set of transfer rules defined on M which is bounded by $\bar{r}$ and type t_i is not indifferent between the pair of messages m_i and m_i' under conjecture σ_{-i}. Define $C_i^{\mathcal{M},\bar{r}} \equiv \bigcap_{t_i,\sigma_{-i}} \bigcap_{m_i \neq m_i'} C_{t_i,\sigma_{-i}}^{\mathcal{M},\bar{r}}(m_i, m_i')$.

Lemma 2 *Let $\mathcal{M} = ((M_i), g)_{i \in I}$ denote a mechanism with zero transfer. Then, the complement of $C_i^{\mathcal{M},\bar{r}}$ has measure zero in $\mathbb{R}^{|M|}$.*

Proof. Observe that the complement of $C_{t_i,\sigma_{-i}}^{\mathcal{M},\bar{r}}(m_i, m_i')$ is the set of solutions of a linear equation in $\mathbb{R}^{|M|}$. Hence, the complement of $C_{t_i,\sigma_{-i}}^{\mathcal{M},\bar{r}}(m_i, m_i')$ is a hyperplane of $\mathbb{R}^{|M|}$ with dimension lower than $|M|$ and thus has measure zero (see p. 52 of Rudin (1987)). Since there are only finitely many types in $\bar{T}_i$, functions $\sigma_{-i} : \bar{T}_{-i} \to M_{-i}$, and messages in M_i, it follows that $C_i^{\mathcal{M},\bar{r}}$ also has measure zero. ■

A.1.2 A Generic Maximally Revealing Mechanism

First, Lemma 3 shows that we can add (arbitrarily) small transfers to the maximally revealing mechanism $\mathcal{M}^{\text{BM}}$ in Bergemann and Morris (2009b) so that it satisfies a generic condition, namely, for any type and against any degenerate belief over the other players' announcements (i.e., a mapping $\sigma_{-i}^* : \bar{T}_{-i} \to M_{-i}^*$), any two distinct messages must result in distinct payoffs. We call such a mechanism $\mathcal{M}^*$ a generic maximally revealing mechanism which we fix hereafter.

Lemma 3 *For any $\tilde{\tau} > 0$, there exists a maximally revealing mechanism $\mathcal{M}^* = ((M_i^*), g^*, (\tau_i^*))_{i \in I}$ with the following properties: for each player i,*

(a) $|\tau_i^*(\cdot)|$ is bounded by $\tilde{\tau}$;

(b) for any $t_i \in \bar{T}_i$, any $m_i, m_i' \in M_i^*$ with $m_i \neq m_i'$, and any $\sigma_{-i}^* : \bar{T}_{-i} \to M_{-i}^*$, we have $V_i(\left(m_i, \sigma_{-i}^*\right), t_i) \neq V_i(\left(m_i', \sigma_{-i}^*\right), t_i)$;

(c) $\hat{S}_i^\infty(\theta_i | \mathcal{M}^*) \cap \hat{S}_i^\infty(\theta_i' | \mathcal{M}^*) = \varnothing$ if $\theta_i \not\sim \theta_i'$.

Proof. Recall that $\mathcal{M}^{\text{BM}} = (M^*, g^*)$ is the maximally revealing mechanism proposed by Bergemann and Morris (2009b). Pick some $\bar{r} < \tilde{\tau}$. By Lemma 2, the complement of $C_i^{\mathcal{M}^{\text{BM}}, \bar{r}}$ has measure zero in $\mathbb{R}^{|M^*|}$. For any transfer rule $(\tau_i)_{i \in I}$ with $\tau_i : M^* \to \mathbb{R}$, denote by $\mathcal{M}^{\text{BM}}(\tau) = \left((M_i^*)_{i \in I}, g^*, (\tau_i)_{i \in I}\right)$ the mechanism which has the same sets of messages and outcome function as the maximally revealing mechanism $\mathcal{M}^{\text{BM}}$ but is augmented by the transfer rule $(\tau_i)_{i \in I}$. Fix any player i. Define

$$C_i = \left\{ \tau_i \in \mathbb{R}^{|M^*|} : \hat{S}^\infty\left(\theta_i | \mathcal{M}^{\text{BM}}(\tau)\right) \cap \hat{S}^\infty\left(\theta_i' | \mathcal{M}^{\text{BM}}(\tau)\right) = \varnothing \text{ whenever } \theta_i \not\sim \theta_i' \right\}.$$

It follows that C_i is a nonempty open set in $\mathbb{R}^{|M^*|}$. Therefore, $C_i \cap C_i^{\mathcal{M}^{\text{BM}}, \bar{r}}$ has positive measure in $\mathbb{R}^{|M^*|}$. Thus, we can find a transfer rule $\tau_i^* \in C_i \cap C_i^{\mathcal{M}^{\text{BM}}, \bar{r}}$. Then, $\mathcal{M}^* = ((M_i^*)_{i \in I}, g^*, (\tau_i)_{i \in I})$ is the desired maximally revealing mechanism. ■

A.1.3 Generic Scoring Rules

Second, the transfer rule in the augmented mechanism consists of a number of proper scoring rules. We prove Lemma 4 below to construct these proper scoring rules. For ease of stating the lemma, denote by Ψ_i^2 the partition over $\bar{T}_i$ jointly induced by $\chi_i^1(\cdot)$ and Ψ_i^1, i.e., $\Psi_i^2 =$

$\left\{\psi_i^2(t_i) : t_i \in \bar{T}_i\right\}$ in which for any types t_i and t_i' in $\bar{T}_i$, we have $t_i' \in \psi_i^2(t_i)$ if and only if $\chi_i^1(t_i) = \chi_i^1(t_i')$ and t_i' belongs to $\psi_i^1(t_i)$. It follows from Assumption 1 that Ψ_i^2 is the finest partition over $\bar{T}_i$, namely that $\Psi_i^2 = \left\{\{t_i\} : t_i \in \bar{T}_i\right\}$. Hence, $\bar{\pi}_i(t_i)$ can be identified with the belief over Ψ_{-i}^2 which we denote by $\chi_i^2(t_i)$.

Indeed, Condition (c) in Lemma 4 says that for $k = 0, 1, 2$, if player i's opponents report the atom in Ψ_{-i}^k which contains their true types, by the transfer rule d_i^k, each player i must report truthfully his belief over Ψ_{-i}^k. As in Lemma 3, Condition (b) in Lemma 4 says that the proper scoring rules are generic so that against any $\sigma_{-i} : \bar{T}_{-i} \to \bar{T}_{-i}$, there is a unique best response.

Lemma 4 *Suppose that Assumption 1 holds. For any $\tilde{\tau} > 0$, any player $i \in I$, and any $k = 0, 1, 2$, there exist $\gamma > 0$ and a function $d_i^k : \bar{T}_i \times \Psi_{-i}^k \to \mathbb{R}$, satisfying the following properties:*

(a) $|d_i^k|$ is bounded by $\tilde{\tau}/\left(\bar{l}+5\right)$;

(b) for any $t_i \in \bar{T}_i$, t_i' and t_i'' in $\bar{T}_i$ with $t_i' \neq t_i''$, and $\sigma_{-i} : \bar{T}_{-i} \to \bar{T}_{-i}$, we have

$$\left| \sum_{t_{-i} \in \bar{T}_{-i}} \left[d_i^k\left(t_i', \sigma_{-i}(t_{-i})\right) - d_i^k\left(t_i'', \sigma_{-i}(t_{-i})\right) \right] \bar{\pi}_i(t_i)[t_{-i}] \right| > \gamma; \tag{8}$$

(c) for every pair of types t_i' and t_i in $\bar{T}_i$ with $\chi_i^k(t_i') \neq \chi_i^k(t_i)$, we have

$$\sum_{t_{-i} \in \bar{T}_{-i}} \left[d_i^k\left(t_i, \psi_{-i}^k(t_{-i})\right) - d_i^k\left(t_i', \psi_{-i}^k(t_{-i})\right) \right] \bar{\pi}_i(t_i)[t_{-i}] > \gamma. \tag{9}$$

Proof. Fix any player i and $k = 0, 1, 2$. We first prove the existence of transfer rules d_i^k which satisfies Condition (c). Consider a mechanism $\mathcal{M} = ((M_j)_{j \in I}, g)$ with zero transfer such that $M_i = \bar{T}_i$ and $M_j = \Psi_j^k$ for every $j \neq i$ and moreover, for some fixed outcome a, we set $g(m) = a$ for every $m \in M$. Pick some $\bar{r} < \tilde{\tau}/\left(\bar{l}+5\right)$. By Lemma 2, we note that $\mathbb{R}^{|M|} \backslash C_i^{\mathcal{M},\bar{r}}$ has measure zero in $\mathbb{R}^{|M|}$. Define

$$D_i^k = \left\{ d_i \in [-\bar{r}, \bar{r}]^{|M|} : \begin{array}{c} \sum_{t_{-i} \in \bar{T}_{-i}} \left[d_i\left(t_i, \psi_{-i}^k(t_{-i})\right) - d_i\left(t_i', \psi_{-i}^k(t_{-i})\right) \right] \bar{\pi}_i(t_i)[t_{-i}] > 0 \\ \text{whenever } \chi_i^k(t_i') \neq \chi_i^k(t_i) \end{array} \right\}.$$

Since each proper scoring rule defined on $\bar{T}_i \times \Psi_{-i}^k$ belongs to D_i^k, it follows that D_i^k is nonempty (and open in $\mathbb{R}^{|M|}$).[23] Therefore, $D_i^k \cap C_i^{\mathcal{M},\bar{r}}$ has positive measure in $\mathbb{R}^{|\bar{T}|}$. Thus,

[23] For example, a proper quadratic scoring rule for $k = 2$ can be defined as $2\bar{\pi}_i(t_i)[t_{-i}] - \bar{\pi}_i(t_i) \cdot \bar{\pi}_i(t_i)$ where "$\cdot$" stands for the inner product of two vectors.

we can find a transfer rule $d_i^k \in D_i^k \cap C_i^{\mathcal{M},\bar{r}}$ which satisfies (9). Then, d_i^k satisfies Conditions (a) and (b) since $d_i^k \in C_i^{\mathcal{M},\bar{r}}$ and d_i^k satisfies Condition (c) since $d_i^k \in D_i^k$. ■

A.2 Proof of Proposition 2

Proposition 2: Suppose that Assumption 1 holds. Let $\mathcal{M}^*$ be a maximally revealing mechanism with transfer rule bounded by $\hat{\tau}/3$ and $\bar{\mathcal{M}}$ an extended direct mechanism. Then, there exists an augmented mechanism $\bar{\mathcal{M}}^* = ((M_i), g, (\tau_i))_{i\in I}$ *such that (a) the transfer rule* $\tau_i(\cdot)$ *is bounded by* $\hat{\tau}$*; (b) if the transfer size in* $\bar{\mathcal{M}}$ *is sufficiently small, then for each* $i \in I$*, each* $t_i \in \bar{T}_i$*, and each* $m_i \in S_i^\infty \hat{W}^\infty(t_i|\bar{\mathcal{M}}^*, \bar{\mathcal{T}})$*, we have* $m_i^l = t_i$ *for* $l = 2, ..., \bar{l}+4$.

The proof of Proposition 2 is divided into three main steps. First, we specify the transfer rule and parameters in $\bar{\mathcal{M}}^*$ which make use of Lemma 4. We then turn to prove Lemma 1 in Section 4.3.2. Finally, we make use of Lemma 1 to prove Proposition 2.

A.2.1 Choice of Parameters for the Augmented Mechanism

Equipped with the functions constructed in Lemma 4, we are now ready to define the transfer rules in our mechanism of Section 4.3.1.

First, recall that each m_{-i}^0 which survives iterative elimination of strictly dominated messages ($\hat{S}_{-i}^\infty$) uniquely identifies an atom in $\mathcal{P}_i$ and hence an element in Ψ_{-i}^0. We denote this atom in Ψ_{-i}^0 by $\psi_{-i}^0\left(m_{-i}^0\right)$. Then, we define

$$\tau_i^0(m_i^1, m_{-i}^0) = \begin{cases} d_i^0(m_i^1, \psi_{-i}^0(t_{-i})), & \text{if } m_{-i}^0 \in \hat{S}_{-i}^\infty(\hat{\theta}_{-i}(t_{-i})|\mathcal{M}^*) \text{ for some } t_{-i} \in \bar{T}_{-i}; \\ 0, & \text{otherwise.} \end{cases}$$

Second, we define

$$\tau_i^1(m_i^2, m_{-i}^1, m_{-i}^0) = d_i^1(m_i^2, \psi_{-i}^1\left(m_{-i}^0, m_{-i}^1\right)),$$

where $\psi_{-i}^1\left(m_{-i}^0, m_{-i}^1\right)$ denotes the unique atom ψ_{-i}^1 in Ψ_{-i}^1 such that $\psi_{-i}^1 \subset \psi_{-i}^0\left(m_{-i}^0\right)$ and $\chi_{-i}^0(t_{-i}) = \chi_{-i}^0\left(m_{-i}^1\right)$ for every $t_{-i} \in \psi_{-i}^1$.

Finally, let

$$\tau_i^2(m_i^l, m_{-i}^{l-1}) = d_i^2(m_i^l, m_{-i}^{l-1}), \forall l \geq 3.$$

Now given $\tilde{\tau} = \hat{\tau}/3$, we set γ as the minimum of the γ given by Lemmas 3 and 4. We denote by E the maximal payoff difference between an outcome resulted from the maximally

revealing mechanism $\mathcal{M}^*$ and that resulted from the extended direct mechanism $\bar{\mathcal{M}}$, i.e.,

$$E \equiv \max_{m^* \in M^*, \bar{m} \in \bar{M}, \theta \in \Theta, i \in I} \left| u_i \left(g^*(m^*), \theta \right) - u_i \left(\bar{g} \left(\bar{m} \right), \theta \right) \right|. \tag{10}$$

We choose $\epsilon > 0$ small enough so that $\gamma > \epsilon E$. Moreover, let $\bar{\tau} > 0$ be the bound of the transfer rule in the extended direct mechanism $\bar{\mathcal{M}}$. In the proof of Property (b) of Proposition 2, we set $\bar{\tau}$ sufficiently small so that

$$\gamma > \epsilon E + \bar{\tau}; \tag{11}$$

furthermore, $\bar{\tau} < \hat{\tau}/3$. Then, Property (a) of Proposition 2 holds, since

$$\begin{aligned} |\tau_i(m)| &\leq \left| \tau_i^* \left(m^0 \right) \right| + \left| \tau_i^0 (m_i^1, m_{-i}^0) \right| + \left| \tau_i^1 (m_i^2, m_{-i}^1, m_{-i}^0) \right| + \sum_{l=3}^{\bar{l}+4} \left| \tau_i^2 (m_i^l, m_{-i}^{l-1}) \right| + |\bar{\tau}_i(\bar{m})| \\ &\leq \frac{\hat{\tau}}{3} + \frac{\tilde{\tau}}{\bar{l}+5} + \frac{\tilde{\tau}}{\bar{l}+5} + \frac{\bar{l}+3}{\bar{l}+5}\tilde{\tau} + \frac{\hat{\tau}}{3} \\ &\leq \frac{2}{3}\hat{\tau} + \tilde{\tau} \leq \hat{\tau}. \end{aligned} \tag{12}$$

We then proceed to prove Property (b) of Proposition 2 in the next two steps.

A.2.2 Proof of Lemma 1

The proof of Lemma 1 will make use of Lemmas 3 and 4 as well as the following lemma as the building blocks. Specifically, the lemma below shows that under the transfer rule τ_i^2 which we identify in Lemma 4, for each misreported type t_i', there always exists a conjecture σ_{-i} which rationalizes this misreported t_i' as the unique maximizer of transfers for the true type t_i.

Lemma 5 *For any $t_i, t_i' \in \bar{T}_i$, there exists $\sigma_{-i} : \bar{T}_{-i} \to \Delta\left(\bar{T}_{-i}\right)$ such that*

$$\sum_{t_{-i} \in \bar{T}_{-i}} \bar{\pi}_i \left(t_i \right) \left[t_{-i} \right] \sum_{\tilde{t}_{-i} \in \bar{T}_{-i}} \left[\tau_i^2 \left(t_i', \tilde{t}_{-i} \right) - \tau_i^2 \left(\tilde{t}_i, \tilde{t}_{-i} \right) \right] \sigma_{-i} \left(t_{-i} \right) \left[\tilde{t}_{-i} \right] > \gamma, \forall \tilde{t}_i \neq t_i'.$$

Proof. By Lemma 4, for type t_i', we have

$$\sum_{\tilde{t}_{-i} \in \bar{T}_{-i}} \left[\tau_i^2 \left(t_i', \tilde{t}_{-i} \right) - \tau_i^2 \left(\tilde{t}_i, \tilde{t}_{-i} \right) \right] \bar{\pi}_i \left(t_i' \right) \left[\tilde{t}_{-i} \right] > \gamma, \forall \tilde{t}_i \neq t_i'. \tag{13}$$

We construct type t_i's conjecture denoted by $\sigma_{-i} : \bar{T}_{-i} \to \Delta\left(\bar{T}_{-i}\right)$ such that

$$\sigma_{-i} \left(t_{-i} \right) \left[\tilde{t}_{-i} \right] = \bar{\pi}_i \left(t_i' \right) \left[\tilde{t}_{-i} \right], \forall t_{-i}, \tilde{t}_{-i} \in \bar{T}_{-i}. \tag{14}$$

Thus, we have

$$\sum_{t_{-i}\in\bar{T}_{-i}} \bar{\pi}_i\left(t_i\right)\left[t_{-i}\right]\sigma_{-i}\left(t_{-i}\right)\left[\tilde{t}_{-i}\right] = \bar{\pi}_i\left(t_i'\right)\left[\tilde{t}_{-i}\right]$$

where the equality follows from (14). Thus, the lemma follows from (13). ■

We now turn to the proof of Lemma 1, which is restated below.

Lemma 1: *Suppose that Assumption 1 holds. For any player i of type $t_i \in \bar{T}_i$ and any $l = 0, 1, ..., \bar{l}$, the following two statements, denoted by $P^1(l)$ and $P^2(l)$, hold:*

- *$P^1(l)$: for any $\hat{m}_i \in M_i$, $\hat{m}_i \in \hat{W}_i^l(\hat{\theta}_i(t_i)|\bar{\mathcal{M}}^*)$ implies $\hat{m}_i^0 \in \hat{S}_i^l(\hat{\theta}_i(t_i)|\mathcal{M}^*)$;*
- *$P^2(l)$: there is some $\left(m_i^0, ..., m_i^l, m_i^{l+1}\right) \in \times_{k=0}^{l+1} M^k$ such that for every $t_i' \in \bar{T}_i$,*

$$\left(m_i^0, ..., m_i^l, m_i^{l+1}, t_i', t_i, ..., t_i\right) \in \hat{W}_i^l(\hat{\theta}_i(t_i)|\bar{\mathcal{M}}^*).$$

Proof of Lemma 1. We prove Lemma 1 by induction. We observe that $P^1(0)$ and $P^2(0)$ hold trivially, since for any $i \in I$, we have $\hat{S}_i^0(\theta_i|M^*) = M_i^*$ for any $\theta_i \in \Theta_i$ and $\hat{W}_i^0(\hat{\theta}_i(t_i)|\bar{\mathcal{M}}^*) = M_i$ for any $t_i \in \bar{T}_i$. Next, for each $l \geq 0$, we assume that $P^1(l)$ and $P^2(l)$ hold and prove that $P^1(l+1)$ and $P^2(l+1)$ also hold.

Consider player i of type t_i and a message $m_i^0 \notin \hat{S}_i^{l+1}(\hat{\theta}_i(t_i)|\mathcal{M}^*)$.[24] This implies that there exists some $\alpha_i^* \in \Delta(M_i^*)$ such that

$$u_i(g^*(\alpha_i^*, m_{-i}^*), (\theta_i, \theta_{-i})) > u_i(g^*(m_i^0, m_{-i}^*), (\theta_i, \theta_{-i})) \tag{15}$$

for all $\theta_{-i} \in \Theta_{-i}$ and $m_{-i}^* \in \hat{S}_{-i}^l\left(\theta_{-i}|\mathcal{M}^*\right)$.

Fix $m_i = (m_i^0, m_i^1, \ldots, m_i^{\bar{l}+5}) \in M_i$ such that $m_i^0 \notin \hat{S}_i^{l+1}(\theta_i|\mathcal{M}^*)$. Let $\alpha_i \in \Delta(M_i)$ be a mixed message that induces the same marginal distribution on M_i^0 as α_i^* and is identical to m_i otherwise. Thus, for any $m_{-i} \in \hat{W}_{-i}^l\left(\theta_{-i}|\mathcal{M}\right)$ and θ_{-i}, we have

$$\begin{aligned}
& u_i(g(\alpha_i, m_{-i}), (\theta_i, \theta_{-i})) + \tau_i\left(\alpha_i, m_{-i}\right)) \\
& -u_i(g(m_i, m_{-i}), (\theta_i, \theta_{-i})) + \tau_i(m_i, m_{-i}) \\
= \;& e(m_i, m_{-i})\left[u_i(g^*(\alpha_i^*, m_{-i}^0), (\theta_i, \theta_{-i})) - u_i(g^*(m_i^0, m_{-i}^0)), (\theta_i, \theta_{-i}))\right] \\
\geq \;& 0
\end{aligned} \tag{16}$$

where the equality follows because α_i differs from m_i only in the 0th round announcement and the inequality follows from (15) and the induction hypothesis $P^1\left(l\right)$. Indeed, by $P^1\left(l\right)$,

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

if $m_{-i} \in \hat{W}_{-i}^{l}\left(\theta_{-i}|\bar{\mathcal{M}}^{*}\right)$, then we must have $m_{-i}^{0} \in \hat{S}_{-i}^{l}\left(\theta_{-i}|\mathcal{M}^{*}\right)$. Thus, the inequality in (16) follows from (15).

In addition, by $P^{2}\left(l\right)$, for each $t_{-i} \in \bar{T}_{-i}$, there exists some $\tilde{m}_{-i} \in \hat{W}_{-i}^{l}(\hat{\theta}_{-i}\left(t_{-i}\right)|\bar{\mathcal{M}}^{*})$ such that $\tilde{m}_{-i}^{2} \neq \tilde{m}_{-i}^{k}$ for some $k \in \{3,\ldots,\bar{l}+3\}$. Thus, $e(m_{i},\tilde{m}_{-i}) = \epsilon$ (by the definition of $e\left(\cdot\right)$ in Section 4.3.1). Against $\tilde{m}_{-i}$ together with an arbitrary θ_{-i}, the inequality in (16) becomes strict. Thus, the message m_{i} is weakly dominated by α_{i} so that $m_{i} \notin \hat{W}_{i}^{l+1}(\hat{\theta}_{i}\left(t_{i}\right)|\bar{\mathcal{M}}^{*})$. Therefore, $P^{1}(l+1)$ holds.

Second, we shall prove $P^{2}\left(l+1\right)$. By the induction hypothesis $P^{2}\left(l\right)$, we can define a mapping $\nu_{-i} : \bar{T}_{-i} \to \times_{k=0}^{l+1} M_{-i}^{k}$ such that for any types t_{-i} and t'_{-i} in $\bar{T}_{-i}$, we have

$$\tilde{m}_{-i}(t_{-i},t'_{-i}) \equiv (\nu_{-i}^{0}(t_{-i}),...,\nu_{-i}^{l+1}(t_{-i}),t'_{-i},t_{-i},...,t_{-i}) \in \hat{W}_{-i}^{l}(\hat{\theta}_{-i}\left(t_{-i}\right)|\bar{\mathcal{M}}^{*}). \tag{17}$$

Moreover, for each $t_{i} \in \bar{T}_{i}$, we define the "coordinate-wise" best reply as follows:

$$\left\{b_{i}^{0}(\nu_{-i},t_{i})\right\} = \arg\max_{m_{i}^{*}\in M_{i}^{*}} \sum_{t_{-i}} u_{i}(g^{*}\left(m_{i}^{*},\nu_{-i}^{0}(t_{-i})\right),\hat{\theta}(t_{i},t_{-i}))\bar{\pi}_{i}\left(t_{i}\right)\left[t_{-i}\right]. \tag{18}$$

$$\left\{b_{i}^{1}(\nu_{-i},t_{i})\right\} = \arg\max_{t'_{i}\in\bar{T}_{i}} \sum_{t_{-i}} \tau_{i}^{0}\left(t'_{i},\nu_{-i}^{0}(t_{-i})\right)\bar{\pi}_{i}\left(t_{i}\right)\left[t_{-i}\right]; \tag{19}$$

$$\left\{b_{i}^{2}(\nu_{-i},t_{i})\right\} = \arg\max_{t'_{i}\in\bar{T}_{i}} \sum_{t_{-i}} \tau_{i}^{1}\left(t'_{i},\nu_{-i}^{1}(t_{-i}),\nu_{-i}^{0}(t_{-i})\right)\bar{\pi}_{i}\left(t_{i}\right)\left[t_{-i}\right]; \tag{20}$$

$$\left\{b_{i}^{k+1}\left(\nu_{-i},t_{i}\right)\right\} = \arg\max_{t'_{i}\in\bar{T}_{i}} \sum_{t_{-i}} \tau_{i}^{2}\left(t'_{i},\nu_{-i}^{k}\left(t_{-i}\right)\right)\bar{\pi}_{i}\left(t_{i}\right)\left[t_{-i}\right], \forall k = 2,...,l+1 \tag{21}$$

where the uniqueness of the best reply $\{b_{i}^{k}\}$ for $k \neq 1,2$ follows from Lemmas 3 and 4. We now prove $P^{2}\left(l+1\right)$ by establishing the following claim: for each $t_{i} \in \bar{T}_{i}$,

$$\bar{m}_{i} \equiv \left(b_{i}^{0}(\nu_{-i},t_{i}),b_{i}^{1}(\nu_{-i},t_{i}),\ldots,b_{i}^{l+2}\left(\nu_{-i},t_{i}\right),t'_{i},t_{i},...,t_{i}\right) \in \hat{W}_{i}^{l+1}(\hat{\theta}_{i}(t_{i})|\bar{\mathcal{M}}^{*}). \tag{22}$$

First, by Lemma 5, for any $t'_{i} \in \bar{T}_{i}$, there exists a mapping $\sigma_{-i} : \bar{T}_{-i} \to \Delta\left(\bar{T}_{-i}\right)$ such that

$$\sum_{t_{-i}\in\bar{T}_{-i}} \bar{\pi}_{i}\left(t_{i}\right)\left[t_{-i}\right] \sum_{\tilde{t}_{-i}\in\bar{T}_{-i}} \left[\tau_{i}^{2}\left(t'_{i},\tilde{t}_{-i}\right) - \tau_{i}^{2}\left(\tilde{t}_{i},\tilde{t}_{-i}\right)\right]\sigma_{-i}\left(t_{-i}\right)\left[\tilde{t}_{-i}\right] > \gamma, \forall \tilde{t}_{i} \neq t'_{i}. \tag{23}$$

For each $t_{-i} \in \bar{T}_{-i}$, pick $s_{-i}^{t_{-i}} \in \bar{T}_{-i}$ such that

$$\begin{aligned} s_{-i}^{t_{-i}} &\neq t_{-i}, && \text{if } l = 0; \\ s_{-i}^{t_{-i}} &\neq \nu_{-i}^{1}(t_{-i}), && \text{if } l \geq 1. \end{aligned}$$

We construct a conjecture $\bar{\sigma}_{-i}^{\varsigma} : \bar{T}_{-i} \to \Delta\left(\bar{T}_{-i}\right)$ as:

$$\bar{\sigma}_{-i}^{\varsigma}\left(t_{-i}\right)\left[t'_{-i}\right] \equiv \left(1-\varsigma\right)\sigma_{-i}\left(t_{-i}\right)\left[t'_{-i}\right] + \varsigma\delta_{s_{-i}^{t_{-i}}}\left[t'_{-i}\right], \forall t_{-i},t'_{-i} \in \bar{T}_{-i}$$

where $\varsigma \in (0,1)$ and $\delta_{s_{-i}^{t_{-i}}}$ stands for the Dirac measure which assigns probability one to the type profile $s_{-i}^{t_{-i}}$. In words, $\bar{\sigma}_{-i}^{\varsigma}$ modifies σ_{-i} such that $\bar{\sigma}_{-i}^{\varsigma}(t_{-i})$ is identical to $\sigma_{-i}(t_{-i})$ with probability $1-\varsigma$; moreover, $\bar{\sigma}_{-i}^{\varsigma}(t_{-i})$ assigns probability ς to some type profile $s_{-i}^{t_{-i}}$ which is either distinct from t_{-i} (if $l=0$) or $\nu_{-i}^{1}(t_{-i})$ (if $l \geq 1$). It follows from (23) that for ς sufficiently small, we still have

$$\sum_{t_{-i} \in \bar{T}_{-i}} \bar{\pi}_i(t_i)[t_{-i}] \sum_{\tilde{t}_{-i} \in \bar{T}_{-i}} \left[\tau_i^2\left(t_i', \tilde{t}_{-i}\right) - \tau_i^2\left(\tilde{t}_i, \tilde{t}_{-i}\right)\right] \bar{\sigma}_{-i}^{\varsigma}(t_{-i})\left[\tilde{t}_{-i}\right] > \gamma, \forall \tilde{t}_i \neq t_i'. \tag{24}$$

Second, let $\nu_{-i}^{\varsigma}: \bar{T}_{-i} \to \Delta(M_{-i})$ be type t_i's conjecture defined as

$$\nu_{-i}^{\varsigma}(t_{-i})\left[\tilde{m}_{-i}(t_{-i}, t_{-i}')\right] \equiv \bar{\sigma}_{-i}^{\varsigma}(t_{-i})\left[t_{-i}'\right], \forall t_{-i}, t_{-i}' \in \bar{T}_{-i} \tag{25}$$

where $\tilde{m}_{-i}(t_{-i}, t_{-i}')$ is defined in (17). By (24), we have $b_i^{l+1}(\nu_{-i}, t_i) = t_i'$. Now define $\mu_i^{\varsigma} \in \Delta(\Theta_{-i} \times M_{-i})$ which is induced from ν_{-i} and $\bar{\pi}_i(t_i)$ as follows: for any (θ_{-i}, m_{-i}),

$$\mu_i^{\varsigma}(\theta_{-i}, m_{-i}) = \sum_{t_{-i} \in \bar{T}_{-i}: \hat{\theta}_{-i}(t_{-i}) = \theta_{-i}} \nu_{-i}^{\varsigma}(t_{-i})[m_{-i}] \bar{\pi}_i(t_i)[t_{-i}].$$

By (17) and (25), $\mu_i^{\varsigma}(\theta_{-i}, m_{-i}) > 0$ implies $m_{-i} \in \hat{W}_{-i}^{l}(\theta_{-i}|\mathcal{M})$.

Third, we show that against the belief μ_i^{ς}, message $\bar{m}_i$ defined in (22) is a strictly better reply for $\hat{\theta}_i(t_i)$ than any other message $\tilde{m}_i$ with $\tilde{m}_i^k \neq \bar{m}_i^k$ for some k. This together with the fact that $\mu_i^{\varsigma}(\theta_{-i}, m_{-i}) > 0$ implies $m_{-i} \in \hat{W}_{-i}^{l}(\theta_{-i}|\mathcal{M})$ implies that $\bar{m}_i \in \hat{W}_i^{l+1}(\hat{\theta}_i(t_i)|\bar{\mathcal{M}}^*)$. It remains to show that $\bar{m}_i$ is a strict best response against μ_i^{ς}. We show this by considering the following two cases:

Case A: $\tilde{m}_i^0 \neq \bar{m}_i^0$ and $\tilde{m}_i^k = \bar{m}_i^k$ for any $k \geq 1$.

In this case, we have $\bar{m}_i^2 \neq \tilde{m}_i^2$. Then,

$$\begin{aligned}
&\sum_{\theta_{-i}, m_{-i}} \left[u_i(g(\bar{m}_i, m_{-i}), \hat{\theta}_i(t_i), \theta_{-i}) + \tau_i(\bar{m}, m_{-i}))\right] \mu_i^{\varsigma}(\theta_{-i}, m_{-i}) \\
&- \sum_{\theta_{-i}, m_{-i}} \left[u_i(g(\tilde{m}_i, m_{-i}), \hat{\theta}_i(t_i), \theta_{-i}) + \tau_i(\tilde{m}_i, m_{-i})\right] \mu_i^{\varsigma}(\theta_{-i}, m_{-i}) \\
= &\sum_{\theta_{-i}, m_{-i}} e((\bar{m}_i, m_{-i}))\mu_i^{\varsigma}(\theta_{-i}, m_{-i}) \\
&\times \left[u_i(g^*(\bar{m}_i^0, m_{-i}^0), \hat{\theta}_i(t_i), \theta_{-i}) - u_i(g^*(\tilde{m}_i^0, m_{-i}^0)), \hat{\theta}_i(t_i), \theta_{-i})\right]
\end{aligned} \tag{26}$$

where the equality follows because $\tilde{m}_i^k = \bar{m}_i^k$ for any $k \geq 1$. Moreover, since the belief μ_i^{ς} is

induced from ν_{-i}^{ς} and $\bar{\pi}_i(t_i)$, it follows that

$$\mu_i^{\varsigma}(\theta_{-i}, m_{-i}) > 0 \Rightarrow \tag{27}$$
$$\text{there exist } t_{-i}, t'_{-i} \in \bar{T}_{-i} \text{ such that } m_{-i} = \tilde{m}_{-i}(t_{-i}, t'_{-i}),\ \hat{\theta}_{-i}(t_{-i}) = \theta_{-i},$$
$$\text{and either } t'_{-i} \neq \nu_{-i}^1(t_{-i}) \text{ or } t'_{-i} \neq t_{-i}.$$

Observe that by (18),

$$\begin{aligned} & u_i(g^*(\bar{m}_i^0, m_{-i}^0), \hat{\theta}_i(t_i), \theta_{-i}) - u_i(g^*(\tilde{m}_i^0, m_{-i}^0)), \hat{\theta}_i(t_i), \theta_{-i}) \\ = \ & u_i(g^*(b_i^0(\nu_{-i}^0, t_i), m_{-i}^0), \hat{\theta}_i(t_i), \theta_{-i}) - u_i(g^*(\tilde{m}_i^0, m_{-i}^0)), \hat{\theta}_i(t_i), \theta_{-i}) \\ > \ & 0. \end{aligned}$$

It follows from (27) that there exist $k \in \{3, \ldots, \bar{l}+3\}$ and $m_{-i} \in M_{-i}$ such that $\mu_i^{\varsigma}(\theta_i, m_{-i}) > 0$ and $m_{-i}^2 \neq m_{-i}^k$. Hence, $e(\bar{m}_i, m_{-i}) = \epsilon$ for some m_{-i} with μ_i^{ς}-positive probability. We thus conclude that the payoff difference in (26) is positive. Thus, $\bar{m}_i$ is a strictly better reply than $\tilde{m}_i$ against the belief μ_i^{ς} for $\hat{\theta}_i(t_i)$ so that $\bar{m}_i \in \hat{W}_i^{l+1}(\hat{\theta}_i(t_i)|\bar{\mathcal{M}}^*)$.

Case B: $\tilde{m}_i^k \neq \bar{m}_i^k$ for some $k \geq 1$. We consider the following two subcases:

Case B1: $\tilde{m}_i^k \neq \bar{m}_i^k$ for some k with $1 \leq k \leq \bar{l}+3$.

We argue that for each such k, if $\tilde{m}_i^k \neq \bar{m}_i^k$, then against ν_{-i}^k, $\bar{m}_i^k$ ensures a gain more than γ over $\tilde{m}_i^1$. For $k = 1$, the claim follows from (19) and Property (b) of Lemma 4 for transfer $\tau_i^0(\cdot)$. For $k = 2$, the claim follows from (20), Property (b) of Lemma 4 for the transfer rule $\tau_i^1(\cdot)$. For $3 \leq k \leq l$, the claim follows from (21) and Property (b) of Lemma 4 for the transfer rule $\tau_i^2(\cdot)$. For $k = l+1$, the claim follows from (24). Finally, for $l+2 \leq k \leq \bar{l}+3$, the claim follows from Property (c) of Lemma 4 for the transfer rule $\tau_i^2(\cdot)$.

Since μ_i^{ς} is induced from ν_{-i}^{ς} and $\bar{\pi}_i(t_i)$, for $\varsigma > 0$ sufficiently small, the gain from changing $\tilde{m}_i^k$ to $\bar{m}_i^k$ is at least γ, while the potential loss is at most $\epsilon E + \bar{\tau}$. Since $\gamma > \epsilon E + \bar{\tau}$ by (11), $\bar{m}_i$ is strictly better than $\tilde{m}_i$ against the belief μ_i^{ς} for $\hat{\theta}_i(t_i)$. Hence, $\bar{m}_i \in \hat{W}_i^{l+1}(\hat{\theta}_i(t_i)|\bar{\mathcal{M}}^*)$.

Case B2: $\tilde{m}_i^h = \bar{m}_i^h$ for any h with $1 \leq h \leq \bar{l}+3$ and $\tilde{m}_i^k \neq \bar{m}_i^k$ for some $k \geq \bar{l}+4$. It follows from (17) that every message $\tilde{m}_{-i}(t_{-i}, t'_{-i})$ on the support of $\nu_{-i}^{\varsigma}(t_{-i})$ truthfully reports the type t_{-i} in the $(\bar{l}+3)$th coordinate as well as all announcements in $\bar{\mathcal{M}}$ (from the $(\bar{l}+4)$th coordinate onwards). Since (c) of Lemma 4 holds (for $k = \bar{l}+3$) and truth-telling is a strict Bayes Nash equilibrium in the game $U(\bar{\mathcal{M}}, \bar{\mathcal{T}})$ induced by the extended direct mechanism $\bar{\mathcal{M}}$, it follows that $\bar{m}_i$ is a strictly better reply than $\tilde{m}_i$ against any such $\tilde{m}_{-i}(t_{-i}, t'_{-i})$. Hence,

$\bar{m}_i$ is a strictly better reply than $\tilde{m}_i$ against the belief μ_i^{ς} for $\hat{\theta}_i(t_i)$. This completes the proof of Lemma 1. ■

A.2.3 Proof of Proposition 2

We prove this claim by induction. Consider any $i \in I$, $t_i \in \bar{T}_i$ with $\hat{\theta}_i(t_i) = \theta_i$, and $m_i \in S^0\hat{W}^{\infty}(t_i|\bar{\mathcal{M}}^*, \bar{\mathcal{T}})$. For each $l \geq 0$, we denote by $\{\psi_i^l(t_i)\}_{t_i \in \bar{T}_i}$ the partition over $\bar{T}_i$ induced by $\{m_i^l\}_{m_i^l \in \bar{T}_i}$ where $m_i^l \in S^l\hat{W}^{\infty}(t_i|\bar{\mathcal{M}}^*, \bar{\mathcal{T}})$ for some $t_i \in \bar{T}_i$. First, we show that $m_i \notin S^1\hat{W}^{\infty}\left(t_i|\bar{\mathcal{M}}^*, \bar{\mathcal{T}}\right)$ if $\psi_i^1(m_i^1) \neq \psi_i^1(t_i)$. Indeed, consider an alternative message:

$$\bar{m}_i = (m_i^0, t_i, m_i^2, ..., m_i^{\bar{l}+5}),$$

which is identical to m_i except that $\bar{m}_i^1 \neq m_i^1$. By Lemma 1, we have that $\hat{m}_j^0 \in \hat{S}_j^{\infty}(\hat{\theta}_j(t_j)|\mathcal{M}^*)$ for any $\hat{m}_j \in S^0\hat{W}^{\infty}(t_j|\bar{\mathcal{M}}^*, \bar{\mathcal{T}})$ and any player $j \in I$. Against any conjecture $\nu_{-i} : \bar{T}_{-i} \to M_{-i}$ satisfying $\nu_{-i}(t_{-i}) \in S^0\hat{W}_{-i}^{\infty}\left(t_{-i}|\bar{\mathcal{M}}^*, \bar{\mathcal{T}}\right)$ for every $t_{-i} \in \bar{T}_{-i}$, by Property (c) of Lemma 4, choosing m_i rather than $\bar{m}_i$ induces the loss of at least γ and no gain. Hence, m_i is strictly dominated by $\bar{m}_i$.

Now define $\psi_i^l(t_i) = \psi_i^2(t_i)$ for any $l \geq 3$. Also recall that by Assumption 1, $\psi_i^l(t_i) = \{t_i\}$ for any $l \geq 2$. Now suppose that any l such that $1 \leq l \leq \bar{l}+3$, we have $\psi_i^l(m_i^l) = \psi_i^l(t_i)$ for every $m_i \in S_i^l\hat{W}^{\infty}(t_i|\bar{\mathcal{M}}^*, \bar{\mathcal{T}})$. Then, we show that $\psi_i^{l+1}(m_i^{l+1}) = \psi_i^{l+1}(t_i)$ for every $m_i \in S_i^{l+1}\hat{W}^{\infty}(t_i|\bar{\mathcal{M}}^*, \bar{\mathcal{T}})$. Suppose to the contrary that $\psi_i^{l+1}(m_i^{l+1}) \neq \psi_i^{l+1}(t_i)$ for some $m_i \in S_i^{l+1}\hat{W}^{\infty}(t_i|\bar{\mathcal{M}}^*, \bar{\mathcal{T}})$. We choose $\bar{m}_i$ to be identical to m_i except that $\bar{m}_i^{l+1} = t_i \neq m_i^{l+1}$. By (c) of Lemma 4, choosing m_i rather than $\bar{m}_i$ induces the loss of at least γ; while the possible gain incurred results from outcome changes due to alternating different values of function $e(\cdot)$, which is bounded by ϵE, and possibly different transfers (when $l = \bar{l}+3$) in the extended direct mechanism $\bar{\mathcal{M}}$ whose difference is bounded by $\bar{\tau}$. Hence, the total gain is bounded by $\epsilon E + \bar{\tau}$. Since we have $\gamma > \epsilon E + \bar{\tau}$ by (11), m_i is still strictly dominated by $\bar{m}_i$. This completes the proof of Proposition 2.

A.3 Proof of Proposition 1

Proposition 1: *Suppose that Assumption 1 holds. If an SCF f is incentive compatible, then it is implementable in $S^{\infty}\hat{W}^{\infty}$ with arbitrarily small transfers.*

We prove this proposition by the following steps. In the first step, we construct an extended direct mechanism $\bar{\mathcal{M}} = (\bar{M}_i, \bar{g}, \bar{\tau}_i)_{i \in I}$ such that $|\bar{\tau}_i(m)| < \bar{\tau}$ for any $m \in M$

and $\bar{\tau}$ satisfies (11). In the second step, we show that the augmented mechanism $\bar{\mathcal{M}}^*$, which connects the maximally revealing mechanism $\mathcal{M}^*$ to $\bar{\mathcal{M}}$ and implements the SCF f in $S^\infty\hat{W}^\infty$ with arbitrarily small transfers, if the SCF f is incentive compatible. Let $\hat{\tau}$ be an arbitrary positive number.

A.3.1 The Construction of Mechanism $\bar{\mathcal{M}}$

Recall that we need to construct a mechanism $\bar{\mathcal{M}} = \left(\left(\bar{M}_i\right), \bar{g}, (\bar{\tau}_i)\right)_{i\in I}$. We define the mechanism as follows.

1. **The message space**:

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

   $$\bar{M}_i = \bar{M}_i^0 \times \cdots \times \bar{M}_i^K = \underbrace{\bar{T}_i \times \cdots \times \bar{T}_i}_{K+1 \text{ times}},$$

   where K is an integer to be specified later. Denote

   $$\bar{m}_i = \left(\bar{m}_i^0, ..., \bar{m}_i^K\right) \in \bar{M}_i,\ \bar{m}_i^k \in M_i^k,\ k \in \{0, 1, ..., K\},$$

   and

   $$\bar{m} = \left(\bar{m}^0, ..., \bar{m}^K\right) \in \bar{M},\ \bar{m}^k = \left(\bar{m}_i^k\right)_{i\in I} \in \bar{M}^k = \times_{i\in I}^k \bar{M}_i^k.$$

2. **The outcome function**:

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

   $$\bar{g}(\bar{m}) = \frac{1}{K}\sum_{k=1}^{K} f\left(\bar{m}^k\right). \tag{28}$$

   The outcome function consists of K equally weighted lotteries the kth of which depends only on the I-tuple of the kth announcements.

3. **The transfer rule**:

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

- 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).

$$c_i\left(\bar{m}^0, ..., \bar{m}^K\right) = \begin{cases} \xi & \text{if there exists } k \in \{1, ..., K\} \text{ s.t. } \bar{m}_i^k \neq \bar{m}_i^0, \\ & \text{and } \bar{m}_j^{k'} = \bar{m}_j^0 \text{ for all } k' \in \{1, ..., k-1\} \text{ for all } j \in I; \\ 0 & \text{otherwise.} \end{cases} \tag{29}$$

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

$$c_i^k\left(\bar{m}_i^0, \bar{m}_i^k\right) = \begin{cases} \eta & \text{if } \bar{m}_i^k \neq \bar{m}_i^0; \\ 0 & \text{otherwise.} \end{cases} \tag{30}$$

In total,

$$\bar{\tau}_i\left(\bar{m}\right) = -c_i\left(\bar{m}^0, ..., \bar{m}^K\right) - \sum_{k=1}^{K} c_i^k\left(\bar{m}_i^0, \bar{m}_i^k\right). \tag{31}$$

4. We provide a summary of conditions which we impose on transfers:

Let D be the maximum gain for player i from altering the kth announcement

$$D \equiv \max_{t_i, t_i' \in \bar{T}_i, t_{-i} \in \bar{T}_{-i}, \theta \in \Theta, i \in I} \left\{u_i(f(t_i', t_{-i}), \theta) - u_i(f(t_i, t_{-i}), \theta)\right\}. \tag{32}$$

Given the transfer bound $\bar{\tau}$, we choose K large enough so that there are positive numbers η and ξ satisfying the following conditions:

$$\frac{\bar{\tau}}{2K} > \eta > 0; \tag{33}$$

$$\frac{\bar{\tau}}{2} > \xi > \frac{D}{K}. \tag{34}$$

It then follows from (31), (33), and (34) that

$$\left|\bar{\tau}_i\left(m\right)\right| < \bar{\tau}. \tag{35}$$

A.3.2 Implementation in $S^\infty \hat{W}^\infty$

Recall that in defining our main implementing mechanism $\bar{\mathcal{M}}^*$ in Section 4.3.1, we write $\bar{M}_i = M_i^{\bar{l}+4} \times M_i^{\bar{l}+5}$ where $M_i^{\bar{l}+4} = \bar{T}_i$ and $M_i^{\bar{l}+5} = \left(\bar{T}_i\right)^K$, which consists of K copies of $\bar{T}_i$. For each $m_i \in \bar{M}_i^*$, we denote by $\bar{m}_i$ the projection of m_i in $\bar{M}_i$. By Proposition 2, it follows that $m_i \in S_i^\infty \hat{W}^\infty(t_i | \bar{\mathcal{M}}^*, \bar{\mathcal{T}})$ only if $m_i^{\bar{l}+4} (= \bar{m}_i^0) = t_i$. We now establish implementation in $S^\infty \hat{W}^\infty$ via mechanism $\bar{\mathcal{M}}^*$ by the following claim.

Claim 1 *In the game* $U\left(\bar{\mathcal{M}}^*, \bar{\mathcal{T}}\right)$*, for every nonnegative integer* $k \leq K$*, player* $i \in I$*, and type* $t_i \in \bar{T}_i$*, if* $m_i \in S_i^\infty \hat{W}^\infty\left(t_i | \bar{\mathcal{M}}^*, \bar{\mathcal{T}}\right)$*, then* $\bar{m}_i^k = t_i$*.*

Proof. When $k = 0$, the result follows from Proposition 2. Fix $k \geq 0$. The induction hypothesis is that for every $i \in I$ and $t_i \in \bar{T}_i$, if $m_i \in S_i^\infty \hat{W}^\infty\left(t_i | \bar{\mathcal{M}}^*, \bar{\mathcal{T}}\right)$, then $\bar{m}_i^{k'} = t_i$ for any nonnegative integer $k' \leq k$.

Then, we show that if $m_i \in S_i^\infty \hat{W}^\infty(t_i | \bar{\mathcal{M}}^*, \bar{\mathcal{T}})$, then $\bar{m}_i^{k'} = t_i$ for any nonnegative integer $k' \leq k+1$. By the induction hypothesis, it suffices to prove that $\bar{m}_i^{k+1} = t_i$. The basic idea is similar to Abreu and Matsushima (1994). Suppose instead that $\bar{m}_i^{k+1} \neq t_i$. Let $\tilde{m}_i$ be a message in $\bar{M}_i^*$ which is identical to m_i except that $\tilde{m}_i$ reports the truth in the $(k+1)$st announcement in $\bar{M}_i$. We hereby slightly abuse the notation by writing $\tilde{m}_i^{k+1}$ (as opposed to the heavier notation $\overline{\tilde{m}}_i^{k+1}$) for the $(k+1)$th announcement of $\tilde{m}_i$ in $\bar{M}_i$. We let $\hat{M}_{-i} = \{m_{-i} \in M_{-i} : \bar{m}_{-i}^{k+1} = \bar{m}_{-i}^0\}$. Fix a conjecture $\nu_{-i} : \bar{T}_{-i} \to M_{-i}$ such that for each $t_{-i} \in \bar{T}_{-i}$,

$$\nu_{-i}(t_{-i}) \in S_i^\infty \hat{W}^\infty(t_i | \bar{\mathcal{M}}^*, \bar{\mathcal{T}}).$$

We will show that

$$V_i((\tilde{m}_i, \nu_{-i}), t_i) - V_i((m_i, \nu_{-i}), t_i) > 0. \tag{36}$$

We decompose the left-hand side of this inequality into the following two parts:

$$\begin{aligned} &\sum_{t_{-i}:\nu_{-i}(t_{-i}) \notin \hat{M}_{-i}} \left\{ \begin{array}{c} \{u_i(g(\tilde{m}_i, \nu_{-i}(t_{-i})), \hat{\theta}(t_i, t_{-i})) + \tau_i(\tilde{m}_i, \nu_{-i}(t_{-i}))\} - \\ \{u_i(g(m_i, \nu_{-i}(t_{-i})), \hat{\theta}(t_i, t_{-i})) + \tau_i(m_i, \nu_{-i}(t_{-i}))\} \end{array} \right\} \bar{\pi}_i(t_i)[t_{-i}] \quad (37) \\ &+ \sum_{t_{-i}:\nu_{-i}(t_{-i}) \in \hat{M}_{-i}} \left\{ \begin{array}{c} \{u_i(g(\tilde{m}_i, \nu_{-i}(t_{-i})), \hat{\theta}(t_i, t_{-i})) + \tau_i(\tilde{m}_i, \nu_{-i}(t_{-i}))\} - \\ \{u_i(g(m_i, \nu_{-i}(t_{-i})), \hat{\theta}(t_i, t_{-i})) + \tau_i(m_i, \nu_{-i}(t_{-i}))\} \end{array} \right\} \bar{\pi}_i(t_i)[t_{-i}]. \end{aligned}$$

Then, we prove the inequality in (36) in the following two steps.

Step 1:

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

From the induction hypothesis, for every $i \in I$ and $t_i \in \bar{T}_i$, if $m_i \in S_i^\infty \hat{W}^\infty(t_i | \bar{\mathcal{M}}^*, \bar{\mathcal{T}})$, then $\bar{m}_i^{k'} = t_i$ for any nonnegative integer $k' \leq k$. When $m_{-i} \notin \hat{M}_{-i}$, there exists some $j \neq i$ such

that $\bar{m}_j^{k+1} \neq \bar{m}_j^0$. We compute the expected loss in terms of payments for player i of type t_i when playing m_i rather than $\tilde{m}_i$:

$$\sum_{t_{-i}:\nu_{-i}(t_{-i})\notin \hat{M}_{-i}} \left\{\tau_i\left(\tilde{m}_i, \nu_{-i}(t_{-i})\right) - \tau_i\left(m_i, \nu_{-i}(t_{-i})\right)\right\} \bar{\pi}_i(t_i)[t_{-i}].$$

By choosing $\tilde{m}_i$ rather than m_i, player i will avoid the fine, η according to the transfer rule c_i^{k+1} (see (30)) and ξ according to the transfer rule c_i (see (29)). That is, for any $t_{-i} \in \bar{T}_{-i}$ such that $\nu_{-i}(t_{-i}) \notin \hat{M}_{-i}$,

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

In terms of outcome function $\bar{g}(\cdot)$ of the mechanism $\bar{\mathcal{M}}$, we have

$$\begin{aligned} &\sum_{t_{-i}:\nu_{-i}(t_{-i})\notin \hat{M}_{-i}} \frac{1}{K}\left\{u_i(f(m_i^{k+1}, \nu_{-i}^{k+1}(t_{-i})), \hat{\theta}(t_i, t_{-i}))\right\} \bar{\pi}_i(t_i)[t_{-i}] \\ - &\sum_{t_{-i}:\nu_{-i}(t_{-i})\notin \hat{M}_{-i}} \frac{1}{K}\left\{u_i(f(\tilde{m}_i^{k+1}, \nu_{-i}^{k+1}(t_{-i})), \hat{\theta}(t_i, t_{-i}))\right\} \bar{\pi}_i(t_i)[t_{-i}] \leq \frac{D}{K} \end{aligned} \tag{38}$$

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

Since $\xi > D/K$ by (34), we have

$$\eta + \xi > \frac{D}{K}. \tag{39}$$

This completes Step 1.

Step 2:

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

When $m_{-i} \in \hat{M}_{-i}$, for any $j \neq i$, we have $\bar{m}_j^{k+1} = \bar{m}_j^0$. From the induction hypothesis, for every $j \in I$ and $t_j \in \bar{T}_j$, if $m_j \in S_j^k \hat{W}^\infty(t_i|\bar{\mathcal{M}}^*, \bar{\mathcal{T}})$, then $\bar{m}_j^{k'} = t_j$, for any nonnegative integer $k' \leq k$. We compute the expected loss in terms of payments for player i of type t_i when playing m_i rather than $\tilde{m}_i$:

$$\sum_{t_{-i}:\nu_{-i}(t_{-i})\in \hat{M}_{-i}} \left\{\tau_i\left(\tilde{m}_i, \nu_{-i}(t_{-i})\right) - \tau_i\left(m_i, \nu_{-i}(t_{-i})\right)\right\} \bar{\pi}_i(t_i)[t_{-i}]$$

Consider $\nu_{-i} : \bar{T}_{-i} \to M_{-i}$ such that $\nu_{-i}(t_{-i}) \in \hat{M}_{-i}$ for any $t_{-i} \in \bar{T}_{-i}$. By choosing $\tilde{m}_i$ rather than m_i, player i will avoid the fine, η according to the transfer rule c_i^{k-1}. Note that the

message $\tilde{m}_i$ triggers the fine ξ for player i only if the message m_i also triggers ξ. Hence, the expected loss in terms of payments from choosing m_i rather than $\tilde{m}_i$ in terms of the transfer rule $\tau(\cdot)$ is

$$\tau_i(\tilde{m}_i, m_{-i}) - \tau_i(m_i, m_{-i}) \geq \eta$$

for any $m_{-i} \in \hat{M}_{-i}$. Therefore, when playing m_i rather than $\tilde{m}_i$, the expected loss in terms of payments is bounded below by η.

In terms of outcome function $\bar{g}(\cdot)$ of the mechanism $\bar{\mathcal{M}}$, the possible gain for player i of type t_i to report m_i rather than $\tilde{m}_i$ is

$$\frac{1}{K} \sum_{t_{-i}:\nu_{-i}(t_{-i})\in\hat{M}_{-i}} \left\{ u_i(f(\bar{m}_i^{k+1}, \nu_{-i}^{k+1}(t_{-i})), \hat{\theta}(t_i, t_{-i})) - u_i(f(\tilde{m}_i^{k+1}, \nu_{-i}^{k+1}(t_{-i})), \hat{\theta}(t_i, t_{-i})) \right\} \bar{\pi}_i(t_i)[t_{-i}],$$

because $\tilde{m}_i$ differs from m_i only in the $(k+1)$st announcement. That is, by playing m_i rather than $\tilde{m}_i$, the possible gain for player i of type t_i is bounded above by 0 because the SCF f is incentive compatible and $\nu_{-i}^{k+1}(t_{-i}) = \bar{m}_{-i}^0$ is truthful. This completes Step 2. ■

Note that the argument of Claim 1 also shows that the truth-telling strategy profile $(\sigma_i)_{i\in I}$ with $\sigma_i(t_i) = (t_i, ..., t_i)$ indeed constitutes a strict Bayes Nash equilibrium in the game $U(\bar{\mathcal{M}}, \bar{\mathcal{T}})$. Let m be a message profile in $S^\infty\hat{W}^\infty(t|\bar{\mathcal{M}}^*, \bar{\mathcal{T}})$. To sum up, Claim 1 shows that $\bar{m}^k = t$ for any nonnegative integer $k \leq K$. Moreover, $m^l = t$ for every $l = 1, 2, ..., \bar{l} + 3$ and hence $e(m) = 0$. It follows that $g(m) = f(t)$. Finally, since $|\bar{\tau}_i(m)| < \bar{\tau}$ by (35) and $\bar{\tau}$ satisfies (11), it follows from Proposition 2 that $|\tau_i(m)| \leq \hat{\tau}$. Moreover, as $\hat{\tau}/3$ in Proposition 2 is the bound of transfer rule of $\mathcal{M}^*$, we can make $\hat{\tau}$ arbitrarily small by Lemma 3. Hence, we complete the proof of Proposition 1.

A.4 Proof of Proposition 3

Proposition 3: *Fix any model $\mathcal{T}$ such that $\bar{\mathcal{T}} \subset \mathcal{T}$ and a mechanism $\mathcal{M}$. Then, for any $t \in \bar{T}$ and any sequence $\{t^n\}_{n=0}^\infty$ in $\mathcal{T}$ such that $t^n \to_p t$, we have $S^\infty\hat{W}^\infty(t^n|\mathcal{M}, \mathcal{T}) \subset S^\infty\hat{W}^\infty(t|\mathcal{M}, \mathcal{T})$ for any n large enough.*

Proof. Since $\mathcal{M}$ is finite, there is a nonnegative integer k^* such that $S^k\hat{W}^\infty(t|\mathcal{M}, \mathcal{T}) = S^\infty\hat{W}^\infty(t|\mathcal{M}, \mathcal{T})$ for every $k \geq k^*$ and $t \in \bar{T}$. Thus, it suffices to show that for each nonnegative integer k, type profile $t \in \bar{T}$, and sequence $\{t^n\}_{n=0}^\infty$ in $\mathcal{T}$ such that $t^n \to_p t$ as $n \to \infty$, there exists a natural number $N_k \in \mathbb{N}$ such that, for any $n \geq N_k$, we have $S^k\hat{W}^\infty(t^n|\mathcal{M}, \mathcal{T}) \subset S^k\hat{W}^\infty(t|\mathcal{M}, \mathcal{T})$. We prove this by induction. We observe that

$\hat{W}_i^{\infty}(\hat{\theta}_i(t_i)|\mathcal{M}) = \hat{W}_i^{\infty}(\hat{\theta}_i(t_i')|\mathcal{M})$ whenever $\hat{\theta}_i(t_i) = \hat{\theta}_i(t_i')$ and $\hat{\theta}_i(t_{i,n}) = \hat{\theta}_i(t_i)$ for any n sufficiently large. Hence, the claim is true for $k = 0$. Now suppose that the claim holds for $k \geq 0$ and we will show that the claim is also valid for $k+1$.

Fix $m_i \notin S_i^{k+1}\hat{W}^{\infty}(t_i|\mathcal{M},\mathcal{T})$. Let $\bar{\Sigma}_{-i}$ be the set of conjectures $\bar{\nu}_{-i} : \bar{T}_{-i} \to \Delta(M_{-i})$ such that $\bar{\nu}_{-i}(t_{-i}) \in S_{-i}^{k}\hat{W}^{\infty}(t_{-i}|\mathcal{M},\mathcal{T})$ for every $t_{-i} \in \bar{T}_{-i}$. Then, there is $\alpha_i \in \Delta(M_i)$ such that

$$\beta \equiv \min_{\bar{\nu}_{-i} \in \bar{\Sigma}_{-i}} \{V_i((\alpha_i, \bar{\nu}_{-i}), t_i) - V_i((m_i, \bar{\nu}_{-i}), t_i)\} > 0. \tag{40}$$

where the minimum is attained since $\bar{\Sigma}_{-i}$ is compact. Let $(t_{-i})^{\varepsilon}$ denotes an open ball consisting of the set of types t_{-i}' whose $(k-1)$st order beliefs are ε-close to those of types t_{-i}. Since $\bar{T}_{-i}$ is a finite set, by the induction hypothesis, there is some $\varepsilon_1 > 0$ such that $S^k\hat{W}_{-i}^{\infty}(t_{-i}'|\mathcal{M},\mathcal{T}) \subset S^k\hat{W}_{-i}^{\infty}(t_{-i}|\mathcal{M},\mathcal{T})$ for every $t_{-i}' \in \bigcup_{t_{-i} \in \bar{T}_{-i}} (t_{-i})^{\varepsilon_1}$ and every $t_{-i} \in \bar{T}_{-i}$. Moreover, since $\bar{T}_{-i}$ is a finite set, for all $t_{-i}, s_{-i} \in \bar{T}_{-i}$ with $s_{-i} \neq t_{-i}$, we can also choose $\varepsilon_2 > 0$ so that we have (1) $(t_{-i})^{\varepsilon_2} \cap (s_{-i})^{\varepsilon_2} = \varnothing$; and (2)

$$\varepsilon_2 < \min\left\{ \frac{\beta}{3D\left|\bar{T}_{-i}\right|}, \min_{t_{-i} \in \mathrm{supp}(\bar{\pi}_i(t_i))} \frac{\bar{\pi}_i(t_i)[t_{-i}]}{2} \right\}. \tag{41}$$

Since $t^n \to_p t$, for any $\varepsilon > 0$, there is n sufficiently large such that for any positive $\varepsilon < \min\{\varepsilon_1, \varepsilon_2\}$, we have[25]

$$|\pi_i(t_{i,n})[(t_{-i})^{\varepsilon}] - \bar{\pi}_i(t_i)[t_{-i}]| < \varepsilon, \forall t_{-i} \in \bar{T}_{-i}. \tag{42}$$

Now consider an arbitrary conjecture $\nu_{-i} : T_{-i} \to M_{-i}$ with $\nu_{-i}(t_{-i}') \in S^k\hat{W}_{-i}^{\infty}(t_{-i}'|\mathcal{M},\mathcal{T})$ for every $t_{-i}' \in T_{-i}$. Based on ν_{-i}, if $t_{-i} \in \bar{T}_{-i}$ with $\bar{\pi}_i(t_i)[t_{-i}] > 0$, we define

$$\bar{\nu}_{-i}(t_{-i})[m_{-i}] = \frac{\pi_i(t_{i,n})\left[\left\{t_{-i}' \in (t_{-i})^{\varepsilon} : \nu_{-i}(t_{-i}') = m_{-i}\right\}\right]}{\pi_i(t_{i,n})[(t_{-i})^{\varepsilon}]}; \tag{43}$$

and if $\bar{\pi}_i(t_i)[t_{-i}] = 0$, let $\bar{\nu}_{-i}(t_{-i})$ assign probability one to some $m_{-i} \in S_{-i}^{k}\hat{W}^{\infty}(t_{-i}|\mathcal{M},\mathcal{T})$. It follows from the choice of ε and n that

$$|V_i((\alpha_i, \nu_{-i}), t_{i,n}) - V_i((\alpha_i, \bar{\nu}_{-i}), t_i)| < \beta/3; \tag{44}$$

$$|V_i((m_i, \nu_{-i}), t_{i,n}) - V_i((m_i, \bar{\nu}_{-i}), t_i)| < \beta/3. \tag{45}$$

[25] This follows from the fact that the Prohorov distance between $t_{i,n}$ and t_i converges to 0. See (Dudley, 2002, pp. 398, 411).

Hence, it follows from (40), (44), and (45) that

$$\begin{aligned} & V_i((\alpha_i, \nu_{-i}), t_{i,n}) - V_i((m_i, \nu_{-i}), t_{i,n}) \\ = \; & V_i((\alpha_i, \bar{\nu}_{-i}), t_i) - V_i((m_i, \bar{\nu}_{-i}), t_i) + [V_i((\alpha_i, \nu_{-i}), t_{i,n}) - V_i((\alpha_i, \bar{nu}_{-i}), t_i)] \\ & + [V_i((m_i, \bar{\nu}_{-i}), t_i) - V_i((m_i, \nu_{-i}), t_{i,n})] \\ > \; & \beta - \frac{\beta}{3} - \frac{\beta}{3} = \frac{\beta}{3} > 0. \end{aligned}$$

Since ν_{-i} is chosen arbitrarily, we conclude that $m_i \notin S_i^{k+1}\hat{W}^{\infty}(t_{i,n}|\mathcal{M}, \mathcal{T})$. ■

A.5 Proof of Proposition 4

Proposition 4: *Fix any model T such that $\bar{\mathcal{T}} \subset T$ and any mechanism $\mathcal{M}$. Then, there exists an equilibrium σ in the game $U(\mathcal{M}, \mathcal{T})$ such that for any player i of type t_i, we have $\sigma_i(t_i) \in \hat{W}_i^{\infty}(\hat{\theta}_i(t_i)|\mathcal{M})$.*

We start from providing two definitions which we use in Lemma 6 below. We then invoke Lemma 6 to prove Proposition 4. Here, we identify $U(\mathcal{M}, \mathcal{T})$ with an agent-normal form game where each $t_i \in T_i$ is a player and M_i the set of actions of t_i; moreover, given any $\sigma_{-i} : T_{-i} \to M_{-i}$, the payoff of t_i of playing a message m_i is denoted by $V_i((m_i, \sigma_{-i}), t_i)$.

Definition 11 *A ζ-**perturbation** of $U(\mathcal{M}, \mathcal{T})$, which we denote by $U^{\zeta}(\mathcal{M}, \mathcal{T})$, is another agent-normal form game with $|V_i^{\zeta}((m_i, \sigma_{-i}), t_i) - V_i((m_i, \sigma_{-i}), t_i)| \leq \zeta$ for every $m_i \in M_i$, every $\sigma_{-i} : T_{-i} \to M_{-i}$, and every $t_i \in T_i$.*

The following definition is a restatement of Property (10.3.1) of Van Damme (1991) in the agent-normal form game $U(\mathcal{M}, \mathcal{T})$.

Definition 12 *Let $U(\mathcal{M}, \mathcal{T})$ be an incomplete information game induced from mechanism $\mathcal{M}$ and model $\mathcal{T}$. We say that a set of (Nash) equilibria F of game $U(\mathcal{M}, \mathcal{T})$ is **quasi-stable** if, for every $\varepsilon > 0$, there exists $\zeta > 0$ such that for every ζ-perturbation $U^{\zeta}(\mathcal{M}, \mathcal{T})$ of $U(\mathcal{M}, \mathcal{T})$, there is an equilibrium of $U^{\zeta}(\mathcal{M}, \mathcal{T})$ which is within ε-distance from the set F.*

Proposition 4 follows from Lemma 6 below. The lemma below essentially restates a well known result that each Kohlberg-Mertens stable set contains a stable set of any truncated game obtained by eliminating a weakly dominated strategy. Indeed, the argument in Kohlberg and Mertens (1986) remains valid in the agent normal-form of any game such

that there are countably many players (where each player corresponds to a type) and each player has finitely many pure messages. We reproduce the proof to make the argument self-contained.

Lemma 6 *Let F be a quasi-stable set of equilibria in the game $U(\mathcal{M},\mathcal{T})$. Assume that $m_i' \notin \hat{W}_i^1(\theta_i|\mathcal{M})$. Then, there is a quasi-stable set of equilibria $F' \subset F$ such that $\sigma_i(m_i') = 0$ for every equilibrium σ in F'.*

Proof. Let $F' = \{\sigma \in F : \sigma_i(t_i)[m_i'] = 0\}$. We shall show that F' is a quasi-stable set of equilibria in $U(\mathcal{M},\mathcal{T})$. Since $m_i' \notin \hat{W}_i^1(\theta_i|\mathcal{M})$, there is some $\alpha_i \in \Delta(M_i)$ such that α_i weakly dominates m_i' in the game $U(\mathcal{M},\mathcal{T})$.

Fix $\varepsilon > 0$. Let $U^{\zeta}(\mathcal{M},\mathcal{T})$ be a ζ-perturbation of the game $U(\mathcal{M},\mathcal{T})$ for some $\zeta > 0$. In addition, we add $\zeta' > 0$ to the corresponding payoff from player i's messages other than m_i'. That is, for any conjecture $\sigma_{-i} : T_{-i} \to M_{-i}$, we satisfy the following two properties: (1) $V_i^{\zeta,\zeta'}((m_i,\sigma_{-i}),t_i) = V_i^{\zeta}((m_i,\sigma_{-i}),t_i) + \zeta'$, for all $m_i \neq m_i'$; and (2) $V_i^{\zeta,\zeta'}((m_i',\sigma_{-i}),t_i) = V_i^{\zeta}((m_i',\sigma_{-i}),t_i)$. Thus, we obtain $U^{\zeta,\zeta'}(\mathcal{M},\mathcal{T})$ as a ζ'-perturbation of the game $U^{\zeta}(\mathcal{M},\mathcal{T})$. Since F is quasi-stable in $U(\mathcal{M},\mathcal{T})$, there exist $\zeta > 0$ and $\zeta' > 0$ small enough so that the game $U^{\zeta,\zeta'}(\mathcal{M},\mathcal{T})$ has a Bayes Nash equilibrium $\sigma^{\zeta,\zeta'}$ which is within ε-distance from F. Moreover, in the game $U^{\zeta,\zeta'}(\mathcal{M},\mathcal{T})$, for any type t_i with conjecture $\sigma : T_{-i} \to M_{-i}$, we have

$$V_i^{\zeta,\zeta'}((\alpha_i,\sigma_{-i}),t_i) > V_i^{\zeta,\zeta'}((m_i',\sigma_{-i}),t_i).$$

Therefore, m_i' cannot be a best response to $\sigma_{-i}^{\zeta,\zeta'}$ for player i of type t_i, i.e., $\sigma_i^{\zeta,\zeta'}(t_i)[m_i'] = 0$. For any $\zeta' > 0$, $\sigma^{\zeta,\zeta'}$ is within ε-distance from F. Thus, we have that $\sigma^{\zeta,0}$ is a Bayes Nash equilibrium in the game $U^{\zeta}(\mathcal{M},\mathcal{T})$ such that $\sigma^{\zeta,0}$ is within ε-distance from F'. In other words, F' is also quasi-stable. ■

We now turn to prove Proposition 4.

Proof of Proposition 4. It follows from the closed graph property of the Nash equilibrium correspondence that the set of Nash equilibria in the agent normal-form game of $U(\mathcal{M},\mathcal{T})$ is quasi-stable (see Van Damme (1991)). Hence, the proposition is proved by repeatedly applying Lemma 6 after we remove each of the (finitely many) weakly dominated message in deriving $\hat{W}^l$ for each l (where within round l, the order of removal does not matter). ■

A.6 Example for Section 5.1

Example 1 *Suppose that there are two agents:* $\{1,2\}$*; two states* $\{\alpha,\beta\}$*; and three pure alternatives:* $\{a,b,c\}$*. Define an SCF such that* $f(\alpha)=a$ *and* $f(\beta)=b$*. Agents' utilities across different states are described in the following table:*

v_1	α	β
a	2	1
b	2	1
c	-1	2
d	-1	-1

v_2	α	β
a	1	1
b	1	1
c	2	2
d	-1	-1

The information is complete, namely that the agents have common knowledge about the state, whether it is α *or* β*. We identify agent 1's payoff type with the states and omit agent 2's payoff type since agent 2 has state-independent preference.*

Claim 2 *f is incentive compatible.*

Proof. To see f is incentive compatible, consider a direct revelation mechanism, $\tilde{f}:\Theta^2\to\{a,b,c,d\}$ such that $\tilde{f}(\alpha,\alpha)=a$, $\tilde{f}(\beta,\beta)=b$, and $\tilde{f}(\alpha,\beta)=\tilde{f}(\beta,\alpha)=d$. By construction, $\tilde{f}(\alpha,\alpha)=a$ and the message profile (α,α) is an equilibrium at state α, and likewise, $\tilde{f}(\beta,\beta)=b$ and the message profile (β,β) is an equilibrium at state β. ■

Claim 3 *f is not Maskin-monotonic.*

Proof. Consider $f(\beta)\neq f(\alpha)$. Since agent 2's preference is state-independent, the only possible whistle blower is agent 1. For agent 1, however, at state β the outcomes which are worse than $f(\beta)=b$ are a and d. However, a and d are both worse than b for agent 1 at state α. In addition, with the reference to outcome $f(\beta)=b$, the utility difference for agent 1 at state β is getting larger than at state α. Specifically, $u_1(b,\beta)=u_1(a,\beta)$, $u_1(b,\beta)-u_1(c,\beta)=-1$,and $u_1(b,\beta)-u_1(d,\beta)=1$; while $u_1(a,\alpha)=u_1(b,\alpha)$, $u_1(c,\alpha)-u_1(b,\alpha)=-3$, and $u_1(b,\beta)-u_1(d,\beta)=3$. Hence, any outcome with transfer which is weakly worse than b at state β remains weakly worse than b at state α. Hence, Maskin-monotonicity fails. ■

Claim 4 *f is not implementable in $S^\infty\tilde{W}^\infty$ with transfers.*

Proof. This follows from Proposition 5 and Claim 4. ■

Claim 5 *f is implementable in $S^\infty\hat{W}^\infty$ with arbitrarily small transfers and also continuously implementable with arbitrarily small transfers.*

Proof. Since f is incentive compatible and the agents' values are private and different payoff types θ_i and θ_i' induce different preferences over lottery allocations $\Delta(A)$, by Corollary 1, f is implementable in $S^\infty\hat{W}^\infty$ with arbitrarily small transfers and continuously implementable with arbitrarily small transfers. ■

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.

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.

Bergemann, D. and S. Morris (2005): "Robust Mechanism Design," *Econometrica*, 73, 1771–1813.

——— (2008): "Interim Rationalizable Implementation," *mimeo*.

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

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

——— (2011): "Robust Implementation in General Mechanisms," *Games and Economic Behavior*, 71, 261–281.

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

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

Chen, Y.-C., T. Kunimoto, and Y. Sun (2016): "Implementation with Transfers," *mimeo*.

Chen, Y.-C., T. Kunimoto, Y. Sun, and S. Xiong (2021): "Rationalizable Implementation in Finite Mechanisms," *Games and Economic Behavior*, 129, 181–197.

Chen, Y.-C., S. Takahashi, and S. Xiong (2015): "Robust Refinement of Rationalizability," *mimeo*.

Chen, Y.-C. and S. Xiong (2011): "The Genericity of Beliefs-Determine-Preferences Models Revisited," *Journal of Economic Theory*, 146, 751–761.

Chung, K.-S. and J. C. Ely (2019): "Efficient and Dominance Solvable Auctions with Interdependent Valuations," *The Journal of Mechanism and Institution Design*, 4, 1–38.

Dasgupta, P. and E. Maskin (2000): "Efficient Auctions," *The Quarterly Journal of Economics*, 115, 341–388.

Dekel, E., D. Fudenberg, and S. Morris (2007): "Interim Correlated Rationalizability," *Theoretical Economics*, 2, 15–40.

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

Fudenberg, D., D. M. Kreps, and D. K. Levine (1988): "On the Robustness of Equilibrium Refinements," *Journal of Economic Theory*, 44, 354–380.

Heifetz, A. and Z. Neeman (2006): "On the Generic (Im) possibility of Full Surplus Extraction in Mechanism Design," *Econometrica*, 213–233.

Jehiel, P., M. Meyer-ter-Vehn, and B. Moldovanu (2012): "Locally robust implementation and its limits," *Journal of Economic Theory*, 147, 2439–2452.

Jehiel, P., M. Meyer-ter-Vehn, B. Moldovanu, and W. R. Zame (2006): "The Limits of Ex Post Implementation," *Econometrica*, 74, 585–610.

Kohlberg, E. and J.-F. Mertens (1986): "On the Strategic Stability of Equilibria," *Econometrica: Journal of the Econometric Society*, 1003–1037.

Kunimoto, T., R. Saran, and R. Serrano (2020): "Interim Rationalizable Implementation of Functions," *mimeo*.

McAfee, R. P. and P. J. Reny (1992): "Correlated Information and Mechanism Design," *Econometrica: Journal of the Econometric Society*, 395–421.

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

Meyer-ter-Vehn, M. and S. Morris (2011): "The Robustness of Robust Implementation," *Journal of Economic Theory*, 146, 2093–2104.

Neeman, Z. (2004): "The Relevance of Private Information in Mechanism Design," *Journal of Economic theory*, 117, 55–77.

Ollár, M. and A. Penta (2017): "Full Implementation and Belief Restrictions," *American Economic Review*, 107, 2243–77.

——— (2019): "Implementation via Transfers with Identical but Unknown Distributions," .

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.

Rudin, W. (1987): *Real and Complex Analysis*, McGraw-Hill.

Van Damme, E. (1991): *Stability and Perfection of Nash Equilibria*, vol. 339, Springer.