Robust implementation in rationalizable strategies in general mechanisms

Takashi KUNIMOTO
Singapore Management University, tkunimoto@smu.edu.sg

Rene SARAN

SMU ECONOMICS & STATISTICS

Robust Implementation in Rationalizable Strategies in General Mechanisms

Takashi Kunimoto, Rene Saran

April 2020

Paper No. 10-2020

Robust Implementation in Rationalizable Strategies in General Mechanisms*

Takashi Kunimoto† Rene Saran‡

April 14, 2020

Abstract

A social choice function (SCF) is robustly implementable in rationalizable strategies if every rationalizable strategy profile on every type space results in outcomes consistent with it. First, we establish an equivalence between robust implementation in rationalizable strategies and "weak rationalizable implementation". Second, using the equivalence result, we identify weak robust monotonicity as a necessary and almost sufficient condition for robust implementation in rationalizable strategies. This exhibits a contrast with robust implementation in interim equilibria, i.e., every equilibrium on every type space achieves outcomes consistent with the SCF. Bergemann and Morris (2011) show that strict robust monotonicity is a necessary and almost sufficient condition for robust implementation in interim equilibria. We argue that strict robust monotonicity is strictly stronger than weak robust monotonicity, which further implies that, within general mechanisms, robust implementation in rationalizable strategies is more permissive than robust implementation in interim equilibria. The gap between robust implementation in rationalizable strategies and that in interim equilibria stems from the strictly stronger nonemptiness requirement inherent in the latter concept.

JEL: C72; D78; D80

Keywords: Ex post incentive compatibility, rationalizability, interim equilibrium, robust implementation, weak rationalizable implementation, weak robust monotonicity

*We thank Roberto Serrano for helpful comments.

†School of Economics, Singapore Management University, 90 Stamford Road, Singapore 178903; tkunimoto@smu.edu.sg

‡Department of Economics, University of Cincinnati, 2906 Woodside Dr, Cincinnati, Ohio 45221, USA. *Email*: rene.saran@uc.edu; *Tel:* +1 513 556 1528

1 Introduction

We consider robust (full) implementation of a social choice function (SCF) in (interim correlated) rationalizable strategies. That is, we want the designer to construct a mechanism such that *every* rationalizable strategy profile results in outcomes that are consistent with the SCF *regardless* of the players' beliefs and higher-order beliefs about each other's types. We thus take a global approach to robustness, following the seminal work of Bergemann and Morris (2005, 2009, 2011). We however depart from the work of Bergemann and Morris in terms of the solution concept: They assume interim equilibrium whereas we assume rationalizability as the solution concept. On a given type space, rationalizability is a weaker solution concept than interim equilibrium, and it characterizes outcomes that are consistent with common knowledge of rationality (Dekel et al., 2007)

We first characterize robust implementation in rationalizable strategies as equivalent to weak rationalizable implementation. In the appendix of their working paper, Bergemann and Morris (2010) define weak rationalizable implementation by imposing conditions directly on messages that are "rationalizable" for *payoff types*. These messages survive the following iterative deletion process: At each stage, for each payoff type of each player, delete every message that is never a best response to any belief over the other players' payoff types and their messages that have survived until the previous stage. Weak rationalizable implementation imposes two conditions: Firstly, if a message profile is rationalizable for a payoff-type profile, then it should be consistent with the SCF at that payoff-type profile. Secondly, each *first-order type* (i.e., a pair of payoff type and first-order belief) of each player has a nonempty best response to some belief over the rationalizable messages of the payoff types of the other players.

The equivalence between robust implementation in rationalizable strategies and weak rationalizable implementation proves useful on two counts: First, we use it to identify *weak robust monotonicity* (weak RM) as a necessary and almost sufficient condition for robustly implementing an SCF in rationalizable strategies. Second, as the name suggests, weak rationalizable implementation is weaker than "rationalizable implementation". Bergemann and Morris (2011) show that robust implementation in interim equilibria implies rationalizable implementation – the converse is true under a particular restriction on the mechanisms. It then follows from our equivalence result that if an SCF is robustly implementable in interim equilibria, then it is robustly implementable in rationalizable strategies.

Bergemann and Morris (2011) show that *strict* robust monotonicity (strict RM) is necessary and almost sufficient for robust implementation in interim equilibria – as well as rationalizable implementation. Strict RM implies weak RM. The converse is true for "responsive" SCFs but not more generally. In Example 6.2, we present an SCF that satisfies weak RM but not strict RM. Thus, there exist SCFs that are robustly implementable in rationalizable strategies but not in interim equilibria. Figure 1 summarizes the relationships between different implementation and monotonicity concepts.

We can understand the gap between robust implementation in rationalizable strategies and that in interim equilibria by comparing the strength and weakness of the two desiderata in the respective definitions. For a given solution concept (interim equilibrium or rationalizability), robust implementation requires: (i) *Nonemptiness*: The solution concept must be nonempty on every type space and (ii) *Uniqueness*: Every outcome associated with the solution concept on every type space must be consistent with the SCF. It turns out that the uniqueness requirement for robust implementation is the same regardless of whether we use rationalizability or interim equilibrium as the solution concept. This is because (a) every interim equilibrium action is rationalizable on a given type space and (b) every rationalizable action on a given type space can be obtained in interim equilibrium on another type space. The latter result is known in the literature on epistemic foundations (see, for instance, Remark 2 in Dekel et al., 2007). It therefore follows that the gap between robust implementation in rationalizable strategies and that in interim equilibria is due to the strictly stronger nonemptiness requirement imposed by the latter solution concept. That is, it is possible that rationalizable strategies exist and the SCF is implementable in rationalizable strategies on all type spaces but there exists a type space on which the mechanism has no interim equilibria.

If we restrict the designer to use finite mechanisms, then interim equilibria will exist on all type spaces. Hence, robust implementation in interim equilibria and that in rationalizable strategies are equivalent under this restriction. The designer will thus be compelled to use countably infinite mechanisms if her aim is to robustly implement an SCF in rationalizable strategies which cannot be robustly implemented in interim equilibria. The implementation literature relies on countably infinite mechanisms to obtain tight necessary and sufficient conditions. In that spirit, we too construct a countably infinite mechanism to prove that weak RM is almost sufficient for robust implementation in rationalizable strategies. However, such constructions have been criticized for being impractical (see, for e.g., Jackson, 1992).

In the context of complete information environments, Bergemann et al. (2011) show that the necessary condition for implementation in rationalizable strategies is stronger than Maskin monotonicity, which is necessary and almost sufficient for Nash implementation (Maskin, 1999). In their Section 5, they also give an example of a Nash implementable SCF that is not implementable in rationalizable strategies. Recently, Xiong (2018) has provided a complete characterization of SCFs that are implementable in rationalizable strategies. The implementing mechanism in Xiong (2018) also Nash implements the SCF. Thus, in

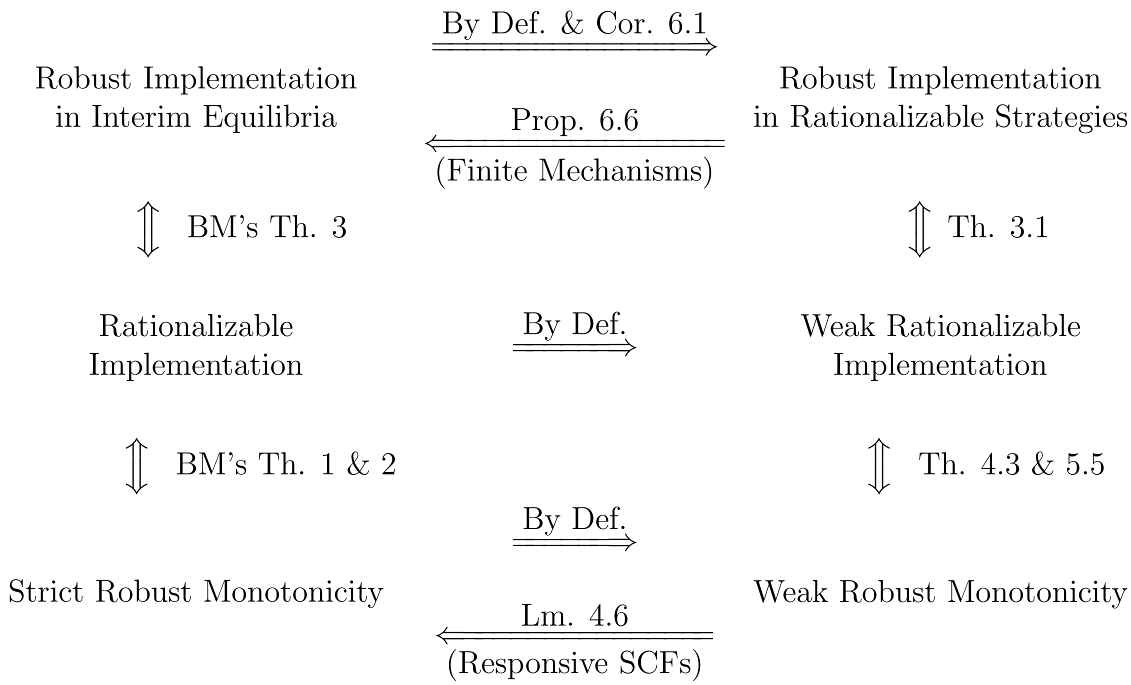

Figure 1: Summary of relationships between different implementation and monotonicity concepts. BM stands for Bergemann and Morris (2011); Cor. stands for Corollary; Def. stands for Definition; Lm. stands for Lemma; and Th. stands for Theorem.

complete information environments, the designer can implement a strictly larger set of SCFs in equilibrium than in rationalizable strategies.[1] In an interesting contrast, we show that the designer can *robustly* implement a strictly larger set of SCFs in rationalizable strategies than in equilibrium.

The rest of the paper is organized as follows. We present the preliminary definitions in Section 2. In Section 3, we show the equivalence between robust implementation in rationalizable strategies and weak rationalizable implementation. In Sections 4 and 5, respectively, we show that weak RM is necessary and almost sufficient for robust implementation in rationalizable strategies. We compare robust implementation in rationalizable strategies and that in interim equilibria in Section 6 before concluding in Section 7.

[1]This is true only for SCFs. For multi-valued social choice correspondences, implementation in rationalizable strategies is strictly weaker than Nash implementation, as shown in Kunimoto and Serrano (2019). Also see Jain (2018).

2 Preliminaries

There is a finite set of players $I = \{1, \ldots, n\}$. A player's *payoff type* is $\theta_i \in \Theta_i$, where we assume that Θ_i is finite. A *payoff state* is $\theta \in \Theta = \times_{i \in N} \Theta_i$. Denote $\Theta_{-i} \equiv \Theta_1 \times \cdots \times \Theta_{i-1} \times \Theta_{i+1} \times \cdots \times \Theta_n$.[2] There is a countable set of alternatives A with at least two elements. We let $\Delta(A)$ to be the set of lotteries over A.[3] We denote an arbitrary lottery by ℓ, and let a be the lottery that puts probability 1 on alternative a. For any lottery ℓ, let $\ell[a]$ be the probability assigned by ℓ to $a \in A$. Let $\mathbb{Z}$ be any countable set of indices. For any countable set of lotteries $\{\ell_z\}_{z \in \mathbb{Z}}$ and corresponding weights $\{\alpha_z\}_{z \in \mathbb{Z}}$ such that $\alpha_z \geq 0, \forall z$, and $\sum_{z \in \mathbb{Z}} \alpha_z = 1$, we let $\sum_{z \in \mathbb{Z}} \alpha_z \ell_z$ be the lottery that is obtained as a reduced form of the compound lottery in which for all $z \in \mathbb{Z}$, lottery ℓ_z is selected with probability α_z.

We endow A with the discrete topology. Thus, A is separable and completely metrizable by the discrete metric, and hence it is a Polish space. As a result, $\Delta(A)$ is also Polish under the weak* topology (Aliprantis and Border, 2006, Theorem 15.15). Therefore, $\Delta(A)$ contains a countable dense subset, which we denote by $\Delta^*(A)$.

Preferences of player i over the set of lotteries are represented by the von Neumann-Morgenstern expected utility function $u_i : \Delta(A) \times \Theta \to \Re$. Thus, for any payoff state θ and lottery ℓ, $u_i(\ell, \theta) = \sum_{a \in A} \ell[a] u_i(a, \theta)$. We assume that utilities are bounded to ensure that expected utility is well defined over the space of lotteries with countable support, i.e., for all $i \in I$ and $\theta \in \Theta$, there exists $\zeta > 0$ such that $|u_i(\ell, \theta)| \leq \zeta$ for all $\ell \in \Delta(A)$.[4]

2.1 Type Space

A type space is a collection $\mathcal{T} = (T_i, \hat{\theta}_i, \hat{\pi}_i)_{i \in I}$ such that for each $i \in I$, T_i is countable, $\hat{\theta}_i : T_i \to \Theta_i$ and $\hat{\pi}_i : T_i \to \Delta(T_{-i})$. A player's *type* $t_i \in T_i$ defines her *payoff type* $\hat{\theta}_i(t_i) \in \Theta_i$ and her *belief type* $\hat{\pi}_i(t_i) \in \Delta(T_{-i})$. For any $t_{-i} \in T_{-i}$, we let $\hat{\pi}_i(t_i)[t_{-i}]$ denote the probability that player i of type t_i assigns to other players having types t_{-i}. We assume that $\hat{\theta}_i : T_i \to \Theta_i$ is surjective for all $i \in I$, i.e., no payoff type is redundant.

For each $i \in I$, let $H_i^0 = \Theta_i$ be the space of *zeroth-order types* of player i. Letting $Z_i^1 = \Delta(H_{-i}^0)$, we can define $H_i^1 = \Theta_i \times Z_i^1$ to be the space of *first-order types* of player i. Thus, the first-order type $h_i^1 = (\theta_i, z_i^1)$ specifies player i's payoff type θ_i and her belief z_i^1 regarding the payoff types of all other individuals. Iterating this way, we define $H_i^k = \Theta_i \times Z_i^k$ to be the space of *k^{th}-order types* of player i, where $Z_i^k = \Delta(H_{-i}^{k-1})$. Finally, we define $H_i^\infty = \times_{k=0}^\infty H_i^k$ as the set of i's all infinite hierarchy of beliefs $(h_i^0, h_i^1, h_i^2, \ldots)$.

[2] Similar notation will be used for products of other sets.

[3] For any set X, we will use $\Delta(X)$ to denote the set of probability measures over X.

[4] See Blackwell and Girshick (1954) for an axiomatization of expected utility over all discrete probability measures on a set, which results in bounded utilities.

There is an infinite hierarchy of beliefs $(h_i^0(t_i), h_i^1(t_i), h_i^2(t_i), \ldots) \in H_i^\infty$ corresponding to each type $t_i \in T_i$: viz., $h_i^0(t_i) = \hat{\theta}_i(t_i)$; $h_i^1(t_i) = (\hat{\theta}_i(t_i), z_i^1(t_i))$, where $z_i^1(t_i)$ is such that $z_i^1(t_i)[\theta_{-i}] = \sum_{t_{-i}:\hat{\theta}_{-i}(t_{-i})=\theta_{-i}} \hat{\pi}_i(t_i)[t_{-i}]$ for all $\theta_{-i} \in \Theta_{-i}$; in general, $h_i^k(t_i) = (\hat{\theta}_i(t_i), z_i^k(t_i))$, where $z_i^k(t_i)$ is such that $z_i^k(t_i)[h_{-i}^{k-1}] = \sum_{t_{-i}:h_{-i}^{k-1}(t_{-i})=h_{-i}^{k-1}} \hat{\pi}_i(t_i)[t_{-i}]$ for all $h_{-i}^{k-1} \in H_{-i}^{k-1}$.

2.2 Social Choice Function and Mechanism

The planner's objective is specified by a *social choice function (henceforth, SCF)* as a function $f : \Theta \to \Delta(A)$.

We say that the SCF f is *responsive to* θ_i *and* θ_i', denoted by $\theta_i' \not\sim_i^f \theta_i$, if $f(\theta_i, \theta_{-i}) \neq f(\theta_i', \theta_{-i})$ for some $\theta_{-i} \in \Theta_{-i}$. Otherwise, f is *non-responsive to* θ_i *and* θ_i', denoted by $\theta_i' \sim_i^f \theta_i$.

The SCF f is *responsive* if for all $i \in I$ and $\theta_i, \theta_i' \in \Theta_i$: $\theta_i \neq \theta_i' \Rightarrow \theta_i \not\sim_i^f \theta_i'$. Otherwise, f is *non-responsive*.

A *mechanism* $\Gamma = ((M_i)_{i\in I}, g)$, where M_i is a countable nonempty set of messages for player i, $M = \times_{i\in I} M_i$, and $g : M \to \Delta(A)$ is the outcome function. The mechanism $\Gamma = ((M_i)_{i\in I}, g)$ is *finite* if M_i is finite for all $i \in I$.

2.3 Rationalizable Strategies

Fix a type space $\mathcal{T}$ and mechanism $\Gamma = ((M_i)_{i\in I}, g)$. A *message correspondence profile* $S = (S_1, \ldots, S_n)$, where each

$$S_i : T_i \to 2^{M_i} \backslash \emptyset.$$

Let $\mathbb{S}$ be the collection of all such message correspondence profiles. The collection $\mathbb{S}$ 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 each $i \in I$ and $t_i \in T_i$. The smallest element is $\underline{S} = (\underline{S}_1, \ldots, \underline{S}_n)$, where $\underline{S}_i(t_i) = \emptyset$ for each $i \in I$ and $t_i \in T_i$.

We define the *best response operator* $b : \mathbb{S} \to \mathbb{S}$ as follows:

$$b_i(S)[t_i] \equiv \left\{ m_i \in M_i : \begin{array}{ll} & \exists \lambda_i \in \Delta(T_{-i} \times M_{-i}) \text{ such that} \\ \text{(i)} & m_i \in \arg\max\limits_{m_i' \in M_i} \sum\limits_{t_{-i}, m_{-i}} \lambda_i(t_{-i}, m_{-i}) u_i(g(m_i', m_{-i}), \hat{\theta}(t_i, t_{-i})) \\ \text{(ii)} & \text{marg}_{T_{-i}} \lambda_i = \hat{\pi}_i(t_i) \\ \text{(iii)} & \lambda_i(t_{-i}, m_{-i}) > 0 \Rightarrow m_{-i} \in S_{-i}(t_{-i}) \end{array} \right\},$$

where $S_{-i}(t_{-i}) = \times_{j\neq i} S_j(t_j)$ for each $t_{-i} \in T_{-i}$.

Starting with the largest element $\bar{S}$ of $\mathbb{S}$, we iteratively eliminate never best responses using the operator b. That is, starting with the message correspondence $B^0 = (B_1^0, \ldots, B_n^0)$ such that $B^0 = \bar{S}$, for each $k \geq 1$, iteratively define the message correspondence $B^k = (B_1^k, \ldots, B_n^k)$ such that for each player $i \in I$ and type $t_i \in T_i$,

$$B_i^k(t_i) = b_i(B^{k-1})[t_i].$$

Observe that b is increasing by definition: i.e., $S \leq S' \Rightarrow b(S) \leq b(S')$. It therefore follows that $B^k \leq B^{k-1}$ for all $k \geq 1$.

As the best response operator b is increasing and $\mathbb{S}$ is a complete lattice, by Tarski's fixed point theorem, there is a largest fixed point of b, which we label B^∞. Thus, (i) $b(B^\infty) = B^\infty$ and (ii) $b(S) \geq S \Rightarrow S \leq B^\infty$.

B^∞ is the (interim correlated) *rationalizable* message correspondence profile (Dekel et al., 2007). For each type of each player, it characterizes the messages that are consistent with common certainty of rationality. In general, $B^\infty \leq B^k$ for all k, i.e., $B_i^\infty(t_i) \subseteq \bigcap_{k=1}^\infty B_i^k(t_i)$ for all $t_i \in T_i$ and $i \in I$. If message sets are finite, $B_i^\infty(t_i) = \bigcap_{k=1}^\infty B_i^k(t_i)$. In this case, we can also construct the fixed point B^∞ by starting with $\bar{S}$ – the largest element of the lattice – and iteratively applying the operator b. But when the mechanism is countably infinite, transfinite induction may be necessary to reach the fixed point. Thus, in this case, $B_i^\infty(t_i)$ are the set of messages surviving (transfinite) iterated deletion of never best responses of type t_i of player i. We refer the reader to Lipman (1994) for the formal treatment.

2.4 $\mathcal{S}^\infty$ Correspondences

As we will see, insisting on implementation that is robust to the underlying type space will force the solution concept to depend only on the payoff types of the individuals. Hence, we need to define strategies that are "rationalizable" for payoff types.

Fix a mechanism $\Gamma = ((M_i)_{i \in I}, g)$. A *message correspondence profile with payoff-type domain* $\mathcal{S} = (\mathcal{S}_1, \ldots, \mathcal{S}_n)$, where each

$$\mathcal{S}_i : \Theta_i \to 2^{M_i} \backslash \emptyset$$

Let $\mathbb{S}^\Theta$ be the collection of such message correspondence profiles with payoff-type domain. The collection $\mathbb{S}^\Theta$ is a complete lattice with the natural ordering of set inclusion: $\mathcal{S} \leq \mathcal{S}'$ if $\mathcal{S}_i(\theta_i) \subseteq \mathcal{S}_i'(\theta_i)$ for all $i \in I$ and $\theta_i \in \Theta_i$. The largest element is $\bar{\mathcal{S}} = (\bar{\mathcal{S}}_1, \ldots, \bar{\mathcal{S}}_n)$, where $\bar{\mathcal{S}}_i(\theta_i) = M_i$ for each $i \in I$ and $\theta_i \in \Theta_i$. The smallest element is $\underline{\mathcal{S}} = (\underline{\mathcal{S}}_1, \ldots, \underline{\mathcal{S}}_n)$, where $\underline{\mathcal{S}}_i(\theta_i) = \emptyset$ for each $i \in I$ and $\theta_i \in \Theta_i$.

We define the *best response operator for payoff types* $b^\Theta : \mathbb{S}^\Theta \to \mathbb{S}^\Theta$ as follows:

$$b_i^\Theta(\mathcal{S})[\theta_i] \equiv \left\{ m_i \in M_i : \begin{array}{ll} & \exists \psi_i \in (\Theta_{-i} \times M_{-i}) \text{ such that} \\ \text{(i)} & m_i \in \arg\max_{m_i'} \sum_{\theta_{-i}, m_{-i}} \psi_i(\theta_{-i}, m_{-i}) u_i(g(m_i', m_{-i}), (\theta_i, \theta_{-i})) \\ \text{(ii)} & \psi_i(\theta_{-i}, m_{-i}) > 0 \Rightarrow m_{-i} \in \mathcal{S}_{-i}(\theta_{-i}) \end{array} \right\},$$

where $\mathcal{S}_{-i}(\theta_{-i}) = \times_{j \neq i} \mathcal{S}_j(\theta_j)$ for each $\theta_{-i} \in \Theta_{-i}$.

As the operator b^Θ is increasing and $\mathbb{S}^\Theta$ is a complete lattice, by Tarski's fixed point theorem, there is a largest fixed point of b^Θ, which is $\mathcal{S}^\infty$. Thus, (i) $b^\Theta(\mathcal{S}^\infty) = \mathcal{S}^\infty$ and (ii) $b^\Theta(\mathcal{S}) \geq \mathcal{S} \Rightarrow \mathcal{S} \leq \mathcal{S}^\infty$.

2.5 Notions of Implementation

In this section, we discuss various notions of implementation. The first one is robust implementation in rationalizable strategies, which is the focus of this paper.

2.5.1 Robust Implementation in Rationalizable Strategies

To define robust implementation in rationalizable strategies, we start by defining what we mean by implementation on a specific type space in rationalizable strategies.

Definition 2.1. A mechanism $\Gamma = ((M_i)_{i \in I}, g)$ *implements the SCF f on the type space $\mathcal{T}$ in rationalizable strategies* if, for all $t \in T$, we have

$$\text{(nonemptiness) } B^\infty(t) \neq \emptyset \quad \text{and} \quad \text{(uniqueness) } g(m) = f(\hat{\theta}(t)), \forall m \in B^\infty(t).$$

We now define robust implementation in rationalizable strategies as implementation over "all type spaces" in rationalizable strategies.

Definition 2.2. A mechanism Γ *robustly implements the SCF f in rationalizable strategies* if, for all type spaces $\mathcal{T}$, the mechanism implements f on $\mathcal{T}$ in rationalizable strategies. The SCF f is *robustly implementable in rationalizable strategies* if there exists a mechanism that robustly implements f in rationalizable strategies.

2.5.2 Robust Implementation in Interim Equilibria

Bergemann and Morris (2011) analyze two notions of implementation: robust implementation in interim equilibria and rationalizable implementation.

To define the former, consider a type space $\mathcal{T}$ and a mechanism $\Gamma = ((M_i)_{i\in I}, g)$. The resulting incomplete information game is denoted by $(\mathcal{T}, \Gamma)$. A *strategy for individual* i in this game is a mapping $\sigma_i : T_i \to \Delta(M_i)$. A strategy profile $\sigma = (\sigma_1, \ldots, \sigma_n)$ is an *interim equilibrium* of the game $(\mathcal{T}, \Gamma)$ if, for all $i \in I$, $t_i \in T_i$, and $m_i \in M_i$ with $\sigma_i(t_i)[m_i] > 0$, we have

$$\sum_{t_{-i}\in T_{-i}} \hat{\pi}_i(t_i)[t_{-i}] \sum_{m_{-i}\in M_{-i}} \sigma_{-i}(t_{-i})[m_{-i}]u_i\big(g(m_i, m_{-i}), \hat{\theta}(t_i, t_{-i})\big)$$
$$\geq \sum_{t_{-i}\in T_{-i}} \hat{\pi}_i(t_i)[t_{-i}] \sum_{m_{-i}\in M_{-i}} \sigma_{-i}(t_{-i})[m_{-i}]u_i\big(g(m_i', m_{-i}), \hat{\theta}(t_i, t_{-i})\big), \forall m_i' \in M_i.$$

We then have the following notion of interim implementation:

Definition 2.3. A mechanism $\Gamma = ((M_i)_{i\in I}, g)$ *interim implements the SCF* f *on the type space* $\mathcal{T}$ if (nonemptiness) the game $(\mathcal{T}, \Gamma)$ has an interim equilibrium and (uniqueness) for every interim equilibrium σ of the game $(\mathcal{T}, \Gamma)$, if $\sigma(t)[m] > 0$, then $g(m) = f(\hat{\theta}(t))$.

Robust implementation in interim equilibria is defined as interim implementation "over all type spaces".

Definition 2.4. A mechanism Γ *robustly implements the SCF* f *in interim equilibria* if, for all type spaces $\mathcal{T}$, the mechanism Γ interim implements f on $\mathcal{T}$. The SCF f is *robustly implementable in interim equilibria* if there exists a mechanism that robustly implements f in interim equilibria.

2.5.3 Rationalizable Implementation

Bergemann and Morris (2011) define rationalizable implementation by directly imposing two conditions on the $\mathcal{S}^\infty$ correspondence, as defined below.

Definition 2.5. A mechanism $\Gamma = ((M_i)_{i\in I}, g)$ *rationalizably implements the SCF* f if

1. (uniqueness) $m \in \mathcal{S}^\infty(\theta) \Rightarrow g(m) = f(\theta)$; and
2. (nonemptyness) For each $i \in I$ and $z_i^1 \in Z_i^1$, there exists a belief $\psi_i \in \Delta(\Theta_{-i} \times M_{-i})$ such that:
   (a) $\arg\max_{m_i'\in M_i} \sum_{\theta_{-i},m_{-i}} \psi_i(\theta_{-i}, m_{-i})u_i\big(g(m_i', m_{-i}), (\theta_i, \theta_{-i})\big) \neq \emptyset$ for all $\theta_i \in \Theta_i$.
   (b) $\psi_i(\theta_{-i}, m_{-i}) > 0 \Rightarrow m_{-i} \in \mathcal{S}^\infty_{-i}(\theta_{-i})$.
   (c) $\text{marg}_{\Theta_{-i}}\psi_i = z_i^1$.

The SCF f is *rationalizably implementable* if there exists a mechanism that rationalizably implements f.

Bergemann and Morris (2011, Theorem 3) prove that if a mechanism robustly implements an SCF in interim equilibria, then the same mechanism also rationalizably implements the SCF. They also prove that the converse is true whenever the mechanism that rationalizably implements the SCF is such that the message correspondence $\mathcal{S}^\infty$ satisfies the *ex post best response property.* The property requires that for all $i \in I$ and $\theta_i \in \Theta_i$, there exist a message $m_i^* \in \mathcal{S}_i^\infty(\theta_i)$ such that

$$m_i^* \in \arg\max_{m_i \in M_i} u_i\big(g(m_i, m_{-i}), (\theta_i, \theta_{-i})\big),$$

for all $\theta_{-i} \in \Theta_{-i}$ and $m_{-i} \in \mathcal{S}_{-i}^\infty(\theta_{-i})$.

2.5.4 Weak Rationalizable Implementation

The nonemptiness requirement in rationalizable implementation (the second condition in Definition 2.5) is strong: It requires that for any individual i and for any belief z_i^1 that i might have over the payoff types of others, there exists a belief ψ_i over $\mathcal{S}_{-i}^\infty(\theta_{-i})$ such that individual i has a best response to this belief *regardless* of her payoff type. In the appendix of their working paper, Bergemann and Morris (2010) define weak rationalizable implementation which weakens the nonemptiness requirement by allowing the belief ψ_i to depend on the payoff type of individual i.

Definition 2.6. A mechanism $\Gamma = ((M_i)_{i\in I}, g)$ *weakly rationalizably implements the SCF f* if

1. (uniqueness) $m \in \mathcal{S}^\infty(\theta) \Rightarrow g(m) = f(\theta)$; and
2. (nonemptyness) For each $i \in I$, $\theta_i \in \Theta_i$ and $z_i^1 \in Z_i^1$, there exists a belief $\psi_i \in \Delta(\Theta_{-i} \times M_{-i})$ such that:
   (a) $\arg\max_{m_i' \in M_i} \sum_{\theta_{-i}, m_{-i}} \psi_i(\theta_{-i}, m_{-i}) u_i\big(g(m_i', m_{-i}), (\theta_i, \theta_{-i})\big) \neq \emptyset$.
   (b) $\psi_i(\theta_{-i}, m_{-i}) > 0 \Rightarrow m_{-i} \in \mathcal{S}_{-i}^\infty(\theta_{-i})$.
   (c) $\text{marg}_{\Theta_{-i}} \psi_i = z_i^1$.

The SCF f is *weakly rationalizably implementable* if there exists a mechanism that weakly rationalizably implements f.

3 Equivalence between Robust Implementation in Rationalizable Strategies and Weak Rationalizable Implementation

We now show that robust implementation in rationalizable strategies is equivalent to weak rationalizable implementation because the former imposes the same conditions on $\mathcal{S}^\infty$ as the latter.

Theorem 3.1. *The SCF f is robustly implementable in rationalizable strategies by the mechanism Γ if and only if f is weakly rationalizably implementable by Γ.*

Proof. We first prove the following lemma:

Lemma 3.2. *Consider any mechanism Γ. The message profile $m \in \mathcal{S}^\infty(\theta)$ if and only if there exists a type space $\mathcal{T}$ such that $m \in \bigcup_{t \in T : \hat{\theta}(t)=\theta} B^\infty(t)$.*

Proof. ($\Rightarrow$) Bergemann and Morris (2011, Proposition 1) show that if $m \in \mathcal{S}^\infty(\theta)$, then there exist a type space $\mathcal{T}$, a pure-strategy interim equilibrium σ, and a type profile t such that $\sigma(t) = m$ and $\hat{\theta}(t) = \theta$. Therefore, $m \in B^\infty(t)$.

($\Leftarrow$) Consider any type space $\mathcal{T}$. Define the message correspondence profile with payoff-type domain $\hat{\mathcal{S}} = (\hat{\mathcal{S}}_1, \ldots, \hat{\mathcal{S}}_n)$ such that for all $i \in I$,

$$\hat{\mathcal{S}}_i(\theta_i') = \bigcup_{t_i : \hat{\theta}_i(t_i)=\theta_i'} B_i^\infty(t_i), \forall \theta_i' \in \Theta_i.$$

If $m_i' \in \hat{\mathcal{S}}_i(\theta_i')$, then there exists $t_i' \in T_i$ such that $\hat{\theta}_i(t_i') = \theta_i'$ and $m_i' \in B_i^\infty(t_i')$. Thus, there exists a belief $\lambda_i \in \Delta(T_{-i} \times M_{-i})$ such that

$$m_i' \in \arg\max_{m_i'' \in M_i} \sum_{t_{-i}, m_{-i}} \lambda_i(t_{-i}, m_{-i}) u_i(g(m_i'', m_{-i}), \hat{\theta}(t_i', t_{-i})),$$

$\text{marg}_{T_{-i}} \lambda_i = \hat{\pi}_i(t_i')$ and $\lambda_i(t_{-i}, m_{-i}) > 0 \Rightarrow m_{-i} \in B_{-i}^\infty(t_{-i})$.

Define $\psi_i \in \Delta(\Theta_{-i} \times M_{-i})$ as follows:

$$\psi_i(\theta_{-i}, m_{-i}) = \sum_{t_{-i} : \hat{\theta}_{-i}(t_{-i})=\theta_{-i}} \lambda_i(t_{-i}, m_{-i}), \forall \theta_{-i}, m_{-i}.$$

Then $\psi_i(\theta_{-i}, m_{-i}) > 0$ implies that $m_{-i} \in \bigcup_{t_{-i} : \hat{\theta}_{-i}(t_{-i})=\theta_{-i}} B_{-i}^\infty(t_{-i}) = \hat{\mathcal{S}}_{-i}(\theta_{-i})$. Moreover, by construction,

$$m_i^{'} \in \arg\max_{m_i^{''}\in M_i} \sum_{\theta_{-i},m_{-i}} \psi_i(\theta_{-i}, m_{-i}) u_i(g(m_i^{''}, m_{-i}), (\theta_i^{'}, \theta_{-i})).$$

Thus, $m_i^{'} \in b_i^{\Theta}(\hat{\mathcal{S}})[\theta_i^{'}]$. Hence, $b^{\Theta}(\hat{\mathcal{S}}) \geq \hat{\mathcal{S}}$. Therefore, $\hat{\mathcal{S}} \leq \mathcal{S}^{\infty}$.

Now suppose there exist $m \in M$ and $\theta \in \Theta$ such that $m \in \bigcup_{t:\hat{\theta}(t)=\theta} B^{\infty}(t)$. Then $m \in \hat{\mathcal{S}}(\theta)$, and hence $m \in \mathcal{S}^{\infty}(\theta)$. This completes the proof of the lemma. □

We prove the necessity part of Theorem 3.1 first.

Suppose the SCF f is robustly implementable in rationalizable strategies by the mechanism Γ. Then the following is true for all type spaces $\mathcal{T}$: For all $t \in T$, we have

$$B^{\infty}(t) \neq \emptyset \quad \text{ and } \quad g(m) = f(\hat{\theta}(t)), \forall m \in B^{\infty}(t).$$

Pick any $\theta \in \Theta$. If $m \in \mathcal{S}^{\infty}(\theta)$, then it follows from Lemma 3.2 that there exists a type space $\mathcal{T}^{'}$ such that $m \in \bigcup_{t\in T^{'}:\hat{\theta}(t)=\theta} B^{\infty}(t)$. Hence, $g(m) = f(\theta)$.

Next, pick any i, θ_i and z_i^1. For each $j \neq i$, pick any $z_j^1 \in Z_j^1$. Define the type space $\mathcal{T}$ such that (i) $T_j = \{t_j^{\tilde{\theta}_j} : \tilde{\theta}_j \in \Theta_j\}$ for all $j \in I$, and (ii) $\hat{\theta}_j(t_j^{\tilde{\theta}_j}) = \tilde{\theta}_j$ and $\hat{\pi}_j(t_j^{\tilde{\theta}_j})[t_{-j}^{\tilde{\theta}_{-j}}] = z_j^1(\tilde{\theta}_{-j})$ for all $t_j^{\tilde{\theta}_j} \in T_j$ and $t_{-j}^{\tilde{\theta}_{-j}} \in T_{-j}$.

By our hypothesis of robust implementation in rationalizable strategies, $B_i^{\infty}(t_i^{\theta_i}) \neq \emptyset$. Therefore, there exists $\lambda_i \in \Delta(T_{-i} \times M_{-i})$ such that

1. $\arg\max_{m_i^{'}} \sum_{t_{-i}^{\theta_{-i}},m_{-i}} \lambda_i(t_{-i}^{\theta_{-i}}, m_{-i}) u_i\big(g(m_i^{'}, m_{-i}), \hat{\theta}(t_i^{\theta_i}, t_{-i}^{\theta_{-i}})\big) \neq \emptyset.$

2. $\text{marg}_{T_{-i}} \lambda_i = \hat{\pi}_i(t_i^{\theta_i})$

3. $\lambda_i(t_{-i}^{\theta_{-i}}, m_{-i}) > 0 \Rightarrow m_{-i} \in B_{-i}^{\infty}(t_{-i}^{\theta_{-i}})$.

Define $\psi_i \in \Delta(\Theta_{-i} \times M_{-i})$ as follows: for any $\theta_{-i} \in \Theta_{-i}$ and $m_{-i} \in M_{-i}$,

$$\psi_i(\theta_{-i}, m_{-i}) = \lambda_i(t_{-i}^{\theta_{-i}}, m_{-i}).$$

Then $\psi_i(\theta_{-i}, m_{-i}) > 0$ implies that $m_{-i} \in B_{-i}^{\infty}(t_{-i}^{\theta_{-i}})$. It follows from Lemma 3.2 that $m_{-i} \in \mathcal{S}_{-i}^{\infty}(\theta_{-i})$. Lastly, by construction, $\text{marg}_{\Theta_{-i}} \psi_i = z_i^1$ and

$$\arg\max_{m_i^{'}\in M_i} \sum_{\theta_{-i},m_{-i}} \psi_i(\theta_{-i}, m_{-i}) u_i(g(m_i^{'}, m_{-i}), (\theta_i, \theta_{-i})).$$

We prove the sufficiency part of Theorem 3.1 next.

Suppose that the SCF f is weakly rationalizably implementable by the mechanism Γ. Consider any type space $\mathcal{T}$. If $m \in B^{\infty}(t)$, then it follows from Lemma 3.2 that $m \in \mathcal{S}^{\infty}(\hat{\theta}(t))$. Hence, $g(m) = f(\hat{\theta}(t))$.

We now show that $B^{\infty}(t) \neq \emptyset$ for all $t \in T$. Define the message correspondence profile $\hat{S} = (\hat{S}_1, \ldots, \hat{S}_n)$ such that, for all $i \in I$ and $t_i \in T_i$,

$$\hat{S}_i(t_i) = \mathcal{S}_i^{\infty}(\hat{\theta}_i(t_i)).$$

Pick any type $t_i \in T_i$. By our hypothesis of weak rationalizable implementability, there exists a belief $\psi_i \in \Delta(\Theta_{-i} \times M_{-i})$ such that

(a) $\arg\max_{m_i' \in M_i} \sum_{\theta_{-i}, m_{-i}} \psi_i(\theta_{-i}, m_{-i}) u_i\big(g(m_i', m_{-i}), (\hat{\theta}_i(t_i), \theta_{-i})\big) \neq \emptyset.$

(b) $\psi_i(\theta_{-i}, m_{-i}) > 0 \Rightarrow m_{-i} \in \mathcal{S}_{-i}^{\infty}(\theta_{-i}).$

(c) $\text{marg}_{\Theta_{-i}} \psi_i = z_i^1(t_i).$

By the definition of $\mathcal{S}_i^{\infty}(\hat{\theta}_i(t_i))$, we have

$$\emptyset \neq \arg\max_{m_i' \in M_i} \sum_{\theta_{-i}, m_{-i}} \psi_i(\theta_{-i}, m_{-i}) u_i\big(g(m_i', m_{-i}), (\hat{\theta}_i(t_i), \theta_{-i})\big) \subseteq \mathcal{S}_i^{\infty}(\hat{\theta}_i(t_i)).$$

Since $\hat{S}_i(t_i) = \mathcal{S}_i^{\infty}(\hat{\theta}_i(t_i))$, we also have $\hat{S}_i(t_i) \neq \emptyset$.

We now show that $\hat{S}_i(t_i) \leq b_i(\hat{S})[t_i]$. Consider any message $\tilde{m}_i \in \hat{S}_i(t_i)$. By our hypothesis of weak rationalizable implementability, we have that for any $\theta \in \Theta$, $m' \in \mathcal{S}^{\infty}(\theta) \Rightarrow g(m') = f(\theta)$. Since $\tilde{m}_i \in \mathcal{S}_i^{\infty}(\hat{\theta}_i(t_i))$ and $\psi_i(\theta_{-i}, m_{-i}) > 0$ implies $m_{-i} \in \mathcal{S}_{-i}^{\infty}(\theta_{-i})$, by weakly rationalizable implementability, we have

$$\sum_{\theta_{-i}, m_{-i}} \psi_i(\theta_{-i}, m_{-i}) u_i\big(g(\tilde{m}_i, m_{-i}), (\hat{\theta}_i(t_i), \theta_{-i})\big) = \sum_{\theta_{-i}, m_{-i}} \psi_i(\theta_{-i}, m_{-i}) u_i\big(f(\hat{\theta}_i(t_i), \theta_{-i}), (\hat{\theta}_i(t_i), \theta_{-i})\big).$$

Thus, either every message in $\hat{S}_i(t_i)$ is a best response to ψ_i or none of the messages in $\hat{S}_i(t_i)$ is a best response to ψ_i. But, as already argued,

$$\hat{S}_i(t_i) = \mathcal{S}_i^{\infty}(\hat{\theta}_i(t_i)) \supseteq \arg\max_{m_i' \in M_i} \sum_{\theta_{-i}, m_{-i}} \psi_i(\theta_{-i}, m_{-i}) u_i\big(g(m_i', m_{-i}), (\hat{\theta}_i(t_i), \theta_{-i})\big) \neq \emptyset.$$

Thus, every message in $\hat{S}_i(t_i)$ is a best response to ψ_i.

Now pick any $m_i \in \hat{S}_i(t_i)$. As argued above,

$$m_i \in \arg\max_{m_i' \in M_i} \sum_{\theta_{-i}, m_{-i}} \psi_i(\theta_{-i}, m_{-i}) u_i\big(g(m_i', m_{-i}), (\hat{\theta}_i(t_i), \theta_{-i})\big).$$

Define the belief $\lambda_i \in \Delta(T_{-i} \times M_{-i})$ such that for all $(t_{-i}, m_{-i}) \in T_{-i} \times M_{-i}$,

$$\lambda_i(t_{-i}, m_{-i}) = \begin{cases} \hat{\pi}_i(t_i)[t_{-i}] \left(\frac{\psi_i(\hat{\theta}_{-i}(t_{-i}), m_{-i})}{z_i^1(t_i)[\hat{\theta}_{-i}(t_{-i})]}\right), & \text{if } \hat{\pi}_i(t_i)[t_{-i}] > 0 \\ 0, & \text{otherwise.} \end{cases}$$

Since $\sum_{m_{-i}} \psi_i(\hat{\theta}_{-i}(t_{-i}), m_{-i}) = z_i^1(t_i)[\hat{\theta}_{-i}(t_{-i})]$, we have $\text{marg}_{T_{-i}} \lambda_i = \hat{\pi}_i(t_i)$. Moreover,

$$\lambda_i(t_{-i}, m_{-i}) > 0 \Rightarrow \psi_i(\hat{\theta}_{-i}(t_{-i}), m_{-i}) > 0 \Rightarrow m_{-i} \in \mathcal{S}_{-i}^{\infty}(\hat{\theta}_{-i}(t_{-i})) = \hat{S}_{-i}(t_{-i}).$$

Finally, for all $m_i' \in M_i$,

$$\begin{aligned} &\sum_{t_{-i}, m_{-i}} \lambda_i(t_{-i}, m_{-i}) u_i\big(g(m_i', m_{-i}), \hat{\theta}(t_i, t_{-i})\big) \\ &= \sum_{\theta_{-i}, m_{-i}} \left(\sum_{t_{-i}: \hat{\theta}_{-i}(t_{-i}) = \theta_{-i}} \hat{\pi}_i(t_i)[t_{-i}] \frac{\psi_i(\theta_{-i}, m_{-i})}{z_i^1(t_i)(\theta_{-i})} u_i\big(g(m_i', m_{-i}), (\hat{\theta}_i(t_i), \theta_{-i})\big) \right) \\ &= \sum_{\theta_{-i}, m_{-i}} \psi_i(\theta_{-i}, m_{-i}) u_i\big(g(m_i', m_{-i}), (\hat{\theta}_i(t_i), \theta_{-i})\big), \end{aligned}$$

where the last equality follows because $\sum_{t_{-i}: \hat{\theta}_{-i}(t_{-i}) = \theta_{-i}} \hat{\pi}_i(t_i)[t_{-i}] = z_i^1(t_i)(\theta_{-i})$. Hence, we must have

$$m_i \in \arg\max_{m_i' \in M_i} \sum_{t_{-i}, m_{-i}} \lambda_i(t_{-i}, m_{-i}) u_i\big(g(m_i', m_{-i}), \hat{\theta}(t_i, t_{-i})\big).$$

We thus conclude that $m_i \in b_i(\hat{S})[t_i]$.

As $b(\hat{S}) \geq \hat{S}$, we have $\hat{S} \leq B^{\infty}$. Pick any $t \in T$. Then $B^{\infty}(t) \neq \emptyset$ because, as already shown, $\hat{S}(t) \neq \emptyset$. This completes the proof of the theorem. □

4 Necessary Condition

We now apply the equivalence result presented in the previous section to present the key necessary condition for robust implementation in rationalizable strategies.

A *deception* is a profile of correspondences $\beta = (\beta_1, \ldots, \beta_n)$ such that $\beta_i : \Theta_i \to 2^{\Theta_i} \setminus \emptyset$ and $\theta_i \in \beta_i(\theta_i)$ for all $\theta_i \in \Theta_i$ and $i \in I$. A deception β is *unacceptable* if there exist $\theta \in \Theta$

and $\theta' \in \beta(\theta)$ for which $f(\theta) \neq f(\theta')$; otherwise, β is *acceptable.*

For each $i \in I$ and $\theta_i \in \Theta_i$, define

$$Y_i[\theta_i] \equiv \left\{ y : \Theta_{-i} \to \Delta(A) : \begin{array}{ll} \multicolumn{2}{c}{\forall \theta_{-i} \in \Theta_{-i},} \\ \text{either} & y(\theta_{-i}) = f(\theta_i, \theta_{-i}) \\ \text{or} & u_i\big(f(\theta_i, \theta_{-i}), (\theta_i, \theta_{-i})\big) > u_i\big(y(\theta_{-i}), (\theta_i, \theta_{-i})\big) \end{array} \right\}.$$

Thus, $Y_i[\theta_i]$ is the collection of all mappings $y : \Theta_{-i} \to \Delta(A)$ such that for every $\theta_{-i} \in \Theta_{-i}$, the lottery $y(\theta_{-i})$ is either equal to $f(\theta_i, \theta_{-i})$ or strictly worse than $f(\theta_i, \theta_{-i})$ for individual i in state (θ_i, θ_{-i}).

Definition 4.1. We say that an unacceptable deception β is *weakly refutable* if there exist $i \in I$, $\theta_i \in \Theta_i$, and $\theta_i' \in \beta_i(\theta_i)$ satisfying $\theta_i' \not\approx_i^f \theta_i$ such that for all $\tilde{\theta}_i \in \Theta_i$ and $\psi_i \in \Delta(\Theta_{-i} \times \Theta_{-i})$ satisfying $\psi_i(\theta_{-i}, \theta_{-i}') > 0 \Rightarrow \theta_{-i}' \in \beta_{-i}(\theta_{-i})$, there exists $y \in Y_i[\tilde{\theta}_i]$ such that

$$\sum_{\theta_{-i}, \theta_{-i}'} \psi_i(\theta_{-i}, \theta_{-i}') u_i\big(y(\theta_{-i}'), (\theta_i, \theta_{-i})\big) > \sum_{\theta_{-i}, \theta_{-i}'} \psi_i(\theta_{-i}, \theta_{-i}') u_i\big(f(\theta_i', \theta_{-i}'), (\theta_i, \theta_{-i})\big).$$

Definition 4.2. The SCF f satisfies *weak robust monotonicity (weak RM)* if every unacceptable deception β is weakly refutable.

Here is the main result of this section:

Theorem 4.3. *If the SCF f is robustly implementable in rationalizable strategies, then f satisfies weak RM.*

Proof. Suppose the mechanism $\Gamma = ((M_i)_{i \in I}, g)$ robustly implements f in rationalizable strategies. It follows from Theorem 3.1 that Γ weakly rationalizably implements f. We now argue that f must satisfy weak RM.

Pick any $i \in I$ and $\theta \in \Theta$. Consider the belief $z_i^1 \in \Delta(\Theta_{-i})$ that puts probability one on θ_{-i}. By weak rationalizable implementability, there exists a belief $\psi_i^\theta \in \Delta(\Theta_{-i} \times M_{-i})$ such that

(a) $\arg\max_{\tilde{m}_i \in M_i} \sum_{\tilde{\theta}_{-i}, \tilde{m}_{-i}} \psi_i^\theta(\tilde{\theta}_{-i}, \tilde{m}_{-i}) u_i\big(g(\tilde{m}_i, \tilde{m}_{-i}), (\theta_i, \tilde{\theta}_{-i})\big) \neq \emptyset.$

(b) $\psi_i^\theta(\tilde{\theta}_{-i}, \tilde{m}_{-i}) > 0 \Rightarrow \tilde{m}_{-i} \in \mathcal{S}_{-i}^\infty(\tilde{\theta}_{-i}).$

(c) $\text{marg}_{\Theta_{-i}} \psi_i^\theta = z_i^1.$

If $\tilde{\theta}_{-i} \neq \theta_{-i}$, then $\psi_i^\theta(\tilde{\theta}_{-i}, \tilde{m}_{-i}) = 0$ because $\mathrm{marg}_{\Theta_{-i}} \psi_i^\theta = z_i^1$ and z_i^1 assigns probability one on θ_{-i}. Therefore, for all $\tilde{m}_i \in M_i$,

$$\begin{aligned}
&\sum_{\tilde{\theta}_{-i}, \tilde{m}_{-i}} \psi_i^\theta(\tilde{\theta}_{-i}, \tilde{m}_{-i}) u_i\big(g(\tilde{m}_i, \tilde{m}_{-i}), (\theta_i, \tilde{\theta}_{-i})\big) \\
&= \sum_{\tilde{m}_{-i} \in \mathcal{S}_{-i}^\infty(\theta_{-i})} \mathrm{marg}_{M_{-i}} \psi_i^\theta(\tilde{m}_{-i}) u_i\big(g(\tilde{m}_i, \tilde{m}_{-i}), \theta\big) \\
&= u_i\left(\sum_{\tilde{m}_{-i} \in \mathcal{S}_{-i}^\infty(\theta_{-i})} \mathrm{marg}_{M_{-i}} \psi_i^\theta(\tilde{m}_{-i}) g(\tilde{m}_i, \tilde{m}_{-i}), \theta\right). \qquad (1)
\end{aligned}$$

Define the set of lotteries

$$L_i(\theta) = \left\{ \sum_{\tilde{m}_{-i} \in \mathcal{S}_{-i}^\infty(\theta_{-i})} \mathrm{marg}_{M_{-i}} \psi_i^\theta(\tilde{m}_{-i}) g(\tilde{m}_i, \tilde{m}_{-i}) : \tilde{m}_i \in M_i \right\}.$$

Pick any $m_i \in \arg\max_{\tilde{m}_i \in M_i} \sum_{\tilde{\theta}_{-i}, \tilde{m}_{-i}} \psi_i^\theta(\tilde{\theta}_{-i}, \tilde{m}_{-i}) u_i\big(g(\tilde{m}_i, \tilde{m}_{-i}), (\theta_i, \tilde{\theta}_{-i})\big)$. Then $m_i \in \mathcal{S}_i^\infty(\theta_i)$ because $\psi_i^\theta(\tilde{\theta}_{-i}, \tilde{m}_{-i}) > 0$ implies $\tilde{m}_{-i} \in \mathcal{S}_{-i}^\infty(\tilde{\theta}_{-i})$. Therefore, by weak rationalizable implementability,

$$\sum_{\tilde{m}_{-i} \in \mathcal{S}_{-i}^\infty(\theta_{-i})} \mathrm{marg}_{M_{-i}} \psi_i^\theta(\tilde{m}_{-i}) g(m_i, \tilde{m}_{-i}) = f(\theta).$$

Moreover, for all $\tilde{m}_i \in M_i$, we have

$$\begin{aligned}
u_i\left(\sum_{\tilde{m}_{-i} \in \mathcal{S}_{-i}^\infty(\theta_{-i})} \mathrm{marg}_{M_{-i}} \psi_i^\theta(\tilde{m}_{-i}) g(m_i, \tilde{m}_{-i}), \theta\right) &= \sum_{\tilde{\theta}_{-i}, \tilde{m}_{-i}} \psi_i^\theta(\tilde{\theta}_{-i}, \tilde{m}_{-i}) u_i\big(g(m_i, \tilde{m}_{-i}), (\theta_i, \tilde{\theta}_{-i})\big) \\
&\geq \sum_{\tilde{\theta}_{-i}, \tilde{m}_{-i}} \psi_i^\theta(\tilde{\theta}_{-i}, \tilde{m}_{-i}) u_i\big(g(\tilde{m}_i, \tilde{m}_{-i}), (\theta_i, \tilde{\theta}_{-i})\big) \\
&= u_i\left(\sum_{\tilde{m}_{-i} \in \mathcal{S}_{-i}^\infty(\theta_{-i})} \mathrm{marg}_{M_{-i}} \psi_i^\theta(\tilde{m}_{-i}) g(\tilde{m}_i, \tilde{m}_{-i}), \theta\right),
\end{aligned}$$

where the first and last equality follows from (1). Hence, $u_i\big(f(\theta), \theta\big) \geq u_i\big(\ell, \theta\big)$ for all $\ell \in L_i(\theta)$.

We next claim that for any $\ell \in L_i(\theta)$, $\ell \neq f(\theta)$ implies $u_i\big(f(\theta), \theta\big) > u_i\big(\ell, \theta\big)$. Suppose not. Then there is some $\ell \in L_i(\theta)$ such that $\ell \neq f(\theta)$ but $u_i\big(\ell, \theta\big) \geq u_i\big(f(\theta), \theta\big)$. By construction of $L_i(\theta)$, there exists a message $\tilde{m}_i$ such that $\sum_{\tilde{m}_{-i} \in \mathcal{S}_{-i}^\infty(\theta_{-i})} \mathrm{marg}_{M_{-i}} \psi_i^\theta(\tilde{m}_{-i}) g(\tilde{m}_i, \tilde{m}_{-i}) = \ell$.

Then, as per the above arguments, $u_i(\ell,\theta) \geq u_i(f(\theta),\theta)$ is equivalent to

$$\sum_{\tilde{\theta}_{-i},\tilde{m}_{-i}} \psi_i^\theta(\tilde{\theta}_{-i},\tilde{m}_{-i}) u_i\big(g(\tilde{m}_i,\tilde{m}_{-i}),(\theta_i,\tilde{\theta}_{-i})\big) \geq \sum_{\tilde{\theta}_{-i},\tilde{m}_{-i}} \psi_i^\theta(\tilde{\theta}_{-i},\tilde{m}_{-i}) u_i\big(g(m_i,\tilde{m}_{-i}),(\theta_i,\tilde{\theta}_{-i})\big),$$

for some $m_i \in \arg\max_{\tilde{m}_i' \in M_i} \sum_{\tilde{\theta}_{-i},\tilde{m}_{-i}} \psi_i^\theta(\tilde{\theta}_{-i},\tilde{m}_{-i}) u_i\big(g(\tilde{m}_i',\tilde{m}_{-i}),(\theta_i,\tilde{\theta}_{-i})\big)$. Therefore, $\tilde{m}_i$ is also a best response to the belief ψ_i^θ when i's payoff type is θ_i. Hence, $\tilde{m}_i \in \mathcal{S}_i^\infty(\theta_i)$. But $g(\tilde{m}_i,\tilde{m}_{-i}) \neq f(\theta)$ for at least one $\tilde{m}_{-i} \in \mathcal{S}_{-i}^\infty(\theta_{-i})$, which contradicts weak rationalizable implementation of f.

We are now ready to prove the theorem. Consider any deception β. Define the message correspondence profile with payoff-type domain $\mathcal{S} = (\mathcal{S}_1,\ldots,\mathcal{S}_n)$ such that

$$\mathcal{S}_i(\theta_i) = \bigcup_{\theta_i' \in \beta_i(\theta_i)} \mathcal{S}_i^\infty(\theta_i').$$

Suppose β is not weakly refutable. Then, by definition of weak refutablility, for all $i \in I$, $\theta_i \in \Theta_i$, and $\theta_i' \in \beta_i(\theta_i)$ satisfying $\theta_i' \not\sim_i^f \theta_i$, there exist $\tilde{\theta}_i$ and $\psi_i \in \Delta(\Theta_{-i} \times \Theta_{-i})$, which satisfies $\psi_i(\theta_{-i},\theta_{-i}') > 0 \Rightarrow \theta_{-i}' \in \beta_{-i}(\theta_{-i})$, such that for all $y \in Y_i[\tilde{\theta}_i]$, we have

$$\sum_{\theta_{-i},\theta_{-i}'} \psi_i(\theta_{-i},\theta_{-i}') u_i\big(f(\theta_i',\theta_{-i}'),(\theta_i,\theta_{-i})\big) \geq \sum_{\theta_{-i},\theta_{-i}'} \psi_i(\theta_{-i},\theta_{-i}') u_i\big(y(\theta_{-i}'),(\theta_i,\theta_{-i})\big). \quad (2)$$

We first show that for any $i \in I$, $\theta_i \in \Theta_i$, and $\theta_i' \in \beta_i(\theta_i)$ satisfying $\theta_i' \sim_i^f \theta_i$, there exist $\tilde{\theta}_i \in \Theta_i$ and $\psi_i \in \Delta(\Theta_{-i} \times \Theta_{-i})$ satisfying $\psi_i(\theta_{-i},\theta_{-i}) > 0 \Rightarrow \theta_{-i}' \in \beta_{-i}(\theta_{-i})$ such that (2) holds for all $y \in Y_i[\tilde{\theta}]$.

Pick any i, θ_i, and $\theta_i' \in \beta_i(\theta_i)$ satisfying $\theta_i' \sim_i^f \theta_i$. We set $\tilde{\theta}_i = \theta_i$ and the belief $\psi_i \in \Delta(\Theta_{-i} \times \Theta_{-i})$ such that $\psi_i(\hat{\theta}_{-i},\hat{\theta}_{-i}) = 1$ for some $\hat{\theta}_{-i} \in \Theta_{-i}$. As $\hat{\theta}_{-i} \in \beta_{-i}(\hat{\theta}_{-i})$, the belief ψ_i satisfies $\psi_i(\theta_{-i},\theta_{-i}') > 0 \Rightarrow \theta_{-i}' \in \beta_{-i}(\theta_{-i})$. Since $\theta_i \sim_i^f \theta_i'$, we have $f(\theta_i',\hat{\theta}_{-i}) = f(\theta_i,\hat{\theta}_{-i})$. Moreover, $Y_i[\tilde{\theta}_i] = Y_i[\theta_i]$ because $\tilde{\theta}_i = \theta_i$. Therefore, for all $y \in Y_i[\tilde{\theta}_i]$, we have

$$\begin{aligned}\sum_{\theta_{-i},\theta_{-i}'} \psi_i(\theta_{-i},\theta_{-i}') u_i\big(f(\theta_i',\theta_{-i}'),(\theta_i,\theta_{-i})\big) &= u_i\big(f(\theta_i,\hat{\theta}_{-i}),(\theta_i,\hat{\theta}_{-i})\big) \\ &\geq u_i\big(y(\hat{\theta}_{-i}),(\theta_i,\hat{\theta}_{-i})\big) \\ &= \sum_{\theta_{-i},\theta_{-i}'} \psi_i(\theta_{-i},\theta_{-i}') u_i\big(y(\theta_{-i}'),(\theta_i,\theta_{-i})\big).\end{aligned}$$

Thus, if we combine the above result with the hypothesis that β is not weakly refutable, then we can hypothesize that for all $i \in I$, $\theta_i \in \Theta_i$, and $\theta_i' \in \beta_i(\theta_i)$, there exist $\tilde{\theta}_i \in \Theta_i$ and

$\psi_i \in \Delta(\Theta_{-i} \times \Theta_{-i})$ satisfying $\psi_i(\theta_{-i}, \theta'_{-i}) > 0 \Rightarrow \theta'_{-i} \in \beta_{-i}(\theta_{-i})$ such that (2) holds for all $y \in Y_i[\tilde{\theta}_i]$.

We next show that $b^{\Theta}(\mathcal{S}) \geq \mathcal{S}$. Pick any $i \in I$, $\theta_i \in \Theta_i$, and $m'_i \in \mathcal{S}_i(\theta_i)$. We now construct a belief $\psi^{\Gamma} \in \Delta(\Theta_{-i} \times M_{-i})$ satisfying $\psi^{\Gamma}_i(\theta_{-i}, m_{-i}) > 0$ implies $m_{-i} \in \mathcal{S}_{-i}(\theta_{-i})$ such that m'_i is a best response for agent i of payoff type θ_i against ψ^{Γ}_i.

By definition of $\mathcal{S}$, we have $m'_i \in \mathcal{S}^{\infty}_i(\theta'_i)$ for some $\theta'_i \in \beta_i(\theta_i)$. Then, by our hypothesis, there exist $\tilde{\theta}_i \in \Theta_i$ and $\psi_i \in \Delta(\Theta_{-i} \times \Theta_{-i})$ satisfying $\psi_i(\theta_{-i}, \theta'_{-i}) > 0 \Rightarrow \theta'_{-i} \in \beta_{-i}(\theta_{-i})$ such that (2) holds for all $y \in Y_i[\tilde{\theta}_i]$. Define the belief $\psi^{\Gamma}_i \in \Delta(\Theta_{-i} \times M_{-i})$ as follows: for any (θ_{-i}, m_{-i}),

$$\psi^{\Gamma}_i(\theta_{-i}, m_{-i}) = \sum_{\theta'_{-i}} \psi_i(\theta_{-i}, \theta'_{-i}) \times \text{marg}_{M_{-i}} \psi_i^{(\tilde{\theta}_i, \theta'_{-i})}(m_{-i}).$$

By construction, $\psi^{\Gamma}_i(\theta_{-i}, m_{-i}) > 0$ implies that there exists $\theta'_{-i} \in \Theta_{-i}$ such that $\psi_i(\theta_{-i}, \theta'_{-i}) > 0$ and $\text{marg}_{M_{-i}} \psi_i^{(\tilde{\theta}_i, \theta'_{-i})}(m_{-i}) > 0$. But $\psi_i(\theta_{-i}, \theta'_{-i}) > 0$ implies $\theta'_{-i} \in \beta_{-i}(\theta_{-i})$. Moreover, $\text{marg}_{M_{-i}} \psi_i^{(\tilde{\theta}_i, \theta'_{-i})}(m_{-i}) > 0$ implies $m_{-i} \in \mathcal{S}^{\infty}_{-i}(\theta'_{-i})$ – recall the definition of $\psi_i^{(\tilde{\theta}_i, \theta'_{-i})}$ from the beginning of this proof. Since $\theta'_{-i} \in \beta_{-i}(\theta_{-i})$ and $m_{-i} \in \mathcal{S}^{\infty}_{-i}(\theta'_{-i})$, it follows from the definition of $\mathcal{S}$ that $m_{-i} \in \mathcal{S}_{-i}(\theta_{-i})$.

For any $m_i \in M_i$, define $y^{m_i} : \Theta_{-i} \to \Delta(A)$ as follows: for all $\theta_{-i} \in \Theta_{-i}$,

$$y^{m_i}(\theta_{-i}) = \sum_{m_{-i}} \text{marg}_{M_{-i}} \psi_i^{(\tilde{\theta}_i, \theta_{-i})}(m_{-i}) g(m_i, m_{-i}).$$

By construction, $y^{m_i}(\theta_{-i}) \in L_i(\tilde{\theta}_i, \theta_{-i})$. Therefore, if $f(\tilde{\theta}_i, \theta_{-i}) \neq y^{m_i}(\theta_{-i})$, then, as argued earlier in the proof, we must have

$$u_i\big(f(\tilde{\theta}_i, \theta_{-i}), (\tilde{\theta}_i, \theta_{-i})\big) > u_i\big(y^{m_i}(\theta_{-i}), (\tilde{\theta}_i, \theta_{-i})\big).$$

So $y^{m_i} \in Y_i[\tilde{\theta}_i]$. By our hypothesis, (1) holds for all $y \in Y_i[\tilde{\theta}_i]$. Hence, for any $m_i \in M_i$,

$$\sum_{\theta_{-i}, \theta'_{-i}} \psi_i(\theta_{-i}, \theta'_{-i}) u_i\big(f(\theta'_i, \theta'_{-i}), (\theta_i, \theta_{-i})\big) \geq \sum_{\theta_{-i}, \theta'_{-i}} \psi_i(\theta_{-i}, \theta'_{-i}) u_i\big(y^{m_i}(\theta'_{-i}), (\theta_i, \theta_{-i})\big). \quad (3)$$

We are ready to show that $m_i^{'}$ is a best response for agent i of payoff type θ_i against ψ_i^{Γ}.

$$
\begin{aligned}
& \sum_{\theta_{-i}, m_{-i}} \psi_i^{\Gamma}(\theta_{-i}, m_{-i}) u_i\big(g(m_i^{'}, m_{-i}), (\theta_i, \theta_{-i})\big) \\
= & \sum_{\theta_{-i}, m_{-i}} \left(\sum_{\theta_{-i}^{'}} \psi_i(\theta_{-i}, \theta_{-i}^{'}) \times \operatorname{marg}_{M_{-i}} \psi_i^{(\tilde{\theta}_i, \theta_{-i}^{'})}(m_{-i}) u_i\big(g(m_i^{'}, m_{-i}), (\theta_i, \theta_{-i})\big) \right) \\
& \quad \text{(by definition of } \psi_i^{\Gamma}) \\
= & \sum_{\theta_{-i}, \theta_{-i}^{'}} \psi_i(\theta_{-i}, \theta_{-i}^{'}) u_i\big(f(\theta_i^{'}, \theta_{-i}^{'}), (\theta_i, \theta_{-i})\big) \\
& \quad \left(\begin{array}{l} \text{by weak rationalizable implementability of } f \text{ because } m_i^{'} \in \mathcal{S}_i^{\infty}(\theta_i^{'}) \\ \text{and } \operatorname{marg}_{M_{-i}} \psi_i^{(\tilde{\theta}_i, \theta_{-i}^{'})}(m_{-i}) > 0 \text{ implies } m_{-i} \in \mathcal{S}_{-i}^{\infty}(\theta_{-i}^{'}) \end{array} \right) \\
\geq & \sum_{\theta_{-i}, \theta_{-i}^{'}} \psi_i(\theta_{-i}, \theta_{-i}^{'}) u_i\big(y^{m_i}(\theta_{-i}^{'}), (\theta_i, \theta_{-i})\big) \\
& \quad (\because \text{ inequality (3) holds for any } m_i \in M_i) \\
= & \sum_{\theta_{-i}, \theta_{-i}^{'}} \psi_i(\theta_{-i}, \theta_{-i}^{'}) \left(\sum_{m_{-i}} \operatorname{marg}_{M_{-i}} \psi_i^{(\tilde{\theta}_i, \theta_{-i}^{'})}(m_{-i}) u_i\big(g(m_i, m_{-i}), (\theta_i, \theta_{-i})\big) \right) \\
& \quad \text{(by definition of } y^{m_i}) \\
= & \sum_{\theta_{-i}, m_{-i}} \psi_i^{\Gamma}(\theta_{-i}, m_{-i}) u_i\big(g(m_i, m_{-i}), (\theta_i, \theta_{-i})\big) \\
& \quad \text{(by definition of } \psi_i^{\Gamma}).
\end{aligned}
$$

Since $m_i^{'}$ is a best response for agent i of payoff type θ_i against ψ_i^{Γ} and $\psi_i^{\Gamma}(\theta_{-i}, m_{-i}) > 0$ implies $m_{-i} \in \mathcal{S}_{-i}(\theta_{-i})$, it follows by definition that $m_i^{'} \in b_i^{\Theta}(\mathcal{S})[\theta_i]$.

As $b^{\Theta}(\mathcal{S}) \geq \mathcal{S}$, we have $\mathcal{S} \leq \mathcal{S}^{\infty}$. For any $\theta \in \Theta$ and $\theta^{'} \in \beta(\theta)$, we obtain $\mathcal{S}^{\infty}(\theta^{'}) \neq \emptyset$ since the mechanism Γ weakly rationalizably implements f. So pick any $m^{'} \in \mathcal{S}^{\infty}(\theta^{'}) \subseteq \mathcal{S}(\theta) \subseteq \mathcal{S}^{\infty}(\theta)$. Then $g(m^{'}) = f(\theta^{'})$ and $g(m^{'}) = f(\theta)$ because, once again, the mechanism Γ weakly rationalizably implements f. Thus, $f(\theta^{'}) = f(\theta)$. So β is acceptable. This completes the proof. $\square$

For responsive SCFs, Bergemann and Morris (2010) identify a necessary condition for weak rationalizable implementation that is a priori stronger than weak RM. We present an equivalent definition below.

Definition 4.4. We say that an unacceptable deception β is *strictly refutable* if there exist $i \in I$, $\theta_i \in \Theta_i$, and $\theta_i^{'} \in \beta_i(\theta_i)$ satisfying $\theta_i^{'} \not\sim_i^f \theta_i$ such that for all $\psi_i \in \Delta(\Theta_{-i} \times \Theta_{-i})$

satisfying $\psi_i(\theta_{-i}, \theta'_{-i}) > 0 \Rightarrow \theta'_{-i} \in \beta_{-i}(\theta_{-i})$, there exists $y \in \bigcap_{\tilde{\theta}_i \in \Theta_i} Y_i[\tilde{\theta}_i]$ such that

$$\sum_{\theta_{-i},\theta'_{-i}} \psi_i(\theta_{-i}, \theta'_{-i}) u_i\big(y(\theta'_{-i}), (\theta_i, \theta_{-i})\big) > \sum_{\theta_{-i},\theta'_{-i}} \psi_i(\theta_{-i}, \theta'_{-i}) u_i\big(f(\theta'_i, \theta'_{-i}), (\theta_i, \theta_{-i})\big).$$

If an unacceptable deception β is strictly refutable, then it is weakly refutable. This is because strict refutability requires us to find a y in $\bigcap_{\tilde{\theta}_i \in \Theta_i} Y_i[\tilde{\theta}_i]$ whereas for weak refutability, we are allowed to find a y in $Y_i[\tilde{\theta}_i]$ that depends on $\tilde{\theta}_i$.

Definition 4.5. The SCF f satisfies *strict robust monotonicity (strict RM)* if every unacceptable deception β is strictly refutable.

Strict RM implies weak RM because strict RM imposes a stronger refutability requirement on every unacceptable deception.

Bergemann and Morris (2010, Lemmata 4, 5, and 6 and Proposition 4) show that strict RM is necessary for weak rationalizable implementation of responsive SCFs. The equivalence between weak rationalizable implementation and robust implementation in rationalizable strategies (Theorem 3.1) immediately implies that for responsive SCFs, strict RM is a necessary condition for robust implementation in rationalizable strategies. Indeed, for responsive SCFs, weak RM implies strict RM.

Lemma 4.6. *If the SCF is responsive and satisfies weak RM, then it must satisfy strict RM.*[5]

However, there are non-responsive SCFs that satisfy weak RM but not strict RM, as shown in Example 6.2.

5 Sufficiency for Robust Implementation in Rationalizable Strategies

In this section, we show that weak RM is sufficient for robust implementation in rationalizable strategies under a mild additional assumption: conditional no total indifference (as discussed below, our definition is weaker than the one appearing in Bergemann and Morris, 2011).

[5]Here is an indirect argument to prove this statement. We show later that weak RM implies semi-strict ex post incentive compatibility (semi-strict EPIC). Bergemann and Morris (2010, Lemma 4) show that for responsive SCFs, semi-strict EPIC implies strict ex post incentive compatibility. Bergemann and Morris (2010, Lemma 5) show that if the SCF satisfies strict ex post incentive compatibility and "strict pairwise robust monotonicity", then it satisfies strict RM. Strict pairwise robust monotonicity is weaker than weak RM. It thus follows that if the SCF is responsive and satisfies weak RM, then it satisfies strict ex post incentive compatibility and strict pairwise robust monotonicity, and hence it satisfies strict RM.

For each $i \in I$ and $\theta_i \in \Theta_i$, define

$$Y_i^w[\theta_i] \equiv \left\{ y : \Theta_{-i} \to \Delta(A) : \forall \theta_{-i},\ u_i\big(f(\theta_i, \theta_{-i}), (\theta_i, \theta_{-i})\big) \geq u_i\big(y(\theta_{-i}), (\theta_i, \theta_{-i})\big) \right\}.$$

Thus, $Y_i^w[\theta_i]$ is the collection of all mappings $y : \Theta_{-i} \to \Delta(A)$ such that for every $\theta_{-i} \in \Theta_{-i}$, the lottery $y(\theta_{-i})$ is weakly worse than $f(\theta_i, \theta_{-i})$ for individual i in state (θ_i, θ_{-i}). Notice that $Y_i[\theta_i]$ (recall the definition from Section 4) is a subset of $Y_i^w[\theta_i]$.

Definition 5.1. The SCF f satisfies *conditional no total indifference* (conditional NTI) if, for all $i \in I$, $\theta_i \in \Theta_i$, and $\psi_i \in \Delta(\Theta_{-i} \times \Theta_{-i})$, there exist $y, y' \in Y_i^w[\theta_i]$ such that

$$\sum_{\theta_{-i}, \theta'_{-i}} \psi_i(\theta_{-i}, \theta'_{-i}) u_i\big(y(\theta'_{-i}), (\theta_i, \theta_{-i})\big) > \sum_{\theta_{-i}, \theta'_{-i}} \psi_i(\theta_{-i}, \theta'_{-i}) u_i\big(y'(\theta'_{-i}), (\theta_i, \theta_{-i})\big).$$

Remark: Bergemann and Morris (2011) also define a "conditional no total indifference" condition which is stronger than our definition. They require the existence of the said y and y' in the set $\bigcap_{\tilde{\theta}_i \in \Theta_i} Y_i^w[\tilde{\theta}_i]$ whereas we only require the existence of y and y' in the set $Y_i^w[\theta_i]$.

In the sufficiency result, we focus on a countable subset of $Y_i^w[\theta_i]$, as defined next. Recall that $\Delta^*(A)$ is a countable dense subset of $\Delta(A)$. For each i and θ_i, define

$$Y_i^*[\theta_i] \equiv \left\{ y : \Theta_{-i} \to \Delta(A) : \begin{array}{ll} & \forall \theta_{-i}, \\ \text{(i)} & y(\theta_{-i}) \in \Delta^*(A) \bigcup_{\theta'_i \in \Theta_i} \{f(\theta'_i, \theta_{-i})\} \text{ and} \\ \text{(ii)} & u_i\big(f(\theta_i, \theta_{-i}), (\theta_i, \theta_{-i})\big) \geq u_i\big(y(\theta_{-i}), (\theta_i, \theta_{-i})\big) \end{array} \right\}$$

Note that $Y_i^*[\theta_i] \subseteq Y_i^w[\theta_i]$. Since Θ_{-i} is finite and $\Delta^*(A)$ is countable, $Y_i^*[\theta_i]$ is also countable. Thus, we denote $Y_i^*[\theta_i]$ by $\{y_i^0[\theta_i], y_i^1[\theta_i], \ldots, y_i^k[\theta_i], \ldots\}$. For each $i \in I$ and $\theta_i \in \Theta_i$, we then define $y_i^{\theta_i}$ such that

$$y_i^{\theta_i}(\theta_{-i}) = (1 - \delta) \sum_{k=0}^{\infty} \delta^k y_i^k[\theta_i](\theta_{-i}), \forall \theta_{-i},$$

where $\delta \in (0, 1)$.

Similarly, since A is countable, we denote it by $\{a_0, a_1, \ldots, a_k, \ldots\}$. Then, we define

$$\bar{\alpha} = (1 - \eta) \sum_{k=0}^{\infty} \eta^k a_k,$$

where $\eta \in (0, 1)$.

The following lemma notes two important consequences of conditional NTI.

Lemma 5.2. *If the SCF f satisfies conditional NTI, then the following statements are true:*

(a) For all $i \in I$, $\theta_i \in \Theta_i$ and $\psi_i \in \Delta(\Theta_{-i} \times \Theta_{-i})$, there exists $y \in Y_i^*[\theta_i]$ such that

$$\sum_{\theta_{-i},\theta'_{-i}} \psi_i(\theta_{-i}, \theta'_{-i}) u_i\big(y(\theta'_{-i}), (\theta_i, \theta_{-i})\big) > \sum_{\theta_{-i},\theta'_{-i}} \psi_i(\theta_{-i}, \theta'_{-i}) u_i\big(y_i^{\theta_i}(\theta'_{-i}), (\theta_i, \theta_{-i})\big).$$

(b) For all $i \in I$, $\theta_i \in \Theta_i$ and $z_i^1 \in \Delta(\Theta_{-i})$, there exists $a \in A$ such that

$$\sum_{\theta_{-i}} z_i^1(\theta_{-i}) u_i\big(a, (\theta_i, \theta_{-i})\big) > \sum_{\theta_{-i}} z_i^1(\theta_{-i}) u_i\big(\bar{\alpha}, (\theta_i, \theta_{-i})\big).$$

Proof. We prove (a) first. Suppose the SCF f satisfies conditional NTI. Pick any $i \in I$, $\theta_i \in \Theta_i$ and $\psi_i \in \Delta(\Theta_{-i} \times \Theta_{-i})$.

Firstly, it follows from the definition of conditional NTI that for all $\theta'_{-i} \in \Theta_{-i}$, there exists $\ell^{\theta'_{-i}} \in \Delta(A)$ such that

$$u_i\big(f(\theta_i, \theta'_{-i}), (\theta_i, \theta'_{-i})\big) > u_i\big(\ell^{\theta'_{-i}}, (\theta_i, \theta'_{-i})\big). \tag{4}$$

To see this, consider the degenerate belief $\tilde{\psi}_i$ such that $\tilde{\psi}_i(\theta'_{-i}, \theta'_{-i}) = 1$. Then there must exist $\tilde{y}, \tilde{y}' \in Y_i^w[\theta_i]$ such that

$$\begin{aligned} u_i\big(f(\theta_i, \theta'_{-i}), (\theta_i, \theta'_{-i})\big) &\geq u_i\big(\tilde{y}(\theta'_{-i}), (\theta_i, \theta'_{-i})\big) \\ &= \sum_{\theta_{-i},\theta''_{-i}} \tilde{\psi}_i(\theta_{-i}, \theta''_{-i}) u_i\big(\tilde{y}(\theta''_{-i}), (\theta_i, \theta_{-i})\big) \\ &> \sum_{\theta_{-i},\theta''_{-i}} \tilde{\psi}_i(\theta_{-i}, \theta''_{-i}) u_i\big(\tilde{y}'(\theta''_{-i}), (\theta_i, \theta_{-i})\big) \\ &= u_i\big(\tilde{y}'(\theta'_{-i}), (\theta_i, \theta'_{-i})\big), \end{aligned}$$

where the first weak inequality follows from the fact that $\tilde{y} \in Y_i^w[\theta_i]$ and the strict inequality follows from conditional NTI. Then $\ell^{\theta'_{-i}} = \tilde{y}'(\theta'_{-i})$ satisfies (4).

Secondly, since f satisfies conditional NTI, there exist $y, y' \in Y_i^w[\theta_i]$ such that

$$\sum_{\theta_{-i},\theta'_{-i}} \psi_i(\theta_{-i}, \theta'_{-i}) u_i\big(y(\theta'_{-i}), (\theta_i, \theta_{-i})\big) > \sum_{\theta_{-i},\theta'_{-i}} \psi_i(\theta_{-i}, \theta'_{-i}) u_i\big(y'(\theta'_{-i}), (\theta_i, \theta_{-i})\big).$$

Pick any $\epsilon \in (0, 1)$ and define $y^\epsilon : \Theta_{-i} \to \Delta(A)$ such that $y^\epsilon(\theta'_{-i}) = (1 - \epsilon)y(\theta'_{-i}) + \epsilon\ell^{\theta'_{-i}}$ for all θ'_{-i}. We similarly define y'^{ϵ}. By construction, y^{ϵ} and y'^{ϵ} are such that for all $\theta'_{-i} \in \Theta_{-i}$,

$$u_i\big(f(\theta_i, \theta'_{-i}), (\theta_i, \theta'_{-i})\big) > u_i\big(y^{\epsilon}(\theta'_{-i}), (\theta_i, \theta'_{-i})\big) \text{ and } u_i\big(f(\theta_i, \theta'_{-i}), (\theta_i, \theta'_{-i})\big) > u_i\big(y'^{\epsilon}(\theta'_{-i}), (\theta_i, \theta'_{-i})\big).$$

For ϵ sufficiently close to 1, we have

$$\sum_{\theta_{-i}, \theta'_{-i}} \psi_i(\theta_{-i}, \theta'_{-i}) u_i\big(y^{\epsilon}(\theta'_{-i}), (\theta_i, \theta_{-i})\big) > \sum_{\theta_{-i}, \theta'_{-i}} \psi_i(\theta_{-i}, \theta'_{-i}) u_i\big(y'^{\epsilon}(\theta'_{-i}), (\theta_i, \theta_{-i})\big).$$

We fix any such sufficiently large ϵ.

Thirdly, since $\Delta^*(A)$ is a dense subset of $\Delta(A)$, for each θ'_{-i}, there exists a sequence of lotteries $\{\ell^z(\theta'_{-i})\}_{z=1}^{\infty} \in \Delta^*(A)$ converging to $y^{\epsilon}(\theta'_{-i})$. For each $z \geq 1$, define $y^z : \Theta_{-i} \to \Delta^*(A)$ such that $y^z(\theta'_{-i}) = \ell^z(\theta'_{-i})$ for all θ'_{-i}. Similarly, we can define $y'^z : \Theta_{-i} \to \Delta^*(A)$ such that $y'^z(\theta'_{-i})$ converges to $y'^{\epsilon}(\theta'_{-i})$ for all θ'_{-i}. As Θ_{-i} is finite, there exists a sufficiently large z such that

$$u_i\big(f(\theta_i, \theta'_{-i}), (\theta_i, \theta'_{-i})\big) > u_i\big(y^{z}(\theta'_{-i}), (\theta_i, \theta'_{-i})\big) \text{ and } u_i\big(f(\theta_i, \theta'_{-i}), (\theta_i, \theta'_{-i})\big) > u_i\big(y'^{z}(\theta'_{-i}), (\theta_i, \theta'_{-i})\big),$$

for all θ'_{-i}, and

$$\sum_{\theta_{-i}, \theta'_{-i}} \psi_i(\theta_{-i}, \theta'_{-i}) u_i\big(y^{z}(\theta'_{-i}), (\theta_i, \theta_{-i})\big) > \sum_{\theta_{-i}, \theta'_{-i}} \psi_i(\theta_{-i}, \theta'_{-i}) u_i\big(y'^{z}(\theta'_{-i}), (\theta_i, \theta_{-i})\big). \tag{5}$$

The first set of inequalities imply that $y^z, y'^z \in Y_i^*[\theta_i]$.

Lastly, since $y_i^{\theta_i}$ assigns a positive weight to all $y \in Y_i^*[\theta_i]$, if

$$\sum_{\theta_{-i}, \theta'_{-i}} \psi_i(\theta_{-i}, \theta'_{-i}) u_i\big(y_i^{\theta_i}(\theta'_{-i}), (\theta_i, \theta_{-i})\big) \geq \sum_{\theta_{-i}, \theta'_{-i}} \psi_i(\theta_{-i}, \theta'_{-i}) u_i\big(y(\theta'_{-i}), (\theta_i, \theta_{-i})\big), \forall y \in Y_i^*[\theta_i],$$

then it must be that

$$\sum_{\theta_{-i}, \theta'_{-i}} \psi_i(\theta_{-i}, \theta'_{-i}) u_i\big(y^{z}(\theta'_{-i}), (\theta_i, \theta_{-i})\big) = \sum_{\theta_{-i}, \theta'_{-i}} \psi_i(\theta_{-i}, \theta'_{-i}) u_i\big(y'^{z}(\theta'_{-i}), (\theta_i, \theta_{-i})\big),$$

which contradicts (5).

We prove (b) next. Suppose the SCF f satisfies conditional NTI. Pick any $i \in I$, $\theta_i \in \Theta_i$

and $z_i^1 \in \Delta(\Theta_{-i})$. As $\bar{\alpha}$ assigns a positive weight to all $a \in A$, if

$$\sum_{\theta_{-i}} z_i^1(\theta_{-i}) u_i\big(\bar{\alpha}, (\theta_i, \theta_{-i})\big) \geq \sum_{\theta_{-i}} z_i^1(\theta_{-i}) u_i\big(a, (\theta_i, \theta_{-i})\big), \forall a \in A,$$

then it must be that

$$\sum_{\theta_{-i}} z_i^1(\theta_{-i}) u_i\big(a, (\theta_i, \theta_{-i})\big) = \sum_{\theta_{-i}} z_i^1(\theta_{-i}) u_i\big(a', (\theta_i, \theta_{-i})\big),$$

for all $a, a' \in A$. Now consider the belief $\tilde{\psi}_i \in \Delta(\Theta_{-i} \times \Theta_{-i})$ such that $\tilde{\psi}_i(\theta_{-i}, \theta_{-i}) = z_i^1(\theta_{-i})$ for all $\theta_{-i} \in \Theta_{-i}$. Then, by conditional NTI, there must exist $\tilde{y}, \tilde{y}' \in Y_i^w[\theta_i]$ such that

$$\sum_{\theta_{-i}, \theta'_{-i}} \psi_i(\theta_{-i}, \theta'_{-i}) u_i\big(\tilde{y}(\theta'_{-i}), (\theta_i, \theta_{-i})\big) > \sum_{\theta_{-i}, \theta'_{-i}} \psi_i(\theta_{-i}, \theta'_{-i}) u_i\big(\tilde{y}'(\theta'_{-i}), (\theta_i, \theta_{-i})\big).$$

But the left-hand side of the above inequality equals $\sum_{\theta_{-i}} z_i^1(\theta_{-i}) u_i\big(\tilde{y}(\theta_{-i}), (\theta_i, \theta_{-i})\big)$ while the right-hand side equals $\sum_{\theta_{-i}} z_i^1(\theta_{-i}) u_i\big(\tilde{y}'(\theta_{-i}), (\theta_i, \theta_{-i})\big)$, which contradicts the fact that type θ_i is indifferent over all alternatives when she holds the belief z_i^1. □

We need one more result before presenting our main sufficiency result for this section. We now show that weak RM implies semi-strict ex post incentive compatibility.

Definition 5.3. The SCF f satisfies *ex post incentive compatibility* (EPIC) if, for all $i \in I$, $\theta_i, \theta'_i \in \Theta_i$, and $\theta_{-i} \in \Theta_{-i}$,

$$u_i\big(f(\theta_i, \theta_{-i}), (\theta_i, \theta_{-i})\big) \geq u_i\big(f(\theta'_i, \theta_{-i}), (\theta_i, \theta_{-i})\big).$$

The SCF f satisfies *semi-strict ex post incentive compatible* (semi-strict EPIC) if the above inequality becomes strict whenever $\theta_i \not\sim_i^f \theta'_i$.

Lemma 5.4. *If the SCF f satisfies weak RM, then it satisfies semi-strict EPIC.*[6]

Proof. Suppose the SCF f satisfies weak RM. Pick any $i \in I$, $\theta_i, \theta'_i \in \Theta_i$. If $\theta_i \sim_i^f \theta'_i$, then trivially $u_i\big(f(\theta_i, \theta_{-i}), (\theta_i, \theta_{-i})\big) \geq u_i\big(f(\theta'_i, \theta_{-i}), (\theta_i, \theta_{-i})\big)$ for all $\theta_{-i} \in \Theta_{-i}$. So suppose

[6]Bergemann and Morris (2010, Lemma 6) show that if f is weakly rationalizable implementable, then it satisfies semi-strict EPIC. It follows from their result and our Theorem 3.1 that semi-strict EPIC is a necessary condition for robust implementation in rationalizable strategies. The above lemma does not immediately follow from Bergemann and Morris's result because weak RM is a necessary condition for weak rationalizable implementation. Moreover, due to this lemma, we do not have to add semi-strict EPIC as an additional condition in our sufficiency result. Bergemann and Morris (2011, Lemma 1) show that "robust monotonicity" implies semi-strict EPIC. Robust monotonicity is a slightly weaker version of strict RM – the only difference is that we need to replace "$y \in \bigcap_{\tilde{\theta}_i} Y_i[\tilde{\theta}_i]$" with "$y \in \bigcap_{\tilde{\theta}_i} Y_i^w[\tilde{\theta}_i]$" in the definition of strict refutability. Strictly speaking, weak RM and robust monotonicity are not comparable.

$\theta_i \not\sim_i^f \theta_i'$. Consider the deception β such that $\beta_j(\theta_j) = \{\theta_j\}$ for all θ_j and $j \neq i$ but

$$\beta_i(\tilde{\theta}_i) = \begin{cases} \{\theta_i, \theta_i'\}, & \text{if } \tilde{\theta}_i = \theta_i \\ \{\tilde{\theta}_i\}, & \text{otherwise.} \end{cases}$$

Since $\theta_i \not\sim_i^f \theta_i'$, the deception β is unacceptable. Hence, it must be weakly refutable. That is, there exist $j \in I$, $\hat{\theta}_j \in \Theta_j$, and $\hat{\theta}_j' \in \beta_j(\hat{\theta}_j)$ satisfying $\hat{\theta}_j' \not\sim_j^f \hat{\theta}_j$ such that for any $\tilde{\theta}_j \in \Theta_j$ and $\psi_j \in \Delta(\Theta_{-j} \times \Theta_{-j})$ satisfying $\psi_j(\theta_{-j}, \theta_{-j}') > 0 \Rightarrow \theta_{-j}' \in \beta_{-j}(\theta_{-j})$, there exists $y \in Y_j[\tilde{\theta}_j]$ such that

$$\sum_{\theta_{-j}, \theta_{-j}'} \psi_j(\theta_{-j}, \theta_{-j}') u_j\big(y(\theta_{-j}'), (\hat{\theta}_j, \theta_{-j})\big) > \sum_{\theta_{-j}, \theta_{-j}'} \psi_j(\theta_{-j}, \theta_{-j}') u_j\big(f(\hat{\theta}_j', \theta_{-j}'), (\hat{\theta}_j, \theta_{-j})\big).$$

Since $\hat{\theta}_j' \not\sim_j^f \hat{\theta}_j$ and $\hat{\theta}_j' \in \beta_j(\hat{\theta}_j)$, it must be that $j = i$, $\hat{\theta}_j = \theta_i$ and $\hat{\theta}_j' = \theta_i'$.

Now pick any $\theta_{-i} \in \Theta_{-i}$. Consider $\tilde{\theta}_i = \theta_i$ and the degenerate belief ψ_i such that $\psi_i(\theta_{-i}, \theta_{-i}) = 1$. Note that $\theta_{-i} \in \beta_{-i}(\theta_{-i})$. Hence, we must have some $y \in Y_i[\tilde{\theta}_i] = Y_i[\theta_i]$ such that

$$u_i\big(y(\theta_{-i}), (\theta_i, \theta_{-i})\big) > u_i\big(f(\theta_i', \theta_{-i}), (\theta_i, \theta_{-i})\big).$$

But $y \in Y_i[\theta_i]$ implies that $u_i\big(f(\theta_i, \theta_{-i}), (\theta_i, \theta_{-i})\big) \geq u_i\big(y(\theta_{-i}), (\theta_i, \theta_{-i})\big)$. Therefore,

$$u_i\big(f(\theta_i, \theta_{-i}), (\theta_i, \theta_{-i})\big) > u_i\big(f(\theta_i', \theta_{-i}), (\theta_i, \theta_{-i})\big).$$

□

For the sufficiency result, we propose the following mechanism $\Gamma = ((M_i)_{i \in I}, g)$: For each individual i, pick any one payoff type from Θ_i. We denote this payoff type as θ_i^*. Each individual i sends a message $m_i = (m_i^1, m_i^2, m_i^3, m_i^4)$, where $m_i^1 = (m_i^1[j])_{j \in I}$ such that $m_i^1[j] \in \Theta_j$ for all $j \in I$, $m_i^2 \in \mathbb{N}, m_i^3 = (m_i^3[\theta_i])_{\theta_i \in \Theta_i}$ such that $m_i^3[\theta_i] \in Y_i^*[\theta_i]$ for all $\theta_i \in \Theta_i$, and $m_i^4 \in A$. Note that each M_i is countable. The outcome function $g : M \to \Delta(A)$ is defined as follows: For each $m \in M$,

Rule 1: $m_i^2 = 1$ for all $i \in I \Rightarrow g(m) = f(m_1^1[1], m_2^1[2], \ldots, m_n^1[n])$.

Rule 2: If there exists $i \in I$ such that $m_i^2 > 1$ but $m_j^2 = 1$ for all $j \in I \backslash \{i\}$, then one of the following sub-rules apply:

Rule 2-1: If there exists $\theta_i \in \Theta_i$ such that $m_j^1[i] = \theta_i$ for all $j \in I \backslash \{i\}$, then

$$g(m) = \begin{cases} m_i^3[\theta_i]\big((m_j^1[j])_{j \neq i}\big) & \text{with probability } m_i^2/(m_i^2+1), \\ y_i^{\theta_i}\big((m_j^1[j])_{j \neq i}\big) & \text{with probability } 1/(m_i^2+1). \end{cases}$$

Rule 2-2: If $m_{j'}^1[i] \neq m_k^1[i]$ for some $j', k \in I \backslash \{i\}$, then

$$g(m) = \begin{cases} m_i^3[\theta_i^*]\big((m_j^1[j])_{j \neq i}\big) & \text{with probability } m_i^2/(m_i^2+1), \\ y_i^{\theta_i^*}\big((m_j^1[j])_{j \neq i}\big) & \text{with probability } 1/(m_i^2+1). \end{cases}$$

Rule 3: In all other cases:

$$g(m) = \begin{cases} m_1^4 & \text{with probability } m_1^2/(1+m_1^2)n, \\ m_2^4 & \text{with probability } m_2^2/(1+m_2^2)n, \\ \vdots & \vdots \\ m_n^4 & \text{with probability } m_n^2/(1+m_n^2)n, \\ \bar{\alpha} & \text{with remaining probability.} \end{cases}$$

Here is our sufficiency result for robust implementation in rationalizable strategies:

Theorem 5.5. *If the SCF f satisfies weak RM and conditional NTI, it is robustly implementable in rationalizable strategies.*

Proof. : We use the mechanism Γ constructed above and prove that Γ weakly robustly implements f, which implies that Γ robustly implements f in rationalizable strategies because of Theorem 3.1. The proof of the theorem consists of Steps 1 through 4.

Step 1: $m_i \in \mathcal{S}_i^\infty(\theta_i) \Rightarrow m_i^2 = 1$.

Proof. Suppose by way of contradiction that $m_i \in \mathcal{S}_i^\infty(\theta_i)$ but $m_i^2 > 1$. Then, m_i is a best response of individual i of payoff type θ_i against some conjecture $\psi_i \in \Delta(\Theta_{-i} \times M_{-i})$.

For each $\theta_i' \neq \theta_i^*$ and $\theta_{-i}' \in \Theta_{-i}$, we define

$$M_{-i}^2(\theta_i', \theta_{-i}') = \left\{ m_{-i} : m_j^2 = 1 \text{ and } m_j^1[i] = \theta_i', \forall j \neq i, \text{and } (m_j^1[j])_{j \neq i} = \theta_{-i}' \right\}.$$

For θ_i^* and each $\theta_{-i}' \in \Theta_{-i}$, we define

$$M_{-i}^2(\theta_i^*, \theta_{-i}') = \left\{ m_{-i} : \begin{array}{l} (m_j^1[j])_{j \neq i} = \theta_{-i}' \text{ and} \\ \text{either } m_j^2 = 1 \text{ and } m_j^1[i] = \theta_i^*, \forall j \neq i, \\ \text{or } m_j^2 = 1, \forall j \neq i, \text{ but } m_{j'}^1[i] \neq m_k^1[i] \text{ for some } j', k \neq i \end{array} \right\}.$$

Also define

$$M_{-i}^3 \;=\; \left\{m_{-i} : \text{there exist one or more } j \neq i \text{ such that } m_j^2 > 1\right\}.$$

Note that $\left((M_{-i}^2(\tilde{\theta}_i, \theta'_{-i}))_{\tilde{\theta}_i \in \Theta_i, \theta'_{-i} \in \Theta_{-i}}, M_{-i}^3\right)$ defines a partition of M_{-i}. As $m_i^2 > 1$, if $m_{-i} \in M_{-i}^2(\tilde{\theta}_i, \theta'_{-i})$, then Rule 2 is used under the profile (m_i, m_{-i}) whereas if $m_{-i} \in M_{-i}^3$, then Rule 3 is used under the profile (m_i, m_{-i}).

For each $\tilde{\theta}_i \in \Theta_i$, define

$$\Psi_i^{2,\tilde{\theta}_i} = \sum_{\theta_{-i}, \theta''_{-i}} \sum_{m_{-i} \in M_{-i}^2(\tilde{\theta}_i, \theta''_{-i})} \psi_i(\theta_{-i}, m_{-i}).$$

Thus, $\Psi_i^{2,\tilde{\theta}_i}$ is the probability of the event that all other individuals report a message profile in $\bigcup_{\theta''_{-i}} M_{-i}^2(\tilde{\theta}_i, \theta''_{-i})$.

Also, define

$$\Psi_i^3 = \sum_{\theta_{-i}, m_{-i} \in M_{-i}^3} \psi_i(\theta_{-i}, m_{-i}).$$

Thus, Ψ_i^3 is the probability of the event that all other individuals report a message profile in M_{-i}^3.

If $\tilde{\theta}_i$ is such that $\Psi_i^{2,\tilde{\theta}_i} > 0$, then define $\psi_i^{2,\tilde{\theta}_i} \in \Delta(\Theta_{-i} \times \Theta_{-i})$ such that for all $\theta_{-i}, \theta'_{-i} \in \Theta_{-i}$,

$$\psi_i^{2,\tilde{\theta}_i}(\theta_{-i}, \theta'_{-i}) = \sum_{m_{-i} \in M_{-i}^2(\tilde{\theta}_i, \theta'_{-i})} \frac{\psi_i(\theta_{-i}, m_{-i})}{\Psi_i^{2,\tilde{\theta}_i}}.$$

Thus, $\psi_i^{2,\tilde{\theta}_i}(\theta_{-i}, \theta'_{-i})$ is the conditional probability of the event that the payoff-type profile of all other individuals is θ_{-i} and they report a message profile in $M_{-i}^2(\tilde{\theta}_i, \theta'_{-i})$ given the event that all other individuals report a message profile in $\bigcup_{\theta''_{-i}} M_{-i}^2(\tilde{\theta}_i, \theta''_{-i})$.

If the payoff-type profile of all other individuals is θ_{-i} and they report a message profile in $M_{-i}^2(\tilde{\theta}_i, \theta'_{-i})$, then when individual i of payoff type θ_i plays m_i, she expects the outcome to be given by the lottery

$$\left(\frac{m_i^2}{1 + m_i^2}\right) m_i^3[\tilde{\theta}_i](\theta'_{-i}) + \left(1 - \frac{m_i^2}{1 + m_i^2}\right) y_i^{\tilde{\theta}_i}(\theta'_{-i}).$$

As a result, conditional on the event that all other individuals report a message profile in

$\bigcup_{\theta''_{-i}} M^2_{-i}(\tilde{\theta}_i, \theta''_{-i})$, the expected payoff of individual i of payoff type θ_i when she plays m_i is

$$\left(\frac{m_i^2}{1+m_i^2}\right) \sum_{\theta_{-i},\theta'_{-i}} \psi_i^{2,\tilde{\theta}_i}(\theta_{-i},\theta'_{-i}) u_i\big(m_i^3[\tilde{\theta}_i](\theta'_{-i}), (\theta_i,\theta_{-i})\big)$$
$$+\left(1-\frac{m_i^2}{1+m_i^2}\right) \sum_{\theta_{-i},\theta'_{-i}} \psi_i^{2,\tilde{\theta}_i}(\theta_{-i},\theta'_{-i}) u_i\big(y_i^{\tilde{\theta}_i}(\theta'_{-i}), (\theta_i,\theta_{-i})\big). \qquad (6)$$

If $\Psi_i^3 > 0$, then define $\psi_i^3 \in \Delta(\Theta_{-i})$ such that, for any $\theta_{-i} \in \Theta_{-i}$,

$$\psi_i^3(\theta_{-i}) = \sum_{m_{-i}\in M^3_{-i}} \frac{\psi_i(\theta_{-i}, m_{-i})}{\Psi_i^3}.$$

Thus, $\psi_i^3(\theta_{-i})$ is the conditional probability of the event that the payoff-type profile of all other individuals is θ_{-i} and they report a message profile in M^3_{-i} given the event that all other individuals report a message profile in M^3_{-i}.

If the payoff-type profile of all other individuals is θ_{-i} and they report a message profile $m_{-i} \in M^3_{-i}$, then when individual i of payoff type θ_i plays m_i, she expects the outcome to be given by the lottery

$$\frac{1}{n}\left(\frac{m_i^2}{1+m_i^2}\right) m_i^4 + \frac{1}{n}\left(1-\frac{m_i^2}{1+m_i^2}\right)\bar{\alpha} + \sum_{j\neq i}\left(\frac{1}{n}\left(\frac{m_j^2}{1+m_j^2}\right) m_j^4 + \frac{1}{n}\left(1-\frac{m_j^2}{1+m_j^2}\right)\bar{\alpha}\right).$$

As a result, conditional on the event that all other individuals report a message profile in M^3_{-i}, the expected payoff of individual i of payoff type θ_i when she plays m_i is

$$\frac{1}{n}\left(\frac{m_i^2}{1+m_i^2}\right)\sum_{\theta_{-i}} \psi_i^3(\theta_{-i}) u_i\big(m_i^4, (\theta_i,\theta_{-i})\big) + \frac{1}{n}\left(1-\frac{m_i^2}{1+m_i^2}\right)\sum_{\theta_{-i}} \psi_i^3(\theta_{-i}) u_i\big(\bar{\alpha}, (\theta_i,\theta_{-i})\big)$$
$$+\sum_{\theta_{-i},m_{-i}\in M^3_{-i}} \frac{\psi_i(\theta_{-i},m_{-i})}{\Psi_i^3} \sum_{j\neq i}\left(\frac{1}{n}\left(\frac{m_j^2}{1+m_j^2}\right) u_i\big(m_j^4, (\theta_i,\theta_{-i})\big) + \frac{1}{n}\left(1-\frac{m_j^2}{1+m_j^2}\right) u_i\big(\bar{\alpha}, (\theta_i,\theta_{-i})\big)\right). \qquad (7)$$

Now let individual i of type θ_i deviate to $\hat{m}_i = (m_i^1, \hat{m}_i^2, \hat{m}_i^3, \hat{m}_i^4)$ such that

- $\hat{m}_i^2 = m_i^2 + 1$.
- $\hat{m}_i^3$ is defined as follows for each $\tilde{\theta}_i$:

▷ If $\Psi_i^{2,\tilde{\theta}_i} > 0$, then let $\hat{m}_i^3[\tilde{\theta}_i] \in Y_i^*[\tilde{\theta}_i]$ be such that

$$\sum_{\theta_{-i},\theta'_{-i}} \psi_i^{2,\tilde{\theta}_i}(\theta_{-i},\theta'_{-i})u_i\big(\hat{m}_i^3[\tilde{\theta}_i](\theta'_{-i}),(\theta_i,\theta_{-i})\big) \geq \sum_{\theta_{-i},\theta'_{-i}} \psi_i^{2,\tilde{\theta}_i}(\theta_{-i},\theta'_{-i})u_i\big(m_i^3[\tilde{\theta}_i](\theta'_{-i}),(\theta_i,\theta_{-i})\big)$$

and

$$\sum_{\theta_{-i},\theta'_{-i}} \psi_i^{2,\tilde{\theta}_i}(\theta_{-i},\theta'_{-i})u_i\big(\hat{m}_i^3[\tilde{\theta}_i](\theta'_{-i}),(\theta_i,\theta_{-i})\big) > \sum_{\theta_{-i},\theta'_{-i}} \psi_i^{2,\tilde{\theta}_i}(\theta_{-i},\theta'_{-i})u_i\big(y_i^{\tilde{\theta}_i}(\theta'_{-i}),(\theta_i,\theta_{-i})\big).$$

Note that such $\hat{m}_i^3[\tilde{\theta}_i]$ exists because of Lemma 5.2.

▷ If $\Psi_i^{2,\tilde{\theta}_i} = 0$, then let $\hat{m}_i^3[\tilde{\theta}_i] = m_i^3[\tilde{\theta}_i]$.

- $\hat{m}_i^4$ is defined as follows:

▷ If $\Psi_i^3 > 0$, then let $\hat{m}_i^4 \in A$ be such that

$$\sum_{\theta_{-i}} \psi_i^3(\theta_{-i})u_i\big(\hat{m}_i^4,(\theta_i,\theta_{-i})\big) \geq \sum_{\theta_{-i}} \psi_i^3(\theta_{-i})u_i\big(m_i^4,(\theta_i,\theta_{-i})\big)$$

and

$$\sum_{\theta_{-i}} \psi_i^3(\theta_{-i})u_i\big(\hat{m}_i^4,(\theta_i,\theta_{-i})\big) > \sum_{\theta_{-i}} \psi_i^3(\theta_{-i})u_i\big(\bar{\alpha},(\theta_i,\theta_{-i})\big).$$

Note that such $\hat{m}_i^4$ exists because of Lemma 5.2.

▷ If $\Psi_i^3 = 0$, then let $\hat{m}_i^4 = m_i^4$.

If $\Psi_i^{2,\tilde{\theta}_i} > 0$, then conditional on the event that all other individuals report a message profile in $\bigcup_{\theta''_{-i}} M_{-i}^2(\tilde{\theta}_i,\theta''_{-i})$, the expected payoff of individual i of payoff type θ_i when she plays $\hat{m}_i$ is

$$\left(\frac{\hat{m}_i^2}{1+\hat{m}_i^2}\right)\sum_{\theta_{-i},\theta'_{-i}} \psi_i^{2,\tilde{\theta}_i}(\theta_{-i},\theta'_{-i})u_i\big(\hat{m}_i^3[\tilde{\theta}_i](\theta'_{-i}),(\theta_i,\theta_{-i})\big)$$
$$+\left(1-\frac{\hat{m}_i^2}{1+\hat{m}_i^2}\right)\sum_{\theta_{-i},\theta'_{-i}} \psi_i^{2,\tilde{\theta}_i}(\theta_{-i},\theta'_{-i})u_i\big(y_i^{\tilde{\theta}_i}(\theta'_{-i}),(\theta_i,\theta_{-i})\big),$$

which is, by construction, greater than her expected payoff in (6) when she plays m_i.

If $\Psi_i^3 > 0$, then conditional on the event that all other individuals report a message profile in M^3_{-i}, the expected payoff of individual i of payoff type θ_i when she plays $\hat{m}_i$ is

$$\frac{1}{n}\left(\frac{\hat{m}_i^2}{1+\hat{m}_i^2}\right)\sum_{\theta_{-i}}\psi_i^3(\theta_{-i})u_i\big(\hat{m}_i^4,(\theta_i,\theta_{-i})\big)+\frac{1}{n}\left(1-\frac{\hat{m}_i^2}{1+\hat{m}_i^2}\right)\sum_{\theta_{-i}}\psi_i^3(\theta_{-i})u_i\big(\bar{\alpha},(\theta_i,\theta_{-i})\big)$$
$$+\sum_{\theta_{-i},m_{-i}\in M^3_{-i}}\frac{\psi_i(\theta_{-i},m_{-i})}{\Psi_i^3}\sum_{j\neq i}\left(\frac{1}{n}\left(\frac{m_j^2}{1+m_j^2}\right)u_i\big(m_j^4,(\theta_i,\theta_{-i})\big)+\frac{1}{n}\left(1-\frac{m_j^2}{1+m_j^2}\right)u_i\big(\bar{\alpha},(\theta_i,\theta_{-i})\big)\right),$$

which is, by construction, greater than her expected payoff in (7) when she plays m_i.

As $\sum_{\tilde{\theta}_i}\Psi_i^{2,\tilde{\theta}_i}+\Psi_i^3=1$ (because $m_i^2>1$), it follows that $\hat{m}_i$ is a better response for individual i of type θ_i against ψ_i, a contradiction. This completes the proof of Step 1. □

Step 2: For each $i\in I$ and $\theta_i\in\Theta_i$, let

$$\beta_i(\theta_i)=\{\theta_i\}\cup\{\theta_i^{'}\in\Theta_i:\exists\ m_i\in\mathcal{S}_i^\infty(\theta_i)\text{ such that }m_i^1[i]=\theta_i^{'}\}.$$

Then, the deception $\beta=(\beta_i)_{i\in I}$ is acceptable.

Proof. Suppose not, that is, β is unacceptable. Then, by weak RM, β must be weakly refutable. That is, there exist $i\in I,\ \theta_i\in\Theta_i$, and $\theta_i^{'}\in\beta_i(\theta_i)$ satisfying $\theta_i^{'}\not\approx_i^f\theta_i$ such that for all $\tilde{\theta}_i\in\Theta_i$ and $\psi_i\in\Delta(\Theta_{-i}\times\Theta_{-i})$ satisfying $\psi_i(\theta_{-i},\theta_{-i}^{'})>0\Rightarrow\theta_{-i}^{'}\in\beta_{-i}(\theta_{-i})$, there exists $y\in Y_i[\tilde{\theta}_i]$ such that

$$\sum_{\theta_{-i},\theta_{-i}^{'}}\psi_i(\theta_{-i},\theta_{-i}^{'})u_i\big(y(\theta_{-i}^{'}),(\theta_i,\theta_{-i})\big)>\sum_{\theta_{-i},\theta_{-i}^{'}}\psi_i(\theta_{-i},\theta_{-i}^{'})u_i\big(f(\theta_i^{'},\theta_{-i}^{'}),(\theta_i,\theta_{-i})\big).$$

As $\theta_i^{'}\not\approx_i^f\theta_i$ and $\theta_i^{'}\in\beta_i(\theta_i)$, we can find a message $m_i\in\mathcal{S}_i^\infty(\theta_i)$ such that $m_i^1[i]=\theta_i^{'}$. Then, m_i is a best response to some belief $\psi_i^\Gamma\in\Delta(\Theta_{-i}\times M_{-i})$ such that $\psi_i^\Gamma(\theta_{-i},m_{-i})>0\Rightarrow m_{-i}\in\mathcal{S}_{-i}^\infty(\theta_{-i})$. From Step 1, it follows that $\psi_i^\Gamma(\theta_{-i},m_{-i})>0$ implies $m_j^2=1$ for all $j\neq i$. We next define a partition of all those message profiles in M_{-i} such that $m_j^2=1$ for all $j\neq i$.

For each $\hat{\theta}_i\neq\theta_i^*$ and $\theta_{-i}^{'}\in\Theta_{-i}$, we define

$$M_{-i}^1(\hat{\theta}_i,\theta_{-i}^{'})\ =\ \left\{m_{-i}:m_j^2=1\text{ and }m_j^1[i]=\hat{\theta}_i,\forall j\neq i,\text{and }(m_j^1[j])_{j\neq i}=\theta_{-i}^{'}\right\}.$$

For θ_i^* and each $\theta_{-i}^{'} \in \Theta_{-i}$, we define

$$M_{-i}^1(\theta_i^*, \theta_{-i}^{'}) \quad = \quad \left\{ m_{-i} : \begin{array}{l} (m_j^1[j])_{j \neq i} = \theta_{-i}^{'} \text{ and} \\ \text{either } m_j^2 = 1 \text{ and } m_j^1[i] = \theta_i^*, \forall j \neq i, \\ \text{or } m_j^2 = 1, \forall j \neq i, \text{ but } m_{j'}^1[i] \neq m_k^1[i] \text{ for some } j', k \neq i \end{array} \right\}.$$

For each $\tilde{\theta}_i \in \Theta_i$, we define

$$\Psi_i^{1,\tilde{\theta}_i} = \sum_{\theta_{-i}, \theta_{-i}^{''}} \sum_{m_{-i} \in M_{-i}^1(\tilde{\theta}_i, \theta_{-i}^{''})} \psi_i^{\Gamma}(\theta_{-i}, m_{-i}).$$

Thus, $\Psi_i^{1,\tilde{\theta}_i}$ is the probability of the event that all other individuals report a message profile in $\bigcup_{\theta_{-i}^{''}} M_{-i}^1(\tilde{\theta}_i, \theta_{-i}^{''})$.

If $\tilde{\theta}_i$ is such that $\Psi_i^{1,\tilde{\theta}_i} > 0$, then define $\psi_i^{1,\tilde{\theta}_i} \in \Delta(\Theta_{-i} \times \Theta_{-i})$ such that for all $\theta_{-i}, \theta_{-i}^{'} \in \Theta_{-i}$,

$$\psi_i^{1,\tilde{\theta}_i}(\theta_{-i}, \theta_{-i}^{'}) = \sum_{m_{-i} \in M_{-i}^1(\tilde{\theta}_i, \theta_{-i}^{'})} \frac{\psi_i^{\Gamma}(\theta_{-i}, m_{-i})}{\Psi_i^{1,\tilde{\theta}_i}}.$$

Thus, $\psi_i^{1,\tilde{\theta}_i}(\theta_{-i}, \theta_{-i}^{'})$ is the conditional probability of the event that the payoff-type profile of all other individuals is θ_{-i} and they report a message profile in $M_{-i}^1(\tilde{\theta}_i, \theta_{-i}^{'})$ given the event that all other individuals report a message profile in $\bigcup_{\theta_{-i}^{''}} M_{-i}^1(\tilde{\theta}_i, \theta_{-i}^{''})$.

If the payoff-type profile of all other individuals is θ_{-i} and they report a message profile in $M_{-i}^1(\tilde{\theta}_i, \theta_{-i}^{'})$, then when individual i of payoff type θ_i plays m_i, she expects the outcome to be $f(\theta_i^{'}, \theta_{-i}^{'})$. As a result, conditional on the event that all other individuals report a message profile in $\bigcup_{\theta_{-i}^{''}} M_{-i}^1(\tilde{\theta}_i, \theta_{-i}^{''})$, the expected payoff of individual i of payoff type θ_i when she plays m_i is

$$\sum_{\theta_{-i}, \theta_{-i}^{'}} \psi_i^{1,\tilde{\theta}_i}(\theta_{-i}, \theta_{-i}^{'}) u_i\big(f(\theta_i^{'}, \theta_{-i}^{'}), (\theta_i, \theta_{-i})\big). \tag{8}$$

Now, $\psi_i^{1,\tilde{\theta}_i}(\theta_{-i}, \theta_{-i}^{'}) > 0$ implies that $\psi_i^{\Gamma}(\theta_{-i}, m_{-i}) > 0$ for some $m_{-i} \in M_{-i}^1(\tilde{\theta}_i, \theta_{-i}^{'})$. But $\psi_i^{\Gamma}(\theta_{-i}, m_{-i}) > 0$ also implies that $m_{-i} \in \mathcal{S}_{-i}^{\infty}(\theta_{-i})$. Hence, due to the construction of β, we have $\theta_{-i}^{'} \in \beta_{-i}(\theta_{-i})$. So, it follows from weak refutability of β that there exists $y[\tilde{\theta}_i] \in Y_i[\tilde{\theta}_i]$ such that

$$\sum_{\theta_{-i}, \theta_{-i}^{'}} \psi_i^{1,\tilde{\theta}_i}(\theta_{-i}, \theta_{-i}^{'}) u_i\big(y[\tilde{\theta}_i](\theta_{-i}^{'}), (\theta_i, \theta_{-i})\big) > \sum_{\theta_{-i}, \theta_{-i}^{'}} \psi_i^{1,\tilde{\theta}_i}(\theta_{-i}, \theta_{-i}^{'}) u_i\big(f(\theta_i^{'}, \theta_{-i}^{'}), (\theta_i, \theta_{-i})\big).$$

It is without loss of generality to assume that $y[\tilde{\theta}_i] \in Y_i^*[\tilde{\theta}_i]$. If not, then consider any sequence $\ell^z : \Theta_{-i} \to \Delta^*(A) \cup \{f(\tilde{\theta}_i, \theta_{-i})\}$ such that (a) if $y[\tilde{\theta}_i](\theta_{-i}) = f(\tilde{\theta}_i, \theta_{-i})$, then $\ell^z(\theta_{-i}) = f(\tilde{\theta}_i, \theta_{-i})$ for all $z \in \mathbb{N}$ and (b) if $y[\tilde{\theta}_i](\theta_{-i}) \neq f(\tilde{\theta}_i, \theta_{-i})$, then $\ell^z(\theta_{-i})$ converges to $y[\tilde{\theta}_i](\theta_{-i})$ for all $\theta_{-i} \in \Theta_{-i}$ as $z \to \infty$. As Θ_{-i} is finite and $u_i(\cdot, \theta)$ is continuous over $\Delta(A)$, we can find a sufficiently large $\hat{z}$ such that

$$\sum_{\theta_{-i}, \theta'_{-i}} \psi_i^{1,\tilde{\theta}_i}(\theta_{-i}, \theta'_{-i}) u_i\big(\ell^{\hat{z}}(\theta'_{-i}), (\theta_i, \theta_{-i})\big) > \sum_{\theta_{-i}, \theta'_{-i}} \psi_i^{1,\tilde{\theta}_i}(\theta_{-i}, \theta'_{-i}) u_i\big(f(\theta'_i, \theta'_{-i}), (\theta_i, \theta_{-i})\big),$$

and, because $y[\tilde{\theta}_i] \in Y_i[\tilde{\theta}_i]$, if $\ell^{\hat{z}}(\theta_{-i}) \neq f(\tilde{\theta}_i, \theta_{-i})$, then

$$u_i\big(f(\tilde{\theta}_i, \theta_{-i}), (\tilde{\theta}_i, \theta_{-i})\big) > u_i\big(\ell^{\hat{z}}(\theta_{-i}), (\tilde{\theta}_i, \theta_{-i})\big).$$

The latter condition implies that $\ell^{\hat{z}} \in Y_i^*[\tilde{\theta}_i]$.

Now, let individual i of type θ_i deviate to $\hat{m}_i = (m_i^1, \hat{m}_i^2, \hat{m}_i^3, m_i^4)$ such that

- $\hat{m}_i^2 > 1$, where the specific value is chosen later.
- $\hat{m}_i^3$ is defined as follows: for each $\tilde{\theta}_i \in \Theta_i$:
  - ▷ If $\Psi_i^{1,\tilde{\theta}_i} > 0$, then let $\hat{m}_i^3[\tilde{\theta}_i] = y[\tilde{\theta}_i]$.
  - ▷ If $\Psi_i^{1,\tilde{\theta}_i} = 0$, then let $\hat{m}_i^3[\tilde{\theta}_i] = m_i^3[\tilde{\theta}_i]$.

If $\Psi_i^{1,\tilde{\theta}_i} > 0$, then conditional on the event that all other individuals report a message profile in $\bigcup_{\theta''_{-i}} M_{-i}^1(\tilde{\theta}_i, \theta''_{-i})$, the expected payoff of individual i of payoff type θ_i when she plays $\hat{m}_i$ is

$$\begin{aligned}&\left(\frac{\hat{m}_i^2}{1+\hat{m}_i^2}\right) \sum_{\theta_{-i}, \theta'_{-i}} \psi_i^{1,\tilde{\theta}_i}(\theta_{-i}, \theta'_{-i}) u_i\big(y[\tilde{\theta}_i](\theta'_{-i}), (\theta_i, \theta_{-i})\big) \\ &+ \left(1 - \frac{\hat{m}_i^2}{1+\hat{m}_i^2}\right) \sum_{\theta_{-i}, \theta'_{-i}} \psi_i^{1,\tilde{\theta}_i}(\theta_{-i}, \theta'_{-i}) u_i\big(y_i^{\tilde{\theta}_i}(\theta'_{-i}), (\theta_i, \theta_{-i})\big).\end{aligned}$$

If $\hat{m}_i^2$ is large enough, then the above expression is greater than her expected payoff in (8) when she plays m_i. Since Θ_i is finite, we can find a sufficiently large $\hat{m}_i^2$ such that the above statement is true for all $\tilde{\theta}_i \in \Theta_i$ such that $\Psi_i^{1,\tilde{\theta}_i} > 0$. As $\sum_{\tilde{\theta}_i} \Psi_i^{1,\tilde{\theta}_i} = 1$ (because $\psi_i^\Gamma(\theta_{-i}, m_{-i}) > 0 \Rightarrow m_{-i} \in \mathcal{S}_{-i}^\infty(\theta_{-i}) \Rightarrow m_j^2 = 1, \forall j \neq i$), it follows that $\hat{m}_i$ is a better response for individual i of type θ_i against ψ_i^Γ, a contradiction. This completes the proof of Step 2. □

It follows from Steps 1 and 2 that $m \in \mathcal{S}^\infty(\theta) \Rightarrow g(m) = f(\theta)$.

Step 3: Define the message correspondence profile with payoff-type domain $\mathcal{S} = (\mathcal{S}_1, \ldots, \mathcal{S}_n)$ such that for all $i \in I$ and $\theta_i \in \Theta_i$,

$$\mathcal{S}_i(\theta_i) = \{(m_i^1, 1, m_i^3, m_i^4) : m_i^1[i] = \theta_i\}.$$

Then, we have $b^\Theta(\mathcal{S}) \geq \mathcal{S}$, which implies that $\mathcal{S} \leq \mathcal{S}^\infty$.

Proof. Pick any $i \in I$, $\theta_i \in \Theta_i$, and $m_i \in \mathcal{S}_i(\theta_i)$. Fix some $\theta_{-i} \in \Theta_{-i}$ and pick any $\tilde{m}_{-i} \in \mathcal{S}_{-i}(\theta_{-i})$ such that $\tilde{m}_j^1[i] = \theta_i$ and $\tilde{m}_j^1[j] = \theta_j$, for all $j \neq i$. Let the belief $\psi_i \in \Delta(\Theta_{-i} \times M_{-i})$ be such that $\psi_i(\theta_{-i}, \tilde{m}_{-i}) = 1$. When individual i of payoff type θ_i holds the belief ψ_i and plays m_i, then she expects the payoff of $u_i\big(f(\theta_i, \theta_{-i}), (\theta_i, \theta_{-i})\big)$. On the one hand, if she deviates to $\hat{m}_i$ such that $\hat{m}_i^1[i] = \theta_i'$ and $\hat{m}_i^2 = 1$, then she expects the payoff of $u_i\big(f(\theta_i', \theta_{-i}), (\theta_i, \theta_{-i})\big)$, which is not improving due to semi-strict EPIC. On the other hand, if she deviates to $\hat{m}_i$ such that $\hat{m}_i^2 > 1$, then she expects the payoff of

$$\left(\frac{\hat{m}_i^2}{1 + \hat{m}_i^2}\right) u_i\big(\hat{m}_i^3[\theta_i](\theta_{-i}), (\theta_i, \theta_{-i})\big) + \left(1 - \frac{\hat{m}_i^2}{1 + \hat{m}_i^2}\right) u_i\big(y_i^{\theta_i}(\theta_{-i}), (\theta_i, \theta_{-i})\big).$$

As $\hat{m}_i^3[\theta_i] \in Y_i^*[\theta_i]$, she cannot improve by any such deviation. Hence, $m_i \in b_i^\Theta(\mathcal{S})[\theta_i]$. This completes the proof of Step 3. □

Step 4: Condition (2) in Theorem 3.1 is satisfied by the constructed mechanism

Proof. Pick $i \in I$, $\theta_i \in \Theta_i$ and $z_i^1 \in Z_i^1$. For each $\theta_{-i} \in \Theta_{-i}$, pick some $\tilde{m}_{-i} \in M_{-i}$ such that $\tilde{m}_j^1[i] = \theta_i$, $\tilde{m}_j^1[j] = \theta_j$, and $\tilde{m}_j^2 = 1$ for all $j \neq i$. From Step 3, it follows that $\tilde{m}_{-i} \in \mathcal{S}_{-i}^\infty(\theta_{-i})$. Define the belief $\psi_i \in \Delta(\Theta_{-i} \times M_{-i})$ such that $\psi_i(\theta_{-i}, \tilde{m}_{-i}) = z_i^1(\theta_{-i})$ for all $\theta_{-i} \in \Theta_{-i}$.

By construction, $\psi_i(\theta_{-i}, m_{-i}) > 0 \Rightarrow m_{-i} \in \mathcal{S}_{-i}^\infty(\theta_{-i})$ and $\text{marg}_{\Theta_{-i}}\psi_i = z_i^1$. When individual i of payoff type θ_i holds the belief ψ_i and plays $m_i = (m_i^1, 1, m_i^3, m_i^4)$ such that $m_i^1[i] = \theta_i$, then she expects the payoff of $\sum_{\theta_{-i}} z_i^1(\theta_{-i}) u_i\big(f(\theta_i, \theta_{-i}), (\theta_i, \theta_{-i})\big)$. On the one hand, if she deviates to $\hat{m}_i$ such that $\hat{m}_i^1[i] = \theta_i'$ and $\hat{m}_i^2 = 1$, then she expects the payoff of $\sum_{\theta_{-i}} z_i^1(\theta_{-i}) u_i\big(f(\theta_i', \theta_{-i}), (\theta_i, \theta_{-i})\big)$, which is not improving due to semi-strict EPIC. On the other hand, if she deviates to $\hat{m}_i$ such that $\hat{m}_i^2 > 1$, then she expects the payoff of

$$\left(\frac{\hat{m}_i^2}{1 + \hat{m}_i^2}\right) \sum_{\theta_{-i}} z_i^1(\theta_{-i}) u_i\big(\hat{m}_i^3[\theta_i](\theta_{-i}), (\theta_i, \theta_{-i})\big) + \left(1 - \frac{\hat{m}_i^2}{1 + \hat{m}_i^2}\right) \sum_{\theta_{-i}} z_i^1(\theta_{-i}) u_i\big(y_i^{\theta_i}(\theta_{-i}), (\theta_i, \theta_{-i})\big).$$

As $\hat{m}_i^3[\theta_i] \in Y_i^*[\theta_i]$, she cannot improve by any such deviation. Hence,

$$\arg\max_{m_i' \in M_i} \sum_{\theta_{-i}, m_{-i}} \psi_i(\theta_{-i}, m_{-i}) u_i\big(g(m_i', m_{-i}), (\theta_i, \theta_{-i})\big) \neq \emptyset,$$

which completes the proof of Step 4. □

Steps 1 through 4 complete the proof of the theorem. □

Remark 5.6. Bergemann and Morris (2011) show that strict RM and their stronger version of conditional NTI are sufficient conditions for both rationalizable implementation and robust implementation in interim equilibria. As rationalizable implementation implies weak rationalizable implementation and the latter is equivalent to robust implementation in rationalizable strategies, we can conclude that strict RM and their stronger conditional NTI are sufficient conditions for robust implementation in rationalizable strategies. Theorem 5.5 weakens both these conditions, thus providing us with a stronger sufficiency result for robust implementation in rationalizable strategies.

6 Robust Implementation: Rationalizable Strategies versus Interim Equilibria

In this section, we compare robust implementation in rationalizable strategies with robust implementation in interim equilibria. *A priori*, robust implementation in rationalizable strategies appears to be neither stronger nor weaker than robust implementation in interim equilibria. On a given type space, any message profile that is supported in interim equilibrium at some type profile is also rationalizable at that type profile. Therefore, the nonemptiness requirement seems weaker whereas the uniqueness requirement seems stronger in robust implementation in rationalizable strategies than in robust implementation in interim equilibria.

It turns out that robust implementation in interim equilibria in fact implies robust implementation in rationalizable strategies by the same mechanism. The nested relationship between different implementation notions offers an indirect way to reach this conclusion: If an SCF is robustly implementable in interim equilibria by a mechanism, then it is rationalizably implementable by the same mechanism (Bergemann and Morris, 2011, Theorem 3). Since rationalizable implementation implies weak rationalizable implementation, it follows from our Theorem 3.1 that the SCF is also robustly implementable in rationalizable strategies by the same mechanism.

But the underlying reason why robust implementation in rationalizable strategies turns out to be weaker than robust implementation in interim equilibria is that, while the nonemptiness requirement in the former concept is weaker than that in the latter concept, the uniqueness requirements in the two concepts are in fact equivalent, as argued next.

Corollary 6.1. *Consider any mechanism* Γ. *There exists a type space* $\mathcal{T}$ *such that* $m \in \bigcup_{t \in T:\hat{\theta}(t)=\theta} B^{\infty}(t)$ *if and only if there exists a type space* $\mathcal{T}'$, *an interim equilibrium* σ *of the game* $(\mathcal{T}', \Gamma)$, *and a type profile* $t' \in T'$ *such that* $\sigma(m|t') > 0$ *and* $\hat{\theta}(t') = \theta$.

Proof. By Lemma 3.2, there exists a type space $\mathcal{T}$ such that $m \in \bigcup_{t \in T:\hat{\theta}(t)=\theta} B^{\infty}(t)$ if and only if $m \in \mathcal{S}^{\infty}(\theta)$. By Proposition 1 in Bergemann and Morris (2011), $m \in \mathcal{S}^{\infty}(\theta)$ if and only if there exists a type space $\mathcal{T}'$, an interim equilibrium σ of the game $(\mathcal{T}', \Gamma)$, and a type profile $t' \in T'$ such that $\sigma(m|t') > 0$ and $\hat{\theta}(t') = \theta$. □

Thus, if a message profile m is rationalizable at some type profile t in some type space $\mathcal{T}$ such that $\hat{\theta}(t) = \theta$, then that message profile is also supported in some interim equilibrium for some type-profile t' in some type space $\mathcal{T}'$ such that $\hat{\theta}(t') = \theta$, and vice versa. Hence, insisting on robust implementation "over all type spaces" makes the uniqueness requirement under the solution concept of interim equilibria equivalent to the uniqueness requirement under the solution concept of rationalizable strategies.

The rest of this section is organized as follows: In Section 6.1, we show by means of an example that robust implementation in rationalizable strategies is strictly more permissive than that in interim equilibria. Section 6.2 establishes the equivalence between robust implementation in rationalizable strategies and that in interim equilibria when we restrict our attention to finite mechanisms.

6.1 An Example

We now present an example with an SCF that is robustly implementable in rationalizable strategies but not in interim equilibria. We do so by exploiting the gap between strict RM and weak RM for non-responsive SCFs. (As already mentioned, strict RM and weak RM are equivalent for responsive SCFs.) To elaborate, Bergemann and Morris (2011) show that strict RM is a necessary condition for rationalizable implementation of any SCF. The non-responsive SCF in the example below fails to satisfy strict RM. Thus, the SCF is not rationalizably implementable, and hence not robustly implementable in interim equilibria. The SCF however satisfies weak RM and conditional NTI. Hence, the SCF is robustly implementable in rationalizable strategies.

Thus robust implementation in interim equilibria is strictly stronger than robust implementation in rationalizable strategies. In light of our above discussion on the equivalence of the uniqueness requirements under the two solution concepts, the explanation for this gap between robust implementation in interim equilibria and robust implementation in rationalizable strategies is that the nonemptiness requirement in the former concept is strictly stronger than the nonemptiness requirement in the latter concept. Any mechanism Γ, in particular the canonical mechanism constructed in the proof of Theorem 5.5, that robustly implements the SCF in the example in rationalizable strategies must fail the nonemptiness requirement for robust implementation in interim equilibria. That is, there must exist some type space in which the set of interim equilibria of the mechanism is empty.

Example 6.2. There are two players $i \in \{1, 2\}$. Player 1 has three payoff types: $\Theta_1 = \{\theta_1, \theta_1', \theta_1''\}$ and player 2 has two payoff types: $\Theta_2 = \{\theta_2, \theta_2'\}$. There are six pure alternatives: $A = \{a, b, c, d, z, z'\}$. The following tables list the payoffs of the two players:

a	θ_2	θ_2'
θ_1	$4, 4$	$4, 0$
θ_1'	$0, 0$	$4, 1$
θ_1''	$1, 1$	$4, 0$

b	θ_2	θ_2'
θ_1	$0, 0$	$3, 3$
θ_1'	$1, 1$	$2, 0$
θ_1''	$0, 0$	$2, 1$

c	θ_2	θ_2'
θ_1	$0, 0$	$3, 1$
θ_1'	$3, 3$	$3, 0$
θ_1''	$3, 3$	$3, 0$

d	θ_2	θ_2'
θ_1	$3, 4$	$2, 0$
θ_1'	$0, 0$	$3, 3$
θ_1''	$0, 0$	$3, 3$

z	θ_2	θ_2'
θ_1	$4, 1$	$2, 0$
θ_1'	$2, 2$	$5, 0$
θ_1''	$2, 2$	$2, 0$

z'	θ_2	θ_2'
θ_1	$4, 0$	$4, 1$
θ_1'	$2, 0$	$2, 2$
θ_1''	$2, 0$	$5, 0$

The SCF f selects the alternative which maximizes the aggregate payoff in each payoff state.

f	θ_2	θ_2'
θ_1	a	b
θ_1'	c	d
θ_1''	c	d

We first show that f fails strict RM.

Claim 6.3. *The SCF f violates strict RM.*

Proof. Consider the deception β such that

$$\beta_1(\theta_1) = \{\theta_1, \theta_1'\}, \quad \beta_1(\theta_1') = \{\theta_1'\}, \quad \beta_1(\theta_1'') = \{\theta_1''\},$$

and

$$\beta_2(\theta_2) = \{\theta_2, \theta_2'\}, \quad \beta_2(\theta_2') = \{\theta_2'\}.$$

Given this deception, there are exactly two tuples (i, θ_i, θ_i') such that $\theta_i' \in \beta_i(\theta_i)$ and $\theta_i' \not\sim_i^f \theta_i$: $(1, \theta_1, \theta_1')$ and $(2, \theta_2, \theta_2')$.

First, consider $(2, \theta_2, \theta_2^{'})$. Fix the degenerate belief $\psi_2 \in \Delta(\Theta_1 \times \Theta_1)$ such that $\psi_2(\theta_1, \theta_1^{'}) = 1$. Then, there does not exist any $y \in \bigcap_{\tilde{\theta}_2 \in \Theta_2} Y_2[\tilde{\theta}_2]$ such that

$$u_2\big(y(\theta_1^{'}), (\theta_1, \theta_2)\big) > u_2\big(f(\theta_1^{'}, \theta_2^{'}), (\theta_1, \theta_2)\big),$$

because $f(\theta_1^{'}, \theta_2^{'}) = d$ is one of the best alternatives for player 2 in the payoff state (θ_1, θ_2).

Second, consider $(1, \theta_1, \theta_1^{'})$. Fix the degenerate belief ψ_1 such that $\psi_1(\theta_2, \theta_2^{'}) = 1$. If there exists $y \in \bigcap_{\tilde{\theta}_1 \in \Theta_1} Y_1[\tilde{\theta}_1]$, then $y(\theta_2^{'})$ must satisfy the following equations

$$u_1\big(f(\theta_1^{'}, \theta_2^{'}), (\theta_1^{'}, \theta_2^{'})\big) \geq u_1\big(y(\theta_2^{'}), (\theta_1^{'}, \theta_2^{'})\big)$$
$$u_1\big(f(\theta_1^{''}, \theta_2^{'}), (\theta_1^{''}, \theta_2^{'})\big) \geq u_1\big(y(\theta_2^{'}), (\theta_1^{''}, \theta_2^{'})\big).$$

These two inequalities imply that

$$2y(\theta_2^{'})[z] + y(\theta_2^{'})[a] \leq y(\theta_2^{'})[z^{'}] + y(\theta_2^{'})[b] \quad \text{and} \quad 2y(\theta_2^{'})[z^{'}] + y(\theta_2^{'})[a] \leq y(\theta_2^{'})[z] + y(\theta_2^{'})[b],$$

where $y(\theta_2^{'})[x]$ is the probability of alternative x in the lottery $y(\theta_2^{'})$. Summing these two inequalities, we obtain $y(\theta_2^{'})[z] + y(\theta_2^{'})[z^{'}] + 2y(\theta_2^{'})[a] \leq 2y(\theta_2^{'})[b]$. In order to satisfy strict RM, we must satisfy the following inequality:

$$u_1\big(y(\theta_2^{'}), (\theta_1, \theta_2)\big) > u_1\big(f(\theta_1^{'}, \theta_2^{'}), (\theta_1, \theta_2)\big).$$

The above inequality is translated into $y(\theta_2^{'})[z] + y(\theta_2^{'})[z^{'}] + y(\theta_2^{'})[a] > 3y(\theta_2^{'})[b] + 3y(\theta_2^{'})[c]$. We then claim that this inequality is impossible to be satisfied. Plugging $y(\theta_2^{'})[z] + y(\theta_2^{'})[z^{'}] + 2y(\theta_2^{'})[a] \leq 2y(\theta_2^{'})[b]$ into $y(\theta_2^{'})[z] + y(\theta_2^{'})[z^{'}] + y(\theta_2^{'})[a] > 3y(\theta_2^{'})[b] + 3y(\theta_2^{'})[c]$, we obtain

$$-y(\theta_2^{'})[a] > y(\theta_2^{'})[b] + 3y(\theta_2^{'})[c].$$

However, this inequality is impossible because $y(\theta_2^{'})[a], y(\theta_2^{'})[b]$, and $y(\theta_2^{'})[c]$ all are nonnegative. We therefore conclude that the SCF f does not satisfy strict RM. $\square$

Next we argue that f satisfies weak RM.

Claim 6.4. *The SCF f satisfies weak RM.*

Proof. First, we consider any unacceptable deception β such that either $\theta_1^{'} \in \beta_1(\theta_1)$ or

$\theta_1^{''} \in \beta_1(\theta_1)$. Pick any belief $\psi_1 \in \Delta(\Theta_2 \times \Theta_2)$. Then

$$\begin{aligned}
&\sum_{\tilde{\theta}_2, \tilde{\theta}_2'} \psi_1(\tilde{\theta}_2, \tilde{\theta}_2') u_1\big(f(\theta_1^{'}, \tilde{\theta}_2^{'}), (\theta_1, \tilde{\theta}_2)\big) \\
= \; & \psi_1(\theta_2, \theta_2) u_1\big(f(\theta_1^{'}, \theta_2), (\theta_1, \theta_2)\big) + \psi_1(\theta_2, \theta_2^{'}) u_1\big(f(\theta_1^{'}, \theta_2^{'}), (\theta_1, \theta_2)\big) \\
& + \psi_1(\theta_2^{'}, \theta_2) u_1\big(f(\theta_1^{'}, \theta_2), (\theta_1, \theta_2^{'})\big) + \psi_1(\theta_2^{'}, \theta_2^{'}) u_1\big(f(\theta_1^{'}, \theta_2^{'}), (\theta_1, \theta_2^{'})\big) \\
= \; & 3\psi_1(\theta_2^{'}, \theta_2) + 3\psi_1(\theta_2, \theta_2^{'}) + 2\psi_1(\theta_2^{'}, \theta_2^{'}).
\end{aligned}$$

Similarly,

$$\sum_{\tilde{\theta}_2, \tilde{\theta}_2'} \psi_1(\tilde{\theta}_2, \tilde{\theta}_2') u_1\big(f(\theta_1^{''}, \tilde{\theta}_2^{'}), (\theta_1, \tilde{\theta}_2)\big) = 3\psi_1(\theta_2^{'}, \theta_2) + 3\psi_1(\theta_2, \theta_2^{'}) + 2\psi_1(\theta_2^{'}, \theta_2^{'}).$$

In what follows, we consider each possible case of $\tilde{\theta}_1 \in \{\theta_1, \theta_1^{'}, \theta_1^{''}\}$.

Case 1: $\tilde{\theta}_1 = \theta_1$.

Define $y : \Theta_2 \to \Delta(A)$ to be such that $y(\theta_2) = a$ and $y(\theta_2^{'}) = \frac{2}{3}z + \frac{1}{3}z^{'}$. It is straightforward to confirm that $y \in Y_1[\theta_1]$. Moreover,

$$\begin{aligned}
&\sum_{\tilde{\theta}_2, \tilde{\theta}_2'} \psi_1(\tilde{\theta}_2, \tilde{\theta}_2') u_1\big(y(\tilde{\theta}_2^{'}), (\theta_1, \tilde{\theta}_2)\big) \\
= \; & \psi_1(\theta_2, \theta_2) u_1\big(y(\theta_2), (\theta_1, \theta_2)\big) + \psi_1(\theta_2, \theta_2^{'}) u_1\big(y(\theta_2^{'}), (\theta_1, \theta_2)\big) \\
& + \psi_1(\theta_2^{'}, \theta_2) u_1\big(y(\theta_2), (\theta_1, \theta_2^{'})\big) + \psi_1(\theta_2^{'}, \theta_2^{'}) u_1\big(y(\theta_2^{'}), (\theta_1, \theta_2^{'})\big) \\
= \; & 4\psi_1(\theta_2, \theta_2) + 4\psi_1(\theta_2^{'}, \theta_2) + 4\psi_1(\theta_2, \theta_2^{'}) + \frac{8}{3}\psi_1(\theta_2^{'}, \theta_2^{'}) \\
> \; & 3\psi_1(\theta_2^{'}, \theta_2) + 3\psi_1(\theta_2, \theta_2^{'}) + 2\psi_1(\theta_2^{'}, \theta_2^{'}) \\
= \; & \sum_{\tilde{\theta}_2, \tilde{\theta}_2'} \psi_1(\tilde{\theta}_2, \tilde{\theta}_2') u_1\big(f(\theta_1^{'}, \tilde{\theta}_2^{'}), (\theta_1, \tilde{\theta}_2)\big).
\end{aligned}$$

Similarly,

$$\sum_{\tilde{\theta}_2, \tilde{\theta}_2'} \psi_1(\tilde{\theta}_2, \tilde{\theta}_2') u_1\big(y(\tilde{\theta}_2^{'}), (\theta_1, \tilde{\theta}_2)\big) > \sum_{\tilde{\theta}_2, \tilde{\theta}_2'} \psi_1(\tilde{\theta}_2, \tilde{\theta}_2') u_1\big(f(\theta_1^{''}, \tilde{\theta}_2^{'}), (\theta_1, \tilde{\theta}_2)\big).$$

Case 2: $\tilde{\theta}_1 = \theta_1^{'}$.

Define $y : \Theta_2 \to \Delta(A)$ to be such that $y(\theta_2) = a$ and $y(\theta_2^{'}) = z^{'}$. It is straightforward to confirm that $y \in Y_1[\theta_1^{'}]$. Moreover,

$$\begin{aligned}
& \sum_{\tilde{\theta}_2, \tilde{\theta}_2^{'}} \psi_1(\tilde{\theta}_2, \tilde{\theta}_2^{'}) u_1\big(y(\tilde{\theta}_2^{'}), (\theta_1, \tilde{\theta}_2)\big) \\
= \ & \psi_1(\theta_2, \theta_2) u_1\big(y(\theta_2), (\theta_1, \theta_2)\big) + \psi_1(\theta_2, \theta_2^{'}) u_1\big(y(\theta_2^{'}), (\theta_1, \theta_2)\big) \\
& + \psi_1(\theta_2^{'}, \theta_2) u_1\big(y(\theta_2), (\theta_1, \theta_2^{'})\big) + \psi_1(\theta_2^{'}, \theta_2^{'}) u_1\big(y(\theta_2^{'}), (\theta_1, \theta_2^{'})\big) \\
= \ & 4\psi_1(\theta_2, \theta_2) + 4\psi_1(\theta_2^{'}, \theta_2) + 4\psi_1(\theta_2, \theta_2^{'}) + 4\psi_1(\theta_2^{'}, \theta_2^{'}) \\
> \ & 3\psi_1(\theta_2^{'}, \theta_2) + 3\psi_1(\theta_2, \theta_2^{'}) + 2\psi_1(\theta_2^{'}, \theta_2^{'}) \\
= \ & \sum_{\tilde{\theta}_2, \tilde{\theta}_2^{'}} \psi_1(\tilde{\theta}_2, \tilde{\theta}_2^{'}) u_1\big(f(\theta_1^{'}, \tilde{\theta}_2^{'}), (\theta_1, \tilde{\theta}_2)\big).
\end{aligned}$$

Similarly,

$$\sum_{\tilde{\theta}_2, \tilde{\theta}_2^{'}} \psi_1(\tilde{\theta}_2, \tilde{\theta}_2^{'}) u_1\big(y(\tilde{\theta}_2^{'}), (\theta_1, \tilde{\theta}_2)\big) > \sum_{\tilde{\theta}_2, \tilde{\theta}_2^{'}} \psi_1(\tilde{\theta}_2, \tilde{\theta}_2^{'}) u_1\big(f(\theta_1^{''}, \tilde{\theta}_2^{'}), (\theta_1, \tilde{\theta}_2)\big).$$

Case 3: $\tilde{\theta}_1 = \theta_1^{''}$.

Define $y : \Theta_2 \to \Delta(A)$ to be such that $y(\theta_2) = a$ and $y(\theta_2^{'}) = \frac{1}{5}c + \frac{4}{5}z$. It is straightforward to confirm that $y \in Y_1[\theta_1^{''}]$. Moreover,

$$\begin{aligned}
& \sum_{\tilde{\theta}_2, \tilde{\theta}_2^{'}} \psi_1(\tilde{\theta}_2, \tilde{\theta}_2^{'}) u_1\big(y(\tilde{\theta}_2^{'}), (\theta_1, \tilde{\theta}_2)\big) \\
= \ & \psi_1(\theta_2, \theta_2) u_1\big(y(\theta_2), (\theta_1, \theta_2)\big) + \psi_1(\theta_2, \theta_2^{'}) u_1\big(y(\theta_2^{'}), (\theta_1, \theta_2)\big) \\
& + \psi_1(\theta_2^{'}, \theta_2) u_1\big(y(\theta_2), (\theta_1, \theta_2^{'})\big) + \psi_1(\theta_2^{'}, \theta_2^{'}) u_1\big(y(\theta_2^{'}), (\theta_1, \theta_2^{'})\big) \\
= \ & 4\psi_1(\theta_2, \theta_2) + 4\psi_1(\theta_2^{'}, \theta_2) + \frac{16}{5}\psi_1(\theta_2, \theta_2^{'}) + \frac{11}{5}\psi_1(\theta_2^{'}, \theta_2^{'}) \\
> \ & 3\psi_1(\theta_2^{'}, \theta_2) + 3\psi_1(\theta_2, \theta_2^{'}) + 2\psi_1(\theta_2^{'}, \theta_2^{'}) \\
= \ & \sum_{\tilde{\theta}_2, \tilde{\theta}_2^{'}} \psi_1(\tilde{\theta}_2, \tilde{\theta}_2^{'}) u_1\big(f(\theta_1^{'}, \tilde{\theta}_2^{'}), (\theta_1, \tilde{\theta}_2)\big).
\end{aligned}$$

Similarly,

$$\sum_{\tilde{\theta}_2, \tilde{\theta}_2^{'}} \psi_1(\tilde{\theta}_2, \tilde{\theta}_2^{'}) u_1\big(y(\tilde{\theta}_2^{'}), (\theta_1, \tilde{\theta}_2)\big) > \sum_{\tilde{\theta}_2, \tilde{\theta}_2^{'}} \psi_1(\tilde{\theta}_2, \tilde{\theta}_2^{'}) u_1\big(f(\theta_1^{''}, \tilde{\theta}_2^{'}), (\theta_1, \tilde{\theta}_2)\big).$$

It follows that any unacceptable deception β satisfying $\theta_1^{'} \in \beta_1(\theta_1)$ is weakly refutable using the tuple $(1, \theta_1, \theta_1^{'})$ whereas any unacceptable deception β satisfying $\theta_1^{''} \in \beta_1(\theta_1)$ is weakly refutable using the tuple $(1, \theta_1, \theta_1^{''})$.

Second, we consider any unacceptable deception β such that $\theta_2' \in \beta_2(\theta_2)$ and $\beta_1(\theta_1) = \{\theta_1\}$. Pick any belief $\psi_2 \in \Delta(\Theta_1 \times \Theta_1)$ such that $\psi_2(\tilde{\theta}_1, \tilde{\theta}_1') > 0 \Rightarrow \tilde{\theta}_1' \in \beta_1(\tilde{\theta}_1)$. Then we have $\psi_2(\theta_1, \theta_1') = \psi_2(\theta_1, \theta_1'') = 0$. Therefore,

$$\begin{aligned}
& \sum_{\tilde{\theta}_1, \tilde{\theta}_1'} \psi_2(\tilde{\theta}_1, \tilde{\theta}_1') u_2\big(f(\tilde{\theta}_1', \theta_2'), (\tilde{\theta}_1, \theta_2)\big) \\
= \; & \psi_2(\theta_1, \theta_1) u_2\big(f(\theta_1, \theta_2'), (\theta_1, \theta_2)\big) + \psi_2(\theta_1, \theta_1') u_2\big(f(\theta_1', \theta_2'), (\theta_1, \theta_2)\big) + \psi_2(\theta_1, \theta_1'') u_2\big(f(\theta_1'', \theta_2'), (\theta_1, \theta_2)\big) \\
& + \psi_2(\theta_1', \theta_1) u_2\big(f(\theta_1, \theta_2'), (\theta_1', \theta_2)\big) + \psi_2(\theta_1', \theta_1') u_2\big(f(\theta_1', \theta_2'), (\theta_1', \theta_2)\big) + \psi_2(\theta_1', \theta_1'') u_2\big(f(\theta_1'', \theta_2'), (\theta_1', \theta_2)\big) \\
& + \psi_2(\theta_1'', \theta_1) u_2\big(f(\theta_1, \theta_2'), (\theta_1'', \theta_2)\big) + \psi_2(\theta_1'', \theta_1') u_2\big(f(\theta_1', \theta_2'), (\theta_1'', \theta_2)\big) + \psi_2(\theta_1'', \theta_1'') u_2\big(f(\theta_1'', \theta_2'), (\theta_1'', \theta_2)\big) \\
= \; & \psi_2(\theta_1', \theta_1).
\end{aligned}$$

Define $y : \Theta_1 \to \Delta(A)$ to be such that $y(\theta_1) = y(\theta_1') = y(\theta_1'') = z$. It is straightforward to confirm that $y \in Y_2[\theta_2] \cap Y_2[\theta_2']$. Moreover, since $\psi_2(\theta_1, \theta_1') = \psi_2(\theta_1, \theta_1'') = 0$, we have

$$\begin{aligned}
& \sum_{\tilde{\theta}_1, \tilde{\theta}_1'} \psi_2(\tilde{\theta}_1, \tilde{\theta}_1') u_2\big(y(\tilde{\theta}_1'), (\tilde{\theta}_1, \theta_2)\big) \\
= \; & \psi_2(\theta_1, \theta_1) u_2\big(y(\theta_1), (\theta_1, \theta_2)\big) + \psi_2(\theta_1, \theta_1') u_2\big(y(\theta_1'), (\theta_1, \theta_2)\big) + \psi_2(\theta_1, \theta_1'') u_2\big(y(\theta_1''), (\theta_1, \theta_2)\big) \\
& + \psi_2(\theta_1', \theta_1) u_2\big(y(\theta_1), (\theta_1', \theta_2)\big) + \psi_2(\theta_1', \theta_1') u_2\big(y(\theta_1'), (\theta_1', \theta_2)\big) + \psi_2(\theta_1', \theta_1'') u_2\big(y(\theta_1''), (\theta_1', \theta_2)\big) \\
& + \psi_2(\theta_1'', \theta_1) u_2\big(y(\theta_1), (\theta_1'', \theta_2)\big) + \psi_2(\theta_1'', \theta_1') u_2\big(y(\theta_1'), (\theta_1'', \theta_2)\big) + \psi_2(\theta_1'', \theta_1'') u_2\big(y(\theta_1''), (\theta_1'', \theta_2)\big) \\
= \; & \psi_2(\theta_1, \theta_1) + 2\big(\psi_2(\theta_1', \theta_1) + \psi_2(\theta_1', \theta_1') + \psi_2(\theta_1', \theta_1'')\big) + 2\big(\psi_2(\theta_1'', \theta_1) + \psi_2(\theta_1'', \theta_1') + \psi_2(\theta_1'', \theta_1'')\big) \\
> \; & \psi_2(\theta_1', \theta_1) \\
= \; & \sum_{\tilde{\theta}_1, \tilde{\theta}_1'} \psi_2(\tilde{\theta}_1, \tilde{\theta}_1') u_2\big(f(\tilde{\theta}_1', \theta_2'), (\tilde{\theta}_1, \theta_2)\big).
\end{aligned}$$

It follows that any unacceptable deception β such that $\theta_2' \in \beta_2(\theta_2)$ and $\beta_1(\theta_1) = \{\theta_1\}$ is weakly refutable using the tuple $(2, \theta_2, \theta_2')$.

Third, we consider any unacceptable deception β such that $\theta_2 \in \beta_2(\theta_2')$ and $\beta_1(\theta_1) = \{\theta_1\}$. Pick any belief $\psi_2 \in \Delta(\Theta_1 \times \Theta_1)$ such that $\psi_2(\tilde{\theta}_1, \tilde{\theta}_1') > 0 \Rightarrow \tilde{\theta}_1' \in \beta_1(\tilde{\theta}_1)$. Then we have that

$\psi_2(\theta_1, \theta_1^{'}) = \psi_2(\theta_1, \theta_1^{''}) = 0$. Therefore,

$$\begin{aligned}
& \sum_{\tilde{\theta}_1, \tilde{\theta}_1^{'}} \psi_2(\tilde{\theta}_1, \tilde{\theta}_1^{'}) u_2\big(f(\tilde{\theta}_1^{'}, \theta_2), (\tilde{\theta}_1, \theta_2^{'})\big) \\
= \; & \psi_2(\theta_1, \theta_1) u_2\big(f(\theta_1, \theta_2), (\theta_1, \theta_2^{'})\big) + \psi_2(\theta_1, \theta_1^{'}) u_2\big(f(\theta_1^{'}, \theta_2), (\theta_1, \theta_2^{'})\big) + \psi_2(\theta_1, \theta_1^{''}) u_2\big(f(\theta_1^{''}, \theta_2), (\theta_1, \theta_2^{'})\big) \\
& + \psi_2(\theta_1^{'}, \theta_1) u_2\big(f(\theta_1, \theta_2), (\theta_1^{'}, \theta_2^{'})\big) + \psi_2(\theta_1^{'}, \theta_1^{'}) u_2\big(f(\theta_1^{'}, \theta_2), (\theta_1^{'}, \theta_2^{'})\big) + \psi_2(\theta_1^{'}, \theta_1^{''}) u_2\big(f(\theta_1^{''}, \theta_2), (\theta_1^{'}, \theta_2^{'})\big) \\
& + \psi_2(\theta_1^{''}, \theta_1) u_2\big(f(\theta_1, \theta_2), (\theta_1^{''}, \theta_2^{'})\big) + \psi_2(\theta_1^{''}, \theta_1^{'}) u_2\big(f(\theta_1^{'}, \theta_2), (\theta_1^{''}, \theta_2^{'})\big) + \psi_2(\theta_1^{''}, \theta_1^{''}) u_2\big(f(\theta_1^{''}, \theta_2), (\theta_1^{''}, \theta_2^{'})\big) \\
= \; & \psi_2(\theta_1^{'}, \theta_1).
\end{aligned}$$

Define $y : \Theta_1 \to \Delta(A)$ to be such that $y(\theta_1) = y(\theta_1^{'}) = y(\theta_1^{''}) = \frac{1}{4}b + \frac{3}{4}z^{'}$. It is straightforward to confirm that $y \in Y_2[\theta_2] \cap Y_2[\theta_2^{'}]$. Moreover, since $\psi_2(\theta_1, \theta_1^{'}) = \psi_2(\theta_1, \theta_1^{''}) = 0$, we have

$$\begin{aligned}
& \sum_{\tilde{\theta}_1, \tilde{\theta}_1^{'}} \psi_2(\tilde{\theta}_1, \tilde{\theta}_1^{'}) u_2\big(y(\tilde{\theta}_1^{'}), (\tilde{\theta}_1, \theta_2^{'})\big) \\
= \; & \psi_2(\theta_1, \theta_1) u_2\big(y(\theta_1), (\theta_1, \theta_2^{'})\big) + \psi_2(\theta_1, \theta_1^{'}) u_2\big(y(\theta_1^{'}), (\theta_1, \theta_2^{'})\big) + \psi_2(\theta_1, \theta_1^{''}) u_2\big(y(\theta_1^{''}), (\theta_1, \theta_2^{'})\big) \\
& + \psi_2(\theta_1^{'}, \theta_1) u_2\big(y(\theta_1), (\theta_1^{'}, \theta_2^{'})\big) + \psi_2(\theta_1^{'}, \theta_1^{'}) u_2\big(y(\theta_1^{'}), (\theta_1^{'}, \theta_2^{'})\big) + \psi_2(\theta_1^{'}, \theta_1^{''}) u_2\big(y(\theta_1^{''}), (\theta_1^{'}, \theta_2^{'})\big) \\
& + \psi_2(\theta_1^{''}, \theta_1) u_2\big(y(\theta_1), (\theta_1^{''}, \theta_2^{'})\big) + \psi_2(\theta_1^{''}, \theta_1^{'}) u_2\big(y(\theta_1^{'}), (\theta_1^{''}, \theta_2^{'})\big) + \psi_2(\theta_1^{''}, \theta_1^{''}) u_2\big(y(\theta_1^{''}), (\theta_1^{''}, \theta_2^{'})\big) \\
= \; & \frac{3}{2}\psi_2(\theta_1, \theta_1) + \frac{3}{2}\big(\psi_2(\theta_1^{'}, \theta_1) + \psi_2(\theta_1^{'}, \theta_1^{'}) + \psi_2(\theta_1^{'}, \theta_1^{''})\big) + \frac{1}{4}\big(\psi_2(\theta_1^{''}, \theta_1) + \psi_2(\theta_1^{''}, \theta_1^{'}) + \psi_2(\theta_1^{''}, \theta_1^{''})\big) \\
> \; & \psi_2(\theta_1^{'}, \theta_1) \\
= \; & \sum_{\tilde{\theta}_1, \tilde{\theta}_1^{'}} \psi_2(\tilde{\theta}_1, \tilde{\theta}_1^{'}) u_2\big(f(\tilde{\theta}_1^{'}, \theta_2), (\tilde{\theta}_1, \theta_2^{'})\big).
\end{aligned}$$

It follows that any unacceptable deception β such that $\theta_2 \in \beta_2(\theta_2^{'})$ and $\beta_1(\theta_1) = \{\theta_1\}$ is weakly refutable using the tuple $(2, \theta_2^{'}, \theta_2)$.

Fourth, we consider any unacceptable deception such that $\beta_1(\theta_1) = \{\theta_1\}$, $\beta_2(\theta_2) = \{\theta_2\}$, and $\beta_2(\theta_2^{'}) = \{\theta_2^{'}\}$. Such a deception involves either $\theta_1 \in \beta_1(\theta_1^{'})$ or $\theta_1 \in \beta_1(\theta_1^{''})$. Then the fact that f satisfies semi-strict EPIC implies that β is weakly refutable. We show this formally for the case when $\theta_1 \in \beta_1(\theta_1^{'})$ as the argument for the case when $\theta_1 \in \beta_1(\theta_1^{''})$ is similar. So suppose $\theta_1 \in \beta_1(\theta_1^{'})$. Pick any belief $\psi_1 \in \Delta(\Theta_2 \times \Theta_2)$ such that $\psi_1(\tilde{\theta}_2, \tilde{\theta}_2^{'}) > 0 \Rightarrow \tilde{\theta}_2^{'} \in \beta_2(\tilde{\theta}_2)$. Then we have that $\psi_1(\theta_2, \theta_2^{'}) = \psi_1(\theta_2^{'}, \theta_2) = 0$. Therefore,

$$\sum_{\tilde{\theta}_2, \tilde{\theta}_2^{'}} \psi_1(\tilde{\theta}_2, \tilde{\theta}_2^{'}) u_1\big(f(\theta_1, \tilde{\theta}_2^{'}), (\theta_1^{'}, \tilde{\theta}_2)\big) = 2\psi_1(\theta_2^{'}, \theta_2^{'})$$

Define $y : \Theta_2 \to \Delta(A)$ to be such that $y(\theta_2) = f(\theta_1^{'}, \theta_2) = c$ and $y(\theta_2^{'}) = f(\theta_1^{'}, \theta_2^{'}) = d$. It is straightforward to confirm that $y \in Y_1[\theta_1] \cap Y_1[\theta_1^{'}] \cap Y_1[\theta_1^{''}]$. Moreover, since $\psi_1(\theta_2, \theta_2^{'}) = \psi_1(\theta_2^{'}, \theta_2) = 0$,

$$\sum_{\tilde{\theta}_2, \tilde{\theta}_2^{'}} \psi_1(\tilde{\theta}_2, \tilde{\theta}_2^{'}) u_1\big(y(\tilde{\theta}_2^{'}), (\theta_1^{'}, \tilde{\theta}_2)\big) = 3\psi_1(\theta_2, \theta_2) + 3\psi_1(\theta_2^{'}, \theta_2^{'}) > \sum_{\tilde{\theta}_2, \tilde{\theta}_2^{'}} \psi_1(\tilde{\theta}_2, \tilde{\theta}_2^{'}) u_1\big(f(\theta_1, \tilde{\theta}_2^{'}), (\theta_1^{'}, \tilde{\theta}_2)\big).$$

It follows that the deception β is weakly refutable using the tuple $(1, \theta_1^{'}, \theta_1)$.

We thus conclude that every unacceptable deception is weakly refutable, and hence f satisfies weak RM. □

We now check that the SCF f satisfies conditional NTI.

Claim 6.5. *The SCF f satisfies conditional NTI.*

Proof. First, we consider player 1 of payoff type θ_1. Let $y : \Theta_2 \to \Delta(A)$ be such that $y(\theta_2) = a$ and $y(\theta_2^{'}) = z$. Also, let $y^{'} : \Theta_2 \to \Delta(A)$ be such that $y^{'}(\theta_2) = b$ and $y^{'}(\theta_2^{'}) = d$. It is straightforward to confirm that $y, y^{'} \in Y_1^w[\theta_1]$. Now,

$$\begin{aligned} & \sum_{\tilde{\theta}_2, \tilde{\theta}_2^{'}} \psi_1(\tilde{\theta}_2, \tilde{\theta}_2^{'}) u_1\big(y(\tilde{\theta}_2^{'}), (\theta_1, \tilde{\theta}_2)\big) \\ = \; & \psi_1(\theta_2, \theta_2) u_1\big(y(\theta_2), (\theta_1, \theta_2)\big) + \psi_1(\theta_2, \theta_2^{'}) u_1\big(y(\theta_2^{'}), (\theta_1, \theta_2)\big) \\ & + \psi_1(\theta_2^{'}, \theta_2) u_1\big(y(\theta_2), (\theta_1, \theta_2^{'})\big) + \psi_1(\theta_2^{'}, \theta_2^{'}) u_1\big(y(\theta_2^{'}), (\theta_1, \theta_2^{'})\big) \\ = \; & 4\psi_1(\theta_2, \theta_2) + 4\psi_1(\theta_2, \theta_2^{'}) + 4\psi_1(\theta_2^{'}, \theta_2) + 2\psi_i(\theta_2^{'}, \theta_2^{'}). \end{aligned}$$

whereas

$$\begin{aligned} & \sum_{\tilde{\theta}_2, \tilde{\theta}_2^{'}} \psi_1(\tilde{\theta}_2, \tilde{\theta}_2^{'}) u_1\big(y^{'}(\tilde{\theta}_2^{'}), (\theta_1, \tilde{\theta}_2)\big) \\ = \; & \psi_1(\theta_2, \theta_2) u_1\big(y^{'}(\theta_2), (\theta_1, \theta_2)\big) + \psi_1(\theta_2, \theta_2^{'}) u_1\big(y^{'}(\theta_2^{'}), (\theta_1, \theta_2)\big) \\ & + \psi_1(\theta_2^{'}, \theta_2) u_1\big(y^{'}(\theta_2), (\theta_1, \theta_2^{'})\big) + \psi_1(\theta_2^{'}, \theta_2^{'}) u_1\big(y^{'}(\theta_2^{'}), (\theta_1, \theta_2^{'})\big) \\ = \; & 3\psi_1(\theta_2, \theta_2^{'}) + 3\psi_1(\theta_2^{'}, \theta_2) + 2\psi_i(\theta_2^{'}, \theta_2^{'}). \end{aligned}$$

We therefore have that

$$\sum_{\tilde{\theta}_2, \tilde{\theta}_2^{'}} \psi_1(\tilde{\theta}_2, \tilde{\theta}_2^{'}) u_1\big(y(\tilde{\theta}_2^{'}), (\theta_1, \tilde{\theta}_2)\big) = \sum_{\tilde{\theta}_2, \tilde{\theta}_2^{'}} \psi_1(\tilde{\theta}_2, \tilde{\theta}_2^{'}) u_1\big(y^{'}(\tilde{\theta}_2^{'}), (\theta_1, \tilde{\theta}_2)\big) \Leftrightarrow \psi_1(\theta_2^{'}, \theta_2^{'}) = 1.$$

Thus, for all $\psi_1 \in \Delta(\Theta_2 \times \Theta_2)$ such that $\psi_1(\theta_2^{'}, \theta_2^{'}) < 1$, we have found $y, y^{'} \in Y_1^w[\theta_1]$ that satisfy the requirement for conditional NTI. If ψ_1 is such that $\psi_1(\theta_2^{'}, \theta_2^{'}) = 1$, then we define $y : \Theta_2 \to \Delta(A)$ such that $y(\theta_2) = y(\theta_2^{'}) = b$ and $y^{'} : \Theta_2 \to \Delta(A)$ such that $y^{'}(\theta_2) = y^{'}(\theta_2^{'}) = d$. It is straightforward to confirm that $y, y^{'} \in Y_1^w[\theta_1]$. Since $\psi_1(\theta_2^{'}, \theta_2^{'}) = 1$, $u_1(y(\theta_2^{'}), (\theta_1, \theta_2^{'})) = u_1(b, (\theta_1, \theta_2^{'})) = 3$ and $u_1(y^{'}(\theta_2^{'}), (\theta_1, \theta_2^{'})) = u_1(d, (\theta_1, \theta_2^{'})) = 2$, we obtain

$$\sum_{\tilde{\theta}_2, \tilde{\theta}_2^{'}} \psi_1(\tilde{\theta}_2, \tilde{\theta}_2^{'}) u_1\big(y(\tilde{\theta}_2^{'}), (\theta_1, \tilde{\theta}_2)\big) > \sum_{\tilde{\theta}_2, \tilde{\theta}_2^{'}} \psi_1(\tilde{\theta}_2, \tilde{\theta}_2^{'}) u_1\big(y^{'}(\tilde{\theta}_2^{'}), (\theta_1, \tilde{\theta}_2)\big).$$

Thus, if ψ_1 is such that $\psi_1(\theta_2^{'}, \theta_2^{'}) = 1$, then too we satisfy the requirement for conditional NTI.

Second, we consider player 1 of payoff type $\theta_1^{'}$. Then we define $y : \Theta_2 \to \Delta(A)$ such that $y(\theta_2) = y(\theta_2^{'}) = c$ and $y^{'} : \Theta_2 \to \Delta(A)$ such that $y^{'}(\theta_2) = y^{'}(\theta_2^{'}) = b$. It is straightforward to confirm that $y, y^{'} \in Y_1^w[\theta_1^{'}]$. Fix $\psi_1 \in \Delta(\Theta_2 \times \Theta_2)$. Now,

$$\begin{aligned}
& \sum_{\tilde{\theta}_2, \tilde{\theta}_2^{'}} \psi_1(\tilde{\theta}_2, \tilde{\theta}_2^{'}) u_1\big(y(\tilde{\theta}_2^{'}), (\theta_1^{'}, \tilde{\theta}_2)\big) \\
= \;& \psi_1(\theta_2, \theta_2) u_1\big(y(\theta_2), (\theta_1^{'}, \theta_2)\big) + \psi_1(\theta_2, \theta_2^{'}) u_1\big(y(\theta_2^{'}), (\theta_1^{'}, \theta_2)\big) \\
& + \psi_1(\theta_2^{'}, \theta_2) u_1\big(y(\theta_2), (\theta_1^{'}, \theta_2^{'})\big) + \psi_1(\theta_2^{'}, \theta_2^{'}) u_1\big(y(\theta_2^{'}), (\theta_1^{'}, \theta_2^{'})\big) \\
= \;& 3\psi_1(\theta_2, \theta_2) + 3\psi_1(\theta_2, \theta_2^{'}) + 3\psi_1(\theta_2^{'}, \theta_2) + 3\psi_i(\theta_2^{'}, \theta_2^{'})
\end{aligned}$$

whereas

$$\begin{aligned}
& \sum_{\tilde{\theta}_2, \tilde{\theta}_2^{'}} \psi_1(\tilde{\theta}_2, \tilde{\theta}_2^{'}) u_1\big(y^{'}(\tilde{\theta}_2^{'}), (\theta_1^{'}, \tilde{\theta}_2)\big) \\
= \;& \psi_1(\theta_2, \theta_2) u_1\big(y^{'}(\theta_2), (\theta_1^{'}, \theta_2)\big) + \psi_1(\theta_2, \theta_2^{'}) u_1\big(y^{'}(\theta_2^{'}), (\theta_1^{'}, \theta_2)\big) \\
& + \psi_1(\theta_2^{'}, \theta_2) u_1\big(y^{'}(\theta_2), (\theta_1^{'}, \theta_2^{'})\big) + \psi_1(\theta_2^{'}, \theta_2^{'}) u_1\big(y^{'}(\theta_2^{'}), (\theta_1^{'}, \theta_2^{'})\big) \\
= \;& \psi_1(\theta_2, \theta_2) + \psi_1(\theta_2, \theta_2^{'}) + 2\psi_1(\theta_2^{'}, \theta_2) + 2\psi_i(\theta_2^{'}, \theta_2^{'}).
\end{aligned}$$

This implies that for any $\psi_1 \in \Delta(\Theta_2 \times \Theta_2)$,

$$\sum_{\tilde{\theta}_2, \tilde{\theta}_2^{'}} \psi_1(\tilde{\theta}_2, \tilde{\theta}_2^{'}) u_1\big(y(\tilde{\theta}_2^{'}), (\theta_1^{'}, \tilde{\theta}_2)\big) > \sum_{\tilde{\theta}_2, \tilde{\theta}_2^{'}} \psi_1(\tilde{\theta}_2, \tilde{\theta}_2^{'}) u_1\big(y^{'}(\tilde{\theta}_2^{'}), (\theta_1^{'}, \tilde{\theta}_2)\big).$$

Thus, we satisfy the requirement for conditional NTI.

Third, we consider player 1 of payoff type $\theta_1^{''}$. Once again, we define $y : \Theta_2 \to \Delta(A)$ such that $y(\theta_2) = y(\theta_2^{'}) = c$ and $y^{'} : \Theta_2 \to \Delta(A)$ such that $y^{'}(\theta_2) = y^{'}(\theta_2^{'}) = b$. It is straightforward to confirm that $y, y^{'} \in Y_1^w[\theta_1^{''}]$. Fix $\psi_1 \in \Delta(\Theta_2 \times \Theta_2)$. Now

$$\begin{aligned}
&\sum_{\tilde{\theta}_2,\tilde{\theta}_2'} \psi_1(\tilde{\theta}_2, \tilde{\theta}_2')u_1\big(y(\tilde{\theta}_2'), (\theta_1^{''}, \tilde{\theta}_2)\big) \\
= \;& \psi_1(\theta_2, \theta_2)u_1\big(y(\theta_2), (\theta_1^{''}, \theta_2)\big) + \psi_1(\theta_2, \theta_2^{'})u_1\big(y(\theta_2^{'}), (\theta_1^{''}, \theta_2)\big) \\
& +\psi_1(\theta_2^{'}, \theta_2)u_1\big(y(\theta_2), (\theta_1^{''}, \theta_2^{'})\big) + \psi_1(\theta_2^{'}, \theta_2^{'})u_1\big(y(\theta_2^{'}), (\theta_1^{''}, \theta_2^{'})\big) \\
= \;& 3\psi_1(\theta_2, \theta_2) + 3\psi_1(\theta_2, \theta_2^{'}) + 3\psi_1(\theta_2^{'}, \theta_2) + 3\psi_i(\theta_2^{'}, \theta_2^{'})
\end{aligned}$$

whereas

$$\begin{aligned}
&\sum_{\tilde{\theta}_2,\tilde{\theta}_2'} \psi_1(\tilde{\theta}_2, \tilde{\theta}_2')u_1\big(y^{'}(\tilde{\theta}_2'), (\theta_1^{''}, \tilde{\theta}_2)\big) \\
= \;& \psi_1(\theta_2, \theta_2)u_1\big(y^{'}(\theta_2), (\theta_1^{''}, \theta_2)\big) + \psi_1(\theta_2, \theta_2^{'})u_1\big(y^{'}(\theta_2^{'}), (\theta_1^{''}, \theta_2)\big) \\
& +\psi_1(\theta_2^{'}, \theta_2)u_1\big(y^{'}(\theta_2), (\theta_1^{''}, \theta_2^{'})\big) + \psi_1(\theta_2^{'}, \theta_2^{'})u_1\big(y^{'}(\theta_2^{'}), (\theta_1^{''}, \theta_2^{'})\big) \\
= \;& 2\psi_1(\theta_2^{'}, \theta_2) + 2\psi_i(\theta_2^{'}, \theta_2^{'}).
\end{aligned}$$

This implies that for any $\psi_1 \in \Delta(\Theta_2 \times \Theta_2)$,

$$\sum_{\tilde{\theta}_2,\tilde{\theta}_2'} \psi_1(\tilde{\theta}_2, \tilde{\theta}_2')u_1\big(y(\tilde{\theta}_2'), (\theta_1^{''}, \tilde{\theta}_2)\big) > \sum_{\tilde{\theta}_2,\tilde{\theta}_2'} \psi_1(\tilde{\theta}_2, \tilde{\theta}_2')u_1\big(y^{'}(\tilde{\theta}_2'), (\theta_1^{''}, \tilde{\theta}_2)\big).$$

Thus, we satisfy the requirement for conditional NTI.

Fourth, we consider player 2 of payoff type θ_2. Then we define $y : \Theta_1 \to \Delta(A)$ such that $y(\theta_1) = y(\theta_1^{'}) = y(\theta_1^{''}) = \frac{1}{2}a + \frac{1}{2}c$ and $y^{'} : \Theta_1 \to \Delta(A)$ such that $y^{'}(\theta_1) = y^{'}(\theta_1^{'}) = y(\theta_1^{''}) = b$. It is straightforward to confirm that $y, y^{'} \in Y_2^w[\theta_2]$. Fix $\psi_2 \in \Delta(\Theta_2 \times \Theta_2)$. Now

$$\begin{aligned}
&\sum_{\tilde{\theta}_1,\tilde{\theta}_1'} \psi_2(\tilde{\theta}_1, \tilde{\theta}_1')u_2(y(\tilde{\theta}_1'), (\tilde{\theta}_1, \theta_2)) \\
= \;& \psi_2(\theta_1, \theta_1)u_2\big(y(\theta_1), (\theta_1, \theta_2)\big) + \psi_2(\theta_1, \theta_1^{'})u_2\big(y(\theta_1^{'}), (\theta_1, \theta_2)\big) + \psi_2(\theta_1, \theta_1^{''})u_2(y(\theta_1^{''}), (\theta_1, \theta_2)) \\
& +\psi_2(\theta_1^{'}, \theta_1)u_2\big(y(\theta_1), (\theta_1^{'}, \theta_2)\big) + \psi_2(\theta_1^{'}, \theta_1^{'})u_2\big(y(\theta_1^{'}), (\theta_1^{'}, \theta_2)\big) + \psi_2(\theta_1^{'}, \theta_1^{''})u_2(y(\theta_1^{''}), (\theta_1^{'}, \theta_2)) \\
& +\psi_2(\theta_1^{''}, \theta_1)u_2\big(y(\theta_1), (\theta_1^{''}, \theta_2)\big) + \psi_2(\theta_1^{''}, \theta_1^{'})u_2\big(y(\theta_1^{'}), (\theta_1^{''}, \theta_2)\big) + \psi_2(\theta_1^{''}, \theta_1^{''})u_2(y(\theta_1^{''}), (\theta_1^{''}, \theta_2)) \\
= \;& 2(\psi_1(\theta_1, \theta_1) + \psi_1(\theta_1, \theta_1^{'}) + \psi_1(\theta_1, \theta_1^{''})) + \frac{3}{2}(\psi_2(\theta_1^{'}, \theta_1) + \psi_2(\theta_1^{'}, \theta_1^{'}) + \psi_2(\theta_1^{'}, \theta_1^{''})) \\
& +2(\psi_2(\theta_1^{''}, \theta_1) + \psi_2(\theta_1^{''}, \theta_1^{'}) + \psi_2(\theta_1^{''}, \theta_1^{''}))
\end{aligned}$$

whereas

$$
\begin{aligned}
&\sum_{\tilde{\theta}_1,\tilde{\theta}_1'} \psi_2(\tilde{\theta}_1,\tilde{\theta}_1')u_2(y'(\tilde{\theta}_1'),(\tilde{\theta}_1,\theta_2)) \\
= \;& \psi_2(\theta_1,\theta_1)u_2\big(y'(\theta_1),(\theta_1,\theta_2)\big) + \psi_2(\theta_1,\theta_1')u_2\big(y'(\theta_1'),(\theta_1,\theta_2)\big) + \psi_2(\theta_1,\theta_1'')u_2(y'(\theta_1''),(\theta_1,\theta_2)) \\
& + \psi_2(\theta_1',\theta_1)u_2\big(y'(\theta_1),(\theta_1',\theta_2)\big) + \psi_2(\theta_1',\theta_1')u_2\big(y'(\theta_1'),(\theta_1',\theta_2)\big) + \psi_2(\theta_1',\theta_1'')u_2(y'(\theta_1''),(\theta_1',\theta_2)) \\
& + \psi_2(\theta_1'',\theta_1)u_2\big(y'(\theta_1),(\theta_1'',\theta_2)\big) + \psi_2(\theta_1'',\theta_1')u_2\big(y'(\theta_1'),(\theta_1'',\theta_2)\big) + \psi_2(\theta_1'',\theta_1'')u_2(y'(\theta_1''),(\theta_1'',\theta_2)) \\
= \;& \psi_2(\theta_1',\theta_1) + \psi_2(\theta_1',\theta_1') + \psi_2(\theta_1',\theta_1'').
\end{aligned}
$$

This implies that for any $\psi_2 \in \Delta(\Theta_1 \times \Theta_1)$,

$$
\sum_{\tilde{\theta}_1,\tilde{\theta}_1'} \psi_2(\tilde{\theta}_1,\tilde{\theta}_1')u_2(y(\tilde{\theta}_1'),(\tilde{\theta}_1,\theta_2)) > \sum_{\tilde{\theta}_1,\tilde{\theta}_1'} \psi_2(\tilde{\theta}_1,\tilde{\theta}_1')u_2(y'(\tilde{\theta}_1'),(\tilde{\theta}_1,\theta_2)).
$$

Thus, we satisfy the requirement for conditional NTI.

Finally, we consider player 2 of payoff type θ_2'. Then we define $y : \Theta_1 \to \Delta(A)$ such that $y(\theta_1) = y(\theta_1') = y(\theta_1'') = \frac{1}{2}b + \frac{1}{2}d$ and $y' : \Theta_1 \to \Delta(A)$ such that $y'(\theta_1) = y'(\theta_1') = y(\theta_1'') = c$. It is straightforward to confirm that $y, y' \in Y_2^w[\theta_2']$. Fix $\psi_2 \in \Delta(\Theta_1 \times \Theta_1)$. Then

$$
\begin{aligned}
&\sum_{\tilde{\theta}_1,\tilde{\theta}_1'} \psi_2(\tilde{\theta}_1,\tilde{\theta}_1')u_2(y(\tilde{\theta}_1'),(\tilde{\theta}_1,\theta_2')) \\
= \;& \psi_2(\theta_1,\theta_1)u_2\big(y(\theta_1),(\theta_1,\theta_2')\big) + \psi_2(\theta_1,\theta_1')u_2\big(y(\theta_1'),(\theta_1,\theta_2')\big) + \psi_2(\theta_1,\theta_1'')u_2(y(\theta_1''),(\theta_1,\theta_2')) \\
& + \psi_2(\theta_1',\theta_1)u_2\big(y(\theta_1),(\theta_1',\theta_2')\big) + \psi_2(\theta_1',\theta_1')u_2\big(y(\theta_1'),(\theta_1',\theta_2')\big) + \psi_2(\theta_1',\theta_1'')u_2(y(\theta_1''),(\theta_1',\theta_2')) \\
& + \psi_2(\theta_1'',\theta_1)u_2\big(y(\theta_1),(\theta_1'',\theta_2')\big) + \psi_2(\theta_1'',\theta_1')u_2\big(y(\theta_1'),(\theta_1'',\theta_2')\big) + \psi_2(\theta_1'',\theta_1'')u_2(y(\theta_1''),(\theta_1'',\theta_2')) \\
= \;& \frac{3}{2}(\psi_1(\theta_1,\theta_1) + \psi_1(\theta_1,\theta_1') + \psi_1(\theta_1,\theta_1'')) + \frac{3}{2}(\psi_2(\theta_1',\theta_1) + \psi_2(\theta_1',\theta_1') + \psi_2(\theta_1',\theta_1'')) \\
& + 2(\psi_2(\theta_1'',\theta_1) + \psi_2(\theta_1'',\theta_1') + \psi_2(\theta_1'',\theta_1''))
\end{aligned}
$$

whereas

$$
\begin{aligned}
&\sum_{\tilde{\theta}_1,\tilde{\theta}_1'} \psi_2(\tilde{\theta}_1,\tilde{\theta}_1')u_2(y'(\tilde{\theta}_1'),(\tilde{\theta}_1,\theta_2')) \\
= \;& \psi_2(\theta_1,\theta_1)u_2\big(y'(\theta_1),(\theta_1,\theta_2')\big) + \psi_2(\theta_1,\theta_1')u_2\big(y'(\theta_1'),(\theta_1,\theta_2')\big) + \psi_2(\theta_1,\theta_1'')u_2(y'(\theta_1''),(\theta_1,\theta_2')) \\
& + \psi_2(\theta_1',\theta_1)u_2\big(y'(\theta_1),(\theta_1',\theta_2')\big) + \psi_2(\theta_1',\theta_1')u_2\big(y'(\theta_1'),(\theta_1',\theta_2')\big) + \psi_2(\theta_1',\theta_1'')u_2(y'(\theta_1''),(\theta_1',\theta_2')) \\
& + \psi_2(\theta_1'',\theta_1)u_2\big(y'(\theta_1),(\theta_1'',\theta_2')\big) + \psi_2(\theta_1'',\theta_1')u_2\big(y'(\theta_1'),(\theta_1'',\theta_2')\big) + \psi_2(\theta_1'',\theta_1'')u_2(y'(\theta_1''),(\theta_1'',\theta_2')) \\
= \;& \psi_2(\theta_1,\theta_1) + \psi_2(\theta_1,\theta_1') + \psi_2(\theta_1,\theta_1'').
\end{aligned}
$$

This implies that for any $\psi_2 \in \Delta(\Theta_1 \times \Theta_1)$,

$$\sum_{\tilde{\theta}_1, \tilde{\theta}_1'} \psi_2(\tilde{\theta}_1, \tilde{\theta}_1') u_2(y(\tilde{\theta}_1'), (\tilde{\theta}_1, \theta_2')) > \sum_{\tilde{\theta}_1, \tilde{\theta}_1'} \psi_2(\tilde{\theta}_1, \tilde{\theta}_1') u_2(y'(\tilde{\theta}_1'), (\tilde{\theta}_1, \theta_2')).$$

Thus, we satisfy the requirement for conditional NTI.

We therefore conclude that f satisfies conditional NTI. □

6.2 Finite Mechanisms

The SCF in the above example satisfies weak RM and conditional NTI, and hence is robustly implementable in rationalizable strategies using the canonical mechanism constructed in Theorem 5.5. The canonical mechanism, however, is countably infinite. If we restrict attention to finite mechanisms, then the set of interim equilibria in every countable type space is nonempty. Hence, robust implementation in interim equilibria becomes equivalent to robust implementation in rationalizable strategies.

Proposition 6.6. *The SCF f is robustly implementable in rationalizable strategies by a finite mechanism Γ if and only if f is robustly implementable in interim equilibria by the same mechanism Γ.*

Proof. We have already argued that if f is robustly implementable in interim equilibria by some mechanism, then it is robustly implementable in rationalizable strategies by the same mechanism. To argue the converse, suppose f is robustly implementable in rationalizable strategies by a finite mechanism Γ. Consider any type space $\mathcal{T}$. First, because the mechanism Γ is finite and T_i is countable for all i, there exists an interim equilibrium for the game $(\mathcal{T}, \Gamma)$ on every type space $\mathcal{T}$.[7] Second, suppose σ is an interim equilibrium of the game $(\mathcal{T}, \Gamma)$. If $\sigma(t)[m] > 0$ such that $\hat{\theta}(t) = \theta$, then by Corollary 6.1 there exists a type space $\mathcal{T}'$ such that $m \in \bigcup_{t' \in T' : \hat{\theta}(t') = \theta} B^\infty(t')$. Since Γ robustly implements f in rationalizable strategies, $g(m) = f(\theta)$. □

Bergemann and Morris (2011) show that an additional "robust measurability" condition is necessary for robust implementation in interim equilibria using finite mechanisms. It thus follows that robust measurability is also necessary for robust implementation in rationalizable strategies using finite mechanisms. Robust measurability is generally not related to weak RM.[8] It is therefore an additional restriction on robust implementation in rationalizable strategies when using finite mechanisms. However, for the class of "single crossing

[7] For a proof of this statement, see Footnote 14 in Bergemann et al. (2017).

[8] We can show this using Examples 1 and 2 in Section 8.3 in Bergemann and Morris (2007).

aggregator" environments, robust measurability is equivalent to strict RM (see section 5 in Bergemann and Morris (2011) for details). In such environments, any responsive SCF satisfying strict RM can be robustly implemented in interim equilibria using a direct mechanism in which players report their payoff types (Bergemann and Morris, 2009). Recall that strict RM is equivalent to weak RM for responsive SCFs. Thus, it follows that for responsive SCFs, weak RM characterizes robust implementation in rationalizable strategies in single crossing aggregator environments even when the designer is restricted to finite mechanisms.

In a complete information environment with lotteries and transfers, Chen, Kunimoto, Sun, and Xiong (2020) show that Maskin monotonicity*, a strengthening of Maskin monotonicity, is a necessary and sufficient condition for implementation in rationalizable strategies by a finite mechanism. They also show that Maskin monotonicity* is strictly stronger than Maskin monotonicity, which is a necessary and sufficient condition for Nash implementation by a finite mechanism in the same class of environments with transfers and lotteries (See Chen, Kunimoto, Sun, and Xiong (2019)). Therefore, if we restrict our attention to finite mechanisms in a complete information setup, implementation in rationalizable strategies is more restrictive than Nash implementation. This exhibits a contrast with our Proposition 6.6, which establishes the equivalence between robust implementation in rationalizable strategies and that in interim equilibria in finite mechanisms. In other words, the robustness consideration within the class of finite mechanisms makes the difference between rationalizable strategies and equilibria moot.

7 Conclusion

We showed that robust implementation in rationalizable strategies is equivalent to weak rationalizable implementation. We utilized this equivalence result to prove that weak RM is necessary and almost sufficient for robust implementation in rationalizable strategies. The equivalence result also helped us establish that if a mechanism robustly implements an SCF in interim equilibria, then the same mechanism robustly implements the SCF in rationalizable strategies too. However, the converse is not necessarily true if the SCF is nonresponsive and the designer is allowed to use countably infinite mechanisms. We argued that this gap between robust implementation in interim equilibria and robust implementation in rationalizable strategies is explained by the more stringent nonemptiness requirement under the former concept because the two concepts impose the same uniqueness requirement. An open question is whether there is any gap between the two concepts for the case of responsive SCFs.

References

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

[2] Bergemann, D., and S. Morris, "Robust Mechanism Design," *Econometrica*, 73, (2005), 1771-1813.

[3] Bergemann, D., and S. Morris, "Strategic Distinguishability with an Application to Robust Virtual Implementation," *Working Paper*, (2007).

[4] Bergemann, D., and S. Morris, "Robust Implementation in Direct Mechanisms," *Review of Economic Studies*, vol. 76, (2009), 1175-1206.

[5] Bergemann, D., and S. Morris, "Robust Implementation in General Mechanisms," *Working Paper*, (2010).

[6] Bergemann, D., and S. Morris, "Robust Implementation in General Mechanisms," *Games and Economic Behavior*, vol. 71, (2011), 261-281.

[7] Bergemann, D., S. Morris, and S. Takahashi, "Interdependent Preferences and Strategic Distinguishability," *Journal of Economic Theory*, vol. 168, (2017), 329-371.

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

[9] Blackwell, D., and M.A. Girshick, "*Theory of Games and Statistical Decisions*," John Wiley & Sons, Inc., 1954.

[10] Chen, Y-C, T. Kunimoto, Y. Sun, and S. Xiong, "Maskin Meets Abreu and Matsushima," *mimeo*, 2019.

[11] Chen, Y-C, T. Kunimoto, Y. Sun, and S. Xiong, "Rationalizable Implementation in Finite Mechanisms," *mimeo*, 2020.

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

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

[14] Jain, R., "Rationalizable Implementation of Social Choice Correspondences," *Working Paper*, (2018).

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

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

[17] Maskin, E., "Nash Equilibrium and Welfare Optimality," *Review of Economic Studies*, 66, (1999), 23-38.

[18] Xiong, S., "Rationalizable Implementation I: Social Choice Functions," *Working Paper*, (2018).