# Rationalizable Incentives: Interim Rationalizable Implementation of Correspondences*

Takashi Kunimoto[†], Rene Saran[‡], and Roberto Serrano[§]

This Version: February 2025

**Abstract**

When the normative goals for a set of agents can be summarized in a set-valued rule and agents take actions that are rationalizable, a new theory of incentives emerges in which standard Bayesian incentive compatibility (BIC) is relaxed significantly. The paper studies the interim rationalizable implementation of social choice sets with a Cartesian product structure, a leading example thereof being ex-post efficiency. Setwise incentive compatibility (setwise IC), much weaker than BIC, is shown to be necessary for implementation. Setwise IC enforces incentives flexibly within the entire correspondence, instead of the pointwise enforcement entailed by BIC. Sufficient conditions, while based on the existence of SCFs in the correspondence that make truthful revelation a dominant strategy, are shown to be permissive to allow the implementation of ex-post efficiency in many settings where equilibrium implementation fails (e.g., bilateral trading, multidimensional signals). Furthermore, this success comes at little cost: all our mechanisms are well behaved, in the sense that best responses always exist.

*JEL Classification:* C72, D78, D82.

*Keywords:* rationalizability; implementation; correspondences; setwise incentive compatibility; setwise dominance; ex-post efficiency.

*This paper subsumes Kunimoto and Serrano (2020). This research is supported by the Singapore Ministry of Education under MOE Academic Research Fund Tier 2 (MOE-T2EP402A20-0007).

[†]School of Economics, Singapore Management University, Singapore; tkunimoto@smu.edu.sg

[‡]Department of Economics, University of Cincinnati, Cincinnati, OH, USA; rene.saran@uc.edu

[§]Department of Economics, Brown University, Providence, RI, USA; roberto_serrano@brown.edu

# 1 Introduction

The theory of incentives is one of the cornerstones of modern economic theory. One of its central conditions is Bayesian incentive compatibility (BIC), viewed as a minimal condition necessary for the implementation of any set of rules or contracts under incomplete information.[1] BIC stipulates that, in the direct mechanism associated with a given rule, truth-telling be a best response for every type to the belief that the others are also telling the truth. Indeed, in the elicitation of the private information held by a group of agents, the mechanism designer should at least hope that, under common knowledge of rationality, if all but one of the agents are going to be truthful, so will be the remaining agent; this is the rationale for BIC as a minimal desideratum. As such, BIC is an equilibrium condition, based on the rational-expectations assumption, by which all agents share exactly the same belief about how the others will play, i.e., the truthful equilibrium belief. The restriction imposed by BIC can sometimes be quite severe, being one of the culprits for impossibility results in some settings (e.g., Myerson and Satterthwaite (1983), Jehiel and Moldovanu (2001)).

Suppose instead that, although the mechanism designer continues to assume that rationality is commonly known by the agents, she does not insist on the rational-expectations assumption. Under incomplete information, this means that she expects the agents to use (interim) rationalizable strategies.[2] In general, if the designer's goals are summarized by a social choice set (SCS), she would seek to design a mechanism whose set of outcomes resulting from agents choosing rationalizable messages will coincide with the SCS of interest.[3] This is the notion of full implementation in interim rationalizable strategies.[4] We contend that, when the normative goals for a set of agents can be summarized in a set-valued rule and agents take actions that are rationalizable, a new theory of incentives emerges in which standard BIC is relaxed significantly. This is the central message of the

[1] See, e.g., Dasgupta, Hammond, and Maskin (1979), Myerson (1979, 1981), d'Aspremont and Gerard-Varet (1979), Green and Laffont (1979), Harris and Townsend (1981), and Holmstrom and Myerson (1983) for seminal contributions to this fundamental idea.

[2] See Bernheim (1984), Pearce (1984), Brandenburger and Dekel (1987), and Lipman (1994) for the formalization of the idea of rationalizability; in this paper, we use the interim extension to games with incomplete information of Dekel, Fudenberg, and Morris (2007).

[3] If her goal is a unique outcome in each state, this is described by a social choice function (SCF).

[4] For full implementation in rationalizable strategies in complete information environments, see Bergemann, Morris, and Tercieux (2011), Xiong (2023), and Jain, Korpela, and Lombardi (2023) for SCFs, and Kunimoto and Serrano (2019) and Jain (2021) for correspondences. For interim rationalizable implementation of SCFs, see Bergemann and Morris (2008) and Oury and Tercieux (2012) for early contributions, Kunimoto, Saran, and Serrano (2024) for an almost characterization, and Jain, Lombardi, and Penta (2024) for a robustness notion building on the double implementation in Kunimoto, Saran, and Serrano's (2024) Section 8.

current work.

In this paper, we confine our attention to SCSs with a Cartesian product structure (we leave the problem of general SCSs to be studied elsewhere). Equivalently, we begin by specifying a social choice correspondence (SCC) and seek to implement *all* its selections. For a leading illustration of our theory in this paper, the reader is invited to think of ex-post Pareto efficiency (as opposed to ex-ante or interim efficiency). We shall also insist on the use of "well behaved" mechanisms, which for us will mean mechanisms satisfying the property that best responses always exist (we shall describe two versions of this property, which we will call BRP and weak BRP).

Our first main result may seem striking at first: BIC is not necessary for implementation in rationalizable strategies in mechanisms satisfying the weak BRP.[5] Rather, the necessary condition is what we call *setwise incentive compatibility* (setwise IC). As opposed to the standard "pointwise" BIC, setwise IC enforces incentives more flexibly through the creation of chains of incentive constraints within the correspondence. Specifically, under the assumption that all other agents are truthful, for every agent $i$ and every $f^{'}$ in the SCS of interest, there exists $f$ in the same SCS such that the truthful report behind $f$ is better for every type $t_i$ of agent $i$ than any manipulated version of $f^{'}$. This seems to make good sense: the designer should not exhibit a strict preference among different outcomes in the SCC, so enforcing incentives in this more flexible manner should suit her normative goal. Of course, setwise IC reduces to BIC in the case of SCFs.

Instead of attempting to close the gap between necessary and sufficient conditions for implementation, which sometimes leads to the use of unrestricted mechanisms, we pursue sufficient conditions that, while being permissive, rely on mechanisms with the best-response property (BRP) or its weaker version (weak BRP). BRP means that best responses always exist against any mixed strategy profile of the other agents, while weak BRP asserts this existence against pure strategy profiles.

Indeed, our sufficiency results rely on a dominance version of setwise IC. The first such condition is termed *setwise dominance*. Setwise dominance requires the existence of an SCF for each agent that strictly dominates every other SCF in the SCC, i.e., where the agent wishes to strictly tell the truth over any misrepresentation of her type, regardless of the type reports of others. We show in our first sufficiency result that setwise dominance is sufficient for interim implementation in rationalizable strategies using mechanisms with the weak BRP. If the dominating SCFs are independent of the types of other agents, setwise dominance is strengthened to *setwise independent dominance*, which is essen-

[5]See also de Clippel, Saran, and Serrano (2019) and Kneeland (2022) for results showing that BIC is not necessary for level-$k$ implementation of an SCS.

tially sufficient for rationalizable implementation with mechanisms satisfying the BRP, as demonstrated in our second sufficiency result.

Our concept of implementation allows for a flexible way to overcome a number of impossibility results, as we illustrate in Section 7. This section contains examples to show how one can get to implement –at least approximately– the ex-post Pareto correspondence in bilateral trading with private values, in a standard exchange economy with interdependent values, or in environments with multidimensional signals, settings in which different impossibility results had been furnished for equilibrium implementation.[6] The flexibility of the approach stems from the use of rationalizability. Rather than insisting on the pointwise implementation of each SCF –as BIC entails even for SCSs–, we allow the designer to come up with institutions where incentives can be provided using different SCFs in the SCC. These are the *rationalizable incentives* enunciated in the paper's title, which sometimes allow for the implementation of SCCs where violations of BIC can be severe.

The rest of the paper is organized as follows. In Section 2, we introduce the general notation. Section 3 proposes the concept of interim implementation in rationalizable strategies. In Section 4, we prove the necessity of setwise IC for interim rationalizable implementation with mechanisms satisfying the weak BRP. Section 5 relies on setwise dominance to prove a sufficiency result for that class of mechanisms. In Section 6, we provide a different sufficiency result, this time for mechanisms satisfying the BRP, based on setwise independent dominance. Section 7 contains the examples showcasing our approach, and Section 8 concludes with several remarks on extensions of our results. While most proofs are in the main text, a couple of them are relegated to the appendix.

# 2 Preliminaries

Let $I = \{1, \dots, n\}$ denote the finite set of agents or players, and $T_i$ be the finite set of types of agent $i \in I$. We endow each $T_i$ with the discrete topology.[7] Let $T \equiv T_1 \times \cdots \times T_n$, and $T_{-i} \equiv T_1 \times \cdots \times T_{i-1} \times T_{i+1} \times \cdots \times T_n$.[8] Let $\Delta(T_{-i})$ denote the set of probability

[6]Thus, the paper continues to develop the theme that rationalizable institutions are more permissive than equilibrium ones (see Kunimoto and Serrano (2019), Kunimoto and Saran (2020), and Kunimoto, Saran, and Serrano (2024) for previous steps in this overall theme).

[7]All finite sets will be endowed the the discrete topology and all topological spaces are endowed with the Borel $\sigma$-algebra.

[8]Similar notation will be used for the products of other sets. All product sets are endowed with the product topology.

distributions on $T_{-i}$.[9] Each agent $i$ has a system of "interim" beliefs that is expressed as a function $\pi_i : T_i \to \Delta(T_{-i})$. Then, we call $(T_i, \pi_i)_{i \in I}$ a *type space*.

Let $A$ denote the set of pure outcomes, which is independent of the information state and assumed to be a compact metric space. Agent $i$'s state dependent von Neumann-Morgenstern utility function is denoted $u_i : \Delta(A) \times T \to \mathbb{R}$, which is assumed to be continuous. Note that the utility function is bounded due to the compactness of its domain and the assumption of continuity. We can now define an *environment* as $\mathcal{E} = (A, \{u_i, T_i, \pi_i\}_{i \in I})$. The environment is one of *private values* if the utility of each player $i \in I$ is independent of the other players' types $t_{-i} \in T_{-i}$. Otherwise, the environment has *interdependent values*. In a private-values environment, we simplify notation and write the expected utility of player $i$ as a function of the lottery and her own type, i.e., $u_i : \Delta(A) \times T_i \to \mathbb{R}$.

The designer's goals are contingent on the realization of the agents' types and, in this paper, are described by a social choice correspondence. Formally, a *social choice correspondence (SCC)* is a nonempty- and compact-valued mapping $F : T \to 2^{\Delta(A)}$. The SCC $F$ is *deterministic* if $F(t) \subseteq A$, for all $t \in T$. Let $T^* \subseteq T$ be the set of states that the designer cares about. Consider any two SCCs $F$ and $F'$. We say that $F$ and $F'$ are *equivalent* (denoted by $F \approx F'$) if $F(t) = F'(t)$, for all $t \in T^*$.

A *social choice function* (SCF) is a single-valued function $f : T \to \Delta(A)$. Notice that a selection of an SCC is an SCF. A *social choice set* (SCS) $H$ is a collection of SCFs.

In environments with incomplete information, it is standard to summarize the designer's goals by a SCS. We plan to undertake the study of general SCS's elsewhere. In the current paper, we take an admittedly more restrictive approach and study the implementability of SCC's, which means effectively that we are considering only SCS's with the Cartesian product property, in the sense that we seek to implement *all* selections of the given SCC. For our purposes, this will suffice in suggesting a first version of the new theory of incentives that the paper proposes.

That is, given the SCC $F$, define the *SCS generated by* $F$ as

$$H_F \equiv \{f : f \text{ is a selection of } F\}.$$

Notice that $H_F = \times_{t \in T} F(t)$. As $F$ is compact-valued, $H_F$ is compact in the product topology.

A *mechanism* (or *game form*) $\Gamma = ((M_i)_{i \in I}, g)$ describes: (i) a nonempty topological

[9] For any topological space $X$, we let $\Delta(X)$ denote the set of Borel probability measures on $X$. Recall that if $X$ is compact and metrizable, then $\Delta(X)$, under the weak$^*$ topology, is also compact and metrizable (Aliprantis and Border, 2006, Theorem 15.11).

space of messages $M_i$ for each agent $i \in I$, and (ii) a measurable outcome function $g : M \to \Delta(A)$, where $M = \times_{i \in I} M_i$. The mechanism $\Gamma$ is *compact* if $M_i$ is compact for all $i \in I$.

Let $\Gamma_f^{DM} = ((T_i)_{i \in I}, f)$ denote the *direct mechanism* associated with an SCF $f$, i.e., a mechanism where $M_i = T_i$, for all $i \in I$, and $g = f$. In the direct mechanism associated with the SCF $f$, the interim expected utility of agent $i$ of type $t_i$ who pretends to be of type $t_i^{'}$, while all other agents truthfully announce their types, is defined as:

$$U_i(f; t_i^{'} | t_i) \equiv \sum_{t_{-i} \in T_{-i}} \pi_i(t_i)[t_{-i}] u_i\big(f(t_i^{'}, t_{-i}), (t_i, t_{-i})\big).$$

Let $U_i(f|t_i) \equiv U_i(f; t_i|t_i)$.

# 3 Implementation in Interim Rationalizable Strategies

Fix a mechanism $\Gamma = ((M)_{i \in I}, g)$ and define a *message correspondence profile* $S = (S_1, \ldots, S_n)$, where each $S_i : T_i \to 2^{M_i}$ is measurable. We write $\mathcal{S}$ for the collection of message correspondence profiles. The collection $\mathcal{S}$ is a complete lattice with the natural ordering of set inclusion: $S \leq S^{'}$ if $S_i(t_i) \subseteq S_i^{'}(t_i)$, for all $i \in I$ and $t_i \in T_i$. The largest element is $\bar{S} = (\bar{S}_1, \ldots, \bar{S}_n)$, where $\bar{S}_i(t_i) = M_i$, for all $t_i \in T_i$ and $i \in I$. The smallest element is $\underline{S} = (\underline{S}_1, \ldots, \underline{S}_n)$, where $\underline{S}_i(t_i) = \emptyset$, for all $t_i \in T_i$ and $i \in I$.

We define the operator $b$ to eliminate never best responses. In what follows, $G(S_{-i})$ denotes the *graph* of $S_{-i}$. The operator $b : \mathcal{S} \to \mathcal{S}$ is thus defined as follows: For every $i \in I$ and $t_i \in T_i$,

$$b_i(S)[t_i] \equiv \left\{ m_i \in M_i : \begin{array}{l} \exists \lambda_i \in \Delta(T_{-i} \times M_{-i}) \text{ such that} \\ (1)\ \lambda_i\big(G(S_{-i})\big) = 1; \\ (2)\ \text{marg}_{T_{-i}} \lambda_i = \pi_i(t_i); \\ (3)\ m_i \in \arg\max_{m_i^{'}} \int_{T_{-i} \times M_{-i}} u_i\big(g(m_i^{'}, m_{-i}), (t_i, t_{-i})\big) d\lambda_i \end{array} \right\},$$

where $\lambda_i\big(G(S_{-i})\big) = 1$ means that $\lambda_i$ assigns probability 1 to a measurable subset of $G(S_{-i})$.

Observe that $b$ is increasing by definition: i.e., $S \leq S^{'} \Rightarrow b(S) \leq b(S^{'})$. By Tarski's fixed-point theorem, there is a largest fixed point of $b$, which we label $S^{\Gamma}$ and refer to as the *interim correlated rationalizability (ICR) correspondence*. That is, (i) $b(S^{\Gamma}) = S^{\Gamma}$ and (ii) $b(S) = S \Rightarrow S \leq S^{\Gamma}$.

A (pure) *strategy* for agent $i$ is a function $\sigma_i : T_i \to M_i$. (Notice that strategies are measurable since type spaces are finite.) For any message correspondence profile $S \in \mathcal{S}$ and agent $i \in I$, we let $\Sigma_i(S)$ denote the set of all strategies $\sigma_i$ of agent $i$ such that $\sigma_i(t_i) \in S_i(t_i)$, for all $t_i \in T_i$. Then $\Sigma(S) \equiv \times_{i\in I}\Sigma_i(S)$.

We have the following definition of implementation in interim rationalizable strategies:

**Definition 1.** A mechanism $\Gamma$ *implements the SCC $F$ in interim rationalizable strategies* if there exists an SCC $F^{'} \approx F$ such that the following two conditions hold:

1. For any $f \in H_{F'}$, there exists $\sigma \in \Sigma(S^\Gamma)$ such that $g(\sigma(t)) = f(t)$, for all $t \in T$.

2. For any $\sigma \in \Sigma(S^\Gamma)$, there exists $f \in H_{F'}$ such that $g(\sigma(t)) = f(t)$, for all $t \in T$.[10]

The SCC $F$ is *implementable in interim rationalizable strategies* if there exists a mechanism $\Gamma$ that implements $F$ in interim rationalizable strategies.

# 4 Necessity

In this section, we present a necessary condition for implementation in rationalizable strategies while restricting the designer to mechanisms that satisfy the following property.

**Definition 2.** The mechanism $\Gamma = ((M)_{i\in I}, g)$ satisfies the *weak best response property (weak BRP)* if for all $i \in I$, $t_i \in T_i$, and $\sigma_{-i} : T_{-i} \to M_{-i}$, we have

$$\arg\max_{m_i\in M_i} \sum_{t_{-i}\in T_{-i}} \pi_i(t_i)[t_{-i}]u_i\big(g(m_i, \sigma_{-i}(t_{-i})), (t_i, t_{-i})\big) \neq \emptyset.$$

Thus, the weak BRP requires that, for each type of each agent, a best response exist regardless of the *pure* strategies played by the other agents. The weak BRP is clearly satisfied by compact and continuous mechanisms because the utility functions are also continuous. However, if a mechanism is not continuous, it might not satisfy the weak BRP, even though it is compact.[11]

We note that the so-called integer and modulo games, often used as part of the implementing mechanisms proposed in the literature, as well as any mechanism involving these

[10] The requirement "for all $t \in T$" makes the statements of the results more transparent, while the requirement "$F^{'} \approx F$" takes care of the designer's preferences for $T^*$.

[11] For instance, in the standard quasilinear setting, the first-price auction with the usual tie-breaking rule (random assignment among the highest bidders) does not satisfy the weak BRP: imagine a bidder whose expected value for the object is 1 but believes that all other bidders are bidding 1/2 regardless of their types. Then the bidder does not have a best response against that belief.

games, satisfy the weak BRP.[12] Bergemann, Morris, and Tercieux (2011) and Serrano and Vohra (2005, 2010), among others, use what we call a stochastic integer game as part of their implementing mechanisms. In a stochastic integer game, each agent has to announce an integer; the higher the integer she announces, the higher the probability with which she can choose her best outcome; but, no matter how high an integer she announces, she cannot obtain her best outcome with probability 1. Thus, the mechanisms involving stochastic integer games do not satisfy the weak BRP.

**Definition 3.** The SCC $F$ satisfies *setwise incentive compatibility (setwise IC)* if for all $i \in I$ and $f^{'} \in H_F$, there exists an $f \in H_F$ such that

$$U_i(f|t_i) \geq U_i(f^{'}; t_i^{'}|t_i), \forall t_i, t_i^{'} \in T_i.$$

Note how setwise IC imposes incentive constraints for each type of each agent under the assumption that all other types of all other agents are truth-telling; but, to discipline each type $t_i$'s behavior, the condition does not fix a single SCF. Rather, for any SCF of interest, setwise IC allows the designer to find another SCF also of interest, i.e., some other selection of the SCC, that provides the incentives to tell the truth. Thus, setwise IC is considerably weaker than Bayesian incentive compatibility –BIC– (this requires BIC on each SCF $f \in H_F$); in principle, it is possible that each SCF $f \in H_F$ may violate BIC, while the entire SCC $F$ satisfies setwise IC. Of course, setwise IC reduces to the standard ("pointwise") BIC if we are focusing only on SCFs.

The main result of this section follows, showing that setwise IC is a necessary condition for implementation in rationalizable strategies by mechanisms satisfying the weak BRP:

**Theorem 1.** *If the SCC $F$ is implementable in rationalizable strategies by a mechanism with the weak BRP, then there exists an $F^{'} \approx F$ that satisfies setwise IC.*

*Proof.* Suppose that the mechanism $\Gamma$ satisfies the weak BRP and implements the SCC $F$ in interim rationalizable strategies. Then there exists an SCC $F^{'} \approx F$ such that the following two conditions hold:

1. For any $f \in H_{F'}$, there exists $\sigma \in \Sigma(S^{\Gamma})$ such that $g(\sigma(t)) = f(t)$, for all $t \in T$.

2. For any $\sigma \in \Sigma(S^{\Gamma})$, there exists $f \in H_{F'}$ such that $g(\sigma(t)) = f(t)$, for all $t \in T$.

[12] In the typical integer game, each agent announces some integer and the person who announces the highest integer gets to name her favorite outcome. When the agents' favorite outcomes differ, an integer game has no pure-strategy equilibria. This feature is also shared by modulo games, regarded as finite versions of the integer game in which agents announce integers from a finite set. The agent who matches the modulo of the sum of the integers gets to name an allocation. However, there is one key difference between the two: the modulo game is compact, whereas the integer game is not.

Pick any $i \in I$ and $f' \in H_{F'}$. By the first requirement of implementation, there exists $\sigma' \in \Sigma(S^\Gamma)$ such that $g(\sigma'(t)) = f'(t)$, for all $t \in T$.

Fix $t_i \in T_i$. Let $\lambda_i \in \Delta(T_{-i} \times M_{-i})$ be the probability distribution corresponding to the belief of agent $i$ that the type profile of all other agents $t_{-i}$ is distributed according to $\pi_i(t_i)$ and they play according to the (pure) strategy profile $\sigma'_{-i}$.

Since $\sigma'_{-i}$ is measurable, $G(\sigma'_{-i})$ is a measurable subset of $T_{-i} \times M_{-i}$ (Aliprantis and Border, 2006, Theorem 4.45). By definition, $\text{marg}_{T_{-i}} \lambda_i = \pi_i(t_i)$ and $\lambda_i\big(G(\sigma'_{-i})\big) = 1$. Since $G(\sigma'_{-i}) \subseteq G(S^\Gamma_{-i})$, it follows that $\lambda_i\big(G(S^\Gamma_{-i})\big) = 1$.

For all $m_i \in M_i$, due to the constructed $\lambda_i$, we have

$$\int_{T_{-i} \times M_{-i}} u_i\big(g(m_i, m_{-i}), (t_i, t_{-i})\big) d\lambda_i = \sum_{t_{-i} \in T_{-i}} \pi_i(t_i)[t_{-i}] u_i\big(g(m_i, \sigma'_{-i}(t_{-i})), (t_i, t_{-i})\big).$$

Then, as $\Gamma$ satisfies the weak BRP, there exists an $m_i(t_i) \in M_i$ such that

$$m_i(t_i) \in \arg\max_{m_i \in M_i} \int_{T_{-i} \times M_{-i}} u_i\big(g(m_i, m_{-i}), (t_i, t_{-i})\big) d\lambda_i.$$

As $\text{marg}_{T_{-i}} \lambda_i = \pi_i(t_i)$ and $\lambda_i\big(G(S^\Gamma_{-i})\big) = 1$, we have $m_i(t_i) \in b_i(S^\Gamma)[t_i]$. Hence, $m_i(t_i) \in S_i^\Gamma(t_i)$.

Define $\sigma_i$ such that $\sigma_i(t_i) = m_i(t_i)$, for all $t_i \in T_i$. Then $(\sigma_i, \sigma'_{-i}) \in \Sigma(S^\Gamma)$. Hence, by the second requirement of implementation, there exists an SCF $f \in H_{F'}$ such that $g\big(\sigma_i(t_i), \sigma'_{-i}(t_{-i})\big) = f(t)$, for all $t \in T$. Then, for all $t_i \in T_i$, we have

$$\begin{aligned}
U_i(f|t_i) &= \sum_{t_{-i} \in T_{-i}} \pi_i(t_i)[t_{-i}] u_i\big(f(t_i, t_{-i}), (t_i, t_{-i})\big) \\
&= \sum_{t_{-i} \in T_{-i}} \pi_i(t_i)[t_{-i}] u_i\big(g(\sigma_i(t_i), \sigma'_{-i}(t_{-i})), (t_i, t_{-i})\big) \\
&= \int_{T_{-i} \times M_{-i}} u_i\big(g(m_i(t_i), m_{-i}), (t_i, t_{-i})\big) d\lambda_i \\
&\geq \int_{T_{-i} \times M_{-i}} u_i\big(g(\sigma'_i(t'_i), m_{-i}), (t_i, t_{-i})\big) d\lambda_i, \forall t'_i \in T_i \\
&= \sum_{t_{-i} \in T_{-i}} \pi_i(t_i)[t_{-i}] u_i\big(g(\sigma'_i(t'_i), \sigma'_{-i}(t_{-i})), (t_i, t_{-i})\big), \forall t'_i \in T_i \\
&= \sum_{t_{-i} \in T_{-i}} \pi_i(t_i)[t_{-i}] u_i\big(f'(t'_i, t_{-i}), (t_i, t_{-i})\big), \forall t'_i \in T_i \\
&= U_i(f'; t'_i|t_i), \forall t'_i \in T_i.
\end{aligned}$$

Thus, $F'$ satisfies setwise IC. $\square$

**Remark 1.** Setwise IC is a necessary condition even when the SCC is not compact-valued or the SCS does not have the Cartesian product property – these two assumptions were not used in the above proof. If, as per our assumptions, the SCC is compact-valued and the SCS has the Cartesian product property, then setwise IC is equivalent to the following condition: For all $i \in I$ there exists an $f_i \in H_F$ such that

$$U_i(f_i|t_i) \geq U_i(f^{'};t_i^{'}|t_i), \forall t_i, t_i^{'} \in T_i \text{ and } f^{'} \in H_F. \quad (1)$$

Thus, for each player $i$, we can find a *single* SCF $f_i \in H_F$ that gives player $i$ the incentives for truthful revelation relative to any other SCF in $H_F$. Clearly, if (1) is true, then the SCC $F$ satisfies setwise IC. To argue the converse, since $F$ is compact-valued, for each $i$ and $t \in T$, there exists an $a_i(t) \in F(t)$ such that $a_i(t) \in \arg\max_{a \in F(t)} u_i(a,t)$. Then define $f_i(t) = a_i(t)$, for all $t \in T$. By the Cartesian product property of $H_F$, we have $f_i \in H_F$. Also, by the construction of $f_i$, we have $U_i(f_i|t_i) \geq U_i(f|t_i)$, for all $f \in H_F$. Hence, if setwise IC holds, then (1) must be true.

Our sufficient condition of setwise dominance, which is presented in the next section, also requires that, for each player, there is a single SCF in $H_F$ that provides that player with the incentives for truthful revelation. However, unlike (1), those incentive make truth-telling a *dominant* action, and whenever $t_i^{'} \neq t_i$, a "strictly" dominant one. $\diamond$

# 5 Sufficiency: Mechanisms Satisfying Weak BRP

In this section, we introduce a sufficient condition for implementation using a mechanism that satisfies the weak BRP. A different sufficient condition will be provided later using a mechanism with the best response property (BRP), to be defined in the sequel.

First, consider a preliminary definition. Given the SCC $F$ and $i \in I$, we say that the SCF $f_i \in H_F$ is *setwise dominant for agent* $i$ if

$$u_i\big(f_i(t_i, t_{-i}^{'}), (t_i, t_{-i})\big) \geq u_i\big(f^{'}(t_i^{'}, t_{-i}^{'}), (t_i, t_{-i})\big),$$

for all $t_i, t_i^{'} \in T_i$, $t_{-i}, t_{-i}^{'} \in T_{-i}$, and $f^{'} \in H_F$, with a strict inequality if $t_i \neq t_i^{'}$.

**Definition 4.** The SCC $F$ satisfies *setwise dominance* if for all $i \in I$, there exists an $f_i \in H_F$ such that $f_i$ is setwise dominant for agent $i$.

It is straightforward to see that if the SCC $F$ satisfies setwise dominance, then $F$ satisfies setwise IC. For a single-valued SCC $F$, setwise dominance is equivalent to truth-telling being a strictly dominant strategy for all $i \in I$ in the direct mechanism associated with the unique $f \in H_F$, that is, $f$ is strictly dominant-strategy incentive compatible.[13] However, setwise dominance is much weaker in the context of multivalued SCCs. We will illustrate this point in the examples below.

A condition related to setwise dominance appears in Jackson (1992). In a private-values environment, Jackson (1992) proposes strategy-resistance as a necessary condition for a deterministic SCC to be implementable in weakly undominated strategies by a bounded mechanism. A deterministic SCC $F$ satisfies *strategy-resistance* if, for each $i \in I$, $t_i, t_i' \in T_i$, $t_{-i}' \in T_{-i}$, and $b \in F(t_i', t_{-i}')$, there exists $a \in F(t_i, t_{-i}')$ such that $u_i(a, t_i) \geq u_i(b, t_i)$, where we use private values.[14] The following lemma helps us clarify the relation between setwise dominance and strategy-resistance (the latter uses private values and does not insist on strict inequalities for distinct types).

**Lemma 1.** *Assume private values. Then, the SCC $F$ satisfies strategy-resistance if and only if, for all $i \in I$, there exists an SCF $f_i \in H_F$ such that $u_i\big(f_i(t_i, t_{-i}'), t_i\big) \geq u_i\big(f'(t_i', t_{-i}'), t_i\big)$, for all $t_i, t_i' \in T_i$, $t_{-i}' \in T_{-i}$, and $f' \in H_F$.*

*Proof.* We argue necessity and leave the simple proof of sufficiency to the reader. So, suppose $F$ satisfies strategy-resistance. Fix an $i \in I$ and define $f_i$ as follows: $f_i(t) \in \arg\max_{a \in F(t)} u_i(a, t_i)$, for all $t \in T$. Now, pick any $t_i, t_i' \in T_i$, $t_{-i}' \in T_{-i}$, and $f' \in H_F$. Then, by strategy-resistance, there exists $a \in F(t_i, t_{-i}')$ such that $u_i(a, t_i) \geq u_i\big(f(t_i', t_{-i}'), t_i\big)$. Next, by the definition of $f_i$, we have $u_i\big(f_i(t_i, t_{-i}'), t_i\big) \geq u_i(a, t_i)$. Thus, $u_i\big(f_i(t_i, t_{-i}'), t_i\big) \geq u_i\big(f'(t_i', t_{-i}'), t_i\big)$. □

Hence, on the one hand, the setwise dominance condition we just stated is more general than strategy resistance because Jackson (1992) only considers private-values environments, while we can also handle interdependent values. On the other hand, setwise dominance is slightly stronger than strategy resistance because we strengthen weak dominance into strict dominance for distinct types of an agent.

**Example 1.** [15] There are two agents $I = \{1, 2\}$, two states $\{\alpha, \beta\}$, and a finite number of pure outcomes $A = \{a_1, a_2, \ldots, a_K\}$, where $K \geq 4$. Assume that agent 1 is uninformed of the state, while agent 2 is fully informed. Accordingly, we define $T_2 = \{t_\alpha, t_\beta\}$ as the set of types for agent 2 such that agent 2 of type $t_\alpha$ knows that the state is $\alpha$ and type $t_\beta$

[13] An SCF $f$ is *dominant-strategy incentive compatible* if $u_i\big(f(t_i, t_{-i}'), (t_i, t_{-i})\big) \geq u_i\big(f(t_i', t_{-i}'), (t_i, t_{-i})\big)$, for all $i \in I$, $t_i, t_i' \in T_i$, and $t_{-i}, t_{-i}' \in T_{-i}$. If the inequality is strict whenever $t_i \neq t_i'$, then the SCF is *strictly* dominant-strategy incentive compatible.

[14] When we focus on SCFs, strategy-resistance is reduced to dominant-strategy incentive compatibility.

[15] This is an elaboration of very useful examples in Bergemann, Morris, and Tercieux (2011) and Kunimoto and Serrano (2019), albeit complete information is assumed in those papers.

knows that it is $\beta$. Assume also that agent 1 believes with probability $q_\alpha$ that the state is $\alpha$ and with probability $(1 - q_\alpha)$ that it is $\beta$, where $q_\alpha \in (0, 1)$. Since agent 1 has only one type, we simplify notation to write $t_\alpha$ or $t_\beta$ for the type profile.

Agent 1's utility function has the following features: (1) $u_1(a_k) \equiv u_1(a_k, t_\alpha) = u_1(a_k, t_\beta)$ for each $a_k \in A$ (state-independence) and (2)

$$u_1(a_K) > u_1(a_1) > u_1(a_2) > \cdots > u_1(a_{K-1}).$$

Agent 2's utility function in state $\alpha$ has the following features:

$$u_2(a_K, t_\alpha) > u_2(a_2, t_\alpha) > u_2(a_{K-1}, t_\alpha) > \cdots > u_2(a_1, t_\alpha).$$

Agent 2's utility function in state $\beta$ has the following features:

$$u_2(a_K, t_\beta) > u_2(a_{K-1}, t_\beta) > \cdots > u_2(a_1, t_\beta) > u_2(a_2, t_\beta).$$

We further assume that $u_2(a_k) \equiv u_2(a_k, t_\alpha) = u_2(a_k, t_\beta)$ for any $a_k \in A \setminus \{a_2\}$ (state-independence, except for $a_2$).

We consider the following SCC $F$: $F(t_\alpha) = \{a_1, a_2, \ldots, a_{K-1}, a_K\}$ and $F(t_\beta) = \{a_{K-1}, a_K\}$. Then, the SCS generated by $F$ is given as $H_F = \{f_{k,K-1}, f_{k,K}\}_{k=1}^{K}$, where $f_{i,j}$ denotes the SCF that assigns alternative $a_i$ in state $\alpha$ and $a_j$ in state $\beta$. It is easy to see that the SCF $f_{K,K}$ dominates any other SCF for each agent $i \in \{1, 2\}$. Thus, the SCC $F$ satisfies setwise dominance. Therefore, $F$ also satisfies setwise IC. However, $F$ violates BIC (in fact, other than the two constant SCFs, $f_{K-1,K-1}$ and $f_{K,K}$, only the SCF $f_{2,K-1}$ satisfies BIC). Hence, if $K$ is large, the SCS $H_F$ constitutes a massive violation of BIC.$\diamond$

The next result shows that setwise dominance is sufficient for implementation in rationalizable strategies by a compact mechanism satisfying the weak BRP. It will be instructive to describe verbally the mechanism employed. The mechanism, of simultaneous moves, asks each agent to report three items: her type, a selection (SCF) of the correspondence $F$, and a real number in the closed unit interval. The latter, if it is less than 1, is interpreted as a bid to become the actor that can potentially change the SCF. Of course, the type report allows for arbitrary manipulation of the selections. As for the outcome function, a fixed arbitrary SCF (selection of $F$) –the status quo– is implemented, unless: either (1) there is an unanimous agreement on an alternative SCF, in which case that is the outcome; or (2) there is a single maximal bid, in which case that person gets to implement her announced SCF.

The statement of the result and its proof follow:

**Theorem 2.** *For any SCC $F$, if there exists an SCC $F' \approx F$ that satisfies setwise dominance, then $F$ is implementable in interim rationalizable strategies by a compact mechanism with the weak BRP.*

*Proof.* Consider the mechanism $\Gamma = ((M_i)_{i\in I}, g)$, where $M_i = T_i \times H_{F'} \times [0,1]$ with a generic element $m_i = (m_i^1, m_i^2, m_i^3)$, for all $i \in I$, and where the outcome function $g$ is defined as follows: for all $m \in M$,

**Rule 1:** If all agents agree on the SCF, i.e., there exists $f \in H_{F'}$ such that $m_i^2 = f$, for all $i \in I$, then $g(m) = f(m^1)$.

**Rule 2:** If there is any disagreement on the SCF, i.e., $m_j^2 \neq m_k^2$ for some $j,k \in I$, then the following subrules apply:

**Rule 2.1:** If $J_m \equiv \{j \in I : m_j^3 < 1\} \neq \emptyset$ and $\arg\max_{j\in J_m} m_j^3 = \{i\}$, then $g(m) = m_i^2(m^1)$.

**Rule 2.2:** In all other cases (i.e., if $J_m = \emptyset$ or $\arg\max_{j\in J_m} m_j^3$ is not a singleton), $g(m) = \check{f}(m^1)$, for any fixed arbitrary $\check{f} \in H_{F'}$.

We endow $[0,1]$ with the Euclidean topology. Since $T_i$ is finite, it is compact. Since $F'$ is compact-valued, $H_{F'}$ is compact in the product topology. It then follows that the mechanism $\Gamma$ is compact. The rest of the argument proceeds in five steps.

**Step 1:** The mechanism $\Gamma$ is measurable.

**Proof of Step 1**: Pick any measurable subset $E \subseteq \Delta(A)$. We show that $g^{-1}(E) \equiv \{m \in M : g(m) \in E\}$ is a measurable subset of $M$. To do so, we partition $g^{-1}(E)$ into the following finite number of subsets of $M$:

Let $M^1$ denote the set of all $m \in M$ such that $m$ induces Rule 1 and the resulting outcome $g(m)$ is in $E$. To argue the measurability of $M^1$, first define $\bar{H} = \bigcup_{f\in H_{F'}} \{f^n\} \subseteq H_{F'}^n$, where $f^n \equiv f \times \cdots \times f$ and $H_{F'}^n \equiv H_{F'} \times \cdots \times H_{F'}$. Let $f^n$ denote a typical element of $\bar{H}$. Next, define the function $K : T \times \bar{H} \to \Delta(A)$ such that $K(t, f^n) = f(t)$, for all $(t, f^n) \in T \times \bar{H}$. For each $f^n \in \bar{H}$, the function $K(\cdot, f^n) = f : T \to \Delta(A)$ is a measurable function, whereas, for each $t \in T$, the function $K(t,\cdot) : \bar{H} \to \Delta(A)$ is a continuous function.[16] Hence, $K$ is a Carathéodory function.[17] Since $\bar{H}$ is a subset of a separable metrizable space $H_{F'}^n$, it follows that $\bar{H}$ is also separable and metrizable.[18] As

[16]Since $H_{F'}$ is endowed with the product topology, a sequence of SCFs in $H_{F'}$ converges to the SCF $f \in H_{F'}$ if it converges pointwise. Likewise, any sequence in $H_{F'}^n$ that converges, it converges pointwise. So, for a fixed $t \in T$, the function $K(t,\cdot)$ is continuous by the definition of pointwise convergence.

[17]A function is Carathéodory if it is continuous in one variable and measurable in another variable.

[18]Since $\Delta(A)$ is compact and metrizable, and $T$ is finite, the finite product $\Delta(A)^T$ is also compact and metrizable. Hence, $\Delta(A)^T$ is separable and metrizable. As $H_{F'}$ is a subset of $\Delta(A)^T$, it too is separable and metrizable. Then its finite product $H_{F'}^n$ is also seperable and metrizable.

$\Delta(A)$ is metrizable, it follows that $K$ is jointly measurable (Aliprantis and Border, 2006, Lemma 4.51).

Now, pick any $m \in M^1$. Since $m$ induces Rule 1, it follows that the profile $m^2$ is in $\bar{H}$ and $g(m) = K(m^1, m^2)$. Hence,

$$M^1 = K^{-1}(E) \times [0,1]^n.$$

Since $K$ is jointly measurable, $K^{-1}(E)$ is a measurable subset of $T \times \bar{H}$, which in turn is a measurable subset of $T \times H^n_{F'}$ because $\bar{H}$ is a closed subset of $H^n_{F'}$. Hence, $M^1$ is a measurable subset of $M$.

Pick any $i \in I$ and let $M^i$ denote the set of all $m \in M$ such that $m$ induces Rule 2.1 and the outcome $g(m) = m_i^2(m^1)$ is in $E$. Pick any $m \in M^i$. Since $m$ induces Rule 2.1, the profile $m^2$ must be such that $m_j^2 \neq m_k^2$ for some $j, k \in I$. Thus, $m^2 \in H^n_{F'} \backslash \bar{H}$.

To argue the measurability of $M^i$, define the function $K_i : T \times (H^n_{F'} \backslash \bar{H}) \to \Delta(A)$ such that $K_i(t, m^2) = m_i^2(t)$, for all $(t, m^2) \in T \times (H^n_{F'} \backslash \bar{H})$. For each $m^2 \in H^n_{F'} \backslash \bar{H}$, the function $K_i(\cdot, m^2) = m_i^2 : T \to \Delta(A)$ is a measurable function whereas, for each $t \in T$, the function $K_i(t, \cdot) : (H^n_{F'} \backslash \bar{H}) \to \Delta(A)$ is a continuous function. Hence, $K_i$ is a Carathéodory function. Since $H^n_{F'} \backslash \bar{H}$ is a subset of a separable metrizable space $H^n_{F'}$, it follows that $H^n_{F'} \backslash \bar{H}$ is also separable and metrizable. As $\Delta(A)$ is metrizable, it follows that $K_i$ is jointly measurable.

Furthermore, since $g(m) = m_i^2(m^1)$, the profile $m^3$ must be such that one of the following two cases is true:

- $J_m = \{i\}$. In this case, $m_i^3 \in [0,1)$ and $m_j^3 = 1$, for all $j \neq i$. The set of all such $m^3$ is equal to $[0,1) \times \{1\}^{n-1}$, which is a measurable subset of $[0,1]^n$.

- There exists $j \neq i$ such that $\{i, j\} \subseteq J_m$. In this case, we must have $m_i^3 \in (0,1)$, $m_j^3 \in [0, m_i^3)$, for all $j \in J_m \backslash \{i\}$, and $m_j^3 = 1$, for all $j \in I \backslash J_m$. Let $\mathcal{J}_i$ denote the set of all $J \subseteq I$ such that $i \in J$ and $|J| \geq 2$. Then, the set of all such $m^3$ is equal to

$$\bigcup_{J \in \mathcal{J}_i} \left( \bigcup_{m_i^3 \in (0,1)} \{m_i^3\} \times [0, m_i^3)^{|J|-1} \times \{1\}^{n-|J|} \right). \tag{2}$$

  For each $J \in \mathcal{J}_i$, the set $\bigcup_{m_i^3 \in (0,1)} \{m_i^3\} \times [0, m_i^3)^{|J|-1}$ is an open subset of $[0,1]^{|J|}$. Hence, the product set $\bigcup_{m_i^3 \in (0,1)} \{m_i^3\} \times [0, m_i^3)^{|J|-1} \times \{1\}^{n-|J|}$ is a measurable subset of $[0,1]^n$. Finally, since $\mathcal{J}_i$ is finite, the set in (2) is a measurable subset of $[0,1]^n$.

Let $M^{i3}$ denote the set of all profiles $m^3$ that fall into one of the two above cases. It follows that $M^{i3}$ is a measurable subset of $[0,1]^n$.

Hence, $M^i = K_i^{-1}(E) \times M^{i3}$. Since $K_i$ is jointly measurable, $K_i^{-1}(E)$ is a measurable subset of $T \times (H^n_{F'} \backslash \bar{H})$, which in turn is a measurable subset of $T \times H^n_{F'}$ because $H^n_{F'} \backslash \bar{H}$ is an open subset of $H^n_{F'}$. Hence, $M^i$ is a measurable subset of $M$.

Let $M^{2.2}$ denote the set of message profiles $m \in M$ such that $m$ induces Rule 2.2 and the outcome $g(m) = \check{f}(m^1)$ is in $E$. Pick any $m \in M^{2.2}$. As $g(m) = \check{f}(m^1) \in E$, the profile $m^1 \in \check{f}^{-1}(E)$. Next, since $m$ induces Rule 2, the profile $m^2$ must be such that $m^2_j \neq m^2_k$ for some $j, k \in I$, i.e., $m_2 \in H^n_{F'} \backslash \bar{H}$. Since $\bar{H}$ is closed, it follows that $H^n_{F'} \backslash \bar{H}$ is open, and hence, a measurable subset of $H^n_{F'}$. Furthermore, since $m$ induces Rule 2.2, the profile $m^3$ is in $[0,1]^n \backslash \left(\bigcup_{i \in I} M^{i3}\right)$. Since $M^{i3}$ is a measurable subset of $[0,1]^n$, for all $i \in I$, it follows that $[0,1]^n \backslash \left(\bigcup_{i \in I} M^{i3}\right)$ is a measurable subset of $[0,1]^n$.

Hence, $M^{2.2} = \check{f}^{-1}(E) \times (H^n_{F'} \backslash \bar{H}) \times ([0,1]^n \backslash (\bigcup_{i \in I} M^{i3}))$, which is a measurable subset of $M$.

Thus, $g^{-1}(E) = M^1 \cup (\bigcup_{i \in I} M^i) \cup M^{2.2}$. Since $g^{-1}(E)$ is a finite union of measurable sets, it is a measurable subset of $M$. ■

**Step 2:** The mechanism $\Gamma$ satisfies the weak BRP.

**Proof of Step 2**: Pick any $i \in I$, $t_i \in T_i$, and $\sigma_{-i} : T_{-i} \to M_{-i}$. For all $j \neq i$ and $t_j \in T_j$, let $\sigma_j(t_j) = \left(\sigma^1_j(t_j), \sigma^2_j(t_j), \sigma^3_j(t_j)\right) \in T_j \times H_{F'} \times [0,1]$.

For all $t_{-i} \in T_{-i}$, let

$$z(t_{-i}) = \begin{cases} \max_{j \neq i:\, \sigma^3_j(t_j) < 1} \sigma^3_j(t_j), & \text{if } \{j \neq i : \sigma^3_j(t_j) < 1\} \neq \emptyset \\ 0, & \text{otherwise.} \end{cases}$$

Next, let $z = \max_{t_{-i} \in T_{-i}} z(t_{-i})$. Clearly, $z < 1$. Let $m_i = (t_i, f_i, z_i)$, where $z_i \in (z, 1)$. Then, we argue that $m_i$ is a best response of type $t_i$ against $\sigma_{-i}$. So pick any $m'_i = (t'_i, f', z'_i) \in M_i$ such that $m'_i \neq m_i$.

Consider any $t_{-i} \in T_{-i}$. On the one hand, if agent $i$ of type $t_i$ plays $m_i$, then the outcome is $g\left(m_i, \sigma_{-i}(t_{-i})\right) = f_i\left(t_i, \sigma^1_{-i}(t_{-i})\right)$, regardless of whether $\left(m_i, \sigma_{-i}(t_{-i})\right)$ induces Rule 1 or 2 .1 – notice that Rule 2.2 never applies at $\left(m_i, \sigma_{-i}(t_{-i})\right)$ because $z_i < 1$, whereas the outcome under Rule 2.1 is determined according to $f_i$ because $z_i > z \geq z(t_{-i})$. On the other hand, if agent $i$ of type $t_i$ plays $m'_i$, then there exists an $f_{t_{-i}} \in H_{F'}$ such that $g\left(m'_i, \sigma_{-i}(t_{-i})\right) = f_{t_{-i}}\left(t'_i, \sigma^1_{-i}(t_{-i})\right)$. By setwise dominance, $u_i\left(f_i\left(t_i, \sigma^1_{-i}(t_{-i})\right), (t_i, t_{-i})\right) \geq u_i\left(f_{t_{-i}}\left(t'_i, \sigma^1_{-i}(t_{-i})\right), (t_i, t_{-i})\right)$. Hence,

$$\sum_{t_{-i} \in T_{-i}} \pi_i(t_i)[t_{-i}] u_i\left(g\left(m_i, \sigma_{-i}(t_{-i})\right), (t_i, t_{-i})\right) = \sum_{t_{-i} \in T_{-i}} \pi_i(t_i)[t_{-i}] u_i\left(f_i\left(t_i, \sigma^1_{-i}(t_{-i})\right), (t_i, t_{-i})\right)$$

$$\geq \sum_{t_{-i}\in T_{-i}} \pi_i(t_i)[t_{-i}]u_i\big(f_{t_{-i}}\big(t_i', \sigma^1_{-i}(t_{-i})\big), (t_i, t_{-i})\big)$$
$$= \sum_{t_{-i}\in T_{-i}} \pi_i(t_i)[t_{-i}]u_i\big(g\big(m_i', \sigma_{-i}(t_{-i})\big), (t_i, t_{-i})\big).$$

Since the above is true for all $m_i' \neq m_i$, it follows that $m_i$ is a best response of type $t_i$ against $\sigma_{-i}$.

**Step 3:** $m_i \in S_i^\Gamma(t_i) \Rightarrow m_i^1 = t_i$, for all $t_i \in T_i$ and $i \in I$.

**Proof of Step 3**: Suppose not, i.e., there exist $i \in I$, $t_i \in T_i$ and $m_i \in S_i^\Gamma(t_i)$ such that $m_i^1 = t_i' \neq t_i$. As $m_i \in S_i^\Gamma(t_i)$, there must exist a belief $\lambda_i \in \Delta(T_{-i} \times M_{-i})$ such that $\lambda_i\big(G(S^\Gamma_{-i})\big) = 1$, $\text{marg}_{T_{-i}}\lambda_i = \pi_i(t_i)$, and

$$m_i \in \arg\max_{m_i' \in M_i} \int_{T_{-i}\times M_{-i}} u_i\big(g(m_i', m_{-i}), (t_i, t_{-i})\big) d\lambda_i.$$

Suppose agent $i$ of type $t_i$ deviates to $m_i' \equiv (t_i, f_i, z_i)$, where $z_i < 1$. Pick any $(t_{-i}, m_{-i})$ in the support of $\lambda_i$. The outcome before the deviation by type $t_i$ is $g(m_i, m_{-i}) = f'(t_i', m^1_{-i})$ for some $f' \in H_{F'}$. The outcome after the deviation by type $t_i$ will depend on one of the following three scenarios:

- First, if $m_j^3 = 1$, for all $j \neq i$, then either Rule 1 (when $m_j^2 = f_i$, for all $j \neq i$) or Rule 2.1 (when $m_j^2 \neq f_i$ for some $j \neq i$) will be induced by $(m_i', m_{-i})$ and the outcome will be $f_i(t_i, m^1_{-i})$. By setwise dominance, we have $u_i\big(f_i(t_i, m^1_{-i}), (t_i, t_{-i})\big) > u_i\big(f'(t_i', m^1_{-i}), (t_i, t_{-i})\big)$.

- Second, if there exists $j \neq i$ such that $m_j^3 < 1$ and $z_i > \max_{k\neq i:\, m_k^3<1} m_k^3$, then again either Rule 1 or Rule 2.1 will be induced by $(m_i', m_{-i})$ and the outcome will be $f_i(t_i, m^1_{-i})$. Hence, again by setwise dominance, we have $u_i\big(f_i(t_i, m^1_{-i}), (t_i, t_{-i})\big) > u_i\big(f'(t_i', m^1_{-i}), (t_i, t_{-i})\big)$.

- Finally, in all other cases, it follows that there exists $f'' \in H_{F'}$ such that $(m_i', m_{-i})$ induces the outcome $f''(t_i, m^1_{-i})$, which may or may not be better for type $t_i$ than $f'(t_i', m^1_{-i})$, the outcome before the deviation.

As $z_i \to 1$, the probability that either the first or the second scenario is realized (i.e., either $m_j = 1$, for all $j \neq i$, or there exists $j \neq i$ such that $m_j^3 < 1$ and $z_i > \max_{k\neq i:\, m_k^3<1} m_k^3$) converges to 1. Furthermore, as argued in the previous paragraph, in either of those two scenarios, agent $i$ of type $t_i$ strictly improves her payoff after the deviation to $(t_i, f_i, z_i)$. Finally, outside of those two scenarios, while individual $i$ of type $t_i$ may suffer a loss after she deviates to $(t_i, f_i, z_i)$, that loss is bounded. Hence, it follows that if $z_i$ is sufficiently high, individual $i$ of type $t_i$ can increase her expected payoff if she were to deviate to $(t_i, f_i, z_i)$, a contradiction. ■

**Step 4:** Define the message correspondence profile $S$ as follows: For all $i \in I$,

$$S_i(t_i) = \{t_i\} \times H_{F'} \times [0,1], \forall t_i \in T_i.$$

Then, $S = S^{\Gamma}$.

**Proof of Step 4**: It follows from Step 3 that $S^{\Gamma} \leq S$. We prove that $S \leq S^{\Gamma}$. To do so, it is sufficient to argue that $S \leq b(S)$.

Fix $i \in I$ and $t_i \in T_i$.

First, pick any $f' \in H_{F'}$ and $z_i \in [0, 1)$. We prove that $(t_i, f', z_i) \in b_i(S)$.

Fix any $j \neq i$ and $z_j \in (z_i, 1)$. Let $\sigma_j : T_j \to M_j$ be such that $\sigma_j(t_j) = (t_j, f_i, z_j)$, for all $t_j \in T_j$. For all $j' \in I\backslash\{i, j\}$, let $\sigma_{j'} : T_{j'} \to M_{j'}$ be such that $\sigma_{j'}(t_{j'}) = (t_{j'}, f_i, 1)$, for all $t_{j'} \in T_{j'}$. Then consider the probability measure $\lambda_i \in \Delta(T_{-i} \times M_{-i})$ corresponding to the belief of agent $i$ that the type profile of all other agents $t_{-i}$ is distributed according to $\pi_i(t_i)$ and they play according to the strategy profile $\sigma_{-i}$. Since $\sigma_{-i}$ is measurable, $G(\sigma_{-i})$ is a measurable subset of $T_{-i} \times M_{-i}$. By definition, $\text{marg}_{T_{-i}} \lambda_i = \pi_i(t_i)$ and $\lambda_i\big(G(\sigma_{-i})\big) = 1$. Since $G(\sigma_{-i}) \subseteq G(S_{-i})$, it follows that $\lambda_i\big(G(S_{-i})\big) = 1$.

Now, given the belief $\lambda_i$, if agent $i$ of type $t_i$ reports $(t_i, f', z_i)$, then either Rule 1 (when $f' = f_i$) or Rule 2.1 (when $f' \neq f_i$) is induced by any $(t_{-i}, m_{-i})$ in the support of $\lambda_i$ and the outcome is $f_i(t_i, t_{-i})$. (When Rule 2.1 applies, the outcome is determined by the SCF $f_i$ announced by agent $j$.) Thus the expected payoff of agent $i$ of type $t_i$ is $U_i(f_i|t_i)$.

Suppose agent $i$ of type $t_i$ deviates to some $m_i \equiv (t_i', f'', z_i') \in M_i$. If $f'' = f_i$, then Rule 1 is induced by any $(t_{-i}, m_{-i})$ in the support of $\lambda_i$ and the outcome is $f_i(t_i', t_{-i})$. If $f'' \neq f_i$ and $z_i' < z_j$ or $z_i' = 1$, then Rule 2.1 is induced by any $(t_{-i}, m_{-i})$ in the support of $\lambda_i$ and the outcome is $f_i(t_i', t_{-i})$. If $f'' \neq f_i$ and $z_i' = z_j$, then Rule 2.2 is induced by any $(t_{-i}, m_{-i})$ in the support of $\lambda_i$ and the outcome is $\check{f}(t_i', t_{-i})$. Finally, if $f'' \neq f_i$ and $z_i' \in (z_j, 1)$, then Rule 2.1 is induced by any $(t_{-i}, m_{-i})$ in the support of $\lambda_i$ and the outcome is $f''(t_i', t_{-i})$. It thus follows that there exists an $f \in H_{F'}$ such that, for all $(t_{-i}, m_{-i})$ in the support of $\lambda_i$, the outcome is $f(t_i', t_{-i})$. Thus, the expected payoff of agent $i$ of type $t_i$ after the deviation is $U_i(f; t_i'|t_i)$. By setwise dominance, we have that $U_i(f_i|t_i) \geq U_i(f; t_i'|t_i)$, for all $t_i' \in T_i$. Hence, agent $i$ of type $t_i$ cannot improve by any deviation. Thus, $(t_i, f', z_i) \in b_i(S)$.

Second, pick any $f^{'} \in H_{F'}$. We prove that $(t_i, f^{'}, 1) \in b_i(S)$.

Fix any $j \neq i$ and $z_j \in (0,1)$. Now let $\sigma_j : T_j \to M_j$ be such that $\sigma_j(t_j) = (t_j, f_i, z_j)$, for all $t_j \in T_j$. And, as before, for all $j^{'} \in I \backslash \{i, j\}$, let $\sigma_{j'} : T_{j'} \to M_{j'}$ be such that $\sigma_{j'}(t_{j'}) = (t_{j'}, f_i, 1)$, for all $t_{j'} \in T_{j'}$. Consider the probability measure $\lambda_i \in \Delta(T_{-i} \times M_{-i})$ corresponding to the belief of agent $i$ that the type profile of all other agents $t_{-i}$ is distributed according to $\pi_i(t_i)$ and they play according to the strategy profile $\sigma_{-i}$. By definition, $\text{marg}_{T_{-i}} \lambda_i = \pi_i(t_i)$ and $\lambda_i\big(G(S_{-i})\big) = 1$.

Now, given the belief $\lambda_i$, if agent $i$ of type $t_i$ reports $(t_i, f^{'}, 1)$, then as in the previous case, the outcome is $f_i(t_i, t_{-i})$ at any $(t_{-i}, m_{-i})$ in the support of $\lambda_i$. Suppose agent $i$ of type $t_i$ deviates to some $m_i \equiv (t_i^{'}, f^{''}, z_i^{'}) \in M_i$. As in the previous case, there exists an $f \in H_{F'}$ such that, for all $(t_{-i}, m_{-i})$ in the support of $\lambda_i$, the outcome is $f(t_i^{'}, t_{-i})$. Thus, the expected payoff of agent $i$ of type $t_i$ after the deviation is $U_i(f; t_i^{'}|t_i)$. By setwise dominance, we have that $U_i(f_i|t_i) \geq U_i(f; t_i^{'}|t_i)$, for all $t_i^{'} \in T_i$. Hence, agent $i$ of type $t_i$ cannot improve by any deviation. Thus, $(t_i, f^{'}, 1) \in b_i(S)$.

It follows that $S \leq b(S)$, as claimed. ■

**Step 5:** The mechanism $\Gamma$ implements $F$ in interim rationalizable strategies.

**Proof of Step 5**: Pick any $f \in H_{F'}$. For each $i \in I$ and $t_i \in T_i$, let $\sigma_i(t_i) = (t_i, f, 0)$. It then follows from Step 4 that $\sigma \in \Sigma(S^{\Gamma})$. So, we have $g(\sigma(t)) = f(t)$, for all $t \in T$. This verifies the first requirement of implementation.

Next, pick any $\sigma \in \Sigma(S^{\Gamma})$. Consider any $t \in T$ and the corresponding message profile $\sigma(t) = \big(\sigma^1(t), \sigma^2(t), \sigma^3(t)\big)$. By the construction of the mechanism, there exists an $f_t \in H_{F'}$ such that $g(\sigma(t)) = f_t(\sigma^1(t))$. From Step 3, we have that $\sigma^1(t) = t$, for all $t \in T$. Hence, $g(\sigma(t)) = f_t(t) \in F^{'}(t)$. Thus, $g \circ \sigma$ is a selection of $F^{'}$, i.e., $g \circ \sigma \in H_{F'}$. This verifies the second requirement of implementation. ■

Steps 1 through 5 complete the proof of the theorem. □

# 6 Sufficiency: Mechanisms Satisfying BRP

In this section, we present a different sufficiency result, using a mechanism with the best response property (BRP), which will be defined shortly.

Given any SCF $f$ and $i \in I$, we say that the SCF $f$ is *independent of the types of all* $j \neq i$ if for all $t, t^{'} \in T$,

$$t_i = t_i^{'} \Rightarrow f(t) = f(t^{'}).$$

**Definition 5.** The SCC $F$ satisfies *setwise independent dominance* if for all $i \in I$, there exists an SCF $f_i \in H_F$ such that $f_i$ is independent of the types of all $j \neq i$ and dominant for agent $i$. That is,

$$u_i\big(f_i(t_i, t'_{-i}), (t_i, t_{-i})\big) \geq u_i\big(f'(t'_i, t'_{-i}), (t_i, t_{-i})\big),$$

for all $t_i, t'_i \in T_i$, $t_{-i}, t'_{-i} \in T_{-i}$, and $f' \in H_F$, with a strict inequality if $t_i \neq t'_i$.

To illustrate this definition, we revisit the environment in Example 1:

**Example 2.** We have already argued that the SCC $F$ in Example 1 satisfies setwise dominance. More specifically, we confirm that the SCF $f_{K,K}$ is a member of $H_F$ and dominates any other SCF for each agent $i \in \{1, 2\}$. Also, $f_{K,K}$ is trivially independent of each agent's type because it is constant. Therefore, the SCC $F$ in this example satisfies setwise independent dominance.

As promised, we now define the best response property as a strengthening of the weak BRP:

**Definition 6.** The mechanism $\Gamma = ((M)_{i \in I}, g)$ satisfies the *best response property (BRP)* if for all $i \in I$, $t_i \in T_i$, and $\lambda_i \in \Delta(T_{-i} \times M_{-i})$ such that $\text{marg}_{T_{-i}} \lambda_i = \pi_i(t_i)$, we have

$$\arg\max_{m_i \in M_i} \int_{T_{-i} \times M_{-i}} u_i\big(g(m_i, m_{-i}), (t_i, t_{-i})\big) d\lambda_i \neq \emptyset.$$

Thus, unlike the weak BRP, which only requires a best response to exist against all pure strategies of the other agents, the BRP requires that a best response exist against *all* beliefs about the behavior of the other agents – including correlated strategies. Unless there is a unanimous agreement on the best outcome, BRP prevents us from using the integer game as part of the implementing mechanisms.[19] While the mechanism we propose in the proof of Theorem 2 does not satisfy BRP, it has nonempty best responses to any belief that has a finite support (not just pure strategies, as required in weak BRP).

The next result shows that setwise independent dominance is, along with an assumption often met in economic environments, sufficient for implementation in interim rationalizable strategies by a compact mechanism satisfying BRP. We construct separate implementing mechanisms for $n \geq 3$ and $n = 2$. For the sake of brevity, we focus on the mechanism for $n \geq 3$ in the main body of the paper. It will again be helpful to convey

[19]The modulo game, being finite, satisfies the BRP. However, the modulo game fails to provide appropriate incentives because an agent may not have a winning strategy in the game when others play a mixed strategy or the environment is one of incomplete information.

the main elements of the implementing mechanism verbally. The mechanism builds on the collection of SCFs $f_i$ that are independent of the types of others and dominant for each agent $i$ (the reader could think of $f_i$ as the maximal element in the ex-post efficient individually rational frontier for agent $i$ in an economy with monotone preferences). The mechanism, of simultaneous moves, asks each agent to report three items: her type, a selection – SCF – of the correspondence $F$, and either the integer 0 or the integer 1. The type report, as is standard, allows for the manipulation of the implemented SCF. The choice of the integer 0 should be viewed, roughly, as wishing to take a passive role in society, while the choice of integer 1 conveys a wish for a more active role. Since the mechanism works for at least three agents, if there is a unanimous agreement on the SCF and the integer 0, that consensus SCF is implemented. If there is an "odd person out" (agent $i$ is the only one either announcing the integer 1 or a different SCF), she gets to implement her dominant $f_i$. In all other cases, the average of the dominant $f_i$'s over those agents who announce the integer 1 (or the average of all $f_i$'s if everyone is passive) is implemented.

The statement and proof of the next result now follow:

**Theorem 3.** *For any SCC $F$, if there exists an SCC $F' \approx F$ that is convex-valued and satisfies setwise independent dominance, then $F$ is implementable in interim rationalizable strategies by a compact mechanism satisfying the BRP.*

*Proof.* We now present the proof of the theorem for the case of $n \geq 3$. The proof for $n = 2$ is relegated to the Appendix.

Since $F'$ satisfies setwise independent dominance, for each $i \in I$, there exists $f_i \in H_{F'}$ such that $f_i$ dominates all other SCFs in $H_{F'}$ for agent $i$ and is independent of the types of all $j \neq i$. The mechanism we propose below utilizes the collection of such SCFs $\{f_i\}_{i \in I}$. Consider the mechanism $\Gamma = ((M_i)_{i \in I}, g)$, where $M_i = T_i \times H_{F'} \times \{0, 1\}$ with a typical element $m_i = (m_i^1, m_i^2, m_i^3)$, for all $i \in I$, and the outcome function $g$ is defined as follows: for all $m \in M$,

**Rule 1:** If $m_i^2 = f$ and $m_i^3 = 0$, for all $i \in I$, then $g(m) = f(m^1)$.

**Rule 2:** If there exists an $i \in I$ such that $m_j^2 = f$ and $m_j^3 = 0$, for all $j \neq i$, but either $m_i^2 \neq f$ or $m_i^3 \neq 0$, then $g(m) = f_i(m^1)$.

**Rule 3:** In all other cases, the following subrules apply: Let $I_m \equiv \{i \in I : m_i^3 = 1\}$.

**Rule 3.1:** If $I_m \neq \emptyset$, then

$$g(m) = \frac{1}{|I_m|} \sum_{i \in I_m} f_i(m^1).$$

**Rule 3.2:** If $I_m = \emptyset$, then

$$g(m) = \frac{1}{n} \sum_{i \in I} f_i(m^1).$$

We endow $\{0,1\}$ with the discrete topology. Then, since $T_i$ is finite, and hence compact for all $i \in I$, and since $H_{F'}$ is compact, it follows that the mechanism $\Gamma$ is compact. The rest of the argument proceeds in five steps.

**Step 1:** The mechanism $\Gamma$ is measurable.

**Proof of Step 1**: Pick any measurable subset $E \subseteq \Delta(A)$. We show that $g^{-1}(E) = \{m \in M : g(m) \in E\}$ is a measurable subset of $M$. To do so, we partition $g^{-1}(E)$ into the following finite number of subsets of $M$:

Let $M^1$ denote the set of all $m \in M$ such that $m$ induces Rule 1 and the outcome $g(m)$ is in $E$. Recall from the proof of Theorem 2 that we introduced $\bar{H} = \bigcup_{f \in H_{F'}} \{f^n\} \subseteq H^n_{F'}$ and $f^n$ is a typical element of $\bar{H}$. Also recall the function $K : T \times \bar{H} \to \Delta(A)$, which was defined as follows: $K(t, f^n) = f(t)$, for all $(t, f^n) \in T \times \bar{H}$. Finally, recall that $K$ is a Carathéodory function that is jointly measurable.

Now, pick any $m \in M^1$. Since $m$ induces Rule 1, it follows that $m^2 \in \bar{H}$, $m^3 \in \{0\}^n$, and $g(m) = K(m^1, m^2)$. Hence,

$$M^1 = K^{-1}(E) \times \{0\}^n.$$

Since $K$ is jointly measurable, $K^{-1}(E)$ is a measurable subset of $T \times \bar{H}$, which in turn is a measurable subset of $T \times H^n_{F'}$. Hence, $M^1$ is a measurable subset of $M$.

Pick any $i \in I$ and let $M^i$ denote the set of all $m \in M$ such that $m$ induces Rule 2, agent $i$ is the "odd-one-out", and the outcome $g(m) = m^2_i(m^1) \in E$. Pick any $m \in M^i$. Since agent $i$ is the "odd-one-out" at $m$, the profile $(m^2, m^3)$ must be such that one of the following two cases applies:

- There exists an $f \in H_{F'}$ such that $m^2_j = f$, for all $j \neq i$, and $m^2_i \neq f$. Let $\bar{H}_i = \bigcup_{f \in H_{F'}} \{f^{n-1}\}$. Furthermore, the profile $m^3$ must be such that $m^3_j = 0$, for all $j \neq i$, and $m^3_i \in \{0,1\}$. Thus,

  $$m^2 \in (\bar{H}_i \times H_{F'}) \backslash \bar{H} \quad \text{and} \quad m^3 \in \{0\}^{n-1} \times \{0,1\}.$$

  Finally, as $g(m) = f_i(m^1) \in E$, the profile $m^1$ is in $f_i^{-1}(E)$.

  Let $M^{i1} \equiv f_i^{-1}(E) \times \big((\bar{H}_i \times H_{F'}) \backslash \bar{H}\big) \times \{0\}^{n-1} \times \{0,1\}$. Since $f_i$ is measurable, $f_i^{-1}(E)$ is a measurable subset of $T$. Moreover, $(\bar{H}_i \times H_{F'}) \backslash \bar{H}$ is a measurable subset of $H^n_{F'}$ because $\bar{H}_i$ and $\bar{H}$ are closed, and hence, measurable subsets of $H^{n-1}_{F'}$ and $H^n_{F'}$, respectively. Hence, $M^{i1}$ is a measurable subset of $M$.

- There exists an $f \in H_{F'}$ such that $m^2_j = f$, for all $j \in I$. In this case, for agent $i$ to be the "odd-one-out", the profile $m^3$ must be such that $m^3_j = 0$, for all $j \neq i$, and $m^3_i = 1$. Thus,
$$m^2 \in \bar{H} \quad \text{and} \quad m^3 \in \{0\}^{n-1} \times \{1\}.$$
Finally, as $g(m) = f_i(m^1) \in E$, the profile $m^1$ is in $f_i^{-1}(E)$.
Let $M^{i2} \equiv f_i^{-1}(E) \times \bar{H} \times \{0\}^{n-1} \times \{1\}$. Clearly, $M^{i2}$ is a measurable subset of $M$.

Hence, $M^i = M^{i1} \cup M^{i2}$, which is a measurable subset of $M$.

Pick any $i \in I$ and let $M^{i,3.1}$ denote the set of message profiles $m \in M$ such that $m$ induces Rule 3.1, $I_m = \{i\}$, and the outcome $g(m) = f_i(m^1)$ is in $E$. Pick any $m \in M^{i,3.1}$. The profile $m^3$ must be such that $m^3_j = 0$, for all $j \neq i$, and $m^3_i = 1$. Since $m$ induces Rule 3.1, the profile $m^2$ must be such that $m^2_j \neq m^2_k$, for some $j, k \neq i$. Hence, $m^2 \in H^n_{F'} \backslash (\bar{H}_i \times H_{F'})$. Finally, as $g(m) = f_i(m^1) \in E$, the profile $m^1$ is in $f_i^{-1}(E)$. Thus, $M^{i,3.1} = f_i^{-1}(E) \times \big(H^n_{F'} \backslash (\bar{H}_i \times H_{F'})\big) \times \{0\}^{n-1} \times \{1\}$. Clearly, $M^{i,3.1}$ is a measurable subset of $M$.

Let $\mathcal{J} = \{J \subseteq I : |J| \geq 2\}$. Pick any $J \in \mathcal{J}$ and let $M^{J,3.1}$ denote the set of message profiles $m \in M$ such that $m$ induces Rule 3.1, $I_m = J$, and the outcome $g(m) = (1/|J|) \sum_{j \in J} f_j(m^1)$ is in $E$. Pick any $m \in M^{J,3.1}$. The profile $m^3$ must be such that $m^3_j = 1$, for all $j \in J$, and $m^3_i = 0$, for all $i \notin J$. Then, the profile $m^2$ can be any element of $H^n_{F'}$. Finally, letting the function $f_J \equiv (1/|J|) \sum_{j \in J} f_j$, the profile $m^1$ is in $f_J^{-1}(E)$. Thus, $M^{J,3.1} = f_J^{-1}(E) \times H^n_{F'} \times \{0\}^{n-|J|} \times \{1\}^{|J|}$. Clearly, $M^{J,3.1}$ is a measurable subset of $M$.

Let $M^{3.2}$ denote the set of message profiles $m \in M$ such that $m$ induces Rule 3.2 and the outcome $g(m) = (1/n) \sum_{i \in I} f_i(m^1) \equiv f_I(m^1)$ is in $E$. Pick any $m \in M^{3.2}$. The profile $m^3$ must be such that $m^3_i = 0$, for all $i \in I$. Then, the profile $m^2$ must not be in $\bar{H}_i \times H_{F'}$, for all $i \in I$. Thus, $m^2 \in H^n_{F'} \backslash (\bigcup_{i \in I} \big(\bar{H}_i \times H_{F'}\big))$. Finally, the profile $m^1$ is in $f_I^{-1}(E)$. Thus, $M^{3.2} = f_I^{-1}(E) \times (H^n_{F'} \backslash (\bigcup_{i \in I} \big(\bar{H}_i \times H_{F'}\big))) \times \{0\}^n$. Clearly, $M^{3.2}$ is a measurable subset of $M$.

Thus, $g^{-1}(E) = M^1 \cup \bigcup_{i \in I} (M^i \cup M^{i,3.1}) \cup \bigcup_{J \in \mathcal{J}} M^{J,3.1} \cup M^{3.2}$. Since $g^{-1}(E)$ is a finite union of measurable sets, it is a measurable subset of $M$. ■

**Step 2:** Define the message correspondence profile $S$ as follows: For all $i \in I$,

$$S_i(t_i) = \{t_i\} \times H_{F'} \times \{0\}, \forall t_i \in T_i.$$

Then, $S \leq S^\Gamma$.

**Proof of Step 2**: It is sufficient to argue that $S \leq b(S)$. Fix $i \in I$ and $t_i \in T_i$. Pick any $f' \in H_{F'}$. We prove that $(t_i, f', 0) \in b_i(S)$.

For all $j \neq i$, let $\sigma_j : T_j \to M_j$ be such that $\sigma_j(t_j) = (t_j, f_i, 0)$, for all $t_j \in T_j$. Then consider the probability measure $\lambda_i \in \Delta(T_{-i} \times M_{-i})$ corresponding to the belief of agent $i$ that the type profile of all other agents $t_{-i}$ is distributed according to $\pi_i(t_i)$ and they play according to the strategy profile $\sigma_{-i}$. Since $\sigma_{-i}$ is measurable, $G(\sigma_{-i})$ is a measurable subset of $T_{-i} \times M_{-i}$. By definition, $\text{marg}_{T_{-i}} \lambda_i = \pi_i(t_i)$ and $\lambda_i\big(G(\sigma_{-i})\big) = 1$. Since $G(\sigma_{-i}) \subseteq G(S_{-i})$, it follows that $\lambda_i\big(G(S_{-i})\big) = 1$.

Now, given the belief $\lambda_i$, if agent $i$ of type $t_i$ reports $(t_i, f', 0)$, then either Rule 1 (when $f' = f_i$) or Rule 2 (when $f' \neq f_i$) is applied at any $(t_{-i}, m_{-i})$ in the support of $\lambda_i$ and the outcome is $f_i(t_i, t_{-i})$. Thus, the expected payoff of agent $i$ of type $t_i$ when she reports $(t_i, f', 0)$ is $U_i(f_i|t_i)$.

Suppose agent $i$ of type $t_i$ deviates to some $m_i = (t_i', f'', z_i) \in M_i$. On the one hand, if $f'' = f_i$ and $z_i = 0$, then Rule 1 is applied at any $(t_{-i}, m_{-i})$ in the support of $\lambda_i$ and the outcome is $f_i(t_i', t_{-i})$. On the other hand, if $f'' \neq f_i$ or $z_i \neq 0$, then Rule 2 is applied at any $(t_{-i}, m_{-i})$ in the support of $\lambda_i$ and, again, the outcome is $f_i(t_i', t_{-i})$. Thus, in either case, the expected payoff of agent $i$ of type $t_i$ equals $U_i(f_i; t_i'|t_i)$. By setwise independent dominance, we have $U_i(f_i|t_i) \geq U_i(f_i; t_i'|t_i)$. Hence, agent $i$ of type $t_i$ cannot improve by any deviation. Thus, $(t_i, f', 0) \in b_i(S)$. ■

**Step 3:** $(t_i', f, z_i) \in S_i^\Gamma(t_i) \Rightarrow t_i' = t_i$.

**Proof of Step 3**: Suppose not, i.e., there exists $m_i \equiv (t_i', f, z_i) \in S_i^\Gamma(t_i)$ such that $t_i' \neq t_i$. Then there must exist a belief $\lambda_i \in \Delta(T_{-i} \times M_{-i})$ such that $\lambda_i\big(G(S_{-i}^\Gamma)\big) = 1$, $\text{marg}_{T_{-i}} \lambda_i = \pi_i(t_i)$, and

$$m_i \in \arg\max_{m_i'' \in M_i} \int_{T_{-i} \times M_{-i}} u_i\big(g(m_i'', m_{-i}), (t_i, t_{-i})\big) d\lambda_i. \tag{3}$$

Suppose that instead of $m_i$, agent $i$ of type $t_i$ were to report $m_i' = (t_i, f_i, z_i')$, where $z_i' = 1$. Pick any $(t_{-i}, m_{-i})$ in the support of $\lambda_i$. Let us denote $m_j = (t_j', m_j^2, z_j)$, for all $j \neq i$.

First, suppose $(m_i, m_{-i})$ induces Rule 1. Then it must be that $z_i = 0$, and $m_j^2 = f$

and $z_j = 0$, for all $j \neq i$. Hence, the outcome is $f(t_i^{'}, t_{-i}^{'})$. Since $z_i^{'} = 1$, $(m_i^{'}, m_{-i})$ induces Rule 2 so that the outcome is $f_i(t_i, t_{-i}^{'})$. By setwise independent dominance, we have $u_i\big(f_i(t_i, t_{-i}^{'}), (t_i, t_{-i})\big) > u_i\big(f(t_i^{'}, t_{-i}^{'}), (t_i, t_{-i})\big)$.

Second, suppose $(m_i, m_{-i})$ induces Rule 2. There are two possibilities to consider:

1. Agent $i$ is the "odd-one-out" at $(m_i, m_{-i})$: Suppose there is some $f^{'} \in H_{F^{'}}$ such that $m_j^2 = f^{'}$ and $z_j = 0$, for all $j \neq i$, but either $f^{'} \neq f$ or $z_i \neq 0$. Then $g(m_i, m_{-i}) = f_i(t_i^{'}, t_{-i}^{'})$.

   Rule 2 is induced by $(m_i^{'}, m_{-i})$ because $z_i^{'} = 1$. Then, as per Rule 2, $g(m_i^{'}, m_{-i}) = f_i(t_i, t_{-i}^{'})$. By setwise independent dominance, we have $u_i\big(f_i(t_i, t_{-i}^{'}), (t_i, t_{-i})\big) > u_i\big(f_i(t_i^{'}, t_{-i}^{'}), (t_i, t_{-i})\big)$.

2. Agent $k \neq i$ is the "odd-one-out" at $(m_i, m_{-i})$: Suppose that either $m_k^2 \neq f$ or $z_k = 1$, $m_j^2 = f$ and $z_j = 0$, for all $j \neq k, i$, and $z_i = 0$. Then $g(m_i, m_{-i}) = f_k(t_i^{'}, t_{-i}^{'})$. Now $(m_i^{'}, m_{-i})$ induces Rule 3 because agent $i$ announces $z_i^{'} = 1$ and either $m_k^2 \neq f = m_j^2$, for all $j \neq k, i$, or $z_k = 1$. Recall that $z_j = 0$, for all $j \neq k, i$. Thus, either $I_{(m_i^{'}, m_{-i})} = \{i\}$ or $I_{(m_i^{'}, m_{-i})} = \{i, k\}$ and Rule 3.1 is induced by $(m_i^{'}, m_{-i})$. On the one hand, if $I_{(m_i^{'}, m_{-i})} = \{i\}$, then, as per Rule 3.1, $g(m_i^{'}, m_{-i}) = f_i(t_i, t_{-i}^{'})$. By setwise independent dominance, we have $u_i\big(f_i(t_i, t_{-i}^{'}), (t_i, t_{-i})\big) > u_i\big(f_k(t_i^{'}, t_{-i}^{'}), (t_i, t_{-i})\big)$. On the other hand, if $I_{(m_i^{'}, m_{-i})} = \{i, k\}$, then, as per Rule 3.1, we have

   $$g(m_i^{'}, m_{-i}) = (1/2) f_i(t_i, t_{-i}^{'}) + (1/2) f_k(t_i, t_{-i}^{'}).$$

   We have $u_i\big(f_k(t_i, t_{-i}^{'}), (t_i, t_{-i})\big) = u_i\big(f_k(t_i^{'}, t_{-i}^{'}), (t_i, t_{-i})\big)$ since $f_k$ is independent of agent $i$'s type. By setwise independent dominance,[20] we have $u_i\big(f_i(t_i, t_{-i}^{'}), (t_i, t_{-i})\big) > u_i\big(f_k(t_i^{'}, t_{-i}^{'}), (t_i, t_{-i})\big)$. Hence,

   $$\begin{aligned} u_i\big(g(m_i^{'}, m_{-i}), (t_i, t_{-i})\big) &= \frac{1}{2} u_i\big(f_i(t_i, t_{-i}^{'}), (t_i, t_{-i})\big) + \frac{1}{2} u_i\big(f_k(t_i, t_{-i}^{'}), (t_i, t_{-i})\big) \\ &> \frac{1}{2} u_i\big(f_k(t_i^{'}, t_{-i}^{'}), (t_i, t_{-i})\big) + \frac{1}{2} u_i\big(f_k(t_i, t_{-i}^{'}), (t_i, t_{-i})\big) \\ &= \frac{1}{2} u_i\big(f_k(t_i^{'}, t_{-i}^{'}), (t_i, t_{-i})\big) + \frac{1}{2} u_i\big(f_k(t_i^{'}, t_{-i}^{'}), (t_i, t_{-i})\big) \\ &= u_i\big(f_k(t_i^{'}, t_{-i}^{'}), (t_i, t_{-i})\big) \\ &= u_i\big(g(m_i, m_{-i}), (t_i, t_{-i})\big). \end{aligned}$$

Finally, suppose $(m_i, m_{-i})$ induces Rule 3. There are three possibilities to consider:

[20] We remark that all applications of setwise independent dominance in this proof, prior to this one, did not use the fact that the SCFs $f_i$ are independent of the types of others.

1. $I_{(m_i,m_{-i})} \neq \emptyset$ and $z_i = 1$. Then, $(m_i, m_{-i})$ induces Rule 3.1, and hence,

$$g(m_i, m_{-i}) = \frac{1}{|I_{(m_i,m_{-i})}|} \sum_{j \in I_{(m_i,m_{-i})}} f_j(t_i^{'}, t_{-i}^{'}).$$

Then, $(m_i^{'}, m_{-i})$ also induces Rule 3.1 because $z_i^{'} = 1$. Since $z_i = 1$, we have $I_{(m_i,m_{-i})} = I_{(m_i^{'},m_{-i})}$. Hence,

$$g(m_i^{'}, m_{-i}) = \frac{1}{|I_{(m_i,m_{-i})}|} \sum_{j \in I_{(m_i,m_{-i})}} f_j(t_i, t_{-i}^{'}).$$

Since $f_j$ is independent of agent $i$'s type for all $j \neq i$, we have $u_i\big(f_j(t_i, t_{-i}^{'}), (t_i, t_{-i})\big) = u_i\big(f_j(t_i^{'}, t_{-i}^{'}), (t_i, t_{-i})\big)$. Also, we have $u_i\big(f_i(t_i, t_{-i}^{'}), (t_i, t_{-i})\big) > u_i\big(f_i(t_i^{'}, t_{-i}^{'}), (t_i, t_{-i})\big)$ because of setwise independent dominance. Hence,

$$\begin{aligned}
&u_i\big(g(m_i^{'}, m_{-i}), (t_i, t_{-i})\big) \\
&= \frac{1}{|I_{(m_i,m_{-i})}|} \left( u_i\big(f_i(t_i, t_{-i}^{'}), (t_i, t_{-i})\big) + \sum_{j \in I_{(m_i,m_{-i})} \setminus \{i\}} u_i\big(f_j(t_i, t_{-i}^{'}), (t_i, t_{-i})\big) \right) \\
&> \frac{1}{|I_{(m_i,m_{-i})}|} \left( u_i\big(f_i(t_i^{'}, t_{-i}^{'}), (t_i, t_{-i})\big) + \sum_{j \in I_{(m_i,m_{-i})} \setminus \{i\}} u_i\big(f_j(t_i, t_{-i}^{'}), (t_i, t_{-i})\big) \right) \\
&= \frac{1}{|I_{(m_i,m_{-i})}|} \left( u_i\big(f_i(t_i^{'}, t_{-i}^{'}), (t_i, t_{-i})\big) + \sum_{j \in I_{(m_i,m_{-i})} \setminus \{i\}} u_i\big(f_j(t_i^{'}, t_{-i}^{'}), (t_i, t_{-i})\big) \right) \\
&= u_i\big(g(m_i, m_{-i}), (t_i, t_{-i})\big).
\end{aligned}$$

2. $I_{(m_i,m_{-i})} \neq \emptyset$ and $z_i = 0$. Then, $(m_i, m_{-i})$ induces Rule 3.1, and hence,

$$g(m_i, m_{-i}) = \frac{1}{|I_{(m_i,m_{-i})}|} \sum_{j \in I_{(m_i,m_{-i})}} f_j(t_i^{'}, t_{-i}^{'}).$$

By setwise independent dominance, we have $u_i\big(f_i(t_i, t_{-i}^{'}), (t_i, t_{-i})\big) > u_i\big(f_j(t_i^{'}, t_{-i}^{'}), (t_i, t_{-i})\big)$, for all $j \in I_{(m_i,m_{-i})}$. Hence,

$$u_i\big(f_i(t_i, t_{-i}^{'}), (t_i, t_{-i})\big) > u_i\big(g(m_i, m_{-i}), (t_i, t_{-i})\big). \tag{4}$$

Then, $(m_i^{'}, m_{-i})$ induces Rule 3.1 because $z_i^{'} = 1$. Also, $I_{(m_i^{'},m_{-i})} = I_{(m_i,m_{-i})} \cup \{i\}$.

Hence,

$$g(m_i', m_{-i}) = \frac{1}{|I_{(m_i,m_{-i})}| + 1} \left( f_i(t_i, t_{-i}') + \sum_{j \in I_{(m_i,m_{-i})}} f_j(t_i, t_{-i}') \right).$$

Since $f_j$ is independent of agent $i$'s type for all $j \neq i$, we have $u_i\big(f_j(t_i, t_{-i}'), (t_i, t_{-i})\big) = u_i\big(f_j(t_i', t_{-i}'), (t_i, t_{-i})\big)$ for all $j \neq i$. Hence,

$$\begin{aligned}
&u_i\big(g(m_i', m_{-i}), (t_i, t_{-i})\big) \\
&= \frac{1}{|I_{(m_i,m_{-i})}| + 1} \left( u_i\big(f_i(t_i, t_{-i}'), (t_i, t_{-i})\big) + \sum_{j \in I_{(m_i,m_{-i})}} u_i\big(f_j(t_i, t_{-i}'), (t_i, t_{-i})\big) \right) \\
&= \frac{1}{|I_{(m_i,m_{-i})}| + 1} \left( u_i\big(f_i(t_i, t_{-i}'), (t_i, t_{-i})\big) + \sum_{j \in I_{(m_i,m_{-i})}} u_i\big(f_j(t_i', t_{-i}'), (t_i, t_{-i})\big) \right) \\
&= \frac{1}{|I_{(m_i,m_{-i})}| + 1} \Big( u_i\big(f_i(t_i, t_{-i}'), (t_i, t_{-i})\big) + |I_{(m_i,m_{-i})}| u_i\big(g(m_i, m_{-i}), (t_i, t_{-i})\big) \Big) \\
&> u_i\big(g(m_i, m_{-i}), (t_i, t_{-i})\big) \qquad (\because of (4)).
\end{aligned}$$

3. $I_{(m_i,m_{-i})} = \emptyset$. Then, $(m_i, m_{-i})$ induces Rule 3.2, and hence,

$$g(m_i, m_{-i}) = \frac{1}{n} \sum_{j \in I} f_j(t_i', t_{-i}').$$

Then, $(m_i', m_{-i})$ induces Rule 3.1 because $z_i' = 1$. Thus, $I_{(m_i', m_{-i})} = \{i\}$. Hence,

$$g(m_i', m_{-i}) = f_i(t_i, t_{-i}').$$

By setwise independent dominance, we have $u_i\big(f_i(t_i, t_{-i}'), (t_i, t_{-i})\big) > u_i\big(f_j(t_i', t_{-i}'), (t_i, t_{-i})\big)$, for all $j \in I$. Hence,

$$\begin{aligned}
u_i\big(g(m_i', m_{-i}), (t_i, t_{-i})\big) &= u_i\big(f_i(t_i, t_{-i}'), (t_i, t_{-i})\big) \\
&> \frac{1}{n} \sum_{j \in I} u_i\big(f_j(t_i', t_{-i}'), (t_i, t_{-i})\big) \\
&= u_i\big(g(m_i, m_{-i}), (t_i, t_{-i})\big).
\end{aligned}$$

It follows from the above arguments that $u_i\big(g(m_i', m_{-i}), (t_i, t_{-i})\big) > u_i\big(g(m_i, m_{-i}), (t_i, t_{-i})\big)$, for all $(t_{-i}, m_{-i})$ in the support of $\lambda_i$. However, this contradicts the hypothesis that $m_i$ is a best response against $\lambda_i$, expressed in (3). ■

**Step 4:** The mechanism $\Gamma$ satisfies the BRP.

**Proof of Step 4**: Pick any $i \in I$, $t_i \in T_i$, and $\lambda_i \in \Delta(T_{-i} \times M_{-i})$ such that $\text{marg}_{T_{-i}} \lambda_i = \pi_i(t_i)$. In Step 3, we have in fact shown that for type $t_i$, reporting $(t_i, f_i, 1)$ is strictly better than reporting any $(t_i^{'}, f, z_i)$ such that $t_i^{'} \neq t_i$ in every ex-post realization of $(t_{-i}, m_{-i}) \in T_{-i} \times M_{-i}$. The same arguments can be repeated to show that for type $t_i$, reporting $(t_i, f_i, 1)$ is weakly better than reporting any $(t_i, f, z_i)$ in every ex-post realization of $(t_{-i}, m_{-i}) \in T_{-i} \times M_{-i}$. Thus, $(t_i, f_i, 1)$ is a best response for type $t_i$ against any $\lambda_i$ such that $\text{marg}_{T_{-i}} \lambda_i = \pi_i(t_i)$. ■

**Step 5:** The mechanism $\Gamma$ implements $F$ in interim rationalizable strategies.

**Proof of Step 5**: Pick any $f \in H_{F'}$. Let $\sigma$ be a strategy profile such that $\sigma_i(t_i) = (t_i, f, 0)$, for all $i \in I$. Then, we have $\sigma \in \Sigma(S^{\Gamma})$ so that $g(\sigma(t)) = f(t)$, for all $t \in T$. This verifies the first requirement of implementation.

Next, pick any $\sigma \in \Sigma(S^{\Gamma})$. Consider any $t \in T$ and the corresponding message profile $\sigma(t) = \big(\sigma^1(t), \sigma^2(t), \sigma^3(t)\big)$. By the construction of the mechanism, there exists a finite set of SCFs, say $\{f^1, \ldots, f^Z\} \subseteq H_{F'}$, such that $g(\sigma(t)) = \frac{1}{Z}\sum_{z=1}^{Z} f^z(\sigma^1(t))$. It follows from Step 3 that $\sigma^1(t) = t$, for all $t \in T$. Hence, $g(\sigma(t)) = \frac{1}{Z}\sum_{z=1}^{Z} f^z(t) \in F^{'}(t)$, where the inclusion in $F^{'}$ is due to the assumption that $F^{'}$ is convex-valued. So, $g \circ \sigma$ is a selection of $F^{'}$, i.e., $g \circ \sigma \in H_{F'}$. This verifies the second requirement of implementation. ■

Steps 1 through 5 complete the proof of the theorem. □

The above sufficiency result can be strengthened by weakening two of its assumptions: convex-valuedness of the SCC and the independence of the dominant SCFs $f_i$ from the types of all $j \neq i$. We discuss these issues in the following remarks.

**Remark 2.** [**On Convex-Valuedness**] We use the assumption that the equivalent SCC $F^{'}$ is convex-valued in Step 5 of the proof of Theorem 3. This assumption helps us conclude that $g \circ \sigma \in H_{F'}$, for all $\sigma \in \Sigma(S^{\Gamma})$. Notice that the constructed mechanism is such that a convex combination of lotteries defines the outcome only in Rule 3. Moreover, this convex combination lies in the set of lotteries induced by the dominant SCFs $\{f_1, \ldots, f_n\}$. Thus, the claim in Step 5 will also hold as long as $F^{'}$ satisfies the following condition: For all subsets $J \subseteq I$, we have $\frac{1}{|J|}\sum_{j \in J} f_j(t) \in F^{'}(t)$, for all $t \in T$. The sufficiency result holds even under this condition, which is of course weaker than assuming $F^{'}$ is convex-valued.

One can completely drop the assumption of convex-valuedness as long as the designer is allowed to use a two-stage mechanism, where the second stage uses a random dictatorship when the agents' reports meet certain conditions. To elaborate on this possibility, consider a two-stage mechanism with the same message space as the one for the mechanism in the proof above. If the message profile satisfies the conditions for Rule 1 or 2, then there is no second stage and the outcome is determined according to those rules. However, if the profile satisfies the conditions for Rule 3, then in the second stage, the designer selects a random dictator from the set of those agents who announce 1; if no one announces 1, the designer selects a random dictator from the set of all agents. After selecting the random dictator in this way, say agent $i$, the designer implements the outcome according to the dominant SCF $f_i$ of agent $i$ at the reported type profile. The ICR correspondence in this two-stage mechanism is the same as in the mechanism in the proof. Therefore, every rationalizable strategy profile is such that every agent reports her type truthfully. As a result, the outcome at the end of the second stage is always an element of the SCC at the true type profile. Hence, the two-stage mechanism implements the SCC in interim rationalizable strategies. ◇

**Remark 3 (On Independence).** We say that the SCC $F$ satisfies *extended setwise dominance* if (i) the SCC $F$ satisfies setwise dominance and (ii) for all $i \in I$, the SCF $f_i$ – found in the definition of setwise dominance – is dominant-strategy incentive compatible. It is straightforward to argue that setwise independent dominance implies extended setwise dominance. Furthermore, Theorem 3 holds even if we replace setwise independent dominance with extended setwise dominance of the SCC.[21]

Interestingly, extended setwise dominance is related to another condition that has been proposed recently by Mukherjee, Muto, and Sen (2024). In a private-values environment, a deterministic SCC $F$ satisfies what Mukherjee et al. (2024) call *extended strategy resistance* if the following two conditions hold: (i) $F$ satisfies strategy-resistance of Jackson (1992) and (ii) for each agent $i \in I$, there exists agent $i$'s top selection $f_i \in H_F$ such that $f_i$ is dominant-strategy incentive compatible.[22] Mukherjee et al. (2024) show that, in private-values environments, extended strategy resistance and the "flip condition" are sufficient for a deterministic SCC to be implementable in weakly undominated strategies by a finite mechanism. Recall from the discussion in Section 5 that strategy resistance is almost equivalent to setwise dominance in private-values environments. (The gap is

[21]Independence of each $f_j$ from the type of agent $i$ is used only in Step 3 of the proof of Theorem 3. There, at multiple points, we apply independence to conclude that $u_i\big(f_j(t_i, t'_{-i}), (t_i, t_{-i})\big) = u_i\big(f_j(t'_i, t'_{-i}), (t_i, t_{-i})\big)$. It is easy to see that the argument that follows this conclusion also holds if $u_i\big(f_j(t_i, t'_{-i}), (t_i, t_{-i})\big) \geq u_i\big(f_j(t'_i, t'_{-i}), (t_i, t_{-i})\big)$, which will be the case if $f_j$ is dominant-strategy incentive compatible.

[22]In the context of private-values environments, agent $i$'s top selection is an SCF $f_i \in H_F$ such that $f_i(t) \in \arg\max_{a \in F(t)} u_i(a, t_i)$, for all $t \in T$.

present because the former requires weak whereas the latter requires strict dominance.) Likewise, extended strategy resistance is almost equivalent to extended setwise dominance in private-values environments. Of course, unlike extended strategy resistance, the extended setwise dominance condition applies to interdependent-values environments as well. ◇

# 7 Examples to Showcase the Significance of our Results

**Example 3 (Efficient Bilateral Trade).** This example addresses the impossibility result in Myerson and Satterthwaite (1983) for bilateral trading, i.e., the nonexistence of an SCF that is Bayesian incentive compatible, ex post efficient, and interim individually rational. We present our framework next.

There is a single buyer $b$ and a single seller $s$ of an indivisible object. The buyer's type is equal to her value $v_b$ for the object whereas the seller's type is her cost $v_s$ for the object. Both players' value/cost are elements of the finite grid

$$V \equiv \left\{0, \frac{1}{K}, \frac{2}{K}, \dots, \frac{K-1}{K}, 1\right\},$$

where $K$ is some positive integer.

Thus, we set $T_b = T_s = V$ as the set of types. The traders' beliefs, given by the functions $\pi_b$ and $\pi_s$, do not play any role in what follows. More importantly, we assume that $T^* = T_b \times T_s$, i.e., the designer cares about all states in $T_b \times T_s$.

An alternative is a pair $(p, z)$, where $p$ specifies the allocation of the object and $z$ the payment from the buyer to the seller. Let $p = 1$ ($p = 0$) denote the allocation where the buyer (seller) receives the object. We assume that the payment $z$ can be any amount in the unit interval $[0, 1]$.[23] As before, $A$ is the set of alternatives.

The utility functions of the buyer and sellers are as follows: On the one hand, if the good is allocated to the buyer and she pays $z$ to the seller, then the buyer's utility is $v_b - z$ whereas the seller's utility is $z - v_s$. On the other, if the good is allocated to the seller and the buyer pays $z$ to the seller, then the buyer's utility is $-z$ whereas the seller's utility is $z$.

We define the *ex-post efficient and individually rational* SCC $F$ as follows: For all

[23] The argument works even if the set of payments is finite as long as it is sufficiently fine.

$(v_b, v_s) \in T_b \times T_s$,

$$F(v_b, v_s) = \begin{cases} \{(1, z) : z \in [v_s, v_b]\}, & \text{if } v_b \geq v_s \\ \{(0, 0)\} & \text{if } v_b < v_s. \end{cases}$$

Notice that $F$ is nonempty- and compact-valued. Consider $H_F$, the SCS generated by $F$. The SCS $H_F$ includes two SCFs, $\bar{f}$ and $\tilde{f}$, that are important for our argument. The SCF $\bar{f}$ is such that the buyer pays a price equal to the seller's cost whenever the good is allocated to the buyer, whereas the SCF $\tilde{f}$ is such that the buyer pays a price equal to her value whenever the good is allocated to the buyer. Of course, since $\bar{f}, \tilde{f} \in H_F$, both of these SCFs are ex-post efficient. The SCF $\bar{f}$ is such that it is weakly dominant for the buyer to report her value truthfully. In fact, as $\bar{f}$ awards all the trade surplus to the buyer, it weakly dominates every SCF in $H_F$ for the buyer, i.e., $u_b\big(\bar{f}(v_b, v_s), v_b\big) \geq u_b\big(f(v_b^{'}, v_s), v_b\big)$, for all $v_b, v_b^{'} \in T_b$, $v_s \in T_s$, and $f \in H_F$. However, $\bar{f}$ does not satisfy $BIC$; specifically, it is in the interest of the seller to misreport her cost even though the buyer is reporting truthfully as long as the seller assigns a positive probability for the buyer to have a value greater than her cost. Similarly, the SCF $\tilde{f}$, which allocates all the trade surplus to the seller, weakly dominates every SCF in $H_F$ for the seller but it does not satisfy the Bayesian incentive constraints for the buyer.

The SCC $F$, however, does not satisfy setwise dominance. Recall that setwise dominance requires that, for each agent $i$, the dominant SCF $f_i$ provide strict incentives for truth-telling. This is not the case here for $\bar{f}$ and $\tilde{f}$. Hence, we now construct another SCC that is approximately equal to $F$, and show that it satisfies setwise dominance. To do so, we first construct two perturbed versions of every SCF $f \in H_F$, as follows.

Consider the SCF $\hat{f}$ such that, for all $(v_b, v_s)$, the good is allocated to the buyer with probability $v_b$ and, regardless of the allocation, the buyer pays $v_b^2/2$ to the seller. Next, pick any $\varepsilon \in (0, 1)$ and $f \in H_F$, and define the SCF $f_b^\varepsilon$ as follows: For all $(v_b, v_s) \in T_b \times T_s$,

$$f_b^\varepsilon(v_b, v_s) = (1 - \varepsilon)f(v_b, v_s) + \varepsilon\hat{f}(v_b, v_s).$$

Also, consider the SCF $\check{f}$ such that, for all $(v_b, v_s)$, the good is allocated to the buyer with probability $(1 - v_s)$ and, regardless of the allocation, the buyer pays $1 - v_s^2/2$ to the seller. Next, pick any $\varepsilon \in (0, 1)$ and $f \in H_F$, and define the SCF $f_s^\varepsilon$ as follows: For all $(v_b, v_s) \in T_b \times T_s$,

$$f_s^\varepsilon(v_b, v_s) = (1 - \varepsilon)f(v_b, v_s) + \varepsilon\check{f}(v_b, v_s).$$

For all $\varepsilon \in (0,1)$, define the SCC $F^\varepsilon$ as follows: For all $(v_b, v_s) \in T_b \times T_s$,

$$F^\varepsilon(v_b, v_s) = \{f_b^\varepsilon(v_b, v_s), f_s^\varepsilon(v_b, v_s) : f \in H_F\}.$$

The SCC $F^\varepsilon$ is nonempty- and compact-valued, for all $\varepsilon$.[24]

We then have the following lemma, whose proof can be found in the appendix:

**Lemma 2.** *For all $\varepsilon \in (0,1)$, the SCC $F^\varepsilon$ satisfies setwise dominance.*

It then follows from Theorem 2 that, for all $\varepsilon \in (0,1)$, the SCC $F^\varepsilon$ is implementable in interim rationalizable strategies by a compact mechanism with the weak BRP. Thus, it is possible to design rationalizable incentives in mechanisms with the weak BRP achieving outcomes that are arbitrarily close to ex-post efficiency in bilateral trading.

How about trying to achieve this goal, but insisting on mechanisms satisfying the more demanding BRP? Although $F^\varepsilon$ satisfies setwise dominance, it fails to satisfy setwise independent dominance. Indeed, to attain anything close to ex-post efficiency, the SCF must respond to changes in the traders' values and costs. Thus, if an SCC satisfies setwise independent dominance, the dominant SCFs for each trader, being independent of the other trader's types, cannot be ex-post efficient or come close to attaining that goal. Another way to see this is that the dominant SCFs in the setwise independent dominance condition satisfy dominant-strategy incentive compatibility, which is of course incompatible with ex-post efficiency when the type space is sufficiently fine (Myerson and Satterthwaite, 1983).

But it is worth noting that, for all $\varepsilon \in (0,1)$, there exists an SCC $F^{'\varepsilon} \supset F^\varepsilon$, a different perturbation of $F$, such that $F^{'\varepsilon}$ satisfies setwise independent dominance. Of course, in

[24]To see compact-valuedness, since $F^\varepsilon(v_b, v_s) \subseteq \Delta(A)$, and $\Delta(A)$ is compact in the weak* topology, it is sufficient to argue that $F^\varepsilon(v_b, v_s)$ is closed. If $v_b < v_s$, then $F^\varepsilon(v_b, v_s)$ is finite. Specifically, it has two elements: (i) the probability measure that assigns the probabilities $(1-\varepsilon)$, $\varepsilon v_b$, and $\varepsilon(1-v_b)$ to the alternatives $(0,0)$, $(1, v_b^2/2)$, and $(0, v_b^2/2)$, respectively, and (ii) the probability measure that assigns the probabilities $(1-\varepsilon)$, $\varepsilon(1-v_s)$, and $\varepsilon v_s$ to the alternatives $(0,0)$, $(1, 1-v_s^2/2)$, and $(0, 1-v_s^2/2)$, respectively. So $F^\varepsilon(v_b, v_s)$ is closed if $v_b < v_s$. Now consider the case $v_b \geq v_s$. Pick any sequence of probability measures $\{\ell_k\}_{k=1}^\infty$ in $F^\varepsilon(v_b, v_s)$ such that $\ell_k$ converges to some probability measure $\ell \in \Delta(A)$ as $k \to \infty$. By construction, $\ell_k$ either assigns the probabilities $(1-\varepsilon)$, $\varepsilon v_b$, and $\varepsilon(1-v_b)$ to the alternatives $(1, z_k)$, $(1, v_b^2/2)$, and $(0, v_b^2/2)$, respectively, or assigns the probabilities $(1-\varepsilon)$, $\varepsilon(1-v_s)$, and $\varepsilon v_s$ to the alternatives $(1, z_k)$, $(1, 1-v_s^2/2)$, and $(0, 1-v_s^2/2)$, respectively. Since $z_k \in [v_s, v_b]$, there exists a convergent subsequence $\{z_{k_j}\}_{j=1}^\infty$ that converges to some $z \in [v_s, v_b]$ as $j \to \infty$. Consider any continuous and bounded function $h : A \to \Re$. For each $k_j$, $\int h d\ell_{k_j}$ is equal to either $(1-\varepsilon)h(1, z_{k_j}) + \varepsilon v_b h(1, v_b^2/2) + \varepsilon(1-v_b)h(0, v_b^2/2)$ or $(1-\varepsilon)h(1, z_{k_j}) + \varepsilon(1-v_s)h(1, 1-v_s^2/2) + \varepsilon v_s h(0, 1-v_s^2/2)$. As $h$ is continuous, the subsequence $\int h d\ell_{k_j}$ converges to either $(1-\varepsilon)h(1,z) + \varepsilon v_b h(1, v_b^2/2) + \varepsilon(1-v_b)h(0, v_b^2/2)$ or $(1-\varepsilon)h(1,z) + \varepsilon(1-v_s)h(1, 1-v_s^2/2) + \varepsilon v_s h(0, 1-v_s^2/2)$ as $j \to \infty$. Hence, $\ell$ must equal either the probability measure that assigns the probabilities $(1-\varepsilon)$, $\varepsilon v_b$, and $\varepsilon(1-v_b)$ to the alternatives $(1, z)$, $(1, v_b^2/2)$, and $(0, v_b^2/2)$, respectively, or the probability measure that assigns the probabilities $(1-\varepsilon)$, $\varepsilon(1-v_s)$, and $\varepsilon v_s$ to the alternatives $(1, z)$, $(1, 1-v_s^2/2)$, and $(0, 1-v_s^2/2)$, respectively. In either case, $\ell \in F^\varepsilon(v_b, v_s)$.

light of the above comments, $F^{'\varepsilon}$ must contain outcomes that are far from being ex-post efficient. To construct $F^{'\varepsilon}$, define two SCFs $\bar{h}$ and $\tilde{h}$, as follows: $\bar{h}(v_b, v_s) = (1, 0)$ whereas $\tilde{h}(v_b, v_s) = (1, 1)$, for all $(v_b, v_s) \in T_b \times T_s$. Thus, both $\bar{h}$ and $\tilde{h}$ always allocate the good to the buyer but the price is zero in the former and 1 in the latter. Clearly, $\bar{h}$ and $\tilde{h}$ are far from being ex-post efficient. Now, consider the SCS $H^{'} \equiv H_F \cup \{\bar{h}, \tilde{h}\}$. As before, we define $f_b^\varepsilon = (1 - \varepsilon)f + \varepsilon\hat{f}$ and $f_s^\varepsilon = (1 - \varepsilon)f + \varepsilon\check{f}$, for all SCFs $f \in H^{'}$ and $\varepsilon \in (0, 1)$. Finally, for all $\varepsilon \in (0, 1)$, we define the SCC $F^{'\varepsilon}$ as follows: For all $(v_b, v_s) \in T_b \times T_s$,

$$F^{'\varepsilon}(v_b, v_s) = \left\{ f_b^\varepsilon(v_b, v_s), f_s^\varepsilon(v_b, v_s) : f \in H^{'} \right\}.$$

Notice that $F^{'\varepsilon}(v_b, v_s) = F^\varepsilon(v_b, v_s) \cup \left\{ \bar{h}^\varepsilon(v_b, v_s), \bar{h}^\varepsilon(v_b, v_s), \tilde{h}^\varepsilon(v_b, v_s), \tilde{h}^\varepsilon(v_b, v_s) \right\}$, for all $(v_b, v_s) \in T_b \times T_s$ Thus, $F^{'\varepsilon} \supset F^\varepsilon$ and it is nonempty- and compact-valued.

By applying similar arguments as in the proof of Lemma 2, we can show that the SCC $F^{'\varepsilon}$ satisfies setwise independent dominance, for all $\varepsilon \in (0, 1)$. Specifically, the SCF $\bar{h}_b^\varepsilon$ ($\tilde{h}_s^\varepsilon$) is independent of the seller's (buyer's) type and dominates every other SCF in $H_{F'^\varepsilon}$ for the buyer (seller). Although the SCC $F^{'\varepsilon}$ satisfies setwise independent dominance, it is not necessarily convex-valued. Thus, we cannot apply Theorem 3. However, as discussed in Remark 2, the assumption of convex-valuedness is not essential. The designer can use a two-stage mechanism in which the second stage is basically a random dictatorship. The two-stage mechanism implements the original SCC $F^{'\varepsilon}$ in interim rationalizable strategies, is compact, and has the BRP. (The small price to pay –especially if $F$ consists of a large number of allocations– for this success in the implementation with BRP is that the extreme SCFs $\bar{h}$ and $\tilde{h}$ also get to be implemented.) $\diamond$

**Example 4** (**Pareto Optimal Allocation of Private Goods).** In this example, we address several impossibility results concerning efficiency in terms of their implementability in Bayesian equilibrium (Palfrey and Srivastava (1987), Chakravorti (1992)). We note that the culprit for these impossibilities is Bayesian monotonicity, a condition that is necessary for implementation in Bayesian equilibrium and that is entirely avoided in our approach. The description of the model follows.

There are a finite number of goods, $L$, where we use $L$ to denote both the set and number of goods. Let $e^l \in \Re_{++}$ denote the endowment of good $l$ in the economy. Then, $e = (e^1, \ldots, e^L)$ is the endowment of all goods in the economy.

An alternative $a = (a_1, \ldots, a_n) \in \mathbb{R}_+^{Ln}$ is an allocation of the endowment, i.e., $\sum_{i \in I} a_i^l = e^l$, for all $l \in L$. The set of alternatives is $A$. When endowed with the Euclidean metric, $A$ is a compact metric space.

As in the general model, there are a finite set of types of each agent and we allow for interdependent values. Each agent cares about only her own consumption. That is, for all $i \in I$, $t \in T$, and $a, \hat{a} \in \mathbb{R}^{Ln}_{+}$, if $a_i = \hat{a}_i$, then $u_i(a,t) = u_i(\hat{a},t)$. We assume that the agents' utility functions are continuous and strictly increasing in the quantity of each good, i.e., for all $i \in I$, $t \in T$, and $a, \hat{a} \in \mathbb{R}^{Ln}_{+}$, if $a_i > \hat{a}_i$, then $u_i(a,t) > u_i(\hat{a},t)$.[25]

The agents' beliefs $(\pi_i)_{i \in I}$ do not play any role in what follows. More importantly, we assume that $T^* = T$, i.e., the designer cares about all states in $T$.

We assume that the environment is such that, for all $i \in I$, there exists an SCF $f_i$ such that truth-telling is a strictly dominant strategy for agent $i$ in the direct mechanism associated with $f_i$, i.e.,

$$u_i\big(f_i(t_i, t^{'}_{-i}), (t_i, t_{-i})\big) > u_i\big(f_i(t^{'}_i, t^{'}_{-i}), (t_i, t_{-i})\big), \forall t_i \in T_i, t^{'}_i \in T_i \backslash \{t_i\}, t_{-i}, t^{'}_{-i} \in T_{-i}. \quad (5)$$

This assumption is satisfied, for instance, in private-values environments under the mild condition that, for all agents $i \in I$, distinct types of agent $i$ induce different preference rankings on $\Delta(A)$ (see the lemma in Abreu and Matsushima, 1992).[26]

The above assumption is clearly a necessary condition for any SCC to satisfy setwise dominance in any given environment. Moreover, whenever the assumption is satisfied, it is without loss of generality to assume that, for each $i \in I$, the SCF $f_i$ is independent of the types of all agents $j \neq i$. This is because for each $i \in I$, we can fix $t^{'}_{-i} \in T_{-i}$ and then define another SCF $f^{'}_i$ such that $f^{'}_i(t_i, t_{-i}) = f_i(t_i, t^{'}_{-i})$, for all $(t_i, t_{-i}) \in T$. Then, the SCF $f^{'}_i$ is independent of the types of all $j \neq i$ and satisfies (5).

So, suppose that the above assumption is satisfied, and that, for each $i \in I$, the SCF $f_i$ is independent of the types of all $j \neq i$. Then, for each agent $i \in I$, define $\bar{a}^i$ as the alternative in which agent $i$ is allocated the aggregate endowment, i.e., $\bar{a}^i_i = e$. Pick an $\varepsilon \in (0,1)$ and define the SCF $f^{\varepsilon}_i$ such that $f^{\varepsilon}_i(t) = (1-\varepsilon)\bar{a}^i + \varepsilon f_i(t)$, for each $t \in T$. Then, by construction,

$$u_i\big(f^{\varepsilon}_i(t_i, t^{'}_{-i}), (t_i, t_{-i})\big) > u_i\big(f^{\varepsilon}_i(t^{'}_i, t^{'}_{-i}), (t_i, t_{-i})\big), \forall t_i \in T_i, t^{'}_i \in T_i \backslash \{t_i\}, t_{-i}, t^{'}_{-i} \in T_{-i}. \quad (6)$$

Clearly, for all $i \in I$ and $\varepsilon \in (0,1)$, the SCF $f^{\varepsilon}_i$ is independent of the types of all $j \neq i$.

The alternative $a \in A$ is *Pareto optimal in state* $t \in T$ if there does not exist another alternative $a^{'} \in A$ such that $u_i(a^{'}, t) \geq u_i(a,t)$, for all $i \in I$, and $u_j(a^{'}, t) > u_j(a,t)$, for at least some $j \in I$. The *Pareto Correspondence* is the SCC $F_p$ such that $F_p(t)$ is the set of Pareto optimal allocations in each state $t \in T$. Since the utility functions are

[25] $a_i > \hat{a}_i \Leftrightarrow a^l_i \geq \hat{a}^l_i$, for all $l \in L$, with at least one strict inequality.

[26] Notice that the assumption that the utility functions are strictly monotonic rules out total indifference over all lotteries.

continuous and strictly monotonic, the Pareto correspondence coincides with the weak Pareto correspondence and is nonempty- and compact-valued.

Fix any $\delta > 0$. For all $t \in T$, let $F_p^\delta(t)$ be the set of all $a \in F_p(t)$ such that $|\bar{a}_i^i - a_i| \geq \delta$, for all $i \in I$. Then, given $\varepsilon \in (0,1)$, define the SCC $F^\varepsilon$ as follows:

$$F^\varepsilon(t) = \bigcup_{i \in I} \{f_i^\varepsilon(t)\} \cup F_p^\delta(t).$$

The SCC $F^\varepsilon$ is clearly nonempty- and compact-valued. For all $i \in I$, the lottery $f_i^\varepsilon(t)$ converges to $\bar{a}^i$ as $\varepsilon \to 0$. Moreover, since the utility functions are strictly monotonic, the alternative $\bar{a}^i \in F_p(t)$, for all $t \in T$. Hence, the set of alternatives $F^\varepsilon(t)$ converges to $\{\bar{a}^1, \ldots, \bar{a}^n\} \cup F_p^\delta(t) \subset F_p(t)$, for all $t \in T$. We now prove the following result:

**Lemma 3.** *There exists $\bar{\varepsilon} > 0$ such that the SCC $F^\varepsilon$ satisfies setwise independent dominance for all $\varepsilon \leq \bar{\varepsilon}$.*

*Proof.* Pick any $i \in I$, $t_i, t_i' \in T_i$, and $t_{-i}, t_{-i}' \in T_{-i}$. Since the utility of agent $i$ is strictly monotonic, there must exist $\kappa > 0$ such that $u_i(\bar{a}^i, (t_i, t_{-i})) > u_i(a, (t_i, t_{-i})) + \kappa$, for all $a \in F_p^\delta(t_i', t_{-i}')$.[27] Then, since the utility functions are continuous and $f_i^\varepsilon(t_i, t_{-i}')$ converges to $\bar{a}^i$ as $\varepsilon \to 0$, there exists a positive number $\varepsilon_i(t_i, t_i', t_{-i}, t_{-i}')$ such that, for all $\varepsilon \leq \varepsilon_i(t_i, t_i', t_{-i}, t_{-i}')$,

$$u_i\big(f_i^\varepsilon(t_i, t_{-i}'), (t_i, t_{-i})\big) > u_i(a, (t_i, t_{-i})), \forall a \in F_p^\delta(t_i', t_{-i}'). \tag{7}$$

Let $\bar{\varepsilon} \equiv \min_{i, t_i, t_i', t_{-i}, t_{-i}'} \varepsilon_i(t_i, t_i', t_{-i}, t_{-i}') > 0$, which is well-defined because the set of agents $I$ and the type space $T$ are finite. Fix $\varepsilon \leq \bar{\varepsilon}$. We now argue that the SCC $F^\varepsilon$ satisfies setwise independent dominance.

Pick any $i \in I$. We show that the SCF $f_i^\varepsilon \in H_{F^\varepsilon}$ is setwise independent dominant for agent $i$. By construction, the SCF $f_i^\varepsilon$ is independent of the types of all $j \neq i$. Now, fix $t_i, t_i' \in T_i$, $t_{-i}, t_{-i}' \in T_{-i}$, and $f \in H_{F^\varepsilon}$. By construction, $f(t_i', t_{-i}') \in F^\varepsilon(t_i', t_{-i}')$ implies that either $f(t_i', t_{-i}') = f_i^\varepsilon(t_i', t_{-i}')$ or $f(t_i', t_{-i}') \in F_p^\delta(t_i', t_{-i}')$. If $f(t_i', t_{-i}') = f_i^\varepsilon(t_i', t_{-i}')$, then using (6), whereas if $f(t_i', t_{-i}') \in F_p^\delta(t_i', t_{-i}')$, then using (7), we obtain that

$$u_i\big(f_i^\varepsilon(t_i, t_{-i}'), (t_i, t_{-i})\big) \geq u_i\big(f(t_i', t_{-i}'), (t_i, t_{-i})\big),$$

[27] We can prove this by contradiction. Suppose, on the contrary, that we can find a sequence of alternatives $a(\kappa) \in F_p^\delta(t_i', t_{-i}')$ such that $u_i(\bar{a}^i, (t_i, t_{-i})) \leq u_i(a(\kappa), (t_i, t_{-i})) + 1/\kappa$, for all $\kappa$. Since $a(\kappa) \in \{a \in A : |\bar{a}_i^i - a_i| \geq \delta\}$ – which is a compact set –, taking the limit as $\kappa \to \infty$, we find an $a(\infty) \in A$ such that $|\bar{a}_i^i - a_i(\infty)| \geq \delta$ and, since the utility functions are continuous, $u_i(\bar{a}^i, (t_i, t_{-i})) \leq u_i(a(\infty), (t_i, t_{-i}))$. As the utility of agent $i$ is strictly monotonic and $a(\infty) \in A$ such that $|\bar{a}_i^i - a_i(\infty)| \geq \delta$, we obtain that $u_i(\bar{a}^i, (t_i, t_{-i})) > u_i(a(\infty), (t_i, t_{-i}))$, a contradiction.

with a strict inequality if $t_i \neq t_i^{'}$, completing the argument. □

It follows that, for all $\varepsilon \in (0, \bar{\varepsilon}]$, the SCC $F^{\varepsilon}$ is implementable in interim rationalizable strategies by a compact mechanism with the weak BRP. Although the SCC $F^{\varepsilon}$ satisfies setwise independent dominance, it is not necessarily convex-valued. Thus, we cannot apply Theorem 3. However, as discussed in Remark 2, the assumption of convex-valuedness is not essential. The designer can use a two-stage mechanism in which the second stage is basically a random dictatorship. The two-stage mechanism implements the original SCC $F^{\varepsilon}$ in interim rationalizable strategies, is compact, and has the BRP. Note how, as $\varepsilon \to 0$, one gets to approximately implement the entire ex-post Pareto correspondence because one can also choose values of $\delta$ arbitrarily close to 0. ◇

**Example 5 (Social Choice with Multidimensional Signals).** Also considering interdependent values, Jehiel and Moldovanu (2001) presents a remarkable impossibility result for the social choice problem (i.e., the society aims to choose a social alternative in every state) in interdependent-values environments with multidimensional signals. We base our discussion on Example 4.4 in their paper, which is described next.

There are two agents $i \in \{1, 2\}$ and three social alternatives $k \in \{A, B, C\}$. Suppose that only agent 1 receives a signal, denoted by $s = (s_A, s_B, s_C)$. Let $S \equiv [0, 1]^3$ be the set of signals.

Let $\alpha_{ki} s_k$ be agent $i$'s value for alternative $k$ in state $s$. Assume that agents' valuations are such that $\alpha_{ki} > 0$, for all $i \in \{1, 2\}$ and $k \in \{A, B\}$, whereas $\alpha_{C1} = \alpha_{C2} = 0$. Thus, regardless of the signal $s \in S$, the alternative $C$ has zero value for both agents – we can thus view alternative $C$ as the status quo option.

The above assumptions further imply that agent's valuations are independent of the third signal, $s_C$. We thus define the set of types of agent 1 simply as $T_1 = [0, 1]^2$, i.e., the set of all possible signals $(s_A, s_B)$. Since agent 2 is uninformed, she has only one type. We thus simplify notation and write $(s_A, s_B)$ for the type profile.

Let $v_i\big(k, (s_A, s_B)\big)$ denote agent $i$'s value for alternative $k$ at the type profile $(s_A, s_B)$. That is,

$$v_i\big(k, (s_A, s_B)\big) = \begin{cases} \alpha_{ki} s_k, & \text{if } k \in \{A, B\} \\ 0, & \text{if } k = C. \end{cases}$$

Utilities are quasilinear such that, for each $i \in \{1, 2\}$ and $(s_A, s_B) \in T_1$, agent $i$'s utility is given by $v_i\big(k, (s_A, s_B)\big) + x_i$ when alternative $k$ is chosen and agent $i$ receives a monetary transfer equal to $x_i$.

Pick any $\bar{x} > \max\{\alpha_{A1}, \alpha_{B1}\}$. Let

$$\bar{A} \equiv \{(k, x_1, x_2) : \ k \in \{A, B, C\}, \ x_1 \in [-\bar{x}, \bar{x}], x_2 = -x_1\}$$

be the set of pure outcomes. Then, $\bar{A}$ is a compact metric space because $\{A, B, C\}$ is compact when endowed with the discrete topology and the set of all $(x_1, x_2) \in \Re^2$ such that $x_1 \in [-\bar{x}, \bar{x}]$ and $x_2 = -x_1$ is compact in $\Re^2$ endowed with the Euclidean distance. Notice that every outcome $(k, x_1, x_2) \in \bar{A}$ satisfies *ex-post budget balancedness* since $x_2 = -x_1$.

For each $(s_A, s_B) \in T_1$, let $\kappa^*(s_A, s_B)$ denote the set of alternatives that maximize the utilitarian social welfare at the type profile $(s_A, s_B)$, i.e.,

$$\kappa^*\big(s_A, s_B\big) = \arg \max_{k \in \{A,B,C\}} v_1\big(k, (s_A, s_B)\big) + v_2\big(k, (s_A, s_B)\big).$$

It follows that all alternatives are efficient at the boundary where $s_A = s_B = 0$. But, excluding that boundary, either only alternative $A$ is efficient, only alternative $B$ is efficient, or both alternatives $A$ and $B$ are efficient. We say that an outcome $(k, x_1, x_2) \in \bar{A}$ is *efficient* at $(s_A, s_B)$ if $k \in \kappa^*(s_A, s_B)$.

Let $f : T_1 \to \bar{A}$ be an SCF. We say that the SCF $f$ is *efficient* if, for each $(s_A, s_B) \in T_1$, the outcome $f(s_A, s_B)$ is efficient at $(s_A, s_B)$. Indeed, Jehiel and Moldovanu (2001) show that for generic valuations, no efficient SCF satisfies BIC. To overcome this impossibility, we show that a slight extension of a sub-correspondence of the efficient SCC – the extension only adds an approximately Pareto efficient SCF – is implementable in interim rationalizable strategies.

Let then the type space $\hat{T}_1$ be *any* finite subset of $T_1$, and assume that $\hat{T}_1 = T_1^*$. As in the other examples, the interim beliefs for the uninformed agent (i.e., agent 2) are of no relevance.

Pick any $\delta \in [0, 1)$. Then define the SCC $F^\delta : \hat{T}_1 \to 2^{\bar{A}}$ as follows: for each $(s_A, s_B) \in \hat{T}_1$,

$$F^\delta(s_A, s_B) = \{(k, x_1, x_2) : k \in \kappa^*(s_A, s_B), -\bar{x} \leq x_1 \leq \delta\bar{x}, x_2 = -x_1\},$$

This means that $F^\delta$ is the SCC such that the efficient alternative is always chosen and agent 1 receives at most $\delta\bar{x} < \bar{x}$ amount of monetary transfer, which is paid for by agent 2. Clearly, $F^\delta$ is nonempty- and compact-valued.

For each $(s_A, s_B) \in \hat{T}_1$, define $\kappa_1^*(s_A, s_B)$ as the set of best alternatives for agent 1 of type $(s_A, s_B)$. Thus,

$$\kappa_1^*(s_A, s_B) = \arg \max_{k \in \{A,B,C\}} v_1\big(k, (s_A, s_B)\big).$$

We now define $\bar{\kappa}_1^*$ as a single-valued selection of $\kappa_1^*$ that satisfies the following:

$$\bar{\kappa}_1^*(s_A, s_B) \in \arg\max_{k \in \kappa_1^*(s_A, s_B)} v_2\big(k, (s_A, s_B)\big).$$

Thus, among the best alternatives for agent 1, $\bar{\kappa}_1^*$ selects that one which gives agent 2 the highest value.

We next define the SCF $f$ as follows: for each $(s_A, s_B) \in \hat{T}_1$,

$$f(s_A, s_B) = (\bar{\kappa}_1^*(s_A, s_B), \bar{x}, -\bar{x}).$$

This means that the SCF $f$ always chooses agent 1's best alternative and allocates the maximal amount of monetary transfer to agent 1, which is paid for by agent 2. Notice that, although not necessarily welfare-maximizing, the outcome $f(s_A, s_B)$ is *Pareto efficient* at the type profile $(s_A, s_B)$.[28]

For each alternative $k \in \{A, B\}$, we define the SCF $f_k$ as follows: for each $(s_A, s_B) \in \hat{T}_1$,

$$f_k(s_A, s_B) = \left(k, -\frac{\alpha_{k1} s_k}{2}, \frac{\alpha_{k1} s_k}{2}\right).$$

That is, the SCF $f_k$ always chooses alternative $k$ and allocates to agent 2 the monetary transfer equal to half of agent 1's value for alternative $k$, which is paid for by agent 1. We also define the SCF $f_C$ such that $f_C(s_A, s_B) = (C, 0, 0)$, for all $(s_A, s_B) \in \hat{T}_1$.

Finally, pick any $\varepsilon \in (0, 1)$, and let the SCF $f_1^\varepsilon$ be as follows: for each $(s_A, s_B) \in \hat{T}_1$,

$$f_1^\varepsilon(s_A, s_B) = (1 - \varepsilon)\, f(s_A, s_B) + \frac{\varepsilon}{2} \sum_{k \in \{A,B\}} \Big(s_k f_k(s_A, s_B) + (1 - s_k) f_C(s_A, s_B)\Big).$$

Thus, the SCF $f_1^\varepsilon$ is such that at each profile $(s_A, s_B)$, the outcome is determined according to a lottery in which the outcome $f(s_A, s_B)$ is selected with probability $1-\varepsilon$, and for each $k \in \{A, B\}$, the outcome $f_k(s_A, s_B)$ is selected with probability $\varepsilon s_k/2$ and the outcome $f_C(s_A, s_B)$ is selected with probability $\varepsilon(1 - s_k)/2$. Notice that as $\varepsilon \to 0$, $f_1^\varepsilon(s_A, s_B) \to f(s_A, s_B)$, for all $(s_A, s_B) \in \hat{T}_1$.

[28] This follows from the definition of $\bar{\kappa}_1^*$. Suppose $f(s_A, s_B)$ were not Pareto efficient at $(s_A, s_B)$. Then there exists an outcome $(k, x_1, x_2) \in \bar{A}$ such that $v_1\big(k, (s_A, s_B)\big) + x_1 \geq v_1\big(\bar{\kappa}_1^*(s_A, s_B), (s_A, s_B)\big) + \bar{x}$ and $v_2\big(k, (s_A, s_B)\big) + x_2 \geq v_2\big(\bar{\kappa}_1^*(s_A, s_B), (s_A, s_B)\big) - \bar{x}$, with at least one strict inequality. By the definition of $\bar{\kappa}_1^*$ and the fact that $x_1 \leq \bar{x}$, it must be that $v_1\big(k, (s_A, s_B)\big) + x_1 = v_1\big(\bar{\kappa}_1^*(s_A, s_B), (s_A, s_B)\big) + \bar{x}$, $k \in \kappa_1^*(s_A, s_B)$, and $x_1 = \bar{x}$. Then, we must have $v_2\big(k, (s_A, s_B)\big) + x_2 > v_2\big(\bar{\kappa}_1^*(s_A, s_B), (s_A, s_B)\big) - \bar{x}$. But, by budget balance, $x_2 = -x_1 = -\bar{x}$. So, it must be that $v_2\big(k, (s_A, s_B)\big) > v_2\big(\bar{\kappa}_1^*(s_A, s_B), (s_A, s_B)\big)$. But that leads to a contradiction because $k \in \kappa_1^*(s_A, s_B)$ and hence, by the definition of $\bar{\kappa}_1^*$, we must $v_2\big(\bar{\kappa}_1^*(s_A, s_B), (s_A, s_B)\big) \geq v_2\big(k, (s_A, s_B)\big)$.

We then define the SCC $F^{\varepsilon}$ as the following extension of $F^{\delta}$: for each $(s_A, s_B) \in \hat{T}_1$,

$$F^{\varepsilon}(s_A, s_B) = F^{\delta}(s_A, s_B) \cup \{f_1^{\varepsilon}(s_A, s_B)\}.$$

We now have the following result:

**Proposition 1.** *There exists $\bar{\varepsilon} > 0$ such that the SCC $F^{\varepsilon}$ satisfies setwise dominance for all $\varepsilon \in (0, \bar{\varepsilon}]$.*

*Proof.* For any $\bar{a} \in \bar{A}$ and $(s_A, s_B) \in \hat{T}_1$, let $u_1\big(\bar{a}, (s_A, s_B)\big)$ denote agent 1's utility when the outcome $\bar{a}$ is implemented at profile $(s_A, s_B)$. Now, pick any $(s_A, s_B), (s_A^{'}, s_B^{'}) \in \hat{T}_1$. By the definition of $F^{\delta}(s_A^{'}, s_B^{'})$, agent 1's monetary transfer is bounded above by $\delta\bar{x}$ in any outcome $\bar{a} \in F^{\delta}(s_A^{'}, s_B^{'})$. Since $f(s_A, s_B)$ chooses agent 1's best alternative at profile $(s_A, s_B)$ and allocates the maximum monetary transfer of $\bar{x}$ to agent 1, we have $u_1\big(f(s_A, s_B), (s_A, s_B)\big) > u_1\big(\bar{a}, (s_A, s_B)\big) + (1-\delta)\bar{x}/2$, for all $\bar{a} \in F^{\delta}(s_A^{'}, s_B^{'})$. Then, since agent 1's utility function is continuous and $f_1^{\varepsilon}(s_A, s_B)$ converges to $f(s_A, s_B)$ as $\varepsilon \to 0$, there exists a positive number $\varepsilon\big((s_A, s_B), (s_A^{'}, s_B^{'})\big)$ such that, for all $\varepsilon \in (0, \varepsilon\big((s_A, s_B), (s_A^{'}, s_B^{'})\big)]$,

$$u_1\big(f_1^{\varepsilon}(s_A, s_B), (s_A, s_B)\big) > u_1\big(\bar{a}, (s_A, s_B)\big), \forall \bar{a} \in F^{\delta}(s_A^{'}, s_B^{'}). \tag{8}$$

Let $\bar{\varepsilon} \equiv \min_{(s_A,s_B),(s_A^{'},s_B^{'}) \in \hat{T}_1} \varepsilon\big((s_A, s_B), (s_A^{'}, s_B^{'})\big) > 0$, which is well-defined because the type space $\hat{T}_1$ is finite. Fix $\varepsilon \in (0, \bar{\varepsilon}]$. We now argue that the SCC $F^{\varepsilon}$ satisfies setwise dominance.

Since agent 2's type plays no role in the example, it is sufficient to show that the SCF $f_1^{\varepsilon} \in H_{F^{\varepsilon}}$ is setwise dominant for agent 1. Fix $(s_A, s_B), (s_A^{'}, s_B^{'}) \in \hat{T}_1$ and $f \in H_{F^{\varepsilon}}$. By construction, $f(s_A^{'}, s_B^{'}) \in F^{\varepsilon}(s_A^{'}, s_B^{'})$ implies that either $f(s_A^{'}, s_B^{'}) = f_1^{\varepsilon}(s_A^{'}, s_B^{'})$ or $f(s_A^{'}, s_B^{'}) \in F^{\delta}(s_A^{'}, s_B^{'})$.

Consider the case when $f(s_A^{'}, s_B^{'}) = f_1^{\varepsilon}(s_A^{'}, s_B^{'})$. Then

$$u_1\big(f_1^{\varepsilon}(s_A, s_B), (s_A, s_B)\big) = (1-\varepsilon)\Big(v_1\big(\bar{\kappa}_1^*(s_A, s_B), (s_A, s_B)\big) + \bar{x}\Big) + \frac{\varepsilon}{2} \sum_{k \in \{A,B\}} \frac{\alpha_{k1} s_k^2}{2}.$$

In contrast,

$$u_1\big(f_1^{\varepsilon}(s_A^{'}, s_B^{'}), (s_A, s_B)\big) = (1-\varepsilon)\Big(v_1\big(\bar{\kappa}_1^*(s_A^{'}, s_B^{'}), (s_A, s_B)\big) + \bar{x}\Big) + \frac{\varepsilon}{2} \sum_{k \in \{A,B\}} \left(s_k^{'}\alpha_{k1}s_k - \frac{\alpha_{k1}(s_k^{'})^2}{2}\right).$$

By the definition of $\bar{\kappa}_1^*$, we have $v_1\big(\bar{\kappa}_1^*(s_A, s_B), (s_A, s_B)\big) \geq v_1\big(\bar{\kappa}_1^*(s_A^{'}, s_B^{'}), (s_A, s_B)\big)$. Moreover, for each $k \in \{A, B\}$, the expression $s_k' \alpha_{k1} s_k - \alpha_{k1}(s_k')^2/2$ is uniquely maximized at $s_k' = s_k$, with the maximum value equal to $\alpha_{k1} s_k^2/2$. Hence, in this case,

$$u_1\big(f_1^\varepsilon(s_A, s_B), (s_A, s_B)\big) \geq u_1\big(f_1^\varepsilon(s_A', s_B'), (s_A, s_B)\big) = u_1\big(f(s_A', s_B'), (s_A, s_B)\big),$$

with a strict inequality if $(s_A', s_B') \neq (s_A, s_B)$. We can use (8) to obtain the same conclusion in the other case, viz., $f(s_A', s_B') \in F^\delta(s_A', s_B')$. This completes the argument. □

It follows from our Theorem 2, that for all $\varepsilon \leq \bar{\varepsilon}$, the SCC $F^\varepsilon$ is implementable in interim rationalizable strategies by a mechanism that satisfies the weak BRP. This example exhibits a stark contrast with Jehiel and Moldovanu's (2001) impossibility result showing that no efficient SCFs are implementable in Bayesian equilibrium. To reiterate our main theme, we overturn this negative result by exploiting the flexibility induced by the correspondence as opposed to single-valued functions and by adopting rationalizability rather than Bayesian equilibrium as the model of behavior in strategic settings. ◇

**Example 6.** Next, we revisit Example 1 once again. The SCC $F$ in that example satisfies setwise independent dominance, as argued in Example 2. But notice that $F$ is not convex-valued. Thus, Theorem 3 does not apply here. Nevertheless, we now show that $F$ is implementable in rationalizable strategies using a finite mechanism, which is trivially compact and satisfies the BRP. Hence, convex-valuedness of the SCC is not necessary for implementation in rationalizable strategies using compact mechanisms with the BRP. Consider the following mechanism $\Gamma = ((M_i)_{i \in I}, g)$, where $M_i = \{m_i^1, m_i^2, \ldots, m_i^{K+1}\}$ for each $i = 1, 2$ and the deterministic outcome function $g(\cdot)$ is given in the table below:

| $g(m)$ | | Agent 2 | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | $m_2^1$ | $m_2^2$ | $m_2^3$ | $m_2^4$ | $\cdots$ | $m_2^{K-1}$ | $m_2^K$ | $m_2^{K+1}$ |
| Agent 1 | $m_1^1$ | $a_1$ | $a_1$ | $a_{K-2}$ | $a_{K-3}$ | $\cdots$ | $a_2$ | $a_{K-1}$ | $a_{K-1}$ |
| | $m_1^2$ | $a_2$ | $a_1$ | $a_1$ | $a_{K-2}$ | $\cdots$ | $a_3$ | $a_{K-1}$ | $a_{K-1}$ |
| | $m_1^3$ | $a_3$ | $a_2$ | $a_1$ | $a_1$ | $\cdots$ | $a_4$ | $a_{K-1}$ | $a_{K-1}$ |
| | $m_1^4$ | $a_4$ | $a_3$ | $a_2$ | $a_1$ | $\cdots$ | $a_5$ | $a_{K-1}$ | $a_{K-1}$ |
| | $\vdots$ | $\vdots$ | $\vdots$ | $\vdots$ | $\vdots$ | $\ddots$ | $\vdots$ | $\vdots$ | $\vdots$ |
| | $m_1^{K-1}$ | $a_1$ | $a_{K-2}$ | $a_{K-3}$ | $a_{K-4}$ | $\cdots$ | $a_1$ | $a_{K-1}$ | $a_{K-1}$ |
| | $m_1^K$ | $a_{K-1}$ | $a_{K-1}$ | $a_{K-1}$ | $a_{K-1}$ | $\cdots$ | $a_{K-1}$ | $a_K$ | $a_{K-1}$ |
| | $m_1^{K+1}$ | $a_1$ | $a_1$ | $a_{K-2}$ | $a_{K-3}$ | $\cdots$ | $a_2$ | $a_{K-1}$ | $a_K$ |

Assume that we choose $q_\alpha$ sufficiently close to one. This means that agent 1 believes with sufficiently high probability that the state is $\alpha$. Then, no message of agent 1 in the mechanism $\Gamma$ is eliminated. We elaborate on this argument further: for each $k \in \{1, \ldots, K+1\}$, $m_1^k$ is a best response to the belief that agent 2 chooses $m_2^k$ for sure.

Recall that agent 2 is informed of the state. In state $\alpha$, no message of agent 2 in the mechanism $\Gamma$ is eliminated. Specifically, we argue as follows: for each $k \in \{1, \ldots, K-2\}$, $m_2^k$ can be a best response to the belief that agent 1 chooses $m_1^{k+1}$ for sure. In addition, $m_2^{K-1}$ can be a best response to the belief that agent 1 chooses $m_1^1$ for sure; and for $k \in \{K, K+1\}$, $m_2^k$ is a best response to the belief that agent 1 chooses $m_1^k$ for sure.

Consider now state $\beta$. In that state, message $m_2^K$ strictly dominates all messages $m_2^1, \ldots, m_2^{K-1}$. Thus, in state $\beta$, agent 2 eliminates all messages $m_2^1, \ldots, m_2^{K-1}$ in the first round of deletion of never best responses. But for $k \in \{K, K+1\}$, $m_2^k$ is a best response to the belief that agent 1 chooses $m_1^k$ for sure.

Furthermore, since $q_\alpha$ is high enough, agent 1 is unable to eliminate any message even after agent 2's first round of elimination (all messages except $m_2^K$ and $m_2^{K+1}$ in state $\beta$).[29] It follows that no further eliminations can occur.

The preceding arguments show that $S_1^\Gamma = M_1$, $S_2^\Gamma(t_\alpha) = M_2$, and $S_2^\Gamma(t_\beta) = \{m_2^K, m_2^{K+1}\}$. Notice that if agent 1 plays the rationalizable message $m_1^{K+1}$, then any outcome in $A$ can be obtained in state $\alpha$ by an appropriate choice of rationalizable message by type $t_\alpha$ of agent 2 and any outcome in $a_k \in \{a_{K-1}, a_K\}$ can be obtained in state $\beta$ if type $t_\beta$ of agent 2 plays the rationalizable message $m_2^{k+1}$. Moreover, as $S_2^\Gamma(t_\beta) = \{m_2^K, m_2^{K+1}\}$, it is impossible to rationalize any outcome other than either $a_{K-1}$ or $a_K$ in state $\beta$. Therefore, while the SCC violates BIC – and a severe violation if $K$ is large –, it easily follows that the mechanism $\Gamma$ implements the SCC $F$ in interim rationalizable strategies. $\diamond$

# 8 Concluding Remarks

We conclude the paper with several remarks on extensions of our results.

**Robust Implementation:** A mechanism *robustly* implements the SCC in interim rationalizable strategies if the implementation succeeds for all type spaces $(\tilde{T}_i, \pi_i)_{i \in I}$ that are consistent with the given payoff environment $\big(A, \{T_i, u_i\}_{i \in I}\big)$. In general, the ICR strategies depend on the interim beliefs $(\pi_i)_{i \in I}$ held by the agents. Robust implementation requires that, regardless of those interim beliefs, the set of SCFs achieved as interim

[29] Specifically, suppose $q_\alpha \geq \big(u_1(a_K) - u_1(a_{K-1})\big)/\big(u_1(a_K) + u_1(a_1) - 2u_1(a_{K-1})\big)$. Notice that the expression on the right-hand side of the inequality is greater than 0.5 but less than 1. Then, for each $k \in \{1, \ldots, K+1\}$, $m_1^k$ is a best response to the belief that agent 2 chooses $m_2^k$ in state $\alpha$ and $m_2^K$ in state $\beta$.

rationalizable outcomes coincides with the SCS generated by the SCC. Since the proofs of Theorems 2 and 3 do not rely on the agents' interim beliefs, it is easy to see that our sufficiency results satisfy the requirements of robust implementation.[30]

**Implementation in $\varepsilon$-ICR:** For any $\varepsilon > 0$, the $\varepsilon$-interim correlated rationalizability ($\varepsilon$-ICR) correspondence, defined as the largest fixed point of the operator that iteratively eliminates never $\varepsilon$-best responses, is a weaker solution concept than ICR (Dekel et al., 2006).[31] We can strengthen Theorems 2 and 3 to obtain implementation in $\varepsilon$-ICR under the same conditions. Indeed, since we have a finite set of types and the setwise (independent) dominance condition requires strict incentives for truth-telling, the mechanisms proposed to prove those results also achieve $\varepsilon$-ICR implementation for small values of $\varepsilon$.

**Infinite Type Spaces:** We are able to extend some of the results to infinite type spaces, namely, a class of compact metric spaces. Specifically, we can show that a weaker version of setwise IC is necessary for interim rationalizable implementation, where now the SCF $f \in H_F$ providing the incentives to type $t_i$ of agent $i$ to be truthful can vary with the type of agent $i$. When there are only two agents ($n = 2$), the sufficiency result in Theorem 3 holds for all type spaces that are compact metric spaces. That sufficiency result also extends to the case of $n \geq 3$ under the assumption that $H_{F'}$ is compact and metrizable. However, we are unable to prove the sufficiency result in Theorem 2 when the type space is a compact metric space. The difficulty lies in showing that the mechanism constructed to prove the theorem satisfies weak BRP. With finite types, an agent $i$ can implement her dominant SCF $f_i$ by announcing a real number that is less than 1 but greater than the real numbers announced by the other agents. But if there are an infinite number of types of the other agents, then it is possible that there exists some $\delta < 1$ such that $[\delta, 1]$ is in the support of the distribution of real numbers announced by the other agents. In that case, agent $i$ can keep increasing her chance of implementing $f_i$ by increasing her number announcement but never want to announce 1 because of the discontinuity in the mechanism.

[30] In contrast to the permissive conditions in our sufficiency results, robust implementation in rationalizable strategies of single-valued rules imposes strong ex-post incentive compatibility constraints; see Bergemann and Morris (2009, 2011), Ollár and Penta (2017), and Kunimoto and Saran (2020).

[31] The $\varepsilon$-best response operator $b^\varepsilon : \mathcal{S} \to \mathcal{S}$ is defined similarly to the best response operator $b$ in Section 3 except that $b_i^\varepsilon(S)[t_i]$ includes all messages that are $\varepsilon$ best responses, i.e., messages with expected payoffs within $\varepsilon$ of the best response.

# Appendix

We now provide the proofs of the results omitted from the main body of the paper.

**Proof of Theorem 3 for the case of** $n = 2$**:** Since $F^{'}$ satisfies setwise independent dominance, for each $i \in I$, there exists $f_i \in H_{F'}$ such that $f_i$ dominates all other SCFs in $H_{F'}$ for agent $i$ and is independent of the types of $j \neq i$. The mechanism we propose below utilizes the collection of such SCFs $\{f_i\}_{i \in I}$. Consider the mechanism $\Gamma = ((M_i)_{i\in I}, g)$, where $M_i = T_i \times H_{F'} \times \{0, 1, 2\}$ with a typical element $m_i = (m_i^1, m_i^2, m_i^3)$, for all $i \in I$, and the outcome function $g$ is defined as follows: for all $m \in M$,

**Rule 1:** If $m_i^2 = f$ and $m_i^3 < 2$, for all $i \in I$, then $g(m) = f(m^1)$.

**Rule 2:** If $m_1^2 \neq m_2^2$ and $m_i^3 < 2$, for all $i \in I$, then

| $g(m)$ | | $m_2^3 = 0$ | $m_2^3 = 1$ |
|---|---|---|---|
| $m_1^3$ | 0 | $f_1(m^1)$ | $f_2(m^1)$ |
| | 1 | $f_2(m^1)$ | $f_1(m^1)$ |

**Rule 3:** In all other cases, there exists an $i \in I$ such that $m_i^3 = 2$. Then the outcome is determined as follows: Let $I_m \equiv \{i \in I : m_i^3 = 2\}$.

$$g(m) = \frac{1}{|I_m|} \sum_{i \in I_m} f_i(m^1).$$

We endow $\{0, 1, 2\}$ with the discrete topology. Then, since $T_i$ is finite, and hence compact for all $i \in I$, and since $H_{F'}$ is compact, it follows that the mechanism $\Gamma$ is compact. The rest of the argument proceeds in five steps.

**Step 1:** The mechanism $\Gamma$ is measurable.

**Proof of Step 1**: Pick any measurable subset $E \subseteq \Delta(A)$. We show that $g^{-1}(E) = \{m \in M : g(m) \in E\}$ is a measurable subset of $M$. To do so, we partition $g^{-1}(E)$ into the following finite number of subsets of $M$:

Let $M^1$ denote the set of all $m \in M$ such that $m$ induces Rule 1 and the outcome $g(m)$ is in $E$. Recall from the proof of Theorem 2 that we introduced $\bar{H} = \bigcup_{f \in H_{F'}} \{f^n\} \subseteq H^n_{F'}$ and $f^n$ is a typical element of $\bar{H}$. Also recall the function $K : T \times \bar{H} \to \Delta(A)$, which was defined as follows: $K(t, f^n) = f(t)$, for all $(t, f^n) \in T \times \bar{H}$. Finally, recall that $K$ is a Carathéodory function that is jointly measurable.

Now, pick any $m \in M^1$. Since $m$ induces Rule 1, it follows that $m^2 \in \bar{H}$, $m^3 = \{0,1\}^n$, and $g(m) = K(m^1, m^2)$. Hence,

$$M^1 = K^{-1}(E) \times \{0,1\}^n.$$

Since $K$ is jointly measurable, $K^{-1}(E)$ is a measurable subset of $T \times \bar{H}$, which in turn is a measurable subset of $T \times H^n_{F'}$. Hence, $M^1$ is a measurable subset of $M$.

Let $M^{2.1}$ denote the set of all $m \in M$ such that $m$ induces Rule 2, $m_1^3 = m_2^3$, and the outcome $g(m) = f_1(m^1)$ is in $E$. Since $m$ induces Rule 2, we must have $m^2 \in H^n_{F'} \backslash \bar{H}$. Moreover, $m_1^3 = m_2^3$ means that the profile $m^3 \in \{(0,0),(1,1)\}$. Finally, since $g(m) = f_1(m^1)$, the profile $m^1$ is in $f_1^{-1}(E)$. Hence,

$$M^{2.1} = f_1^{-1}(E) \times \left(H^n_{F'} \backslash \bar{H}\right) \times \{(0,0),(1,1)\}.$$

Since $f_1$ is measurable, $f_1^{-1}(E)$ is a measurable subset of $T$. Moreover, $H^n_{F'} \backslash \bar{H}$ is a measurable subset of $H^n_{F'}$ because $\bar{H}$ is closed, and hence, a measurable subset of $H^n_{F'}$. Hence, $M^{2.1}$ is a measurable subset of $M$.

Let $M^{2.2}$ denote the set of all $m \in M$ such that $m$ induces Rule 2, $m_1^3 \neq m_2^3$, and the outcome $g(m) = f_2(m^1)$ is in $E$. Since $m$ induces Rule 2, we must have $m^2 \in H^n_{F'} \backslash \bar{H}$. Moreover, $m_1^3 \neq m_2^3$ means that the profile $m^3 \in \{(0,1),(1,0)\}$. Finally, since $g(m) = f_2(m^1)$, the profile $m^1$ is in $f_2^{-1}(E)$. Hence,

$$M^{2.2} = f_2^{-1}(E) \times \left(H^n_{F'} \backslash \bar{H}\right) \times \{(0,1),(1,0)\}.$$

Since $f_2$ is measurable, $f_2^{-1}(E)$ is a measurable subset of $T$. Moreover, as already argued, $H^n_{F'} \setminus \bar{H}$ is a measurable subset of $H^n_{F'}$. Hence, $M^{2.2}$ is a measurable subset of $M$.

Pick any $i \in I$ and let $M^{i,3}$ denote the set of message profiles $m \in M$ such that $m$ induces Rule 3, $I_m = \{i\}$, and the outcome $g(m) = f_i(m^1)$ is in $E$. Pick any $m \in M^{i,3}$. The profile $m^3$ must be such that $m_j^3 < 2$, for $j \neq i$, and $m_i^3 = 2$. The profile $m^2$ is unrestricted in the definition of Rule 3; hence, $m^2 \in H^n_{F'}$. Finally, as $g(m) = f_i(m^1) \in E$, the profile $m^1$ is in $f_i^{-1}(E)$. Thus, if $i = 1$, then $M^{1,3} = f_1^{-1}(E) \times H^n_{F'} \times \{(2,0),(2,1)\}$, whereas if $i = 2$, then $M^{2,3} = f_2^{-1}(E) \times H^n_{F'} \times \{(0,2),(1,2)\}$. Clearly, $M^{i,3}$ is a measurable subset of $M$, for all $i \in I$.

Let $M^3$ denote the set of message profiles $m \in M$ such that $m$ induces Rule 3, $I_m = I$, and the outcome $g(m) = (1/2)f_1(m^1) + (1/2)f_2(m^1) \equiv f_I(m^1)$ is in $E$. Pick any $m \in M^3$. The profile $m^3$ must be such that $m_i^3 = 2$, for all $i \in I$. The profile $m^2$ is unrestricted in the definition of Rule 3; hence, $m^2 \in H^n_{F'}$. Finally, as $g(m) = f_I(m^1) \in E$, the profile $m^1$ is in $f_I^{-1}(E)$. Thus, $M^3 = f_I^{-1}(E) \times H^n_{F'} \times \{(2,2)\}$. Clearly, $M^3$ is a measurable subset of $M$.

Thus, $g^{-1}(E) = M^1 \cup M^{2.1} \cup M^{2.2} \cup \bigcup_{i\in I} M^{i,3} \cup M^3$. Since $g^{-1}(E)$ is a finite union of measurable sets, it is a measurable subset of $M$. ■

**Step 2:** Define the message correspondence profile $S$ as follows: For all $i \in I$,

$$S_i(t_i) = \{t_i\} \times H_{F'} \times \{0,1\}, \forall t_i \in T_i.$$

Then, $S \leq S^\Gamma$.

**Proof of Step 2**: It is sufficient to argue that $S \leq b(S)$. First, consider agent 1 of type $t_1 \in T_1$. Pick any $f' \in H_{F'}$ and $z \in \{0,1\}$. We prove that $(t_1, f', z) \in b_1(S)$.

Let $\sigma_2 : T_2 \to M_2$ be such that $\sigma_2(t_2) = (t_2, f_1, z)$, for all $t_2 \in T_2$. Then consider the probability measure $\lambda_1 \in \Delta(T_2 \times M_2)$ corresponding to the belief of agent 1 that the type $t_2$ is distributed according to $\pi_1(t_2)$ and the agent 2 plays according to the strategy $\sigma_2$. Since $\sigma_2$ is measurable, $G(\sigma_2)$ is a measurable subset of $T_2 \times M_2$. By definition, $\text{marg}_{T_2}\lambda_1 = \pi_1(t_2)$ and $\lambda_1\big(G(\sigma_2)\big) = 1$. Since $G(\sigma_2) \subseteq G(S_2)$, it follows that $\lambda_1\big(G(S_2)\big) = 1$.

Now, given the belief $\lambda_1$, if agent 1 of type $t_1$ reports $(t_1, f', z)$, then either Rule 1 (when $f' = f_1$) or Rule 2 (when $f' \neq f_1$) is applied at any $(t_2, m_2)$ in the support of $\lambda_1$ and the outcome is $f_1(t_1, t_2)$. (This is trivial when Rule 1 is applied; whereas, it follows from the definition of Rule 2 because both agents are reporting the same integer $z < 2$.) Thus, the expected payoff of agent 1 of type $t_1$ when she reports $(t_1, f', z)$ is $U_1(f_1|t_1)$.

Suppose agent 1 of type $t_1$ deviates to some $m_1 = (t_1', f'', z_1) \in M_1$. On the one hand, if $f'' = f_1$, then either Rule 1 (when $z_1 < 2$) or Rule 3 (when $z_1 = 2$) is applied at any $(t_2, m_2)$ in the support of $\lambda_1$ and the outcome is $f_1(t_1', t_2)$. (This is trivial when Rule 1 is applied; whereas, it follows from the definition of Rule 3 because agent 2 is reporting $z < 2$ whereas agent 1 is reporting $z_1 = 2$.) Hence, the expected payoff of agent 1 of type $t_1$ equals $U_1(f_1; t_1'|t_1)$. By setwise independent dominance, we have $U_1(f_1|t_1) \geq U_1(f_1; t_1'|t_1)$. On the other hand, if $f'' \neq f_1$, then either Rule 2 (when $z_1 < 2$) or Rule 3 (when $z_1 = 2$) is applied at any $(t_2, m_2)$ in the support of $\lambda_1$ and the outcome is either $f_1(t_1', t_2)$ or $f_2(t_1', t_2)$. (Notice that since agent 2 is reporting $z < 2$, whenever Rule 3 is applied, the outcome is $f_1(t_1', t_2)$.) By setwise independent dominance, $u_1\big(f_1(t_1, t_2), (t_1, t_2)\big) \geq u_1\big(f_j(t_1', t_2), (t_1, t_2)\big)$, for all $j \in I$. Thus, even in this case, $U_1(f_1|t_1)$ is greater than or equal to the expected payoff of agent 1 of type $t_1$ when she deviates to $(t_1', f'', z_1)$. Hence, agent 1 of type $t_1$ cannot improve by any deviation. Thus, $(t_1, f', z) \in b_1(S)$.

Next, consider agent 2 of type $t_2 \in T_2$. Pick any $f' \in H_{F'}$ and $z \in \{0, 1\}$. We want to prove that $(t_2, f', z) \in b_2(S)$. Let $\sigma_1 : T_1 \to M_1$ be such that $\sigma_1(t_1) = (t_1, f_2, |z - 1|)$, for all $t_1 \in T_1$. We can now essentially make the same argument for agent 2 as the one we made for agent 1 to show that $(t_2, f', z) \in b_2(S)$.■

**Step 3:** $(t_i', f, z_i) \in S_i^\Gamma(t_i) \Rightarrow t_i' = t_i$.

**Proof of Step 3**: Suppose not, i.e., there exists $m_i \equiv (t_i', f, z_i) \in S_i^\Gamma(t_i)$ such that $t_i' \neq t_i$. Then there must exist a belief $\lambda_i \in \Delta(T_{-i} \times M_{-i})$ such that $\lambda_i\big(G(S_{-i}^\Gamma)\big) = 1$, $\text{marg}_{T_{-i}}\lambda_i = \pi_i(t_i)$, and

$$m_i \in \arg\max_{m_i'' \in M_i} \int_{T_{-i}\times M_{-i}} u_i\big(g(m_i'', m_{-i}), (t_i, t_{-i})\big) d\lambda_i. \tag{9}$$

Suppose that instead of $m_i$, agent $i$ of type $t_i$ were to report $m_i' = (t_i, f_i, z_i')$, where $z_i' = 2$. For agent $j \neq i$, pick any $(t_j, m_j)$ in the support of $\lambda_i$. Let us denote $m_j = (t_j', m_j^2, z_j)$.

First, suppose $(m_i, m_j)$ induces Rule 1. Then it must be that $m_j^2 = f$, $z_j < 2$, and $z_i < 2$. Hence, the outcome is $f(t_i', t_j')$. Since $z_i' = 2$, $(m_i', m_j)$ induces Rule 3, and since $z_j < 2$, the outcome is $f_i(t_i, t_j')$. By setwise independent dominance, we have $u_i\big(f_i(t_i, t_j'), (t_i, t_j)\big) > u_i\big(f(t_i', t_j'), (t_i, t_j)\big)$.

Second, suppose $(m_i, m_j)$ induces Rule 2. Then $m_j^2 \neq f$, $z_j < 2$, and $z_i < 2$. Depending on the values of $z_i$ and $z_j$, the outcome $g(m_i, m_j)$ is either $f_i(t_i', t_j')$ or $f_j(t_i', t_j')$. Rule 3 is induced by $(m_i', m_j)$ because $z_i' = 2$. Then, as per Rule 3, $g(m_i', m_j) = f_i(t_i, t_j')$. By setwise independent dominance, we have $u_i\big(f_i(t_i, t_j'), (t_i, t_j)\big) > u_i\big(g(m_i, m_j), (t_i, t_j)\big)$.

Finally, suppose $(m_i, m_j)$ induces Rule 3. There are three possibilities to consider:

1. $z_i = 2$ and $z_j < 2$. Then $I_{(m_i, m_j)} = \{i\}$, and hence, $g(m_i, m_j) = f_i(t_i', t_j')$. Then, $(m_i', m_j)$ also induces Rule 3 because $z_i' = 2$. Since $z_j < 2$, we have $I_{(m_i', m_j)} = \{i\}$. Hence, $g(m_i', m_j) = f_i(t_i, t_j')$. By setwise independent dominance, we have $u_i\big(f_i(t_i, t_j'), (t_i, t_j)\big) > u_i\big(f_i(t_i', t_j'), (t_i, t_j)\big)$.

2. $z_i < 2$ and $z_j = 2$. Then $I_{(m_i, m_j)} = \{j\}$, and hence, $g(m_i, m_j) = f_j(t_i', t_j')$. By setwise independent dominance, we have $u_i\big(f_i(t_i, t_j'), (t_i, t_j)\big) > u_i\big(f_j(t_i', t_j'), (t_i, t_j)\big)$. Hence,

$$u_i\big(f_i(t_i, t_j'), (t_i, t_j)\big) > u_i\big(g(m_i, m_j), (t_i, t_j)\big). \tag{10}$$

Then, $(m_i', m_j)$ also induces Rule 3 because $z_i' = z_j = 2$. Since $z_i' = z_j = 2$, we have

$I_{(m_i',m_j)} = \{i,j\}$. Hence,

$$g(m_i', m_j) = \frac{1}{2} f_i(t_i, t_j') + \frac{1}{2} f_j(t_i, t_j').$$

Since $f_j$ is independent of agent $i$'s type, we have $u_i\big(f_j(t_i,t_j'),(t_i,t_j)\big) = u_i\big(f_j(t_i',t_j'),(t_i,t_j)\big)$. Hence,

$$\begin{aligned}
u_i\big(g(m_i',m_j),(t_i,t_j)\big) &= \frac{1}{2}\Big(u_i\big(f_i(t_i,t_j'),(t_i,t_j)\big) + u_i\big(f_j(t_i,t_j'),(t_i,t_j)\big)\Big)\\
&= \frac{1}{2}\Big(u_i\big(f_i(t_i,t_j'),(t_i,t_j)\big) + u_i\big(f_j(t_i',t_j'),(t_i,t_j)\big)\Big)\\
&= \frac{1}{2}\Big(u_i\big(f_i(t_i,t_j'),(t_i,t_j)\big) + u_i\big(g(m_i,m_j),(t_i,t_j)\big)\Big)\\
&> u_i\big(g(m_i,m_j),(t_i,t_j)\big) \qquad (\because (10)).
\end{aligned}$$

3. $z_i = z_j = 2$. Then $I_{(m_i,m_j)} = \{i,j\}$, and hence,

$$g(m_i, m_j) = \frac{1}{2}\Big(f_i(t_i',t_j') + f_j(t_i',t_j')\Big).$$

Then, $(m_i', m_j)$ again induces Rule 3 because $z_i' = z_j = 2$. We also have $I_{(m_i',m_j)} = \{i,j\}$. Hence,

$$g(m_i', m_j) = \frac{1}{2}\Big(f_i(t_i,t_j') + f_j(t_i,t_j')\Big).$$

Since $f_j$ is independent of agent $i$'s type, we have $u_i\big(f_j(t_i,t_j'),(t_i,t_j)\big) = u_i\big(f_j(t_i',t_j'),(t_i,t_j)\big)$. Also, we have $u_i\big(f_i(t_i,t_j'),(t_i,t_j)\big) > u_i\big(f_i(t_i',t_j'),(t_i,t_j)\big)$ because of setwise independent dominance. Hence,

$$\begin{aligned}
u_i\big(g(m_i',m_j),(t_i,t_j)\big) &= \frac{1}{2}\Big(u_i\big(f_i(t_i,t_j'),(t_i,t_j)\big) + u_i\big(f_j(t_i,t_j'),(t_i,t_j)\big)\Big)\\
&> \frac{1}{2}\Big(u_i\big(f_i(t_i',t_j'),(t_i,t_j)\big) + u_i\big(f_j(t_i,t_j'),(t_i,t_j)\big)\Big)\\
&= \frac{1}{2}\Big(u_i\big(f_i(t_i',t_j'),(t_i,t_j)\big) + u_i\big(f_j(t_i',t_j'),(t_i,t_j)\big)\Big)\\
&= u_i\big(g(m_i,m_j),(t_i,t_j)\big).
\end{aligned}$$

It follows from the above arguments that $u_i\big(g(m_i',m_j),(t_i,t_j)\big) > u_i\big(g(m_i,m_j),(t_i,t_j)\big)$, for all $(t_j, m_j)$ in the support of $\lambda_i$. However, this contradicts the hypothesis that $m_i$ is a best response against $\lambda_i$, expressed in (9). ■

**Step 4:** The mechanism $\Gamma$ satisfies the BRP.

**Proof of Step 4**: Pick any $i \in I$, $t_i \in T_i$, and $\lambda_i \in \Delta(T_j \times M_j)$ such that $\text{marg}_{T_j}\lambda_j = \pi_i(t_i)$. In Step 3, we have in fact shown that for type $t_i$, reporting $(t_i, f_i, 2)$ is strictly better than reporting any $(t_i^{'}, f, z_i)$ such that $t_i^{'} \neq t_i$ in every ex-post realization of $(t_j, m_j) \in T_j \times M_j$. The same arguments can be repeated to show that for type $t_i$, reporting $(t_i, f_i, 2)$ is weakly better than reporting any $(t_i, f, z_i)$ in every ex-post realization of $(t_j, m_j) \in T_j \times M_j$. Thus, $(t_i, f_i, 2)$ is a best response for type $t_i$ against any $\lambda_i$ such that $\text{marg}_{T_j}\lambda_i = \pi_i(t_i)$. ■

**Step 5:** The mechanism $\Gamma$ implements $F$ in interim rationalizable strategies.

**Proof of Step 5**: Pick any $f \in H_{F'}$. Let $\sigma$ be a strategy profile such that $\sigma_i(t_i) = (t_i, f, 0)$, for all $i \in I$. Then, we have $\sigma \in \Sigma(S^\Gamma)$ so that $g(\sigma(t)) = f(t)$, for all $t \in T$. This verifies the first requirement of implementation.

Next, pick any $\sigma \in \Sigma(S^\Gamma)$. Consider any $t \in T$ and the corresponding message profile $\sigma(t) = \big(\sigma^1(t), \sigma^2(t), \sigma^3(t)\big)$. It follows from Step 3 that $\sigma^1(t) = t$, for all $t \in T$. By the construction of the mechanism, if either Rule 1 or 2 is applied at $\sigma(t)$, then the outcome $g(\sigma(t)) = f(\sigma^1(t)) = f(t)$, for some $f \in H_{F'}$; whereas if Rule 3 is applied at $\sigma(t)$, then the outcome $g(\sigma(t))$ is either equal to $f_i(\sigma^1(t)) = f_i(t)$, for some $i \in I$, or $(1/2)f_1(\sigma^1(t)) + (1/2)f_2(\sigma^1(t)) = (1/2)f_1(t) + (1/2)f_2(t)$. Hence, in all cases, $g(\sigma(t)) \in F^{'}(t)$, where the inclusion in $F^{'}$ is due to the assumption that $F^{'}$ is convex-valued. So, $g \circ \sigma$ is a selection of $F^{'}$, i.e., $g \circ \sigma \in H_{F'}$. This verifies the second requirement of implementation. ■

Steps 1 through 5 complete the proof of the theorem. □

**Proof of Lemma 2:** By construction, the SCFs $\bar{f}_b^\varepsilon$ and $\tilde{f}_s^\varepsilon$ that are defined by perturbing the SCFs $\bar{f}$ and $\tilde{f}$, respectively, are elements of $H_{F^\varepsilon}$.

We begin with the buyer. Consider the SCF $\bar{f}_b^\varepsilon$. Pick any $v_b, v_b^{'} \in T_b$, $v_s \in T_s$, and $f \in H_{F^\varepsilon}$. There are four cases to consider.

**Case 1**: $\min\{v_b, v_b^{'}\} \geq v_s$

Then, we have

$$u_b\big(\bar{f}_b^\varepsilon(v_b, v_s), v_b\big) = (1-\varepsilon)(v_b - v_s) + \varepsilon(v_b^2 - v_b^2/2) = (1-\varepsilon)(v_b - v_s) + \varepsilon v_b^2/2.$$

In contrast, there are two possibilities for $f(v_b^{'}, v_s)$:

- If $f(v_b^{'}, v_s) = (1-\varepsilon)f^{'}(v_b^{'}, v_s) + \varepsilon\hat{f}(v_b^{'}, v_s)$, for some $f^{'} \in H_F$, then

$$u_b\big(f(v_b^{'}, v_s), v_b\big) = (1-\varepsilon)(v_b - z) + \varepsilon\big(v_b^{'}v_b - (v_b^{'})^2/2\big) \leq (1-\varepsilon)(v_b - v_s) + \varepsilon\big(v_b^{'}v_b - (v_b^{'})^2/2\big),$$

where the weak inequality follows because $z \geq v_s$. Now, since $v_b^{'} v_b - (v_b^{'})^2/2$ is uniquely maximized at $v_b^{'} = v_b$, we conclude that $u_b\big(\bar{f}_b^{\varepsilon}(v_b, v_s), v_b\big) \geq u_b\big(f(v_b^{'}, v_s), v_b\big)$ with a strict inequality if $v_b^{'} \neq v_b$.

- If $f(v_b^{'}, v_s) = (1-\varepsilon) f^{'}(v_b^{'}, v_s) + \varepsilon \check{f}(v_b^{'}, v_s)$, for some $f^{'} \in H_F$, then

$$\begin{aligned} u_b\big(f(v_b^{'}, v_s), v_b\big) &= (1-\varepsilon)(v_b - z) + \varepsilon\big((1-v_s)v_b - (1 - v_s^2/2)\big) \\ &\leq (1-\varepsilon)(v_b - v_s) + \varepsilon\big((1-v_s)v_b - (1 - v_s^2/2)\big) \quad (\because\ z \geq v_s) \\ &< (1-\varepsilon)(v_b - v_s) + \varepsilon v_b^2/2, \end{aligned}$$

  where the last strict inequality follows because, if $v_b = 1$ and $v_s = 0$, we have

$$(1-v_s)v_b - (1 - v_s^2/2) = 0 < 1/2 = v_b^2/2$$

  while, if either $v_b \neq 1$ or $v_s \neq 0$, we have

$$(1-v_s)v_b - (1 - v_s^2/2) < 1 - (1 - v_s^2/2) = v_s^2/2 \leq v_b^2/2.$$

  Thus, $(1-v_s)v_b-(1-v_s^2/2) < v_b^2/2$, for all $v_b$ and $v_s$. We conclude that $u_b\big(\bar{f}_b^{\varepsilon}(v_b, v_s), v_b\big) > u_b\big(f(v_b^{'}, v_s), v_b\big)$.

**Case 2**: $v_b \geq v_s > v_b^{'}$

In this case, we have $u_b\big(\bar{f}_b^{\varepsilon}(v_b, v_s), v_b\big) = (1-\varepsilon)(v_b - v_s) + \varepsilon v_b^2/2$. In contrast, there are two possibilities for $f(v_b^{'}, v_s)$:

- If $f(v_b^{'}, v_s) = (1-\varepsilon) f^{'}(v_b^{'}, v_s) + \varepsilon \hat{f}(v_b^{'}, v_s)$, for some $f^{'} \in H_F$, then

$$u_b\big(f(v_b^{'}, v_s), v_b\big) = \varepsilon\big(v_b^{'} v_b - (v_b^{'})^2/2\big) < \varepsilon v_b^2/2. \quad (\because\ v_b^{'} \neq v_b)$$

  So, we conclude that $u_b\big(\bar{f}_b^{\varepsilon}(v_b, v_s), v_b\big) > u_b\big(f(v_b^{'}, v_s), v_b\big)$.

- If $f(v_b^{'}, v_s) = (1-\varepsilon) f^{'}(v_b^{'}, v_s) + \varepsilon \check{f}(v_b^{'}, v_s)$, for some $f^{'} \in H_F$, then

$$u_b\big(f(v_b^{'}, v_s), v_b\big) = \varepsilon\big((1-v_s)v_b - (1 - v_s^2/2)\big) < \varepsilon v_b^2/2,$$

  where the strict inequality follows, as already argued in Case 1. We again conclude that $u_b\big(\bar{f}_b^{\varepsilon}(v_b, v_s), v_b\big) > u_b\big(f(v_b^{'}, v_s), v_b\big)$.

**Case 3**: $v_b^{'} \geq v_s > v_b$

In this case, we have $u_b\big(\bar{f}_b^\varepsilon(v_b, v_s), v_b\big) = \varepsilon v_b^2/2$. In contrast, there are two possibilities for $f(v_b^{'}, v_s)$:

- If $f(v_b^{'}, v_s) = (1-\varepsilon)f^{'}(v_b^{'}, v_s) + \varepsilon\hat{f}(v_b^{'}, v_s)$, for some $f^{'} \in H_F$, then

$$\begin{aligned} u_b\big(f(v_b^{'}, v_s), v_b\big) &= (1-\varepsilon)(v_b - z) + \varepsilon\big(v_b^{'}v_b - (v_b^{'})^2/2\big) \\ &\leq (1-\varepsilon)(v_b - v_s) + \varepsilon\big(v_b^{'}v_b - (v_b^{'})^2/2\big) \ \ (\because\ z \geq v_s) \\ &< \varepsilon v_b^2/2 \ \ (\because\ v_s > v_b \text{ and } v_b^{'} \neq v_b). \end{aligned}$$

  So, we conclude that $u_b\big(\bar{f}_b^\varepsilon(v_b, v_s), v_b\big) > u_b\big(f(v_b^{'}, v_s), v_b\big)$.

- If $f(v_b^{'}, v_s) = (1-\varepsilon)f^{'}(v_b^{'}, v_s) + \varepsilon\check{f}(v_b^{'}, v_s)$, for some $f^{'} \in H_F$, then

$$\begin{aligned} u_b\big(f(v_b^{'}, v_s), v_b\big) &= (1-\varepsilon)(v_b - z) + \varepsilon\big((1-v_s)v_b - (1 - v_s^2/2)\big) \\ &\leq (1-\varepsilon)(v_b - v_s) + \varepsilon\big((1-v_s)v_b - (1 - v_s^2/2)\big) \quad (\because\ z \geq v_s) \\ &< (1-\varepsilon)(v_b - v_s) + \varepsilon v_b^2/2 \quad \text{(as already argued in Case 1)} \\ &< \varepsilon v_b^2/2 \quad (\because\ v_s > v_b). \end{aligned}$$

  We again conclude that $u_b\big(\bar{f}_b^\varepsilon(v_b, v_s), v_b\big) > u_b\big(f(v_b^{'}, v_s), v_b\big)$.

**Case 4**: $v_s > \max\{v_b, v_b^{'}\}$

In this case, we have $u_b\big(\bar{f}_b^\varepsilon(v_b, v_s), v_b\big) = \varepsilon v_b^2/2$. In contrast, there are two possibilities for $f(v_b^{'}, v_s)$:

- If $f(v_b^{'}, v_s) = (1-\varepsilon)f^{'}(v_b^{'}, v_s) + \varepsilon\hat{f}(v_b^{'}, v_s)$, for some $f^{'} \in H_F$, then

$$u_b\big(f(v_b^{'}, v_s), v_b\big) = \varepsilon\big(v_b^{'}v_b - (v_b^{'})^2/2\big) \leq \varepsilon v_b^2/2,$$

  with a strict inequality whenever $v_b^{'} \neq v_b$. So, we conclude that $u_b\big(\bar{f}_b^\varepsilon(v_b, v_s), v_b\big) \geq u_b\big(f(v_b^{'}, v_s), v_b\big)$ with a strict inequality whenever $v_b^{'} \neq v_b$.

- If $f(v_b^{'}, v_s) = (1-\varepsilon)f^{'}(v_b^{'}, v_s) + \varepsilon\check{f}(v_b^{'}, v_s)$, then

$$u_b\big(f(v_b^{'}, v_s), v_b\big) = \varepsilon\big((1-v_s)v_b - (1 - v_s^2/2)\big) < \varepsilon v_b^2/2,$$

  as already argued in Case 1. We conclude that $u_b\big(\bar{f}_b^\varepsilon(v_b, v_s), v_b\big) > u_b\big(f(v_b^{'}, v_s), v_b\big)$.

Now, consider the seller and the SCF $\tilde{f}_s^\varepsilon$. Pick any $v_s, v_s^{'} \in T_s$, $v_b \in T_b$, and $f \in H_{F^\varepsilon}$. There are four cases to consider.

**Case I**: $v_b \geq \max\{v_s, v_s^{'}\}$

In this case, we have

$$\begin{aligned} u_s\big(\tilde{f}_s^{\varepsilon}(v_b, v_s), v_s\big) &= (1-\varepsilon)(v_b - v_s) + \varepsilon\big(1 - v_s^2/2 - (1-v_s)v_s\big) \\ &= (1-\varepsilon)(v_b - v_s) + \varepsilon\big(1 - v_s + v_s^2/2\big). \end{aligned}$$

In contrast, there are two possibilities for $f(v_b, v_s^{'})$:

- If $f(v_b, v_s^{'}) = (1-\varepsilon)f^{'}(v_b, v_s^{'}) + \varepsilon\hat{f}(v_b, v_s^{'})$, for some $f^{'} \in H_F$, then

$$\begin{aligned} u_s\big(f(v_b, v_s^{'}), v_s\big) &= (1-\varepsilon)(z - v_s) + \varepsilon\big(v_b^2/2 - v_b v_s\big) \\ &\leq (1-\varepsilon)(v_b - v_s) + \varepsilon\big(v_b^2/2 - v_b v_s\big) \quad (\because\ z \leq v_b) \\ &< (1-\varepsilon)(v_b - v_s) + \varepsilon\big(1 - v_s + v_s^2/2\big), \end{aligned}$$

  where the last strict inequality follows because, for all values of $v_b$ and $v_s$, we have

$$\begin{aligned} v_b^2/2 - v_b v_s &= \frac{1}{2}(v_b - v_s)^2 - \frac{1}{2}v_s^2 \\ &< \frac{1}{2}(v_b - v_s)^2 + \frac{1}{2} \\ &\leq \frac{1}{2}(1 - v_s)^2 + \frac{1}{2} = (1 - v_s) + v_s^2/2. \end{aligned}$$

  Hence, we conclude that $u_s\big(\tilde{f}_s^{\varepsilon}(v_b, v_s), v_s\big) > u_s\big(f(v_b, v_s^{'}), v_s\big)$.

- If $f(v_b, v_s^{'}) = (1-\varepsilon)f^{'}(v_b, v_s^{'}) + \varepsilon\check{f}(v_b, v_s^{'})$, for some $f^{'} \in H_F$, then

$$\begin{aligned} u_s\big(f(v_b, v_s^{'}), v_s\big) &= (1-\varepsilon)(z - v_s) + \varepsilon\big(1 - (v_s^{'})^2/2 - (1 - v_s^{'})v_s\big) \\ &\leq (1-\varepsilon)(v_b - v_s) + \varepsilon\big(1 - v_s + v_s^{'}v_s - (v_s^{'})^2/2\big) \quad (\because z \leq v_b) \\ &\leq (1-\varepsilon)(v_b - v_s) + \varepsilon\big(1 - v_s + v_s^2/2\big), \end{aligned}$$

  where the last inequality is strict if $v_s^{'} \neq v_s$ because $v_s^{'}v_s - (v_s^{'})^2/2$ is uniquely maximized at $v_s^{'} = v_s$. We conclude that $u_s\big(\tilde{f}_s^{\varepsilon}(v_b, v_s), v_s\big) \geq u_s\big(f(v_b, v_s^{'}), v_s\big)$ with a strict inequality if $v_s^{'} \neq v_s$.

**Case II**: $v_s^{'} > v_b \geq v_s$

In this case, we have $u_s\big(\tilde{f}_s^{\varepsilon}(v_b, v_s), v_s\big) = (1-\varepsilon)(v_b - v_s) + \varepsilon\big(1 - v_s + v_s^2/2\big)$. In contrast, there are two possibilities for $f(v_b, v_s^{'})$:

- If $f(v_b, v_s^{'}) = (1-\varepsilon)f^{'}(v_b, v_s^{'}) + \varepsilon\hat{f}(v_b, v_s^{'})$, for some $f^{'} \in H_F$, then

$$u_s\big(f(v_b, v_s^{'}), v_s\big) = \varepsilon\big(v_b^2/2 - v_b v_s\big) < \varepsilon\big(1 - v_s + v_s^2/2\big),$$

  as already argued in Case I. So, we conclude that $u_s\big(\tilde{f}_s^{\varepsilon}(v_b, v_s), v_s\big) > u_s\big(f(v_b, v_s^{'}), v_s\big)$.

- If $f(v_b, v_s^{'}) = (1-\varepsilon)f^{'}(v_b, v_s^{'}) + \varepsilon\check{f}(v_b, v_s^{'})$, for some $f^{'} \in H_F$, then

$$u_s\big(f(v_b, v_s^{'}), v_s\big) = \varepsilon\big(1 - (v_s^{'})^2/2 - (1 - v_s^{'})v_s\big) < \varepsilon\big(1 - v_s + v_s^2/2\big)$$

  where the strict inequality follows because $v_s^{'} \neq v_s$. We again conclude that $u_s\big(\tilde{f}_s^{\varepsilon}(v_b, v_s), v_s\big) > u_s\big(f(v_b, v_s^{'}), v_s\big)$.

**Case III**: $v_s > v_b \geq v_s^{'}$

In this case, we have $u_s\big(\tilde{f}_s^{\varepsilon}(v_b, v_s), v_s\big) = \varepsilon\big(1 - v_s + v_s^2/2\big)$. In contrast, there are two possibilities for $f(v_b, v_s^{'})$:

- If $f(v_b, v_s^{'}) = (1-\varepsilon)f^{'}(v_b, v_s^{'}) + \varepsilon\hat{f}(v_b, v_s^{'})$, for some $f^{'} \in H_F$, then

$$\begin{aligned} u_s\big(f(v_b, v_s^{'}), v_s\big) &= (1-\varepsilon)(z - v_s) + \varepsilon\big(v_b^2/2 - v_b v_s\big) \\ &\leq (1-\varepsilon)(v_b - v_s) + \varepsilon\big(v_b^2/2 - v_b v_s\big) \quad (\because\ z \leq v_b) \\ &< (1-\varepsilon)(v_b - v_s) + \varepsilon\big(1 - v_s + v_s^2/2\big) \quad \text{(as already argued in Case I)} \\ &< \varepsilon\big(1 - v_s + v_s^2/2\big) \quad (\because\ v_s > v_b). \end{aligned}$$

  So, we conclude that $u_s\big(\tilde{f}_s^{\varepsilon}(v_b, v_s), v_s\big) > u_s\big(f(v_b, v_s^{'}), v_s\big)$.

- If $f(v_b, v_s^{'}) = (1-\varepsilon)f^{'}(v_b, v_s^{'}) + \varepsilon\check{f}(v_b, v_s^{'})$, for some $f^{'} \in H_F$, then

$$\begin{aligned} u_s\big(f(v_b, v_s^{'}), v_s\big) &= (1-\varepsilon)(z - v_s) + \varepsilon\big(1 - (v_s^{'})^2/2 - (1 - v_s^{'})v_s\big) \\ &< (1-\varepsilon)(v_b - v_s) + \varepsilon\big(1 - v_s + v_s^2/2\big) \quad (\because\ z \leq v_b \text{ and } v_s^{'} \neq v_s) \\ &< \varepsilon\big(1 - v_s + v_s^2/2\big) \quad (\because\ v_s > v_b). \end{aligned}$$

  We again conclude that $u_s\big(\tilde{f}_s^{\varepsilon}(v_b, v_s), v_s\big) > u_s\big(f(v_b, v_s^{'}), v_s\big)$.

**Case IV**: $\min\{v_s, v_s^{'}\} > v_b$

In this case, we have $u_s\big(\tilde{f}_s^{\varepsilon}(v_b, v_s), v_s\big) = \varepsilon\big(1 - v_s + v_s^2/2\big)$. In contrast, there are two possibilities for $f(v_b, v_s^{'})$:

- If $f(v_b, v_s') = (1-\varepsilon) f'(v_b, v_s') + \varepsilon \hat{f}(v_b, v_s')$, for some $f' \in H_F$, then

$$u_s\big(f(v_b, v_s'), v_s\big) = \varepsilon\big(v_b^2/2 - v_b v_s\big) < \varepsilon\big(1 - v_s + v_s^2/2\big),$$

  where the strict inequality follows, as already argued in Case I. So, we conclude that $u_s\big(\tilde{f}_s^{\varepsilon}(v_b, v_s), v_s\big) > u_s\big(f(v_b, v_s'), v_s\big)$.

- If $f(v_b, v_s') = (1-\varepsilon) f'(v_b, v_s') + \varepsilon \check{f}(v_b, v_s')$, for some $f' \in H_F$, then

$$u_s\big(f(v_b, v_s'), v_s\big) = \varepsilon\big(1 - (v_s')^2/2 - (1 - v_s')v_s\big) \leq \varepsilon\big(1 - v_s + v_s^2/2\big)$$

  with a strict inequality whenever $v_s' \neq v_s$. We conclude that $u_s\big(\tilde{f}_s^{\varepsilon}(v_b, v_s), v_s\big) \geq u_s\big(f(v_b, v_s'), v_s\big)$ with a strict inequality if $v_s' \neq v_s$.

This completes the proof of the lemma. □

# References


[1] Abreu, D. and H. Matsushima (1992), "Virtual Implementation in Iteratively Undominated Strategies: Incomplete Information," Mimeo, Princeton University.

[2] Aliprantis, C.D. and K.C. Border (2006), "Infinite Dimensional Analysis: A Hitchhiker's Guide," *Springer-Verlag.*

[3] Bergemann, D. and S. Morris (2008), "Interim Rationalizable Implementation," Working paper, Yale University, New Haven, CT.

[4] Bergemann, D., S. Morris, and O. Tercieux (2011), "Rationalizable Implementation," *Journal of Economic Theory*, vol. 146, 1253-1274.

[5] Bernheim, D. (1984), "Rationalizable Strategic Behavior." *Econometrica*, vol. 52, 1007-1028.

[6] Brandenburger, A. and E. Dekel (1987), "Rationalizability and Correlated Equilibria," *Econometrica*, vol. 55, 1391-1402.

[7] Chakravorti, B. (1992), "Efficiency and Mechanisms with no Regret," *International Economic Review*, vol. 33, 45-59.

[8] Dasgupta, P., P. Hammond, and E. Maskin (1979), "Implementation of Social Choice Rules: Some General Results on Incentive Compatibility," *Review of Economic Studies*, vol. 46, 195-216.

[9] D'Aspremont, C. and P.-A. Gerard-Varet (1979), "Incentives and Incomplete Information," *Journal of Public Economics*, vol. 11, 25-45.

[10] de Clippel, G., R. Saran, and R. Serrano (2019), "Level-$k$ Mechanism Design," *Review of Economic Studies*, vol. 86, 481-501.

[11] Dekel, E., D. Fudenberg, and S. Morris (2006), "Topologies on Types," *Theoretical Economics*, vol. 1, 275-309.

[12] Dekel, E., D. Fudenberg, and S. Morris (2007), "Interim Correlated Rationalizability," *Theoretical Economics*, vol. 2, 15-40.

[13] Green, J. R. and J.-J. Laffont (1979), *Incentives in Public Decision Making*, Amsterdam, North Holland.

[14] Harris, M. and R. Townsend (1981), "Resource Allocation with Asymmetric Information," *Econometrica*, vol. 49, 33-64.

[15] Holmström, B. and R. B. Myerson (1983),, "Efficient and Durable Decision Rules with Incomplete Information," *Econometrica*, vol. 51, 1799-1819.

[16] Jackson, M. (1992), "Implementation in Undominated Strategies: A Look at Bounded Mechanisms," *Review of Economic Studies*, vol. 59(4), 757-775.

[17] Jehiel, P. and B. Moldovanu (2001), "Efficient Design with Interdependent Valuations," *Econometrica*, vol. 69, 1237-1259.

[18] Jain, R. (2021), "Rationalizable Implementation of Social Choice Correspondences," *Games and Economic Behavior*, vol. 127, 47-66.

[19] Jain, R., V. Korpela, and M. Lombardi (2023), "Two-Player Rationalizable Implementation," SSRN: http://dx.doi.org/10.2139/ssrn.4302053

[20] Jain, R., M. Lombardi, and A. Penta (2024), "Strategically Robust Implementation," SSRN: http://dx.doi.org/10.2139/ssrn.4949387

[21] Kneeland, T. (2022), "Mechanism Design with Level-k Types: Theory and an Application to Bilateral Trade," *Journal of Economic Theory*, vol. 201, 105421.

[22] Kunimoto, T. and R. Saran (2020), "Robust Implementation in Rationalizable Strategies in General Mechanisms," Working Paper, Singapore Management University.

[23] Kunimoto, T. and R. Serrano (2019), "Rationalizable Implementation of Correspondences," *Mathematics of Operations Research*, vol. 44, 1326-1344.

[24] Kunimoto, T. and R. Serrano (2020), "Rationalizable Incentives: Interim Implementation of Sets in Rationalizable Strategies," Bravo Center Working Paper 2020-015, Department of Economics, Brown University.

[25] Kunimoto, T., R. Saran, and R. Serrano (2024), "Interim Rationalizable Implementation of Functions," *Mathematics of Operations Research*, vol. 49(3), 1791-1824.

[26] Lipman, B. (1994), "A Note on the Implications of Common Knowledge of Rationality," *Games and Economic Behavior*, vol. 6, 114-129.

[27] Mukherjee, S., S., Muto, and A. Sen (2024), "Implementation in Undominated Strategies with Applications to Auction Design, Public Good Provision and Matching," *Journal of Economic Theory*, vol. 216, 105783.

[28] Myerson, R. B. (1979), "Incentive Compatibility and the Bargaining Problem," *Econometrica*, vol. 47, 61-73.

[29] Myerson, R. B. (1981), "Optimal Auction Design," *Mathematics of Operations Research*, vol. 6, 58-73.

[30] Myerson, R. and M. Satterthwaite (1983), "Efficient Mechanisms for Bilateral Trading," *Journal of Economic Theory*, vol. 19, 265–281.

[31] Ollár, M. and A. Penta (2017), "Full Implementation and Belief Restrictions," *American Economic Review*, vol. 107, 2243-2377.

[32] Oury, M. and O. Tercieux (2012), "Continuous Implementation," *Econometrica*, vol. 80, 1605-1637.

[33] Palfrey, T. R. and S. Srivastava (1987), "On Bayesian Implementable Allocations," *Review of Economic Studies*, vol. 54, 193-208.

[34] Pearce, D. (1984), "Rationalizable Strategic Behavior and the Problem of Perfection," *Econometrica*, vol. 52, 1029-1050.

[35] Xiong, S. (2023), "Rationalizable Implementation of Social Choice Functions: Complete Characterization," *Theoretical Economics*, vol. 18, 197-230.