UNIVERSITY OF LIVERPOOL | Management School

Working Paper in Economics

# 202204

On the Relationship between Robust and Rationalizable Implementation

R Jain
M Lombardi

On the Relationship between Robust and Rationalizable Implementation[*]

R Jain[†] M Lombardi[‡]

March 8, 2022

Abstract

We introduce a notion of rationalizable implementation for social choice functions, termed s-rationalizable implementation, and show that it is equivalent to robust implementation.

Keywords: Robust Implementation, Rationalizable Implementation, Social Choice Functions, Interim Best Response Property.
JEL classification: C79, D82.

[*] Ritesh Jain greatly acknowledges financial support from the Ministry of Science and Technology Grant, Taiwan under the grant: 107- 2410-H-001-113-MY2. All remaining errors are ours.

[†] Institute of Economics, Academia Sinica, e-mail: ritesh@econ.sinica.edu.tw.
[‡] University of Liverpool Management School. E-mail: michele.lombardi@liverpool.ac.uk.

I. Introduction

A social choice function (SCF) is robustly implementable if every equilibrium on every type space achieves outcomes consistent with it. A seminal paper on robust implementation in general environments is Bergemann and Morris (2011). They show that the conditions for robust implementation can be derived as an implication of rationalizable implementation. Therefore, this approach depends on deriving a notion of rationalizable implementation that is equivalent to robust implementation.

Although Bergemann and Morris (2011)'s notion of rationalizable implementation is "almost" equivalent to robust implementation, the literature lacks the full equivalence. In this paper, we establish it. This characterization result is significant because it allows deriving necessary and sufficient conditions for robust implementation using the iterative deletion procedure associated with rationalizability.

Bergemann and Morris (2011) use as a solution concept a robust version of rationalizability, which in their set-up is equivalent to the solution concept of belief free rationalizability (Bergemann and Morris (2017)).[1] However, their notion of rationalizable implementation imposes restrictions on the class of implementing mechanisms. Indeed, Bergemann and Morris (2011)'s definition of rationalizable implementation consists of two parts. The first part requires that every rationalizable message profile be consistent with the SCF. The second part requires that rationalizable messages exist. The latter requirement is satisfied when the implementing mechanism is finite. However, the existence is non-trivial when infinite mechanisms are allowed. The reason is that best responses may not exist for all conjectures. The existence condition requires that best responses exist for some conjectures. Specifically, it requires that for each belief that player i may have over the payoff types of his opponents, player i has a belief over the rationalizable strategies that his opponents might play such that he has a best response, whatever his payoff type.

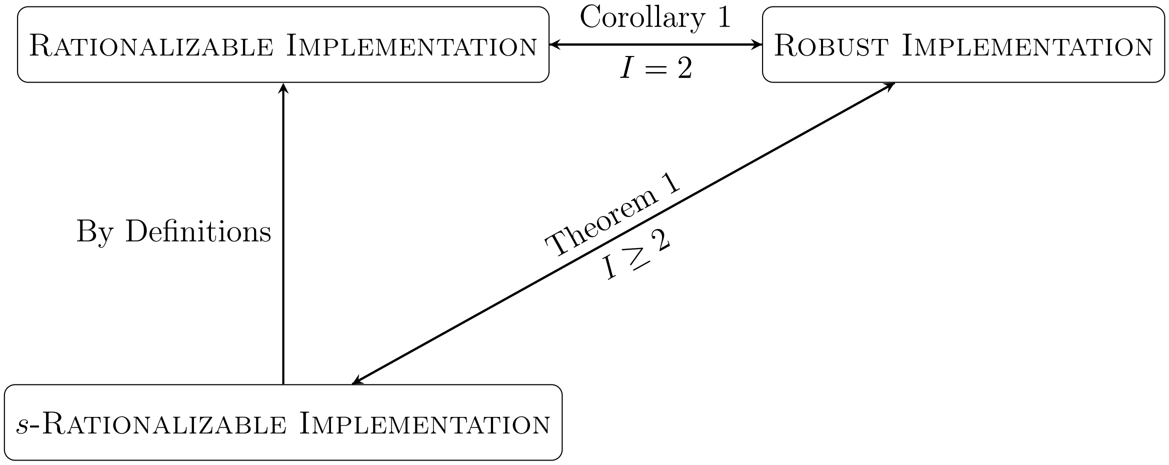
We show that Bergemann and Morris (2011)'s definition is equivalent to robust im-

[1] See Section 3.1 of Bergemann and Morris (2017).

plementation in a two-player society. When there are three or more players, we introduce a notion of rationalizable implementation, which we refer to as s-rationalizable implementation. This notion is obtained by strengthening the existence requirement of Bergemann and Morris (2011); that is, by further restricting the class of implementing mechanisms. However, we use the same solution concept employed by Bergemann and Morris (2011).

The notion of s-rationalizable implementation is derived in the following way. First, we show that the existence requirement of Bergemann and Morris (2011)'s definition is equivalent to the following ex post existence requirement. For each profile of payoff types of player i's opponents, player i has an ex post belief over the rationalizable strategies that his opponents might play such that he has a best response, whatever his payoff type. Second, we add to the ex post existence requirement two properties, which are reminiscent of the familiar epistemic assumptions required to characterize the Nash equilibrium (see, e.g., Dekel and Siniscalchi (2015)). The first property requires players' ex post beliefs over opponents' rationalizable strategies are independent. The second property requires that all opponents of player i have the same ex post beliefs over the rationalizable strategies that player i might play when he is of a given payoff type. Finally, we show that s-rationalizable implementation is equivalent to robust implementation. The figure below summarizes our contribution.

The rest of the paper is organized as follows. Section II presents the theoretical framework and outlines the basic model, with the equivalence presented in Section III. Section IV concludes and discusses the related literature.



II. Model

Payoff Environment and Social Choice Function

We consider a finite set of players $\mathcal{I} = \{1, ..., I\}$. Player i's payoff type is denoted by θ_i. The set of admissible payoff types for player i is denoted by Θ_i. A payoff type profile is described by an I-tuple of payoff types $\theta \in \prod\limits_{i \in \mathcal{I}} \Theta_i = \Theta$. Z is the set of (pure) outcomes. We assume that Θ_i and Z are countable sets. The set of all lotteries over Z is denoted by Y. Player i's preferences over lotteries is described by a continuous and bounded utility function $u_i : Y \times \Theta \to \mathbb{R}$, where $u_i\left(y, \theta\right)$ is player i's utility of the lottery y when θ is the true payoff type profile. For each $\theta \in \Theta$, $u_i\left(\cdot, \theta\right)$ satisfies the expected utility hypothesis.

A social choice function (SCF) is a mapping $f : \Theta \to Y$ such that $f\left(\theta\right) \in Y$ for all $\theta \in \Theta$. Therefore, the planner would like to attain the social outcome $f\left(\theta\right)$ when θ is the true payoff type profile.

The planner must choose a game form or mechanism for the players to play in order to determine the social outcome. Let M_i be the countably infinite set of messages available to player i. A player i's message is denoted by $m_i \in M_i$. A message profile is denoted by $m \in M \triangleq \prod\limits_{i \in \mathcal{I}} M_i$. Let $g(m)$ be the distribution over outcomes when players play m. A mechanism is a collection $\mathcal{M} = (M, g)$, where $g : M \to Y$ is the outcome function.

Θ-Based Type Space

We consider an implementation model with interdependent values. The set of admissible types for player i is assumed to be countable and it is denoted by T_i. A type of player i must include a description of his payoff type. Thus, there is a function $\hat{\theta}_i : T_i \to \Theta_i$ with $\hat{\theta}_i(t_i)$ being player i's payoff type when his type is t_i. A type of player i must also include a description of his beliefs about the types of the other agents; that is, there is a function $\hat{\pi}_i : T_i \to \Delta(T_{-i})$, with $\hat{\pi}_i(t_i)$ being player i's belief type when his type is t_i. Thus, $\hat{\pi}_i(t_i)[t_{-i}]$ is the probability that type t_i of player i assigns to other players having types $t_{-i} \in T_{-i} \triangleq \prod\limits_{j \in \mathcal{I}\setminus\{i\}} T_j$. A type space is a collection $\mathcal{T} \triangleq \left(T_i, \hat{\theta}_i, \hat{\pi}_i\right)_{i\in\mathcal{I}}$. We assume throughout that for each type space $\mathcal{T} \triangleq \left(T_i, \hat{\theta}_i, \hat{\pi}_i\right)_{i\in\mathcal{I}}$ and each $i \in \mathcal{I}$, the function $\hat{\theta}_i : T_i \to \Theta_i$ is onto.[2]

A special type space is the complete information type space, defined by $\mathcal{T}^{CI} = (\Theta_i, \hat{\theta}_i^{CI}, \hat{\pi}_i^{CI})_{i\in\mathcal{I}}$, where $\hat{\theta}_i^{CI}(\theta_i) = \theta_i$ for all $i \in \mathcal{I}$ and all $\theta_i \in \Theta_i$, and where $\hat{\pi}^{CI}(\theta_i)[\theta_{-i}] = 1$ for all $\theta \in \Theta$ and all $i \in \mathcal{I}$.

Solution Concepts

Interim Equilibrium and ex post equilibrium

A type space $\mathcal{T}$ and a mechanism define an incomplete information game, which is denoted by $(\mathcal{T}, \mathcal{M})$. The payoff of player i when players play m and the realized type profile is t is given by

$$u_i\left(g(m), \hat{\theta}(t)\right).$$

A pure strategy for player i in the incomplete information game $(\mathcal{T}, \mathcal{M})$ is given by $s_i : T_i \to M_i$. A (behavioral) strategy is given by $\sigma_i : T_i \to \Delta(M_i)$.

[2] This ensures that the set of types is at least as large as the payoff types.

Definition 1 (Interim equilibrium)**.** A strategy profile $\sigma = (\sigma_i)_{i \in \mathcal{I}}$ is an interim equilibrium for $(\mathcal{T}, \mathcal{M})$ if, for all $i \in \mathcal{I}$, all $t_i \in T_i$ and all $m_i \in M_i$ such that $\sigma_i(m_i|t_i) > 0$, it holds that

$$\sum_{t_{-i} \in T_{-i}} \sum_{m_{-i} \in M_{-i}} \left(\prod_{j \in \mathcal{I} \setminus \{i\}} \sigma_j(m_j|t_j) \right) u_i\left(g(m_i, m_{-i}), \hat{\theta}(t) \right) \hat{\pi}_i(t_{-i}|t_i) \geq$$

$$\sum_{t_{-i} \in T_{-i}} \sum_{m_{-i} \in M_{-i}} \left(\prod_{j \in \mathcal{I} \setminus \{i\}} \sigma_j(m_j|t_j) \right) u_i\left(g(m_i', m_{-i}), \hat{\theta}(t) \right) \hat{\pi}_i(t_{-i}|t_i)$$

for all $m_i' \in M_i$.

Definition 2 (Ex post equilibrium)**.** A strategy profile $\sigma = (\sigma_i)_{i \in I}$ is an ex post equilibrium for $\mathcal{M}$ if, for all $i \in \mathcal{I}$, for all $\theta_i \in \Theta_i$ and all $\bar{m}_i \in M_i$ such that $\sigma_i(\overline{m}_i|\theta_i) > 0$, it holds that

$$\bar{m}_i \in \arg\max_{m_i \in M_i} \left(\sum_{(m_{-i}, \theta_{-i}) \in M_{-i} \times \Theta_{-i}} \sigma_{-i}(m_{-i}|\theta_{-i}) u_i\left(g(m_i, m_{-i}), (\theta_i, \theta_{-i}) \right) \right)$$

for all $\theta_{-i} \in \Theta_{-i}$.

(Robust) Rationalizability

For any mechanism $\mathcal{M}$, let $S^{\mathcal{M}} = \left(S_i^{\mathcal{M}}\right)_{i \in \mathcal{I}}$ denote a profile of message correspondences, where each $S_i^{\mathcal{M}} : \Theta_i \to 2^{M_i}$. We write $\mathcal{S}^{\mathcal{M}}$ for the collection of message correspondence profiles. The collection $S^{\mathcal{M}}$ is a lattice with the natural ordering of set inclusion: $S^{\mathcal{M}} \leq S'^{\mathcal{M}}$ if $S_i^{\mathcal{M}}(\theta_i) \subseteq S_i'^{\mathcal{M}}(\theta_i)$ for all $i \in \mathcal{I}$ and all $\theta_i \in \Theta_i$. The largest element is $\bar{S}^{\mathcal{M}} = \left(\bar{S}_i^{\mathcal{M}}\right)_{i \in \mathcal{I}}$ where $S_i^{\mathcal{M}}(\theta_i) = M_i$ for all $i \in \mathcal{I}$ and all $\theta_i \in \Theta_i$. The smallest element is $\underline{S}^{\mathcal{M}} = \left(\underline{S}_i^{\mathcal{M}}\right)_{i \in \mathcal{I}}$, where $\underline{S}^{\mathcal{M}}(\theta_i) = \emptyset$ for all $i \in \mathcal{I}$ and all $\theta_i \in \Theta_i$.

We define an operator $b^{\mathcal{M}} = (b_1^{\mathcal{M}}, ..., b_I^{\mathcal{M}})$ to iteratively eliminate never best responses. To this end, we denote the belief of player i over message and payoff type profiles of the remaining players by $\lambda_i \in \Delta(M_{-i} \times \Theta_{-i})$. The operator $b^{\mathcal{M}} : \mathcal{S} \to \mathcal{S}$ is defined by

$$b_i^{\mathcal{M}}(S)[\theta_i] = \left\{ m_i \in M_i \;\middle|\; \begin{array}{c} \text{There exists } \lambda_i \in \Delta(M_{-i} \times \Theta_{-i}) \text{ such that:} \\ 1)\ \lambda_i(m_{-i}, \theta_{-i}) > 0 \implies m_j \in S_j(\theta_j) \text{ for all } j \in \mathcal{I} \backslash \{i\}; \\ 2)\ \sum_{m_{-i}, \theta_{-i}} \lambda_i(m_{-i}, \theta_{-i}) u_i(g(m_i, m_{-i}), (\theta_i, \theta_{-i})) \geq \\ \sum_{m_{-i}, \theta_{-i}} \lambda_i(m_{-i}, \theta_{-i}) u_i(g(m_i', m_{-i}), (\theta_i, \theta_{-i})) \\ \text{for all } m_i' \in M_i. \end{array} \right\}$$

Let us note that $b^{\mathcal{M}}$ is increasing by definition, that is, $S^{\mathcal{M}} \leq S'^{\mathcal{M}} \implies b^{\mathcal{M}}(S) \leq b^{\mathcal{M}}(S')$. By Tarski's fixed point theorem, there is a largest fixed point of $b^{\mathcal{M}}$, which is denoted by $S^{\mathcal{M}}$. Thus, (i) $b^{\mathcal{M}}\left(S^{\mathcal{M}}\right) = S^{\mathcal{M}}$ and (ii) $b^{\mathcal{M}}(S) = S \implies S = S^{\mathcal{M}}$. We can also construct the fixed point $S^{\mathcal{M}}$ by starting with $\bar{S}^{\mathcal{M}}$ – the largest element of the lattice – and iteratively applying the operator $b^{\mathcal{M}}$. If the message sets and types are finite, we have

$$S_i^{\mathcal{M}}(\theta_i) = \bigcap_{n \geq 1} b_i^{\mathcal{M}}\left(b^{\mathcal{M},n}\left(\bar{S}\right)\right)[\theta_i].$$

Since the mechanism $\mathcal{M}$ may be infinite, transfinite induction may be necessary to reach the fixed point (Lipman (1994)). It is useful to define

$$S_i^{\mathcal{M},k}(\theta_i) = b_i^{\mathcal{M}}\left(b^{\mathcal{M},k-1}\left(\bar{S}\right)\right)[\theta_i],$$

again using transfinite induction if necessary. Thus $S_i^{\mathcal{M}}(\theta_i)$ is the set of messages surviving (transfinite) iterated deletion of never best responses; equivalently, $S_i^{\mathcal{M}}(\theta_i)$ is the set of messages that player i with the payoff type θ_i might send consistent with common certainty of rationality. We refer to $S_i^{\mathcal{M}}(\theta_i)$ as the rationalizable messages of payoff type θ_i of player i in mechanism $\mathcal{M}$.

If message sets are finite (or compact), a well-known duality argument implies that never best responses are equivalent to strictly dominated actions. However, the equivalence does not hold with infinite (non-compact) message sets. The solution concept defined through the iterative application of the operator $b^{\mathcal{M}}$ is tightly connected to the notion of interim rationalizability for a given type space $\mathcal{T}$, as defined by Battigalli and Siniscalchi (2003) and Dekel et al. (2007). In particular, for any fixed type space $\mathcal{T}$, $S^{\mathcal{M}}$ would be equal to the union of all interim rationalizable actions of player i over all types $t_i \in T_i$ whose payoff type coincides with θ_i; that is, $\hat{\theta}_i(t_i) = \theta_i$.

The following epistemic result highlights the relationship between $S^{\mathcal{M}}$ and interim equilibrium on all type spaces.

Lemma 1 (Bergemann and Morris (2011), Proposition 1, p. 272)**.** $m_i \in S_i^{\mathcal{M}}(\theta_i)$ if and only if there exist a type space $\mathcal{T}$, an interim equilibrium σ of the game $(\mathcal{M}, \mathcal{T})$ and a type $t_i \in T_i$ such that (i) $\sigma_i(m_i|t_i) > 0$ and (ii) $\hat{\theta}_i(t_i) = \theta_i$.

Let us now define the notions of robust and rationalizable implementation.

Definition 3 (Robust implementation)**.** A mechanism $\mathcal{M}$ robustly implements $f : \Theta \to Y$ if, for every type space $\mathcal{T}$, the game $(\mathcal{T}, \mathcal{M})$ (i) has an interim equilibrium and (ii) every interim equilibrium σ of the game $(\mathcal{T}, \mathcal{M})$ satisfies

$$\sigma(m|t) > 0 \implies g(m) = f\left(\hat{\theta}(t)\right).$$

f is robustly implementable if there exists a mechanism $\mathcal{M}$ such that f is robustly implemented by $\mathcal{M}$.

Bergemann and Morris (2011) introduce the notion of (robust) rationalizable implementation, which can be defined as follows.

Definition 4. A mechanism $\mathcal{M}$ implements f in (robust) rationalizable strategies if the following requirements are satisfied.

1. **Full implementation property**: For all $\theta \in \Theta$, $m \in S^{\mathcal{M}}(\theta) \implies g(m) = f(\theta)$.

2. **Interim best response property**: For all $i \in \mathcal{I}$ and all $\psi_i \in \Delta(\Theta_{-i})$, there exists $\lambda_i \in \Delta(M_{-i} \times \Theta_{-i})$ such that:

   (a) $\lambda_i(m_{-i}, \theta_{-i}) > 0 \implies m_{-i} \in S_{-i}^{\mathcal{M}}(\theta_{-i})$.

   (b) $marg_{\Theta_{-i}} \lambda_i = \psi_i$.

   (c) For every $\theta_i \in \Theta_i$,

$$\arg\max_{m_i \in M_i} \left(\sum_{(m_{-i},\theta_{-i}) \in M_{-i} \times \Theta_{-i}} \lambda_i \left(m_{-i}, \theta_{-i}\right) u_i \left(g \left(m_i, m_{-i}\right), \left(\theta_i, \theta_{-i}\right)\right) \right) \neq \emptyset$$

f is rationalizably implementable if there exists a mechanism $\mathcal{M}$ such that f is (robust) rationalizably implemented by $\mathcal{M}$.

Part (1), termed full implementation, requires that every rationalizable message profile must lead to an outcome consistent with f. Part (2), termed interim best response property,[3] requires that for every conjecture over the payoff type space, there exists some beliefs over messages consistent with the correspondence $S^{\mathcal{M}}$ such that player i's best response consists of messages selected by $S_i^{\mathcal{M}}$. This implies that the set of rationalizable strategies is never empty. Also, note that the interim best response property does not require that a best response exists for all possible beliefs over message profiles and, moreover, it is a restriction on the class of implementing mechanisms.

Bergemann and Morris (2011) introduce a strengthening of the interim best response property, termed ex post best response property, which requires that for each payoff type of player i, there is a single message which is rationalizable whatever player i's belief about other players' payoff types. The condition can be stated as follows.

[3] See Bergemann and Morris (2008), Definition 13, p. 23.

Definition 5. Given $\mathcal{M}$, the correspondence $S^{\mathcal{M}}$ satisfies the ex post best response property if, for all $i \in \mathcal{I}$ and all $\theta_i \in \Theta_i$, there exists $m_i^* \in S_i^{\mathcal{M}}(\theta_i)$ such that:

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

for all $\theta_{-i} \in \Theta_{-i}$ and all $m_{-i} \in S_{-i}^{\mathcal{M}}(\theta_{-i})$.

The almost equivalence result between robust and rationalizable implementation of Bergemann and Morris (2011) can stated as follows.

Proposition 1 (Bergemann and Morris (2011), Theorem 3, p. 273)**.**

1. If f is rationalizably implementable by mechanism $\mathcal{M}$ and $S^{\mathcal{M}}$ satisfies the ex post best response property, then f is robustly implementable by $\mathcal{M}$.
2. If f is robustly implementable by mechanism $\mathcal{M}$, then f is rationalizably implementable by $\mathcal{M}$.

Bergemann and Morris (2011) also provide an example which shows that the ex post best response property is not a necessary requirement for robust implementation.

III. Equivalence

Before stating our main result, we first show that the restriction on the correspondence $S^{\mathcal{M}}$ imposed by the interim best response property reduces to the requirement that, for each θ_{-i}, player i has an ex post conjecture $\xi_{-i}\left(\theta_{-i}\right)$ over his opponents' rationalizable strategies such that his best response is not empty, irrespective of his own payoff type θ_i. Formally:

Lemma 2. Suppose $\mathcal{M}$ implements f in rationalizable strategies. The following statements are equivalent.

(i) $S^{\mathcal{M}}$ satisfies the interim best response property.

(ii) For all i and all $\psi_i \in \Delta(\Theta_{-i})$, there exist $\lambda_i \in \Delta(M_{-i} \times \Theta_{-i})$ and $\xi_{-i}(\theta_{-i}) \in \Delta(S^{\mathcal{M}}_{-i}(\theta_{-i}))$ for all $\theta_{-i} \in \Theta_{-i}$ such that:

1. For all $(m_{-i}, \theta_{-i}) \in M_{-i} \times \Theta_{-i}$,

$$\lambda_i(m_{-i}, \theta_{-i}) = \xi_{-i}(m_{-i}|\theta_{-i})\psi_i(\theta_{-i}).$$

2. For all $\theta_i \in \Theta_i$,

$$\arg\max_{m_i \in M_i} \left(\sum_{(m_{-i},\theta_{-i}) \in M_{-i} \times \Theta_{-i}} \xi_{-i}\left(m_{-i}|\theta_{-i}\right) \psi_i\left(\theta_{-i}\right) u_i\left(g\left(m_i, m_{-i}\right), \left(\theta_i, \theta_{-i}\right)\right) \right) \neq \emptyset.$$

(iii) For all $i \in \mathcal{I}$, there exists $\xi_{-i}(\theta_{-i}) \in \Delta(S^{\mathcal{M}}_{-i}(\theta_{-i}))$ for all $\theta_{-i} \in \Theta_{-i}$ such that for all $\theta_i \in \Theta_i$:

$$\arg\max_{m_i \in M_i} \left(\sum_{(m_{-i},\theta_{-i}) \in M_{-i} \times \Theta_{-i}} \xi_{-i}\left(m_{-i}|\theta_{-i}\right) u_i\left(g\left(m_i, m_{-i}\right), \left(\theta_i, \theta_{-i}\right)\right) \right) \neq \emptyset.$$

Proof. Suppose that f is rationalizable implemented by $\mathcal{M}$. Since it is plain that $(ii) \implies (i)$, by definitions, we show below that $(i) \implies (iii)$ and that $(iii) \implies (ii)$.

$(i) \implies (iii)$. Suppose that $S^{\mathcal{M}}$ satisfies the interim best response property. Fix any $i \in \mathcal{I}$ and any $\theta_{-i} \in \Theta_{-i}$. Let ψ_i be a degenerate distribution putting probability 1 on θ_{-i}. Since $marg_{\Theta_{-i}}\lambda_i = \delta_{\theta_{-i}}$, by part (b) of the interim best response property, part (a) of the interim best response property implies that $\lambda_i(\theta_{-i}) \in \Delta(S^{\mathcal{M}}_{-i}(\theta_{-i}))$. Let us define $\xi_{-i}(\theta_{-i}) = \lambda_i(\theta_{-i})$, so that $\xi_{-i}(\theta_{-i}) \in \Delta(S^{\mathcal{M}}_{-i}(\theta_{-i}))$. Since the choice of $\theta_{-i} \in \Theta_{-i}$ was arbitrary, we have that there exists $\xi_{-i}(\theta_{-i}) \in \Delta(S^{\mathcal{M}}_{-i}(\theta_{-i}))$ for all $\theta_{-i} \in \Theta_{-i}$. Finally, part (c) of the interim best response property implies that

$$\arg\max_{m_i \in M_i} \left(\sum_{(m_{-i},\theta_{-i}) \in M_{-i} \times \Theta_{-i}} \xi_{-i}\left(m_{-i}|\theta_{-i}\right) u_i\left(g\left(m_i, m_{-i}\right), \left(\theta_i, \theta_{-i}\right)\right) \right) \neq \emptyset.$$

for all $\theta_i \in \Theta_i$.

Since $i \in \mathcal{I}$ was arbitrary, we conclude that (iii) is implied by (i).

$(iii) \implies (ii)$. Suppose that $\mathcal{M}$ is such that (iii) is satisfied. We show that $\mathcal{M}$ satisfies (ii). Fix any $i \in \mathcal{I}$ and any $\psi_i \in \Delta(\theta_{-i})$. Since $\mathcal{M}$ satisfies (iii), it follows that there exists $\xi_{-i}(\theta_{-i}) \in \Delta(S^{\mathcal{M}}_{-i}(\theta_{-i}))$ for all $\theta_{-i} \in \Theta_{-i}$.

Let us define $\lambda_i(m_{-i}, \theta_{-i})$ by

$$\lambda_i(m_{-i}, \theta_{-i}) = \xi_{-i}(m_{-i}|\theta_{-i}) \cdot \psi_i(\theta_{-i}). \tag{1}$$

for all $(m_{-i}, \theta_{-i}) \in M_{-i} \times \Theta_{-i}$. Thus, $\lambda_i \in \Delta(M_{-i} \times \Theta_{-i})$, and so part (1) of (ii) is satisfied.

To show that $\mathcal{M}$ also satisfies part (2) of (ii), note that for all $\theta_i \in \Theta_i$, it holds that:

$$\begin{aligned}
& \arg\max_{m_i \in M_i} \left[\sum_{\theta_{-i} \in \Theta_{-i}} \left(\sum_{m_{-i} \in M_{-i}} \xi_{-i}\left(m_{-i}|\theta_{-i}\right) \psi_i\left(\theta_{-i}\right) u_i\left(g\left(m_i, m_{-i}\right), \left(\theta_i, \theta_{-i}\right)\right) \right) \right] \\
= \; & \arg\max_{m_i \in M_i} \left[\left(\sum_{\theta_{-i} \in \Theta_{-i}} \psi_i\left(\theta_{-i}\right) \right) \left(\sum_{m_{-i} \in M_{-i}} \xi_{-i}\left(m_{-i}|\theta_{-i}\right) u_i\left(g\left(m_i, m_{-i}\right), \left(\theta_i, \theta_{-i}\right)\right) \right) \right] \\
= \; & \left(\sum_{\theta_{-i} \in \Theta_{-i}} \psi_i\left(\theta_{-i}\right) \right) \arg\max_{m_i \in M_i} \left[\sum_{m_{-i} \in M_{-i}} \xi_{-i}\left(m_{-i}|\theta_{-i}\right) \cdot u_i\left(g\left(m_i, m_{-i}\right), \left(\theta_i, \theta_{-i}\right)\right) \right] \\
\neq \; & \emptyset,
\end{aligned}$$

where the non-emptiness requirement follows from the fact that $\sum_{\theta_{-i} \in \Theta_{-i}} \psi_i\left(\theta_{-i}\right) \neq \emptyset$ and the fact that $\mathcal{M}$ satisfies (iii).

Since i and $\psi_i \in \Delta(\Theta_{-i})$ were arbitrary, the statement follows.

□

To present our main result, we need additional notation. For all $i, j \in \mathcal{I}$ with $i \neq j$

and all $\theta_i \in \Theta_i$, let $\Theta_{-j}(\theta_i) \subseteq \Theta_{-j}$ be defined by:

$$\Theta_{-j}(\theta_i) = \left\{ \theta_{-j} \in \Theta_{-j} \middle| proj_{\Theta_i}(\theta_{-j}) = \{\theta_i\} \right\}.$$

In words, $\Theta_{-j}(\theta_i)$ consists of all payoff type profiles θ_{-j} which list the payoff type θ_i for player i. Note that when $I = 2$, $\Theta_{-j}(\theta_i) = \{\theta_i\}$. We can now introduce our notion of rationalizable implementation, which is shown to be equivalent to robust implementation.

Definition 6. A mechanism $\mathcal{M}$ implements f in (robust) s-rationalizable strategies if the following requirements are satisfied.

1. **Full implementation property**: For every $\theta \in \Theta$, $m \in S^{\mathcal{M}}(\theta) \implies g(m) = f(\theta)$.

2. s**-Interim best response property**: For all $i \in \mathcal{I}$ and all $\theta_{-i} \in \Theta_{-i}$, there exists $\xi_{-i}(\theta_{-i}) \in \Delta(S^{\mathcal{M}}_{-i}(\theta_{-i}))$ such that the following conditions are satisfied.

   (a) *Independence*: For all $\theta_{-i} \in \Theta_{-i}$,

   $$\xi_{-i}(\theta_{-i}) \in \prod_{j \in \mathcal{I} \setminus \{i\}} \Delta\left(S^{\mathcal{M}}_j(\theta_j)\right).$$

   (b) *Consistency*: For all $i \in \mathcal{I}$ and all $\theta_i \in \Theta_i$,

   $$marg_{S^{\mathcal{M}}_i(\theta_i)}\left(\xi_{-j}(\theta_{-j})\right) = marg_{S^{\mathcal{M}}_i(\theta_i)}\left(\xi_{-k}(\theta_{-k})\right)$$

   for all $j, k \in \mathcal{I} \setminus \{i\}$, all $\theta_{-j} \in \Theta_{-j}(\theta_i)$ and all $\theta_{-k} \in \Theta_{-k}(\theta_i)$.

   (c) For all $\theta_i \in \Theta_i$,

   $$\arg\max_{m_i \in M_i} \left(\sum_{m_{-i} \in M_{-i}} \xi_{-i}(m_{-i}|\theta_{-i}) \cdot u_i\left(g(m_i, m_{-i}), (\theta_i, \theta_{-i})\right) \right) \neq \emptyset.$$

f is s-rationalizably implementable if there exists a mechanism $\mathcal{M}$ such that f is (robust) s-rationalizably implemented by $\mathcal{M}$.

The only difference between s-rationalizable implementation and rationalizable implementation concerns part (2) of their definitions. Indeed, part (2) of s-rationalizable implementation is termed s-interim best response property and it consists of three parts. Part (a) requires that every player i's beliefs $\xi_{-i}(\theta_{-i})$ over opponents' rationalizable strategies are independent. Part (b) requires that all opponents of player i have the same ex post beliefs over the rationalizable strategies that player i might play when he is of a given payoff type. Part (c) requires that given player i's beliefs $\xi_{-i}(\theta_{-i})$, his best response is never empty, irrespective of his own payoff type. Note that when there are only two players, part (a) and part (b) do not have any bite. Indeed, when $I = 2$, we show below that s-rationalizable implementation is equivalent to rationalizable implementation.[4]

We have the following characterization result.

Theorem 1. *f is s-rationalizably implementable if and only if f is robustly implementable.*

Proof. Let us first show the "only if" part. Suppose that f is s-rationalizable implemented by $\mathcal{M}$. Fix any type space $\mathcal{T}$, any interim equilibrium $\hat{\sigma}$ of $(\mathcal{T}, \mathcal{M})$, and any $m \in M$ such that $\hat{\sigma}(m|t) > 0$. Since $\hat{\theta}_i(t_i) \in \Theta_i$ and $\hat{\sigma}_i(m_i|t_i) > 0$ for all $i \in \mathcal{I}$, Lemma 1 implies that $m_i \in S_i^{\mathcal{M}}\left(\hat{\theta}_i(t_i)\right)$ for all $i \in \mathcal{I}$. Since f is s-rationalizable implemented by $\mathcal{M}$, it follows that $g(m) = f\left(\hat{\theta}(t)\right)$. Since an ex-post equilibrium is an interim equilibrium of $(\mathcal{M}, \mathcal{T})$ for any $\mathcal{T}$, to show that f is robustly implemented by $\mathcal{M}$, it suffices to show that there exists an ex post equilibrium.

To this end, since $S^{\mathcal{M}}$ satisfies the s-interim best response property, it follows that for all $i \in \mathcal{I}$ and all $\theta_{-i} \in \Theta_{-i}$, there exists $\xi_{-i}(\theta_{-i}) \in \Delta\left(S_{-i}^{\mathcal{M}}(\theta_{-i})\right)$ satisfying conditions (a)-(c) of part (2) of Definition 6. Since f is s-rationalizable implemented

[4] It is clear that if f is s-rationalizable implementable, then it is rationalizale implementable. We still do not know whether s-rationalizable implementation is equivalent to rationalizale implementation. This is left as an open question.

by $\mathcal{M}$ and since $\xi_{-i}(\theta_{-i})$ satisfies part (a) and part (c) for all $i \in \mathcal{I}$ and all $\theta_{-i} \in \Theta_{-i}$, we have that for all $i \in \mathcal{I}$ and all $\theta_{-i} \in \Theta_{-i}$,

$$\arg\max_{m_i \in M_i} \left[\sum_{m_{-i} \in M_{-i}} \xi_{-i}(m_{-i}|\theta_{-i}) \, u_i(g(m_i, m_{-i}), (\theta_i, \theta_{-i})) \right] = S_i^{\mathcal{M}}(\theta_i) \neq \emptyset \tag{2}$$

for all $\theta_i \in \Theta_i$.

By part (a) (Independence) of Definition 6, we have that for all $i \in \mathcal{I}$ and all $\theta_{-i} \in \Theta_{-i}$, $\xi_{-i}(\theta_{-i}) = \prod_{j \in \mathcal{I} \setminus \{i\}} \sigma_{j,i}(\theta_j)$. Fix any $i \in \mathcal{I}$ and any $\theta_i \in \Theta_i$. Part (b) (Consistency) of Definition 6 implies that $\sigma_{i,j}(\theta_i) = \sigma_{i,k}(\theta_i)$ for all $j, k \in \mathcal{I} \setminus \{i\}$. Thus, for all $i \in \mathcal{I}$ and all $\theta_{-i} \in \Theta_{-i}$, $\xi_{-i}(\theta_{-i}) = \prod_{j \in \mathcal{I} \setminus \{i\}} \sigma_{j,i}(\theta_j)$ can be written as $\xi_{-i}(\theta_{-i}) = \prod_{j \in \mathcal{I} \setminus \{i\}} \sigma_j(\theta_j)$.

Since for all $i \in \mathcal{I}$ and all $m_{-i} \in M_{-i}$,

$$\xi_{-i}(m_{-i}|\theta_{-i}) = \left(\prod_{j \in \mathcal{I} \setminus \{i\}} \sigma_j(m_j|\theta_j) \right),$$

(2) can be rewritten as follows: for all $i \in \mathcal{I}$ and all $\theta_{-i} \in \Theta_{-i}$,

$$\arg\max_{m_i \in M_i} \left[\sum_{m_{-i} \in M_{-i}} \left(\prod_{j \in \mathcal{I} \setminus \{i\}} \sigma_j(m_j|\theta_j) \right) u_i(g(m_i, m_{-i}), (\theta_i, \theta_{-i})) \right] = S_i^{\mathcal{M}}(\theta_i) \tag{3}$$

for all $\theta_i \in \Theta_i$. Since for all $i \in \mathcal{I}$ and all $\theta_i \in \Theta_i$, $supp(\sigma_i(\theta_i)) \subseteq S_i^{\mathcal{M}}(\theta_i)$, it follows from (3) that σ is an ex post equilibrium, as we sought.

Next, we prove the "if" part of the statement. Suppose that $\mathcal{M}$ robustly implements f. To show that the full implementation property of Definition 6 is satisfied, take any $\theta \in \Theta$ and any $m \in S^{\mathcal{M}}(\theta)$. Lemma 1 implies that there exist $\mathcal{T}$, an interim equilibrium σ of $(\mathcal{M}, \mathcal{T})$ and a type profile $t \in T$ such that $\sigma(m|t) > 0$ and $\hat{\theta}(\theta) = \theta$. Since $\mathcal{M}$ robustly implements f, it follows that $g(m) = f(\theta)$. Since the choice of $\theta \in \Theta$ and $m \in S^{\mathcal{M}}(\theta)$ were arbitrary, we have that the full implementation property is satisfied.

In the remaining part of the proof, we show that the s-interim best response property of Definition 6 is satisfied. Let $\mathcal{T}$ be a type space such that

$$T_i = \Theta_i \times \Delta\left(\Theta_{-i}\right)$$

for all $i \in \mathcal{I}$, and for all $t_i \in T_i$, $\hat{\theta}\left(t_i\right)$ and $\hat{\pi}_i\left(t_{-i}|t_i\right)$ are defined as follows:

$$\hat{\theta}\left(t_i\right) = proj_{\Theta_i} t_i,$$

$$\sum_{t_{-i} \in \hat{\theta}_{-i}^{-1}(\theta_{-i})} \hat{\pi}_i\left(t_{-i}|t_i\right) = \psi_i\left(\theta_{-i}|t_i\right),$$

where $\psi_i\left(\theta_{-i}|t_i\right) = proj_{\Delta(\Theta_{-i})} t_i\left(\theta_{-i}\right)$ for all $\theta_{-i} \in \Theta_{-i}$. Since $\mathcal{M}$ robustly implements f, let σ be an interim equilibrium of $(\mathcal{T}, \mathcal{M})$; that is, for all $i \in \mathcal{I}$ and all $t_i \in T_i$, it holds that

$$\emptyset \neq \arg\max_{m_i \in M_i} \left[\sum_{t_{-i} \in T_{-i}} \hat{\pi}_i\left(t_{-i}|t_i\right) \sum_{m_{-i} \in M_{-i}} \sigma_{-i}\left(m_{-i}|\theta_{-i}\right) u_i\left(g\left(m_i, m_{-i}\right), \left(\hat{\theta}_i\left(t_i\right), \hat{\theta}_{-i}\left(t_{-i}\right)\right)\right) \right]. \tag{4}$$

Since f is robustly implemented by M, we can assume, without loss of generality, that for all $i \in \mathcal{I}$ and all $t_i, t_i' \in T_i$ such that $\hat{\theta}_i\left(t_i\right) = \hat{\theta}_i\left(t_i'\right)$, it holds that $\sigma_i\left(t_i\right) = \sigma_i\left(t_i'\right)$. Thus, we can restrict the domain of player i's strategy σ_i to Θ_i; that is, for all $i \in \mathcal{I}$,

$$\sigma_i : \Theta_i \to \Delta\left(M_{-i}\right).$$

This allows us to rewrite and simplify (4) as follows. For all $i \in \mathcal{I}$ and all $t_i \in T_i$,

$$\emptyset \neq \arg\max_{m_i \in M_i} \left[\sum_{\theta_{-i} \in \Theta_{-i}} \psi_i\left(\theta_{-i}|t_i\right) \sum_{m_{-i} \in M_{-i}} \sigma_{-i}\left(m_{-i}|\theta_{-i}\right) u_i\left(g\left(m_i, m_{-i}\right), \left(\hat{\theta}_i\left(t_i\right), \theta_{-i}\right)\right) \right]. \tag{5}$$

For all $i \in \mathcal{I}$ and all $\theta_{-i} \in \Theta_{-i}$, let

$$T_i(\theta_{-i}) = \{t_i \in T_i | \psi_i(\theta_{-i}|t_i) = 1\}.$$

Since, by definition of T_i, $T_i(\theta_{-i})$ is not empty, it follows that (5) can be simplified as follows. For all $i \in \mathcal{I}$, all $\theta_{-i} \in \Theta_{-i}$ and all $t_i \in T_i(\theta_{-i})$,

$$\emptyset \neq \arg\max_{m_i \in M_i} \left[\sum_{m_{-i} \in M_{-i}} \sigma_{-i}(m_{-i}|\theta_{-i}) u_i \left(g(m_i, m_{-i}), \left(\hat{\theta}_i(t_i), \theta_{-i} \right) \right) \right] \tag{6}$$

Since for all $i \in \mathcal{I}$ and all $\theta_i \in \Theta_i$, there exists $t_i \in T_i$ such that $\hat{\theta}(t_i) = \theta_i$, (6) simplifies as follows. For all $i \in \mathcal{I}$, all $\theta_i \in \Theta_i$ and all $\theta_{-i} \in \Theta_{-i}$,

$$\emptyset \neq \arg\max_{m_i \in M_i} \left[\sum_{m_{-i} \in M_{-i}} \sigma_{-i}(m_{-i}|\theta_{-i}) u_i(g(m_i, m_{-i}), (\theta_i, \theta_{-i})) \right]. \tag{7}$$

To see that the s-interim best response property is satisfied, for all $i \in \mathcal{I}$ and all $\theta_{-i} \in \Theta_{-i}$, let

$$\xi_{-i}(m_{-i}|\theta_{-i}) = \prod_{j \in \mathcal{I} \setminus \{i\}} \sigma_j(m_j|\theta_j).$$

Clearly, by definition and the initial supposition that σ is an interim equilibrium of $(\mathcal{T}, \mathcal{M})$, conditions (a)-(b) of the s-interim best response property are satisfied. Moreover, (7) verifies condition (c) of the s-interim best response property. Thus, $S^{\mathcal{M}}$ satisfies the s-interim best response property. □

An immediate corollary of Theorem 1 is that rationalizable implementation is equivalent to robust implementation when there are only two players.

Corollary 1. Suppose that $I = 2$. f is rationalizably implementable if and only if f is robustly implementable.

IV. Conclusion and Related Literature

We show that Bergemann and Morris (2011)'s definition of rationalizable implementation is equivalent to robust implementation in a two-player society (Corollary 1). Bergemann and Morris (2011)'s definition of rationalizable implementation includes a restriction on the class of implementing mechanisms, which is termed interim best response property (Part (2) of Definition 4). We show that the existence requirement reduces to the following ex post existence requirement: For each θ_{-i}, player i has an ex post conjecture $\xi_{-i}\left(\theta_{-i}\right)$ over his opponents' rationalizable strategies such that his best response is not empty, irrespective of player i's payoff type (Lemma 2).

To extend the equivalence result to n-player societies, we add to the ex post existence requirement two properties reminiscent of the epistemic characterizations of the Nash equilibrium: *independence* and *consistency*. Roughly speaking, independence requires that each player's ex post conjecture over his opponents' rationalizable strategies is independent, whereas consistency requires that all opponents of player i have the same ex post beliefs over the rationalizable strategies that player i might play when he is of a given payoff type. Based on these two extra requirements, we introduce the notion of s-rationalizable implementation and show that it is equivalent to robust implementation. This is an important characterization because the class of robustly implementable SCFs can be derived by using the iterative deletion procedure associated with rationalizability. This critical exercise is left for future research.

Before closing the paper, let us discuss how our result relates to a recent interesting contribution of Kunimoto and Saran (2020). Bergemann and Morris (2010) introduce a notion of "weak rationalizable implementation". This notion is derived by relaxing the restrictions on the message correspondence $S^{\mathcal{M}}$ imposed by the notion of rationalizable implementation. Kunimoto and Saran (2020) introduce a notion of robust implementation in rationalizable strategies. An SCF is robustly implementable in rationalizable strategies if every interim correlated rationalizable strategy profile on every type space achieves outcomes consistent with it. Kunimoto and Saran (2020) shows that their notion of robust implementation in rationalizable strategies is equivalent to weak rationalizable implementation. Moreover, Kunimoto and Saran (2020) show that robust implementation in rationalizable strategies does not imply robust implementation, though the converse statement is true (by Lemma 1). Our result implies that robust implementation is equivalent to robust implementation in rationalizable strategies via mechanisms satisfying the s-interim best response property.

References

Battigalli, P. and Siniscalchi, M. (2003). Rationalization and incomplete information. *Advances in Theoretical Economics*, 3(1).

Bergemann, D. and Morris, S. (2008). Robust implementation in general mechanisms. Cowles Foundation Discussion Paper No. 1666.

Bergemann, D. and Morris, S. (2010). Robust implementation in general mechanisms. Cowles Foundation Discussion Paper No. 1666R.

Bergemann, D. and Morris, S. (2011). Robust implementation in general mechanisms. *Games and Economic Behavior*, 71(2):261–281.

Bergemann, D. and Morris, S. (2017). Belief-free rationalizability and informational robustness. *Games and Economic Behavior*, 104:744–759.

Dekel, E., Fudenberg, D., and Morris, S. (2007). Interim correlated rationalizability. *Theoretical Economics*.

Dekel, E. and Siniscalchi, M. (2015). Epistemic game theory. In *Handbook of Game Theory with Economic Applications*, volume 4, pages 619–702. Elsevier.

Kunimoto, T. and Saran, R. (2020). Robust implementation in rationalizable strategies in general mechanisms. Economics and Statistics Working Papers 10-2020, Singapore Management University, School of Economics.

Lipman, B. L. (1994). A note on the implications of common knowledge of rationality. *Games and Economic Behavior*, 6(1):114–129.