Strategically Robust Implementation

BSE Working Paper 1461| September 2024

Ritesh Jain, Michele Lombardi, Antonio Penta

Strategically Robust Implementation.[*]

R Jain[†] M Lombardi[‡] A Penta[§]

September 10, 2024

Abstract

We put forward a notion of implementation for Social Choice Functions (SCF) that is *robust* with respect to the solution concept used to model agents' strategic interaction. Formally, we define *implementation in Interim Correlated Rationalizability and its Refinements* (ICRR implementation) as implementation in Interim Correlated Rationalizability (ICR), with the extra requirement that it be achieved by a mechanism in which all selections from ICR have the best-reply property. We provide a tight characterization in terms of a novel notion of monotonicity, *Iterative Interim Monotonicity (IIM)*. Our condition relates the possibility of ICRR-implementation with a specific way in which the SCF is constrained by agents' preference reversals. We provide several alternative formulations of IIM, that clarify both its connection with various parts of the literature (such as Oury and Tercieux (2012)'s Interim Rationalizable Monotonicity,

[*] This paper subsumes and supersedes Jain and Lombardi (2022). Special thanks go to John (Yehuda) Levy and Arunava Sen, whose comments and suggestions have led to substantial improvements in the article. We are grateful to Olivier Tercieux, Siyang Xiong, Olivier Bochet, Bhaskar Dutta and audiences at the University of York, the University of Glasgow, ISI Delhi, Ashoka University, IIM Bangalore, the National Taiwan University, the 22nd annual SAET Conference (France), and the 2024 Conference on Mechanism and Institution Design (Hungary). Antonio Penta acknowledges the financial support of the Spanish Ministry of Economy and Competitiveness, through the Severo Ochoa Programme for Centres of Excellence in R&D (CEX2019-000915-S).

[†] University of Liverpool Management School. E-mail: ritesh.jain@liverpool.ac.uk.

[‡] University of Liverpool Management School. E-mail: michele.lombardi@liverpool.ac.uk. DISES and CSEF, University of Napoli Federico II.

[§] ICREA-UPF, BSE and TSE. E-mail: antonio.penta@upf.edu.

and others), and the source of IIM's ability to overcome several limitations of the previous conditions in the literature.

JEL classification: C79, D82

1. Introduction

The main objective of implementation theory is to characterize under what conditions it is possible to specify a mechanism where agents' strategic interaction results in the outcome that the designer wishes to induce, as a function of agents' *types*. Given a model of the environment, which specifies agents' information and beliefs about everyone's preferences over the feasible allocations, the mode of agents' interaction is captured by game theoretic solution concepts, each giving rise to a distinct notion of implementation. The characterizations that ensue, as well as the implementing mechanisms, are often very sensitive to the fine the details of the model's components.

In recent years, several notions of robustness have been considered, to take into account various forms of possible model mis-specifications (e.g., about agents' preferences, information, and beliefs).[1] With few exceptions, however, little attention has been paid to the possibility that it is the very mode of agents' interaction that could be mis-specified.[2] Obviously, if a mechanism is designed having a specific solution concept in mind, but agents interact in a different way, the outcomes that ensue may be very different from the desired ones. But how can we ensure that a mechanism would perform 'well', even if the designer adopts the 'wrong' solution concept?

Somewhat contrary to standard game-theoretic intuition, resorting to a weaker solution concept would not be enough here. That is because implementation with respect to the weaker solution concept does not ensure that its refinements are non-empty

[1] See, for instance, Bergemann and Morris (2005, 2009, 2011); Bergemann et al. (2010), Oury and Tercieux (2012), Ollár and Penta (2017, 2023, 2024), Börgers and Li (2019), Penta (2015), Müller (2016, 2020), Jain et al. (2022), Xiong (2023).

[2] Albeit from a very different perspective from that of the present paper, a few papers have taken into account the possibility of mis-specifications of the model of agents' interaction. See, for instance, Eliaz (2002), Bochet and Tumennasan (2023b,a), and Gavan and Penta (2023).

valued, and the theory is silent about what happens when agents face a mechanism that admits no solution for the 'true' model of strategic interaction. Indeed, matters of existence play a key role in this literature, as the implementing mechanisms are typically tailored to a specific solution concept, and fine-tuned so as to ensure its existence, but not that of its refinements.

For instance, consider a designer who adopts Interim Correlated Rationalizability (ICR; Dekel et al. (2007)) to model agents' strategic interaction. ICR accommodates agents' strategic uncertainty, it has nice properties, it is weak, and it is supported by compelling epistemic foundations (cf. Dekel et al. (2007); Battigalli et al. (2011)). But suppose that agents' behavior is best captured by (say) Interim *Independent* Rationalizability (IIR; Ely and Peski (2006)). If the designer achieves ICR-implementation with a mechanism that also admits IIR solutions, then he automatically attains his objectives, since IIR is contained within ICR and hence all such profiles would also induce desirable outcomes. But if no IIR profile exists in the mechanism, it is unclear how the agents would behave, and undesirable outcomes may ensue. It may be tempting, then, to pursue *double* implementation with respect to both ICR and IIR. But what if the 'true' solution concept is not IIR either, but a refinement of IIR, say BNE? The problem would remain, just at a different level in the hierarchy of refinements.

To address this issue, one would need to pursue implementation not only with respect to ICR, but for a wide range of refinements. Formalising this idea presents its challenges, and requires making some non-obvious modelling choices. We define implementation in *Interim Correlated Rationalizability and its Refinements* (ICRR implementation) as ICR-implementation, with the extra requirement that it be achieved by a mechanism in which *all* selections from ICR have the *best-reply property*.[3] That is to say, should players' conjectures be concentrated on any subset of the ICR pro-

[3] This formulation speaks directly to the thought-provoking construction by Kunimoto et al. (2023), which we discuss in Section 3. The requirement that the mechanism induces 'well-behaved' best-response correspondences is most closely related to Bergemann et al. (2010) and Bergemann and Morris (2008, 2011).

files (for instance because, for reasons that are beyond the designer's control, certain profiles attain a focality that others lack), then they would still find a best-reply to such conjectures that is within that subset. Note that whenever such a subset is a singleton, this extra restriction implies that it forms an equilibrium, and hence our notion also implies BNE-implementation.[4]

Besides addressing head-on the issue of strategic robustness, it turns out that our approach enables us to provide a tight characterization in terms of a novel notion of monotonicity – which we name *Iterative Interim Monotonicity (IIM)* – that relates the possibility of achieving ICRR-implementation with a specific way in which the SCF is constrained by agents' preference reversals. Up to date, similarly tight characterizations are lacking for both ICR and BNE-implementation, in that existing results combine notions of monotonicity with auxiliary assumptions that restrict agents' preferences or the set of alternatives (see, e.g., Oury and Tercieux (2012), Kunimoto et al. (2023)). Our main result dispenses with all such extra assumptions, and provides a full characterization purely in terms of a condition on preference reversals.

We provide several alternative formulations of IIM, that clarify both its connection with various parts of the literature (such as Oury and Tercieux (2012)'s Interim Rationalizable Monotonicity, which is closest in spirit, as well as other notions from the literature on both rationalizable and BNE-implementation), and the source of IIM's ability to overcome the limitations of the previous conditions in the literature. The latter is achieved thanks to a novel formulation of the conditioning event that identifies the relevant preference reversals, which is based on an *iterative* construction.[5] As we will argue, this structure provides a template to unify key concepts in

[4] Since ICRR-implementation implies ICR&BNE-double implementation, our results are closely related to the influential notion of *continuous implementation* (Oury and Tercieux (2012)). As we will discuss, this also uncovers a connection between our approach to *strategic robustness* and a distinct notion of robustness, that concerns possible mis-specifications of agents' belief hierarchies.

[5] The iterative construction is reminiscent of the AM-measurability condition of Abreu and Matsushima (1992b), which however focus on strategic distinguishability of types using the notion of *virtual* implementation by finite mechanisms. In contrast, we study strategic distinguishability using the notion of *exact* implementation in general mechanisms (cf. Bergemann et al. (2017)).

the literature, as well as suggest extensions in several directions.

That our condition is directly expressed in terms of *preference reversals* is also a novelty with respect to the literature on rationalizable implementation. We see this as a valuable aspect of our results, since preference reversals are the standard language for notions of implementation that are based on equilibrium concepts, but not for rationalizable implementation. Identifying a condition that attains a characterization in terms of preference reversals therefore enables a more direct comparison of the equilibrium and non-equilibrium approaches, and favors the integration of rationalizable implementation within the classical literature.

2. The Environment

2.1. Preliminaries

Preferences. We consider environments with a finite set of agents, $I = \{1, ..., n\}$, and a countable space of allocations, A. The set of lotteries over allocations, or *outcomes*, is denoted by $X \equiv \Delta(A)$.[6] To represent a situation of incomplete information, we parameterize as usual agents' preferences over outcomes by a set of states of nature, Θ. For each $i \in I$, we let $u_i : X \times \Theta \to \mathbb{R}$ denote agent i's state-dependent preferences about lotteries over allocations, and assume that $u(\cdot, \theta)$ is consistent with expected utility theory for each $\theta \in \Theta$

Information. We assume that agents' information about the state of the world has a product structure, and hence without loss of generality we can write the set of states as $\Theta = \Theta_0 \times \Theta_1 \ldots \times \Theta_n$, with the understanding that each agent i only observes the i-the component; $\theta_0 \in \Theta_0$ denotes the residual uncertainty that is left after pooling all agents' information. As standard, for each $i \in I$, we let $\Theta_{-i} := \times_{j \in I \setminus \{i\}} \Theta_i$, so

[6] Throughout the paper, if S is a topological space, we treat it as a measurable space with its Borel sigma field, and the space of Borel probability measures on S is denoted by $\Delta(S)$. Spaces $\Delta(S)$ are endowed with the topology of weak convergence of measures. Also, we treat each countable set as a topological space endowed with the discrete topology. A subset Y of a topological space S is a dense subset of S if the closure of Y in S is equal to S.

that $\Theta = \Theta_0 \times \Theta_i \times \Theta_{-i}$.[7] This model accommodates general environments, including the following special cases: (i) *complete information* corresponds to the case when $|\Theta_k| = 1$ for all $k = 0, 1, .., n$; (ii) *private values* obtain if each u_i is constant in (θ_0, θ_{-i}); (iii) *pure common values* is obtained letting $|\Theta_i| = 1$ for all $i \in I$; (iv) *distributed knowledge* holds if $|\Theta_0| = 1$; (v) *interdependent values* are present whenever u_i depends on θ_{-i} for some $i \in I$; etc. (cf. Battigalli et al. (2011); Penta (2012)).

Beliefs. Agents' beliefs about the state of the world and others' beliefs are represented via a (Harsanyi) type space. Formally, a *type space* is a tuple $\mathcal{T} = \left(T_i, \hat{\theta}_i, \kappa_i\right)_{i \in I}$, where for each $i \in I$, T_i is a countable set of types, $\hat{\theta}_i : T_i \to \Theta_i$ is a function assigning to each type t_i a payoff type in Θ_i, and $\kappa_i : T_i \to \Delta\left(\Theta_0 \times T_{-i}\right)$ assigns to each type his beliefs about Θ_0 and the profile of types of the other players, where $T_{-i} := \times_{j \neq i} T_j$. As usual, we also define $T := \times_{i \in I} T_i$.

Social Choice Functions. The designer's objective is represented by a *social choice function* (SCF, henceforth), $f : T \to X$. We do not impose any restriction on the SCF. In particular, we do not assume that it be responsive (cf. Bergemann and Morris (2009) and Ollár and Penta (2017)).

Mechanisms. A *mechanism* is a tuple $\mathcal{M} = ((M)_{i \in I}, g)$, where M_i is a non-empty and countable set of messages for player i, and $g : M \to X$ is the outcome function, which assigns to every profile of messages an outcome. A message profile $m \in M$ is often written as (m_i, m_{-i}), where $m_{-i} \in M_{-i}$. A mechanism is *direct* if $M_i = T_i$ for all i and $g = f$. This mechanism is denoted by $\mathcal{M}^*$.

Bayesian Games. Given an environment (i.e., with a set of agents with state-dependent preferences), and a type space $\mathcal{T}$, any mechanism $\mathcal{M}$ induces a Bayesian game, with strategies $\sigma_i : T_i \to \Delta\left(M_i\right)$ for each i. We write $\sigma_i\left(t_i\right)\left[m_i\right]$ for the

[7] We will adopt a similar notation throughout. That is, if there is a space or a correspondence S_i for each $i \in I$, we define the set of profiles of all agents and of i's opponents, respectively, as $S := \times_{i \in I} S_i$ and $S_{-i} := \times_{j \neq i} S_j$, with typical elements s and s_{-i}, respectively.

probability that σ_i assigns to m_i when player i is of type t_i. We say that σ_i is *pure* if $\sigma_i(t_i)$ is a degenerate lottery for all t_i.

For any payoff type $\theta_i \in \Theta_i$ and probabilistic conjectures $\mu^i \in \Delta(\Theta_0 \times \Theta_{-i} \times M_{-i})$, we let $r_i(\mu^i, \theta_i)$ denote the set of player i's best responses in the mechanism:[8]

$$r_i\left(\mu^i, \theta_i\right) := \arg\max_{m_i \in M_i} \left(\sum_{(\theta_0, \theta_{-i}, m_{-i}) \in \Theta_0 \times \Theta_{-i} \times M_{-i}} \mu^i\left[\theta_0, \theta_{-i}, m_{-i}\right]\left[u_i\left(g\left(m_i, m_{-i}\right), \left(\theta_0, \theta_i, \theta_{-i}\right)\right)\right] \right).$$

Finally, it will be useful to introduce special notation for the conjectures that are induced by type t_i's beliefs of player i over the state of nature, $\Theta_0 \times \Theta_{-i}$, and the opponents' types (as specified in $\mathcal{T}$) when he thinks that the opponents' play follows the strategy profile σ_{-i}. Such conjectures will be denoted by $\mu^i\left(t_i, \sigma_{-i}\right) \in \Delta(\Theta_0 \times \Theta_{-i} \times M_{-i})$. Formally, for all $i \in I$, all $t_i \in T_i$ and all strategy profiles $\sigma_{-i} = (\sigma_j)_{j \neq i}$, we let $\mu^i\left(t_i, \sigma_{-i}\right)$ be s.t., for each $(\theta_0, \theta_{-i}, m_{-i}) \in \Theta_0 \times \Theta_{-i} \times M_{-i}$,

$$\mu^i\left(t_i, \sigma_{-i}\right)\left[\theta_0, \theta_{-i}, m_{-i}\right] = \sum_{t_{-i} \in \hat{\theta}_{-i}^{-1}(\theta_{-i})} \kappa\left(t_i\right)\left[\theta_0, t_{-i}\right] \sigma_{-i}(t_{-i})\left[m_{-i}\right]. \qquad (1)$$

2.2. Solution Concepts

In this section, we introduce our main solution concepts, starting from Bayes-Nash Equilibrium (BNE). Exploiting the notation introduced above, the BNE for the Bayesian game $(\mathcal{M}, \mathcal{T})$ can be defined as follows:

Definition 1. For any $(\mathcal{M}, \mathcal{T})$, a profile of strategies $\sigma = \left(\sigma_i\right)_{i \in I}$ is a *Bayes-Nash equilibrium* if, for all $i \in I$ and all $t_i \in T_i$, $m_i \in \text{Supp}\left(\sigma_i\left(t_i\right)\right)$ only if $m_i \in r_i\left(\mu^i\left(t_i, \sigma_{-i}\right), \hat{\theta}_i(t_i)\right)$.

BNE is obviously a central concept in the implementation literature, but our analysis will be mainly concerned with *Interim Correlated Rationalizability* (ICR, hence-

[8] Since the reference to the mechanism will be clear from the context, we keep the notation simple and omit the dependence of the best reply correspondence on $\mathcal{M}$.

forth), which was introduced by Dekel et al. (2007), and the notion of *best-reply set.* To introduce these concepts, we need additional notation.

Fix any $(\mathcal{M}, \mathcal{T})$. Let $\Sigma = (\Sigma_i)_{i \in I}$ denote a profile of correspondences $\Sigma_i : T_i \to 2^{M_i} \backslash \{\emptyset\}$ that assign to each type of player i a (non-empty) *set* of messages in M_i. Σ_i can be thought of as a set-valued counterpart of a pure strategy, $\sigma_i : T_i \to M_i$. Hence, if a pure strategy profile σ_{-i} can be thought of as a player i's theory of what message profile would be sent by each type profile of the opponents, Σ_{-i} similarly represents i's view that, for each type profile t_{-i}, the opponents would send messages in the set $\Sigma_{-i}(t_{-i}) \subseteq M_{-i}$. In analogy with the notation we introduced above, $\mu^i(t_i, \sigma_{-i})$, we let $C_i(t_i, \Sigma_{-i}) \subseteq \Delta(\Theta_0 \times \Theta_{-i} \times M_{-i})$ denote the *set of type* t_i*'s correlated conjectures that are concentrated on* Σ_{-i}. To this end, let $\gamma_{-i} : \Theta_0 \times T_{-i} \to \Delta(M_{-i})$ be a 'correlated strategy profile' of i's opponents, and let $\mu^i(t_i, \gamma_{-i})$ denote the conjectures induced by type t_i's beliefs, given γ_{-i}. Formally, for all t_i and all $\gamma_{-i} : \Theta_0 \times T_{-i} \to \Delta(M_{-i})$, we let $\mu^i(t_i, \gamma_{-i}) \in \Delta(\Theta_0 \times \Theta_{-i} \times M_{-i})$ be s.t., for each $(\theta_0, \theta_{-i}, m_{-i})$,

$$\mu^i(t_i, \gamma_{-i})[\theta_0, \theta_{-i}, m_{-i}] = \sum_{t_{-i} \in \hat{\theta}^{-1}(\theta_{-i})} \kappa(t_i)[\theta_0, t_{-i}]\, \gamma_{-i}(\theta_0, t_{-i})[m_{-i}] \tag{2}$$

Then, the *set of type* t_i*'s correlated conjectures that are concentrated on* Σ_{-i} is

$$C_i(t_i, \Sigma_{-i}) = \left\{ \mu^i : \begin{array}{l} \mu^i = \mu^i(t_i, \gamma_{-i}) \in \Delta(\Theta_0 \times \Theta_{-i} \times M_{-i}) \\ \text{for some } \gamma_{-i} : \Theta_0 \times T_{-i} \to \Delta(M_{-i}) \text{ s.t.} \\ \gamma_{-i}(\theta_0, t_{-i}) \in \Delta(\Sigma_{-i}(t_{-i})) \text{ for all } (\theta_0, t_{-i}) \in \Theta_0 \times T_{-i} \end{array} \right\}. \tag{3}$$

With this, we define the best-reply property as follows:

Definition 2. Fix any $(\mathcal{M}, \mathcal{T})$. A collection $\Sigma = (\Sigma_i)_{i \in I}$ of correspondences $\Sigma_i : T_i \to 2^{M_i} \backslash \{\emptyset\}$ has the *best-reply property* if, for all $i \in I$, all $t_i \in T_i$, and all $m_i \in \Sigma_i(t_i)$, there exists a conjecture $\mu^i \in C_i(t_i, \Sigma_{-i})$ s.t. $m_i \in r_i(\mu^i, \hat{\theta}_i(t_i))$.

Def. 2 extends the standard definition of best-reply sets to a Bayesian game. The only non necessarily trivial aspect of this extension is the way that the set of relevant conjectures could be defined. The definition above allows for correlated conjectures, and it is the right one to connect this notion of best-reply property to *Interim Correlated Rationalizability*, which will be introuced shortly. Just like (pure and mixed) Nash Equilibria are connected with best-reply sets in complete information games, it is useful to observe that the notion in Def. 2 is connected with BNE in the following way:

Remark 1. Fix any $(\mathcal{M}, \mathcal{T})$. The following statements hold:

(i) A collection Σ of single-valued correspondences (i.e., s.t., for all $i \in I$, there exists $\sigma_i : T_i \to M_i$ s.t. $\Sigma_i(t_i) = \{\sigma(t_i)\}$ for all $t_i \in T_i$ has the best-reply property if and only if $\sigma = (\sigma_i)_{i \in I}$ is a pure BNE.

(ii) If σ is a (possibly mixed) BNE, then the collection Σ s.t. $\Sigma_i(t_i) := Supp(\sigma_i(t_i))$ for all i and all t_i, has the best-reply property.

We can now introduce *Interim Correlated Rationalizability*, which essentially consists of the 'largest' collection Σ of correspondences $\Sigma_i : T_i \to 2^{M_i} \backslash \{\emptyset\}$ that have the best-reply property. That is:

Definition 3 (Interim Correlated Rationalizability)**.** Fix any $(\mathcal{M}, \mathcal{T})$. For each $t_i \in T_i$, the set of *Interim Correlated Rationalizable* (ICR) messages for type t_i, denoted by $ICR(t_i)$, is the set of all messages $m_i \in M_i$ s.t. $m_i \in \Sigma_i(t_i)$ for some collection Σ with the best-reply property.

This definition of ICR is analogous to the standard 'fixed point' definition of Rationalizability for complete information games (cf. (Bernheim (1984); Pearce (1984)). As usual, the concept may also be given a recursive definition, in terms of an iterated deletion procedure. Since we impose no restrictions on the mechanism, however, the resulting game need not be well-behaved, and hence such a recursive definition is equivalent to the 'fixed point' version above only provided that we allow for transfinite induction (cf. Arieli (2010)). To this end, following Aliprantis and Border (2006), we let Ω denote a set of 'ordinals', which are ordered by $\leq$, and assume that Ω is (i) uncountable and (ii) has a greatest element ω_1.[9]

Definition 4 (ICR: Recursive Formulation)**.** Fix any $(\mathcal{M}, \mathcal{T})$. For all $i \in I$ and all $t_i \in T_i$, let $R_i^0(t_i) = M_i$, and for all ordinals $\alpha \in \Omega \backslash \{0\}$, define $R_i^\alpha(t_i)$ as follows:

- If α is a successor ordinal, then

$$R_i^\alpha(t_i) = \left\{ m_i \in R_i^{\alpha-1}(t_i) : m_i \in r_i(\mu^i, \hat{\theta}_i(t_i)) \text{ for some } \mu^i \in C_i(t_i, R_{-i}^{\alpha-1}) \right\}$$

- If α is a limit ordinal, then $R_i^\alpha(t_i) = \bigcap_{\alpha' < \alpha} R_i^{\alpha'}(t_i)$

Finally, define the set $R_i(t_i) := \bigcap_{\alpha \in \Omega} R_i^\alpha(t_i)$.

The next result shows that, for any $(\mathcal{M}, \mathcal{T})$, the correspondence $R_i : T_i \rightrightarrows M_i$ is well-defined and the collection $R = (R_i)_{i \in I}$ has the best-reply property. Furthermore, for all i and all t_i, we have that $ICR_i(t_i) = R_i(t_i)$.[10]

Lemma 1. For all $(\mathcal{M}, \mathcal{T})$, the net $\{R^\alpha\}_{\alpha \in \Omega}$ is monotonically decreasing with respect to set inclusion. Its limit, which we denote by R, exists and has the best-reply property. Moreover, there exists $\alpha^* \in \Omega$ s.t., for all $\alpha \in \Omega$ with $\alpha \geq \alpha^*$, $R_i^\alpha(t_i) = R_i^{\alpha+1}(t_i) = R_i(t_i) = ICR_i(t_i)$ for all $i \in I$ and all $t_i \in T_i$.

3. Implementation

In this section, we introduce the main notions of implementation that we consider, starting with BNE- and ICR-implementation:

Definition 5 (BNE-Implementation)**.** A mechanism $\mathcal{M}$ *implements* $f : T \to X$ in *Bayes-Nash Equilibrium* (BNE-implements, henceforth) if (i) the set of BNE is non-empty, and (ii) if σ is a BNE, then $g(\sigma(t)) = f(t)$ for all $t \in T$. If such a mechanism exists, then we say that f is *BNE-implementable.*

[9] The existence of this set Ω is proved in Theorem 1.14 of Aliprantis and Border (2006) p. 19.

[10] Lemma 1 generalizes the results of Arieli (2010) to the case of incomplete information.

Definition 6 (ICR-Implementation)**.** A mechanism *implements* $f : T \to X$ *in Interim Correlated Rationalizability* (ICR-implements, henceforth) if (i) $ICR(t) \neq \emptyset$ for all $t \in T$, and (ii) for all $t \in T$, $m \in ICR\,(t) \to g\,(m) = f\,(t)$. If such a mechanism exists, then we say that f is *ICR-implementable.*

ICR-implementation was first studied by Oury and Tercieux (2012), who related it to *continuous implementation*: The requirement that an SCF be interim incentive compatible (IIC) on a neighborhood of the belief-hierarchies associated with agents' types in that type space.[11] More recently, Kunimoto et al. (2023) showed that ICR-implementation may be achieved by a mechanism in which, in fact, *all* refinements of ICR are empty-valued, and hence with the feature that the SCF is not implemented in any refinement of ICR, including BNE. We use their example below to illustrate the point.

Example 1. There are two players $i \in \{1, 2\}$. There is no residual payoff state (i.e., Θ_0 is a singleton), player 1 has three types, $T_1 = \{t_1, t_1', t_1''\}$, and player 2 has two types, $T_2 = \{t_2, t_2'\}$. Beliefs are as follows: $\kappa_1(t_1)[t_2] = .99, \kappa_1(t_1')[t_2] = \kappa_1(t_1'')[t_2] = 0$, and $\kappa_2(t_2)[t_1] = \kappa_2(t_2)[t_1'] = \kappa_2(t_2)[t_1''] = \frac{1}{3}, \pi_2(t_2')[t_i'] = 1$.

There are six pure alternatives $X = \{a, b, c, d, z, z'\}$, and agents' state-dependent utilities are depicted in the following tables:

a	t_2	t_2'
t_1	$4, 4$	$4, 0$
t_1'	$0, 0$	$4, 1$
t_1''	$1, 1$	$4, 0$

b	t_2	t_2'
t_1	$0, 0$	$3, 3$
t_1'	$1, 1$	$2, 0$
t_1''	$0, 0$	$2, 1$

c	t_2	t_2'
t_1	$0, 0$	$3, 1$
t_1'	$3, 3$	$3, 0$
t_1''	$3, 3$	$3, 0$

d	t_2	t_2'
t_1	$3, 4$	$2, 0$
t_1'	$0, 3$	$3, 3$
t_1''	$0, 3$	$3, 3$

z	t_2	t_2'
t_1	$4, 1$	$2, 0$
t_1'	$2, 2$	$5, 0$
t_1''	$2, 2$	$2, 0$

z'	t_2	t_2'
t_1	$4, 0$	$4, 1$
t_1'	$2, 0$	$2, 2$
t_1''	$2, 0$	$5, 0$

z''	t_2	t_2'
t_1	$-1, -1$	$-1, -1$
t_1'	$-1, -1$	$-1, -1$
t_1''	$-1, -1$	$-1, -1$

[11] Continuous implementation being the main focus of Oury and Tercieux (2012), ICR-implementation in that paper was studied in combination with other properties implied by continuous implementation, which include, in particular, the existence of a BNE.

The SCF is as in the following table:

f	t_2	t_2'
t_1	a	b
t_1'	c	d
t_1''	c	d

First, it can be checked that SCF f is Bayesian incentive compatible, and hence truthful revelation is a BNE in the direct mechanism. In this sense, f is *partially* implementable. The direct mechanism, however, fails full implementation, as it admits both BNE and ICR profiles that *do not* induce f.[12] Nonetheless, as the results in Kunimoto et al. (2023) show, this SCF can be ICR-implemented, by some properly constructed *'augmented'* mechanism. To this end, let Y_i^f denote the set of all mappings $y_i : T_i \times T_{-i} \to \Delta(X)$ with the property that each type $t_i \in T_i$ prefers the allocation to be chosen according to f than y_i.[13]

Then, consider the following mechanism, which is an adaptation of the one provided in Kunimoto et al. (2023). For each i, the message space is

$$M_i = T_i \times T_{-i} \times \mathbb{N} \times Y_i^f \times \Delta(X),$$

with generic element $m_i = (m_{i,1}^1, m_{i,2}^1, m_i^2, m_i^3, m_i^4) \in T_i \times T_{-i} \times \mathbb{N} \times Y_i^f \times \Delta(X)$, and the outcome function is as follows:

Rule 1. If $m_i^2 = m_{-i}^2 = 1$, then

$$g(m) = f(m_{i,1}^1, m_{-i,1}^1). \tag{4}$$

[12] For instance, the profile (σ_1, σ_2), where $\sigma_1(t_1) = t_1', \sigma_1(t_1') = t_1', \sigma_1(t_1'') = t_1''$ and $\sigma_1(t_2) = t_2', \sigma_1(t_2') = t_2'$ is a BNE, which does not implement f.

[13] Formally, Y_i^f consists of all mappings $y_i : T_i \times T_{-i} \to \Delta(X)$ s.t.

$$\sum_{t_{-i} \in T_{-i}} \kappa(t_i)[t_{-i}] u_i(f(t_i, t_{-i}), t_i, t_{-i}) \geq \sum_{t_{-i} \in T_{-i}} \kappa(t_i)[t_{-i}] u_i(y(t_i, t_{-i}), t_i, t_{-i}) \text{ for all } t_i \in T_i$$

.

Rule 2. If there is an $i \in I$ s.t. $m_i^2 \neq 1$ $m_{-i}^2 = 1$, then

$$g(m) = \frac{m_i^2}{1+m_i^2} m_i^3(m_{-i,2}^1, m_{-i,1}^1) + (1 - \frac{m_i^2}{1+m_i^2})z''. \tag{5}$$

Rule 3. if $m_i^2 \neq 1$ and $m_{-i}^2 \neq 1$, then

$$g(m) = \frac{m_i^2}{1+m_i^2} m_i^4 + (1 - \frac{m_i^2}{1+m_i^2})z''. \tag{6}$$

As we show in Appendix D, in this mechanism the ICR strategies are s.t., for each i and each $t_i \in T_i$, $ICR_i(t_i) = \{m_i \in M_i | m_{i,1}^1 \in \beta_i^f(t_i) \text{ and } m_i^2 = 1\}$, where $\beta_i^f(t_i) = \{t_i' \in T_i | f(t_i', \cdot) = f(t_i, \cdot)\}$. Such an augmented mechanism therefore does achieve ICR-implementation. But, despite all selections from the ICR set induce the same outcomes, here no ICR-profile forms a mutual best reply, and hence a BNE. For instance, suppose that player 2 plays according to strategy $\sigma_2(\tilde{t}_2) = ((\tilde{t}_2, t_1), 1, m_2^3, m_2^4)$ for each $\tilde{t}_2 \in T_2$, and let $\hat{m}_1^3$ be a message s.t. $\hat{m}_1^3(t_1, t_2) = \hat{m}_1^3(t_1, t_2') = \kappa_1(t_1)[t_2]f(t_1, t_2) + \kappa_1(t_1)[t_2']f(t_1, t_2')$. Then, for type t_1' of agent 1, announcing $\hat{m}_1 = ((t_1', t_2), m_1^2, \hat{m}_1^3, m_1^4)$, instead of any of the ICR-messages $((t_1', , t_2), 1, m_1^3, m_1^4)$, is profitable if m_1^2 is large enough. Indeed, no strategy of player 1 is a best response to σ_2, and hence this mechanism is not 'well-behaved' in the sense of Bergemann et al. (2010); Bergemann and Morris (2008), and Bergemann and Morris (2011). In fact, as it turns out, despite being ICR-implemented by the above mechanism, this SCF is not BNE-implementable, by *any* mechanism. ■

Hence, by cleverly tailoring the mechanism to the specific solution concept (ICR in this case), Kunimoto et al. (2023) obtained the insightful and thought-provoking result that ICR-implementation may sometimes be more permissive than BNE-implementation. As we discussed in the introduction, our notion of implementation limits such 'fine tuning' possibilities, by insisting that the mechanism be well-behaved not only for ICR, but also for its refinements. At a minimum, this should include BNE. But an adequate notion of robustness should also include ‘all concepts *in between*’. We formalize this idea as follows:

Definition 7 (ICRR-Implementation)**.** A mechanism *implements* $f : T \to X$ *in Interim Correlated Rationalizability and its Refinements* (ICRR-implements) if: (i) it ICR-implements f; and (ii) it is *strategically robust*, i.e. it is s.t. all collections of non-empty valued correspondences Σ s.t. $\Sigma_i(t_i) \subseteq ICR_i(t_i)$ for all t_i have the best-reply property. If such a mechanism exists, then f is *ICRR-implementable.*

Clearly, ICRR-implementation implies both ICR and BNE-implementation, but the converse is not true.

4. Iterative Interim Monotonicity

In this section, we introduce the key condition, Iterative Interim Monotonicity (IIM), and our main result, that is that IIM fully characterizes ICRR-Implementation.

4.1. Deceptions

A standard concept within the literature on BNE- and ICR-implementation is the concept of *deception.* Formally, for each $i \in I$, let us call any map $\beta_i : T_i \to 2^{T_i} \backslash \{\emptyset\}$ as player i’s *deception.* Deceptions can be thought of as the non-empty valued correspondences $\Sigma_i : T_i \to 2^{M_i} \setminus \{\emptyset\}$ defined in the previous section (see, e.g., Def. 2), for the special case in which the mechanism is *direct*, $\mathcal{M}^*$.

A special deception for player i is the *truth-telling deception*, β_i^{id}, defined by $\beta_i^{id}(t_i) := \{t_i\}$ for all $t_i \in T_i$. Another special deception is the *babbling deception*, which is denoted by $\bar{\beta}_i$ and defined by $\bar{\beta}_i(t_i) := T_i$ for all $t_i \in T_i$.

For any β_i and any β_i', we write $\beta_i \subseteq \beta_i'$ if $\beta_i(t_i) \subseteq \beta_i'(t_i)$ for all $t_i \in T_i$. Let $\mathcal{B}_i$ denote the set of player i’s deceptions containing the truth-telling deception:

$$\mathcal{B}_i = \left\{\beta_i : T_i \to 2^{T_i} \backslash \{\emptyset\} \,|\beta_i^{id} \subseteq \beta_i\right\}.$$

The set of deception profiles is $\mathcal{B} := \times_{i\in I}\mathcal{B}_i$, with typical element $\beta = (\beta_i)_{i\in I}$. Note that, by definition, the truthtelling deception, β^{id}, is minimal with respect to inclusion within this set. (Formally: $\beta^{id} \in \mathcal{B}$ and $\nexists \beta' \in \mathcal{B} \setminus \{\beta^{id}\}$ s.t. $\beta' \subseteq \beta^{id}$.)

4.2. IIM: Definition

The logic of IIM is similar to several classical notions of monotonicity, including Maskin (1999)'s, which restricts the SCF to being constant over states of the world in which agents' preferences do not display *preferences reversals* of a special kind. This is expressed with respect to the preferences that a given type, t_i, induces over a specific collection of (Anscombe-Aumann) acts, $y_i : T_{-i} \to X$. Hence, we introduce next the familiar notion of lower counter sets in this space.

To this end, we let type t_i's expected utility of act $y_i : T_{-i} \to X$, given the belief $\mu^i \in \Delta(\Theta_0 \times \Theta_{-i} \times T_{-i})$, be denoted as:

$$U_i(y_i, \mu^i, t_i) := \sum_{(\theta_0,\theta_{-i},t_{-i})\in\Theta_0\times\Theta_{-i}\times T_{-i}} \mu^i\left[\theta_0, \theta_{-i}, t_{-i}\right] u_i\left(y_i(t_{-i}), (\theta_0, \hat{\theta}_i(t_i), \theta_{-i})\right)$$

Note that, for each $t_i' \in T_i$, the function $f(t_i', \cdot) : T_{-i} \to X$ is one such act. Then, we can define the lower contour set of $f(t_i', \cdot)$, for type t_i and given beliefs μ^i, as the set of acts $y_i : T_{-i} \to X$ that, given μ^i, yield an expected utility that is not higher than that produced by $f(t_i', \cdot)$. Formally:

$$\mathcal{L}_i(f(t_i', \cdot), \mu^i, t_i) := \left\{ \; y_i : T_{-i} \to \Delta\left(X\right) \;\middle|\; U_i(f(t_i', \cdot), \mu^i, t_i) \geq U_i(y_i, \mu^i, t_i) \; \right\}$$

Now, fix a deception profile $\beta \in \mathcal{B}$. The lower contour set of $(f(t_i, \cdot), \beta)$ for type t_i, denoted $\mathcal{L}_i(f(t_i, \cdot), \beta, t_i)$, consists of all acts $y_i : T_{-i} \to X$ that induce a weakly lower expected utility than $f(t_i, \cdot)$, for all beliefs $\mu^i \in C_i(t_i, \beta_{-i})$.[14] Formally:

[14] The set of consistent conjectures concentrated on β_{-i}, $C_i(t_i, \beta_{-i})$, is formally defined in eq. (3).

$$\mathcal{L}_i(f(t_i,\cdot),\beta,t_i) := \bigcap_{\mu^i \in C_i(t_i,\beta)} \mathcal{L}_i(f(t_i,\cdot),\mu^i,t_i)$$

Finally, we let $Y_i(\beta)$ denote the set of all acts that are in the lower contour set of $(f(t_i,\cdot),\beta)$ *for all* types of player i:

$$Y_i(\beta) := \bigcap_{t_i \in T_i} \mathcal{L}_i(f(t_i,\cdot),\beta,t_i).$$

Intuitively, $Y_i(\beta)$ contains all acts $y_i : T_{-i} \to X$ that no type t_i with conjectures concentrated on β would prefer over the act that is induced by truth-telling under the SCF (namely, $f(t_i,\cdot)$). In that sense, it can be thought of as the set of *credible reward functions*, when the sole information about agents' beliefs is that they are concentrated on β.[15] Note that if $\hat{\beta} \subseteq \beta$, then $Y_i(\beta) \subseteq Y_i(\hat{\beta})$. Hence, $Y_i(\beta)$ is largest when $\beta = \beta^{id}$, since the truthtelling profile is minimal within $\mathcal{B}$.

Next, let the deception β^f be defined so that $t' \in \beta(t)$ if and only if $f(t) = f(t')$. By definition, $\beta^{id} \subseteq \beta^f$, with equality if f is *responsive.* Also note that, for a general deception β', requiring it to be s.t. $\beta' \subseteq \beta^f$ amounts to stating that $f(t) = f(t')$ whenever $t' \in \beta'(t)$.

Definition 8. $f : T \to X$ satisfies IIM if and only if $\beta' \subseteq \beta^f$ whenever β' is s.t. for each $i \in I$, each $t_i \in T_i$ and each $t'_i \in \beta'_i(t_i)$, there exists $\mu^i \in C_i(t_i,\beta'_{-i})$ s.t. $Y_i(\beta^f) \subseteq \mathcal{L}_i(f(t'_i,\cdot),\mu^i,t_i)$.

In words, IIM states that whenever a deception β' is s.t., for each agent i and type t_i, and for each deceiving report $t'_i \in \beta'_i(t_i)$, it is possible to find a conjecture μ^i according to which *all* the *credible rewards* $Y(\beta^f)$ are in the lower contour set of $f(t'_i,\cdot)$ for type t_i, then it must be case that $f(t) = f(t')$ for all t and t' s.t. $t' \in \beta(t)$.

[15] $Y_i(\beta)$ is a metrizable separable space. To see it, observe that $\Delta(X)$ is a separable metric space under the Prohokorov metric given that X is a separable metric space (Aliprantis and Border (2006); Theorem 14.15). Moreover, a countable product of the space $\Delta(X)$ is separable metric space under the standard metric (see, e.g., Ok (2011), p. 196). Thus, since $Y_i(\beta)$ is a subset of a separable metric space, it is also a separable metric space.

Equivalently, the SCF may chose different allocations at states t and t' only if, moving from t' to t, there exists at least some agent i and a deception β' with the property that for each of t_i's consistent conjectures on β', at least one of the acts in $Y_i(\beta^f)$ has *climbed up* from being ranked below to above $f(t_i', \cdot)$.

4.3. ICRR-Implementation: A full characterization

We can now present our main characterization result:

Theorem 1. $f : T \to X$ is ICRR-implementable if and only if f satisfies IIM.

The proof of this result is in Appendix A, and it is based on the characterization of IIM that we provide in Theorem 3 below. Here, it is useful to discuss a necessary condition for IIM (and, hence, for ICRR-implementability of an SCF), which is easier to check and may hence be used to easily identify SCFs that do not satisfy IIM.

Remark 2. If $f : T \to X$ satisfies IIM, then whenever t_i and t_i' are s.t. there exists $\mu^i \in C_i(t_i, \beta^f_{-i})$ s.t. $Y_i(\beta^f) \subseteq \mathcal{L}_i(f(t_i', \cdot), \mu^i, t_i)$, it must be that $f(t_i, \cdot) = f(t_i', \cdot)$.

Hence, for any two types, if there exists at least one feasible conjecture concentrated on β^f under which 'no reversals' occur from one type to the other, then IIM forces the SCF to induce the same act under the two types. The convenience of this condition stems from the fact that, for any f, it suffices to check properties of the corresponding β^f deception. On its own, however, it need not suffice to ensure IIM, since it may be that some $\beta' \not\subseteq \beta^f$ also satisfies the same 'no reversal' condition, when conjectures concentrated on β' are taken into account. As per Def.8, IIM requires that no such deceptions exist (in that all β' with such property must be already 'inside' β^f, which need not be for an arbitrary f, even if the condition in Remark 2 holds.)

5. IIM, Measurability and Non-Refutability

We present next a few equivalent formulations of IIM, which help clarify the connection with other conditions in the literature. Similar to the measurability conditions in

Abreu and Matsushima (1992a), Bergemann et al. (2017), such alternative formulations of IIM require the SCF to be constant over types that cannot be 'separated' from one another. This perspective also enables a direct comparison with Bayesian Incentive Compatibility, as well as other notions of monotonicity, such as Jackson (1991)'s Bayesan Monotonicity, and Oury and Tercieux (2012)'s IRM. First we introduce the following standard notion:

Definition 9. A deception $\beta \in \mathcal{B}$ is *acceptable* for $f : T \to X$ if $f(t) = f(t')$ whenever $t' \in \beta(t)$. A deception is unacceptable otherwise.

By definition, the truthful deception β^{id} is trivially acceptable. Requiring acceptability for some $\beta \neq \beta^{id}$ is essentially imposing a measurability restriction on f.

Fix any deception β, and any conjectures $\mu^i \in \Delta(\Theta_0 \times \Theta_{-i} \times T_{-i})$. We say that μ^i separates type t_i' from t_i with respect to β (or β-**separates** t_i' **from** t_i), if there exists a reward function $y_i \in Y_i(\beta)$ that is both *credible* with respect to β and that type t_i *strictly* prefers over the act $f(t_i', \cdot)$. The **non-separability correspondence**, ρ_i^β, that assigns to each μ^i and each t_i the set of types that are not β-separated from t_i under the conjectures μ^i, is thus defined as follows: $\rho_i^\beta : \Delta(\Theta_0 \times \Theta_{-i} \times T_{-i}) \times T_i \rightrightarrows T_i$ s.t.

$$\rho_i^\beta(\mu^i, t_i) := \Big\{ t_i' \in T_i : Y_i(\beta) \subseteq \mathcal{L}_i(f(t_i', \cdot), \mu^i, t_i) \Big\}. \tag{7}$$

A deception β' is *non-refutable* with respect to β, if there exists no type t_i and $t_i' \in \beta_i'(t_i)$ which can be β-separated from t_i under all consistent conjectures that are concentrated on β'. Formally:

Definition 10. Fix any $\beta, \beta' \in \mathcal{B}$. We say that β' is **non-refutable w.r.t.** β if for all $i \in I$, all $t_i \in T_i$, and all $t_i' \in \beta_i'(t_i)$, there exists $\mu^i \in C_i(t_i, \beta_{-i}')$ s.t. $t_i' \in \rho_i^\beta(\mu^i, t_i)$.

It makes sense to define a deception β^* as maximally non-refutable with respect to itself if it is s.t. $\beta \subseteq \beta^*$ whenever β is non-refutable with respect to β^*. With this, we define the following:

Definition 11. A deception β^* is **tightly non-refutable** if it satisfies the following:

(i) β^* is maximally non-refutable w.r.t. itself;

(ii) $\beta^* \subseteq \beta$ for all β that are maximally non-refutable w.r.t. themselves.

As it will be shown, a tightly non-refutable deception always exists.[16] Note that, by definition, if β^* is tightly non-refutable, then $t_i \in \beta_i^*(t_i')$ if and only if $t_i' \in \beta_i^*(t_i)$. That is, any tightly non-refutable deception β^* induces a partition over the set of types. Let $\sim^*$ denote the corresponding equivalence relation. Also, by definition, $t_i' \in \beta^*(t_i)$ if and only if there $\exists \mu^i \in C_i(t_i, \beta_{-i}^*)$ s.t. $t_i' \in \rho^{\beta^*}(\mu^i, t_i)$. Hence, $t_i \sim^* t_i'$ if and only if t_i' cannot be β^*-separated from t_i under all consistent conjectures concentrated on β^*.

The next result provides a few equivalent formulations of IIM that help connect the notion in Def.8 with other concepts in the literature.

Theorem 2. The following are equivalent:

1. f satisfies IIM.

2. β^f is tightly non-refutable.

3. Every deception that is tightly non-refutable is acceptable.

This theorem follows directly from the proof of Theorem 3, in the next section. The characterization in point 3, in particular, enables a direct comparison with the main notions of monotonicity that have been provided for BNE and ICR. We start with the notions from Jackson (1991) and Oury and Tercieux (2012), respectively, Bayesian Monotonicity and Interim Rationalizable Monotonicity.[17]

Definition 12. $f : T \to X$ satisfies Bayesian Monotonicity (BM) if every *single-valued* deception that is non-refutable w.r.t. β^{id} is acceptable.

[16] In the next section we will show that it is unique, and provide an explicit algorithm to identify it for any SCF.

[17] Our formulation of Bayesian Monotonicity favors a direct comparison with both IRM and IIR, but it is easy to show that it is equivalent to Jackson (1991)'s original definition.

Definition 13. $f : T \to X$ satisfies Interim Rationalizable Monotonicity (IRM) if every deception that is non-refutable w.r.t. β^{id} is acceptable.

Oury and Tercieux (2012) introduced IRM as a condition for ICR-implementation by a mechanism with a BNE. By dropping the BNE requirement, on the other hand, Kunimoto et al. (2023) showed that a weaker form of IRM, called weak-IRM, is necessary for implementation in ICR. Weak-IRM is also sufficient when combined with an auxiliary condition. Although IRM and weak-IRM share the same logical structure, the latter relies on a weaker notion of non-refutability. To introduce it, we need to introduce the notion of *weak non-separability correspondence.*

The **weak non-separability correspondence**, $\tilde{\rho}_i^{\beta}$, that assigns to each μ^i and each t_i the set of types that are not β-weakly separated from t_i under the conjectures μ^i, is defined as follows: $\tilde{\rho}_i^{\beta} : \Delta(\Theta_0 \times \Theta_{-i} \times T_{-i}) \times T_i \rightrightarrows T_i$ s.t.

$$\tilde{\rho}_i^{\beta}(\mu^i, t_i) := \Big\{ t_i' \in T_i : \mathcal{L}_i(f(t_i, \cdot), \beta, t_i) \subseteq \mathcal{L}_i(f(t_i', \cdot), \mu^i, t_i) \Big\}. \tag{8}$$

Definition 14. Fix any $\beta, \beta' \in \mathcal{B}$. We say that β' is **weakly non-refutable w.r.t.** β if for all $i \in I$, all $t_i \in T_i$, and all $t_i' \in \beta_i'(t_i)$, there exists $\mu^i \in C_i(t_i, \beta_{-i}')$ s.t. $t_i' \in \tilde{\rho}_i^{\beta}(\mu^i, t_i)$.

Definition 15. $f : T \to X$ satisfies weak-IRM if every deception that is weakly non-refutable w.r.t. β^{id} is acceptable.

These monotonicity notions are necessary for the corresponding notions of implementation, but unlike IIM, they may not also sufficient on their own. Indeed, by adapting the arguments used here to develop the iterative version of IIM, iterative variants for both IRM and weak-IRM can be developed to identify the class of SCFs that are ICR-implementable by a mechanism with a BNE, as well as the class of SCFs that are ICR-implementable.[18] Nonetheless, the next Corollary follows immediately from the observation that $\beta^{id} \subseteq \beta^*$.

[18] The interested reader can consult the subsumed paper by Jain and Lombardi (2022) for details.

Corollary 1. IIM $\Rightarrow$ IRM $\Rightarrow$ BM. (The converse relations do not hold)

Corollary 2. IRM $\Rightarrow$ weak-IRM $\nRightarrow$ BM and weak-IRM$\nRightarrow$ IRM.

Next, consider the following definition of incentive compatibility:

Definition 16. Fix a deception $\beta \in \mathcal{B}$. SCF $f : T \to X$ is *Interim β-Incentive Compatible* (β-IC) if for all $i \in I$ and for all $t_i, t_i' \in T_i$, $f(t_i', \cdot) \in \mathcal{L}_i(f(t_i, \cdot), \beta, t_i)$.

Note that if $\beta = \beta^{id}$, then β-IC becomes the standard notion of Interim Incentive Compatibility (IIC). At the opposite extreme, for $\beta = \bar{\beta}$, β-IC coincides with *interim* Dominant-Strategy Incentive Compability (cf. Ollár and Penta (2017)). Again, the next result follows directly from the definitions and the observation that $\beta^{id} \subseteq \beta^f$:

Corollary 3. IIM $\Rightarrow$ β^f-IC $\Rightarrow$ IIC. (The converse relations do not hold)

6. A Recursive Algorithm

In this section, we introduce a recursive algorithm that identifies the *tightly non-refutable deception* β^* that we introduced above. To this end, for each $\beta \in \mathcal{B}$, we introduce a correspondence $W^\beta : T_i \rightrightarrows T_i$ that effectively identifies the largest deception that is non-refutable with respect to β. This is obtained by an iterated elimination procedure, similar to the one we provided for ICR in Def. 4. In fact, the definition of W^β is obtained by replacing the best-response correspondence $r_i : \Delta(\Theta_0 \times \Theta_{-i} \times M_{-i}) \times \Theta_i \rightrightarrows M_i$ in Def. 4 with the ρ_i^β correspondence we defined in equation (7), for a mechanism in which $M_i = T_i$ for all i. Formally:

Definition 17 (Iterated Deletion of β-Separated Types)**.** For all $\mathcal{T}$, all $\beta \in \mathcal{B}$, and all $(i, t_i) \in I \times T_i$, let $W_i^{0,\beta}(t_i) = T_i$, and for each ordinal $\alpha \in \Omega \backslash \{0\}$, define $W_i^{\alpha,\beta}(t_i) = \overline{\beta}_i(t_i)$ if β is unacceptable, otherwise, if β is acceptable, then $W_i^{\alpha,\beta}(t_i)$ is defined as follows:

- If α is a successor ordinal, then

$$W_i^{\alpha,\beta}(t_i) = \left\{ t_i' \in W_i^{\alpha-1,\beta}(t_i) : t_i' \in \rho_i^{\beta}(\mu^i, t_i) \text{ for some } \mu^i \in C_i(t_i, W_{-i}^{\alpha-1,\beta}) \right\} \tag{9}$$

- If α is a limit ordinal, then

$$W_i^{\alpha,\beta}(t_i) = \bigcap_{\alpha' < \alpha} W_i^{\alpha',\beta}(t_i) \tag{10}$$

Finally, define the set $W_i^{\beta}(t_i) := \bigcap_{\alpha \in \Omega} W_i^{\alpha,\beta}(t_i)$.

The next lemma ensures that W^{β} is non-empty valued for all $\mathcal{T}$ and all $\beta \in \mathcal{B}$.

Lemma 2. For all $\mathcal{T}$ and all $\beta \in \mathcal{B}$, the net $\{W^{\alpha,\beta}\}_{\alpha \in \Omega}$ is monotone decreasing w.r.t. set inclusion and there exists $\alpha^* \in \Omega$ s.t. $W^{\alpha,\beta} = W^{\alpha+1,\beta}$ for all $\alpha \geq \alpha^*$. Hence, W^{β} exists and is s.t. $W_i^{\beta}(t_i) \neq \emptyset$ for all $t_i \in T_i$ and all $i \in I$.

Since W^{β} is well-defined and non-empty valued for all $\beta \in \mathcal{B}$, it can effectively be envisioned as map $W : \beta \mapsto W^{\beta}$ from $\mathcal{B}$ to itself. We can thus define the set of *fixed-points* of W by

$$\mathcal{E}(W) = \{\beta \in \mathcal{B} : W^{\beta} = \beta\} \tag{11}$$

Such fixed-points of W effectively consist of the deceptions $\beta \in \mathcal{B}$ that are *maximally non-refutable with respect to themselves.* Since W is monotone, it follows from Tarski's fixed-point theorem that $\mathcal{E}(W)$ is a non-empty lattice, with unique minimal and maximal (with respect to set-inclusion) elements, β^{min} and β^{max}. By definition, such minimal element is *tightly non refutable.* Hence, the next result follows:

Lemma 3. For any $\mathcal{T}$ and any $f : T \to X$, there exists one, and only one, *tightly non-refutable deception*, β^*

Next, let $\{\beta_\alpha\}_{\alpha \in \Omega} \subseteq \mathcal{B}$ be a net recursively defined as follows: (i) $\beta_0 = \beta^{id}$; (ii) for each successor ordinal $\alpha \in \Omega \setminus \{0\}$, $\beta_\alpha = W^{\beta_{\alpha-1}}$; and (iii) for each limit ordinal

$\alpha \in \Omega \setminus \{0\}$, $\beta_\alpha = \bigcup\limits_{\gamma<\alpha} W^{\beta_\gamma}$. The limit of the net $\{\beta_\alpha\}_{\alpha\in\Omega}$ exists if there exists $\alpha^* \in \Omega$ s.t. for all $\alpha \geq \alpha^*$, $\beta_\alpha = \beta_{\alpha+1}$. If the limit of $\{\beta_\alpha\}_{\alpha\in\Omega}$ exists, we denote it by β^{lim} and write $\lim_{\alpha\in\Omega} \beta_\alpha = \beta^{lim}$. The next result states important properties of the algorithm.

Lemma 4. Fix any $\mathcal{T}$. Then:

1. for all $\beta \in \mathcal{B}$, $\beta \subseteq W^\beta$;

2. the net $\{\beta_\alpha\}_{\alpha\in\Omega}$ is monotone increasing w.r.t. set inclusion and $\lim_{\alpha\in\Omega} \beta_\alpha = \beta^{lim} \in \mathcal{E}(W)$;

3. if $\tilde{\beta} \in \mathcal{E}(W)$, then $\beta^{lim} \subseteq \tilde{\beta}$;

4. for all $\beta \in \mathcal{B}$, if β is non-refutable w.r.t. β^{lim}, then $\beta \subseteq \beta^{lim}$.

Proof. See Appendix B. ■

With this, we provide the main result of this section:

Theorem 3. f satisfies IIM if and only if $\beta^{lim} = \beta^* = \beta^f$

Proof. Assume f satisfies IIM on $\mathcal{T}$. By Lemma 3 and Part 3) of Lemma 4, it holds that $\beta^{lim} = \beta^*$. It follows from the definition of the W operator that $\beta^f \subseteq \beta^{lim} = \beta^*$. To complete the proof, by Part 3) of Lemma 4, it suffices to show that $\beta^f = W^{\beta^f}$. Part 1) of Lemma 4 implies that $\beta^f \subseteq W^{\beta^f}$. From the definition of W^{β^f}, it holds that for every $(i, t_i, t_i') \in I \times T_i \times W_i^{\beta^f}(t_i)$, there exists $\mu^i \in C_i(t_i, W_{-i}^{\beta^f})$ s.t. $t_i' \in \rho_i^{\beta^f}(\mu^i, t_i)$. IIM implies that $W^{\beta^f} \subseteq \beta^f$. Thus, $\beta^{lim} = \beta^* = \beta^f$.

For the converse, assume that $\beta^{lim} = \beta^* = \beta^f$. We show that f satisfies IIM on $\mathcal{T}$. Take any $\beta \in \mathcal{B}$ s.t. the premises of IIM is satisfied. Then, β is non-refutable w.r.t. β^f. Since $\beta^f = \beta^{lim}$, part 4) of Lemma 4 implies that $\beta \subseteq \beta^f$. Thus, f satisfies IIM. ■

7. Discussion and Extensions

We have formulated a notion of implementation that is robust to the mis-specification of the solution concept. The logic of strategic robustness gives rise to a more demanding notion than plain rationalizable implementation, as it imposes extra restrictions on the implementing mechanism. Namely, that all refinements of ICR also be non-empty valued. Indirectly, these restrictions limit the use of mechanisms with a tail-chasing structure, which has been at the core of a classical critique of implementation theory (Jackson (1992)). Thus, our approach indirectly speaks to this famous critique by Jackson, and relates it to a specific notion of robustness, that naturally limits the designer's ability to fine-tune the mechanism to a specific solution concept.

The methodology developed in this paper can also be used to provide full characterizations in other related solution concepts. Indeed, it provides a template to unify existing results and formulate new ones. We discuss some of them next.

Robust Implementation: Our analysis is based on the assumption that we are dealing with a fixed, arbitrary type space. Following Bergemann and Morris (2012), it may be interesting to explore which set of SCFs can be implemented in a robust manner, i.e. 'across' *all* type spaces. In this context, the relevant solution concept is belief-free rationalizability. Jain et al. (2023) introduces a concept of implementation that is based on belief-free rationalizability, and that is equivalent to requiring implementation on every type space through a mechanism with an ex-post equilibrium. A version of our paper that is also robust in the sense of Bergemann and Morris (2005, 2009, 2012), etc., can thus be developed by adapting the techniques that here we used for a fixed, arbitrary type space. In that context, a robust version of IIM would be necessary and sufficient for robust implementation and all its refinements. We are currently exploring this idea in our ongoing work Jain et al. (2024).

Implementation via extensive form games: Müller (2020) studies a strong form of robust implementation by a dynamic mechanism which is both belief-free and belief-revision-free is studied. Specifically, the study focuses on full implementation problems in weak Perfect Bayesian equilibrium across all type spaces. The author formulates a necessary condition for implementation called dynamic robust monotonicity, which is weaker than the robust monotonicity condition proposed by Bergemann and Morris (2011). Furthermore, the study shows that under the conditional no total indifference condition, ex-post incentive compatibility and dynamic robust monotonicity are sufficient for implementation in weak Perfect Bayesian equilibrium by general dynamic mechanisms. Following Müller (2020), one can study implementation problems in weak Perfect Bayesian equilibrium by a general dynamic mechanism on a fixed, arbitrary type space. By using our template, a dynamic IRM condition, which is weaker than IRM, can be formulated as a necessary condition for implementation, and it is expected to be sufficient as well.

References

Abreu, D. and Matsushima, H. (1992a). Virtual implementation in iteratively undominated strategies: complete information. *Econometrica: Journal of the Econometric Society*, pages 993–1008.

Abreu, D. and Matsushima, H. (1992b). Virtual implementation in iteratively undominated strategies II: Incomplete information. Princeton University working paper.

Aliprantis, C. D. and Border, K. C. (2006). *Infinite dimensional analysis.* Springer.

Arieli, I. (2010). Rationalizability in continuous games. *Journal of Mathematical Economics*, 46:912–924.

Battigalli, P., Di Tillio, A., Grillo, E., and Penta, A. (2011). Interactive epistemology and solution concepts for games with asymmetric information. *The BE Journal of Theoretical Economics*, 11(1).

Bergemann, D. and Morris, S. (2005). Robust mechanism design. *Econometrica*, pages 1771–1813.

Bergemann, D. and Morris, S. (2008). Interim rationalizable implementation. *Mimeo: Available upon request.*

Bergemann, D. and Morris, S. (2009). Robust implementation in direct mechanisms. *The Review of Economic Studies*, 76(4):1175–1204.

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

Bergemann, D. and Morris, S. (2012). *Robust mechanism design: The role of private information and higher order beliefs*, volume 2. World Scientific.

Bergemann, D., Morris, S., and Takahashi, S. (2017). Interdependent preferences and strategic distinguishability. *Journal of Economic Theory*, 168:329–371.

Bergemann, D., Morris, S., Tercieux, O., et al. (2010). Rationalizable implementation. Technical report, David K. Levine.

Bernheim, B. D. (1984). Rationalizable strategic behavior. *Econometrica: Journal of the Econometric Society*, pages 1007–1028.

Bochet, O. and Tumennasan, N. (2023a). Defaults and benchmarks in mechanism design. *Working Paper.*

Bochet, O. and Tumennasan, N. (2023b). Resilient mechanisms. *Working Paper.*

Börgers, T. and Li, J. (2019). Strategically simple mechanisms. *Econometrica*, 87(6):2003–2035.

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

Eliaz, K. (2002). Fault tolerant implementation. *The Review of Economic Studies*, 69(3):589–610.

Ely, J. C. and Peski, M. (2006). Hierarchies of belief and interim rationalizability. *Theoretical Economics*, 1(1):19–65.

Gavan, M. J. and Penta, A. (2023). Safe implementation. *BSE working paper.*

Jackson, M. O. (1991). Bayesian implementation. *Econometrica: Journal of the Econometric Society*, pages 461–477.

Jackson, M. O. (1992). Implementation in undominated strategies: A look at bounded mechanisms. *The Review of Economic Studies*, 59(4):757–775.

Jain, R., Korpela, V., and Lombardi, M. (2022). Two-player rationalizable implementation. *Available at SSRN 4302053.*

Jain, R. and Lombardi, M. (2022). On interim rationalizable monotonicity. *Available at SSRN 4106795.*

Jain, R., Lombardi, M., and Müller, C. (2023). An alternative equivalent formulation for robust implementation. *Games and Economic Behavior*, 142:368–380.

Jain, R., Lombardi, M., and Penta, A. (2024). On robust implementation. *mimeo.*

Kunimoto, T., Saran, R., and Serrano, R. (2023). Interim rationalizable implementation of functions. *Mathematics of Operations Research, forthcoming.*

Maskin, E. (1999). Nash equilibrium and welfare optimality. *The Review of Economic Studies*, 66(1):23–38.

Müller, C. (2016). Robust virtual implementation under common strong belief in rationality. *Journal of Economic Theory*, 162:407–450.

Müller, C. (2020). Robust implementation in weakly perfect bayesian strategies. *Journal of Economic Theory*, 189:105038.

Ok, E. A. (2011). *Real analysis with economic applications.* Princeton University Press.

Ollár, M. and Penta, A. (2017). Full implementation and belief restrictions. *American Economic Review*, 107(8):2243–77.

Ollár, M. and Penta, A. (2023). A network solution to robust implementation: the case of identical but unknown distributions. *Review of Economic Studies*, 90(5):2517–2554.

Ollár, M. and Penta, A. (2024). Incentive compatibility and belief restrictions. *BSE working paper.*

Oury, M. and Tercieux, O. (2012). Continuous implementation. *Econometrica*, 80(4):1605–1637.

Pearce, D. G. (1984). Rationalizable strategic behavior and the problem of perfection. *Econometrica: Journal of the Econometric Society*, pages 1029–1050.

Penta, A. (2012). Higher order uncertainty and information: Static and dynamic games. *Econometrica*, 80(2):631–660.

Penta, A. (2015). Robust dynamic implementation. *Journal of Economic Theory*, 160:280–316.

Xiong, S. (2023). Rationalizable implementation of social choice functions: complete characterization. *Theoretical Economics.*

A. Proof of Theorem 1

Fix any $\mathcal{T}$. Before proving this result, we need to connect IIM with a condition that we call β^{lim}-Recursive No Worst Alternative Condition.

A.1. β^{lim}-Recursive No Worst Alternative Condition and its relationship with IIM

For all $\mathcal{T}$, all $\beta \in \mathcal{B}$, all $\alpha \in \Omega$ and all $i \in I$, let $T_i^*\left(W^{\alpha,\beta}\right) \subseteq T_i$ be defined by

$$T_i^*\left(W^{\alpha,\beta}\right) := \left\{ t_i \in T_i \;\middle|\; \begin{array}{l} \text{there exists } y_i : C_i(t_i, W_{-i}^{\alpha,\beta}) \to Y_i(\beta) \text{ s.t.} \\ \bigcap\limits_{\mu^i \in C_i(t_i, W_{-i}^{\alpha,\beta})} \mathcal{SL}_i(y_i(\mu^i), \mu^i, t_i) \neq \emptyset, \end{array} \right\} \tag{12}$$

where $\mathcal{SL}_i(y_i(\mu^i), \mu^i, t_i)$ is the strict lower contour set of player i at $(y_i(\mu^i), \mu^i, t_i)$.

Definition 18. For all $\mathcal{T}$ and all $\beta \in \mathcal{B}$, $f : T \to X$ satisfies β-Recursive No Worst Alternative (β-RNWA) on $\mathcal{T}$ provided that for all $(i, t_i) \in I \times T_i$, either $t_i \in T_i^*\left(W^{0,\beta}\right)$ or there exists $\hat{\alpha}(t_i) \in \Omega \setminus \{0\}$ s.t. $t_i \in T_i^*\left(W^{\hat{\alpha}(t_i),\beta}\right)$ and $t_i \in T_i^{*c}\left(W^{\gamma,\beta}\right)$ for all $\gamma \in \Omega$ s.t. $\gamma < \hat{\alpha}(t_i)$.[19]

The recursive characterization of IIM provides a common language to connect IIM with the recursive-NWA condition. This section provides a precise relationship between IIM and β^{lim}-RNWA. The following properties are consequences of IIM.

Lemma 5. Suppose that $f : T \to X$ satisfies IIM on $\mathcal{T}$. For all $(i, t_i, \alpha) \in I \times T_i \times \Omega$:

1. If $W_i^{\alpha+1,\beta^{lim}}(t_i) \neq W_i^{\alpha,\beta^{lim}}(t_i)$, then $t_i \in T_i^*(W^{\alpha,\beta^{lim}})$.
2. If $t_i \in T_i^{*c}\left(W^{\alpha,\beta^{lim}}\right)$, then $W_i^{\alpha,\beta^{lim}}(t_i) = W_i^{\alpha+1,\beta^{lim}}(t_i) = T_i$.
3. If $T_i^*(\beta^{lim}) \neq T_i$, then $\beta_i^{lim}(t_i) = T_i$ for some $t_i \in T_i$.

Proof. See Appendix C. ■

Theorem 4. If $f : T \to X$ satisfies IIM on $\mathcal{T}$, then either f satisfies β^{lim}-RNWA or there exists $i \in I$ s.t. for all $t_i, t_i' \in T_i$, $f(t_i, t_{-i}) = f(t_i', t_{-i})$ for all $t_{-i} \in T_{-i}$.

Proof. See Appendix C. ■

[19] $T_i^{*c}\left(W^{\gamma,\beta}\right)$ denotes the complement of $T_i^*\left(W^{\gamma,\beta}\right)$.

To avoid trivialities, in what follows, we say that $f : T \to X$ is *non-trivial* provided that for all $i \in I$, there exists $t_i, t_i' \in T_i$ s.t. $f(t_i, t_{-i}) \neq f(t_i', t_{-i})$ for some $t_{-i} \in T_{-i}$.

The following result is useful in defining *Rule 3* of the mechanism.

Lemma 6. For all $i \in I$ and all $t_i \in T_i$, there exists $\hat{y}_i \in X$ s.t. for all $\phi_i \in \Delta(\Theta_0 \times \Theta_{-i})$, there exists $y_i \in X$ s.t.

$$\sum_{(\theta_0,\theta_{-i}) \in \Theta_0 \times \Theta_{-i}} \phi_i(\theta_0, \theta_{-i})\, u_i\left(y_i, \theta_0, \hat{\theta}_i(t_i), \theta_{-i}\right) > \sum_{(\theta_0,\theta_{-i}) \in \Theta_0 \times \Theta_{-i}} \phi_i(\theta_0, \theta_{-i})\, u_i\left(\hat{y}_i, \theta_0, \hat{\theta}_i(t_i), \theta_{-i}\right). \tag{13}$$

Proof. Fix any $i \in I$ and any $t_i \in T_i(\beta^*)$. Lemma 13 (stated and proved below) implies that $t_i \in T_i^*(\beta^*)$. (12) implies that there exists $\bar{y}_i \in Y_i(\beta^{lim})$ s.t. for all $\mu^i \in C_i(t_i, \beta_{-i}^{lim})$, there exists $y_i \in Y_i(\beta^{lim})$ s.t. the inequality in (12) holds. Since $\beta^* \subseteq \beta^{lim}$, it follows that there exists $\bar{y}_i \in Y_i(\beta^{lim})$ s.t. for all $\mu^i \in C_i(t_i, \beta_{-i}^{id})$, there exists $y_i \in Y_i(\beta^{lim})$ such that (12) holds. Fix any $t_i \in T_i$. Observe that $\phi_i \circ \left(\text{marg}_{T_{-i}} \kappa(t_i)\right) \in C_i(t_i, \beta_{-i}^{id})$ for all $\phi_i \in \Delta(\Theta_0 \times \Theta_{-i})$. We denote by $\phi_i \circ \left(\text{marg}_{T_{-i}} \kappa(t_i)\right) \equiv \zeta(t)$. Therefore, it holds that

$$\sum \zeta(t_i)[\theta_0, \theta_{-i}, \hat{t}_{-i}] \left[u_i\left(y_i\left(\hat{t}_{-i}\right), \theta_0, \hat{\theta}_i(t_i), \theta_{-i}\right) - u_i\left(\bar{y}_i\left(\hat{t}_{-i}\right), \theta_0, \hat{\theta}_i(t_i), \theta_{-i}\right)\right] > 0.$$

By setting

$$y_i = \sum_{(\theta_0, \hat{t}_{-i}) \in \Theta_0 \times \hat{T}_{-i}} \left(\text{marg}_{T_{-i}} \kappa(t_i)\left[\theta_0, \hat{t}_{-i}\right]\right) y_i\left(\hat{t}_{-i}\right)$$

and

$$\hat{y}_i = \sum_{(\theta_0, \hat{t}_{-i}) \in \Theta_0 \times \hat{T}_{-i}} \left(\text{marg}_{T_{-i}} \kappa(t_i)\left[\theta_0, \hat{t}_{-i}\right]\right) \bar{y}_i\left(\hat{t}_{-i}\right),$$

and by noting that $y_i, \hat{y}_i \in X$, the inequality in (13) follows for i. Since the choice of $i \in I$ s.t. $T_i(\beta^{lim}) \neq \emptyset$ was arbitrary, the statement follows. ■

Since $T_i(\beta^{lim}) = T_i$ for all $i \in I$ and since Lemma 6 guarantees the existence of the lottery $\hat{y}_i \in \Delta\left(A\right)$ for all $i \in I$, let us define the lottery $\hat{y}$ by

$$\hat{y} = \frac{1}{I}\sum_{i\in I}\hat{y}_i.$$

The following result is useful in defining *Rule 2* of the mechanism.

Lemma 7. If $f : T \to X$ satisfies β^{lim}-RNWA and it is non-trivial, then for all $(i,t_i) \in I \times T_i$, there exist $y_i(t_i,\cdot) : C_i(t_i, W_{-i}^{\hat{\alpha}(t_i),\beta^*}) \to Y_i(\beta^*)$ and an allocation $\bar{y}_i[\hat{\alpha}(i)] \in Y_i(\beta^*)$ s.t. $\hat{y} \in \text{Supp}(\bar{y}_i[\hat{\alpha}(i)])$ and

$$\bar{y}_i[\hat{\alpha}(i)] \in \left(\bigcap_{t_i\in T_i}\left(\bigcap_{\mu_i\in C_i\left(t_i,W^{\hat{\alpha}(t_i),\beta^*}\right)}\mathcal{SL}_i\left(y_i\left(t_i,\mu^i\right),\mu^i,t_i\right)\right)\right) \tag{14}$$

Proof. Suppose that $f : T \to X$ satisfies β^{lim}-RNWA on $\mathcal{T}$ and it is non-trivial. Since it is non-trivial, it holds that $T_i^*\left(\beta^*\right) = T_i$ for all $i \in I$.

Fix any $(i,t_i) \in I \times T_i$. Since f satisfies β^{lim}-RNWA, it follows that either $t_i \in T_i^*\left(W^{0,\beta}\right)$ or there exists $\alpha(t_i) \in \Omega\backslash\{0\}$ s.t. $t_i \in T_i^*\left(W^{\alpha(t_i),\beta}\right)$ and $t_i \in T_i^{*c}\left(W^{\gamma,\beta}\right)$ for all $\gamma \in \Omega$ s.t. $\gamma < \hat{\alpha}(t_i)$. Thus, let $\hat{\alpha}(t_i) = \alpha(t_i)$ if $\alpha(t_i) \neq 0$, otherwise, let $\hat{\alpha}(t_i) = 0$. Since $t_i \in T^*\left(W^{\hat{\alpha}(t_i),\beta}\right)$, it follows from (12) that there exists $y_i : C_i(t_i, W_{-i}^{\hat{\alpha}(t_i),\beta^{lim}}) \to Y_i(\beta^{lim})$ s.t.

$$\bigcap_{\mu^i\in C_i(t_i,W_{-i}^{\hat{\alpha}(t_i),\beta^{lim}})}\mathcal{SL}_i(y_i(\mu^i),\mu^i,t_i) \neq \emptyset. \tag{15}$$

Since the choice of $t_i \in T_i$ was arbitrary and f satisfies β^{lim}-RNWA on $\mathcal{T}$, it follows that for all $t_i \in T_i = T^*(W^{\hat{\alpha}(t_i),\beta^*})$, there exists $y_i : C_i(t_i, W_{-i}^{\hat{\alpha}(t_i),\beta^{lim}}) \to Y_i(\beta^{lim})$ s.t. (15) holds. For all $t_i \in T_i$, let $\bar{y}_i\left[\hat{\alpha}\left(t_i\right)\right] \in Y_i(\beta^{lim})$ be s.t.

$$\bar{y}_i\left[\hat{\alpha}\left(t_i\right)\right] \in \left(\bigcap_{\mu_i\in C_i\left(t_i,W^{\hat{\alpha}(t_i),\beta^{lim}}\right)}\mathcal{SL}_i\left(y_i(\mu^i),\mu^i,t_i\right)\right)$$

Let us define the allocation $\tilde{y}_i[\hat{\alpha}(i)]$ by

$$\tilde{y}_i[\hat{\alpha}(i)] = \sum_{t_i \in T_i} p(t_i) \cdot \bar{y}_i[\hat{\alpha}(t_i)].$$

where $p \in \Delta(T_i)$ is a probability distribution with full support. Since $\bar{y}_i\left[\hat{\alpha}\left(t_i\right)\right] \in Y_i(\beta^{lim})$ for all $t_i \in T_i$ and since $Y_i(\beta^{lim})$ is a convex set, it follows that $\tilde{y}_i[\hat{\alpha}(i)] \in Y_i(\beta^{lim})$.

Fix any $t_i \in T_i$. Let us define $y_i(t_i, \cdot) : C_i(t_i, W_{-i}^{\hat{\alpha}(t_i),\beta^*}) \to Y_i(\beta^*)$ by

$$y_i(t_i, \mu^i) = \sum_{t_i' \in T_i \setminus \{t_i\}} p(t_i') \cdot \bar{y}_i[\hat{\alpha}(t_i)] + p(t_i) \cdot y_i(\mu^i).$$

Since $\bar{y}_i[\alpha(t_i)] \in \text{Supp}(\bar{y}_i[\alpha(i)])$, it follows that

$$\tilde{y}_i[\hat{\alpha}(i)] \in \left(\bigcap_{t_i \in T_i} \left(\bigcap_{\mu_i \in C_i\left(t_i, W^{\hat{\alpha}(t_i),\beta^*}\right)} \mathcal{SL}_i\left(y_i\left(t_i, \mu^i\right), \mu^i, t_i\right) \right) \right). \tag{16}$$

Since (16) holds, we can choose an $\varepsilon > 0$ sufficiently small s.t. the allocation $\bar{y}_i[\hat{\alpha}(i)]$ defined by

$$\bar{y}_i[\hat{\alpha}(i)] = (1 - \varepsilon)\, \tilde{y}_i[\hat{\alpha}(i)] + \varepsilon \hat{y} \tag{17}$$

is s.t. the statement follows.

■

A.2. Proof of the "IF" part of Theorem 1

Proof. Assume that $\mathcal{M}$ ICRR-implements f on $\mathcal{T}$. Let us show that f satisfies IIM on $\mathcal{T}$. For all $i \in I$ and all $t_i \in T_i$, let $\tilde{\beta}_i(t_i)$ be defined by

$$\tilde{\beta}_i(t_i) = \left\{ t_i' \in T_i \middle| R_i(t_i') \bigcap R_i(t_i) \neq \emptyset \right\}. \tag{18}$$

Since f is ICRR-implementable, we have that $\tilde{\beta}$ is an acceptable deception for f

on $\mathcal{T}$; that is, for all $t, t' \in T$, if $t' \in \tilde{\beta}(t)$, then $f(t) = f(t')$. Moreover, ICRR-implementation implies that *every* pure strategy $\sigma \in R$ is a pure BNE. Fix any $\sigma \in R$. It follows that $\sigma \circ \beta'$ is a pure BNE for all $\beta' \in \mathcal{B}(\tilde{\beta}) = \{\hat{\beta} \in \mathcal{B} : \hat{\beta}(t) \in \tilde{\beta}(t)$ for all $t \in T\}$. The following claim delivers the result.

Claim 1. For all $\beta \in \mathcal{B}$, if β is non-refutable w.r.t. $\tilde{\beta}$, then $\beta \subseteq \tilde{\beta}$.

Proof. Take any $\beta \in \mathcal{B}$ s.t. $\beta \nsubseteq \tilde{\beta}$. For all $i \in I$ and all $t_i \in T_i$, let $\Sigma_i^{\beta}[\sigma](t_i) = \{\sigma_i(t_i') \in M_i | t_i' \in \beta_i(t_i)\}$. Then, $\Sigma_i^{\beta}[\sigma]$ is a correspondence from T_i to $2^{M_i} \backslash \{\emptyset\}$. Since $\mathcal{M}$ ICRR-implements f, it follows that $\Sigma^{\beta}[\sigma]$ cannot be a best-reply set in $(\mathcal{M}, \mathcal{T})$. To see it, assume, to the contrary, that $\Sigma^{\beta}[\sigma]$ is a best-reply set in $(\mathcal{M}, \mathcal{T})$. Since $\beta \nsubseteq \tilde{\beta}$, it follows that there exists $(i, t_i, t_i') \in I \times T_i \times T_i$ s.t. $t_i' \in \beta_i(t_i)$ and $t_i' \notin \tilde{\beta}_i(t_i)$. Thus, $R_i(t_i) \bigcap R_i(t_i') = \emptyset$, by (18). Since $\Sigma^{\beta}[\sigma]$ is a best-reply set in $(\mathcal{M}, \mathcal{T})$ and since σ is a pure BNE, it holds that $\sigma_i(t_i') \in R_i(t_i) \bigcap R_i(t_i')$, which is a contradiction. Thus, $\Sigma^{\beta}[\sigma]$ is not a best-reply set in $(\mathcal{M}, \mathcal{T})$. Then, for some $\left(i, t_i, \hat{t}_i\right) \in I \times T_i \times \beta_i(t_i)$, it holds that $\sigma_i\left(\hat{t}_i\right) \notin r_i(\mu^i, \hat{\theta}_i(t_i))$ for all $\mu^i \in C_i(t_i, \Sigma_{-i}^{\beta}[\sigma])$, and so

$$\sum_{(\theta_0, \theta_{-i}, m_{-i}) \in \Theta_0 \times \Theta_{-i} \times M_{-i}} \mu^i[\theta_0, \theta_{-i}, m_{-i}] \left[u_i\left(g(m_i, m_{-i}), \theta_0, \hat{\theta}_i(t_i), \theta_{-i}\right)\right] > \sum_{(\theta_0, \theta_{-i}, m_{-i}) \in \Theta_0 \times \Theta_{-i} \times M_{-i}} \mu^i[\theta_0, \theta_{-i}, m_{-i}] \left[u_i\left(g\left(\sigma_i\left(\hat{t}_i\right), m_{-i}\right), \theta_0, \hat{\theta}_i(t_i), \theta_{-i}\right)\right] \tag{19}$$

for some $m_i \in M_i$. Let $\tau^i : C_i(t_i, \Sigma_{-i}^{\beta}[\sigma]) \rightarrow C_i(t_i, \beta_{-i})$ be defined as follows:

$$\tau^i(\mu^i)[\theta_0, \theta_{-i}, \hat{t}_{-i}] = \mu^i(\theta_0, \theta_{-i}, \sigma_{-i}(\hat{t}_{-i})) \tag{20}$$

for all $(\theta_0, \theta_{-i}, \hat{t}_{-i}) \in \Theta_0 \times \Theta_{-i} \times \hat{T}_{-i}$. It can be shown that τ^i is a surjection.[20]

Fix any $\tau^i \in C_i(t_i, \beta_{-i})$. Then, there exists $\mu^i \in C_i(t_i, \Sigma_{-i}^{\beta}[\sigma])$ s.t. (19) holds, and

[20] To see it, suppose that $\tau^i \in C_i(t_i, \beta_{-i})$. Then, for all $(\theta_0, \theta_{-i}, \hat{t}_{-i}) \in Supp(\tau^i)$, let $\mu^i[\theta_0, \theta_{-i}, m_{-i}] = \delta_{\sigma_{-i}(\hat{t}_{-i})}$. By definition, it follows that $\mu^i \in C_i(t_i, \Sigma_{-i}^{\beta}[\sigma])$.

so

$$\begin{aligned}&\sum\tau^i(\mu^i)\left[\theta_0,\theta_{-i},\hat{t}_{-i}\right]\left[u_i\left(g\left(m_i,\sigma_{-i}(\hat{t}_{-i})\right),\theta_0,\hat{\theta}_i(t_i),\theta_{-i}\right)\right] \qquad > \\ &\sum\tau^i(\mu^i)\left[\theta_0,\theta_{-i},\hat{t}_{-i}\right]\left[u_i\left(g\left(\sigma_i(\hat{t}_i),\sigma_{-i}(\hat{t}_{-i})\right),\theta_0,\hat{\theta}_i(t_i),\theta_{-i}\right)\right] = U_i(\hat{t}_i,\tau^i(\mu^i);f)\end{aligned} \tag{21}$$

for some $m_i \in M_i$. Let us define $y_i^*(\cdot)$ by $y_i^*(\cdot) = g(m_i, \sigma_{-i} \circ \beta^*_{-i}(\cdot))$.

We are left to show that $y_i^* \in Y_i(\tilde{\beta})$. To see it, take any $\beta' \in \mathcal{B}(\tilde{\beta})$. Then, $\sigma \circ \beta'$ is a pure BNE by ICRR-implementability of f. This implies that for all $t_i \in T_i$, it holds that

$$\begin{aligned}&\sum_{(\theta_0,\theta_{-i},t_{-i})\in\Theta_0\times\theta_{-i}\times T_{-i}}\kappa\left(t_i\right)\left[\theta_0,t_{-i}\right]u_i\left(g\left(\sigma\circ\beta'\left(t\right)\right),\theta_0,\hat{\theta}_i(t_i),\theta_{-i}\right) \qquad \geq \\ &\sum_{(\theta_0,\theta_{-i},t_{-i})\in\Theta_0\times\Theta_{-i}\times T_{-i}}\kappa\left(t_i\right)\left[\theta_0,t_{-i}\right]u_i\left(g\left(m_i,\sigma_{-i}\circ\beta'_{-i}\left(t_{-i}\right)\right),\theta_0,\hat{\theta}_i(t_i),\theta_{-i}\right)\end{aligned}$$

for all $m_i \in M_i$. Since $\mathcal{M}$ ICRR-implements f, it follows that for all $t_i \in T_i$,

$$\begin{aligned}&\sum_{(\theta_0,\theta_{-i},t_{-i})\in\Theta_0\times\Theta_{-i}\times T_{-i}}\kappa\left(t_i\right)\left[\theta_0,t_{-i}\right]u_i\left(f\left(t_i,\beta'_{-i}(t_{-i})\right),\theta_0,\hat{\theta}_i(t_i),\theta_{-i}\right) \qquad \geq \\ &\sum_{(\theta_0,\theta_{-i},t_{-i})\in\Theta_0\times\Theta_{-i}\times T_{-i}}\kappa\left(t_i\right)\left[\theta_0,t_{-i}\right]u_i\left(g\left(\left(m_i,\sigma_{-i}\circ\beta'_{-i}\left(t_{-i}\right)\right)\right),\theta_0,\hat{\theta}_i(t_i),\theta_{-i}\right)\end{aligned} \tag{22}$$

for all $m_i \in M_i$, where $g\left(m_i, \sigma_{-i} \circ \beta'_{-i}\left(t_{-i}\right)\right) = y_i^* \circ \beta'_{-i}\left(t_{-i}\right)$. Since the choice of $\beta' \in \mathcal{B}(\tilde{\beta})$ was arbitrary, it follows that $y_i^* \in Y_i(\tilde{\beta})$. ■

■

A.3. Proof of the "ONLY IF" part of Theorem 1

Suppose that $f : T \to X$ satisfies IIM on $\mathcal{T}$. Theorem 3 implies that β^{lim} is acceptable for f on $\mathcal{T}$. Let us suppose that f is non-trivial.[21] It follows from Theorem 4 that

[21] $f : T \to X$ is *non-trivial* provided that for all $i \in I$, there exists $t_i, t_i' \in T_i$ s.t. $f\left(t_i, t_{-i}\right) \neq f\left(t_i', t_{-i}\right)$ for some $t_{-i} \in T_{-i}$.

β^{lim}-RNWA. Let us construct our implementing mechanism $\mathcal{M}$. For all $i \in I$, let

$$M_i = M_i^1 \times M_i^2 \times M_i^3 \times M_i^4,$$

where

$$M_i^1 = T_i,\ M_i^2 = \mathbb{N},\ M_i^3 = Y_i^*(\beta^{lim}) \text{ and } M_i^4 = X^*,$$

where $\mathbb{N}$ is the set of natural numbers, $Y_i^*(\beta^{lim})$ is a countable, dense subset of $Y_i(\beta^{lim})$, and X^* is a countable, dense subset of X. For all $m \in M$, let $g : M \to \Delta(A)$ be defined as follows.

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

Rule 2: For all $i \in I$, if $m_j^2 = 1$ for all $j \in I \backslash \{i\}$ and $m_i^2 > 1$, then

$$g(m) = m_i^3\left(m_{-i}^1\right)\left(1 - \frac{1}{1+m_i^2}\right) \oplus \bar{y}_i[\alpha(i)]\left(m_{-i}^1\right)\left(\frac{1}{1+m_i^2}\right), \tag{23}$$

where the existence of the allocation $\bar{y}_i[\alpha(i)] \in Y_i(\beta^{lim})$ is guaranteed by Lemma 7.

Rule 3: Otherwise, for each $i \in I$, m_i^4 is picked with probability $\frac{1}{I}\left(1 - \frac{1}{1+m_i^2}\right)$ and $\hat{y}_i$ is picked with probability $\frac{1}{I}\left(\frac{1}{1+m_i^2}\right)$; that is,

$$g(m) = \frac{1}{I}\left[m_i^4\left(1 - \frac{1}{1+m_i^2}\right) \oplus \hat{y}_i\left(\frac{1}{1+m_i^2}\right)\right], \tag{24}$$

where the existence of $\hat{y}_i$ is guaranteed by Lemma 6.

In what follows, we prove that $\mathcal{M}$ ICRR-implements f on $\mathcal{T}$. The following lemmata will help us to complete the proof. To state these results, let us introduce the following definitions. For all $\beta \subseteq \beta^{lim}$ and all $i \in I$, define $\Sigma_i^{\beta_i} : T_i \to 2^{M_i} \backslash \{\emptyset\}$ by

$$\Sigma_i^{\beta_i}(t_i) = \left\{m_i \in M_i | m_i^1 \in \beta_i(t_i)\right\}, \tag{25}$$

and $\tilde{\Sigma}_i^{\beta_i} : T_i \to 2^{M_i} \backslash \{\emptyset\}$ by

$$\tilde{\Sigma}_i^{\beta_i}(t_i) = \left\{ m_i \in \Sigma_i^{\beta_i}(t_i) \, | m_i^2 = 1 \right\}. \tag{26}$$

Lemma 8. For all $\beta \subseteq \beta^{lim}$, $\tilde{\Sigma}^{\beta}$ is a best-reply set in $(\mathcal{M}, \mathcal{T})$.

Proof. Fix any $\beta \subseteq \beta^{lim}$. Let $\tilde{\beta}$ be any single-valued deception profile selected from β. For all $i \in I$, let $\sigma_i : T_i \to M_i$ be defined by $\sigma_i(t_i) = \left(\tilde{\beta}_i(t_i), 1, \cdot, \cdot \right)$. As a first step, we show that $m_i = (t_i, 1, \cdot, \cdot) \in r_i(\mu^i(t_i, \sigma_{-i}), \hat{\theta}_i(t_i))$ for all $(i, t_i) \in I \times T_i$. Thus, fix any $(i, t_i) \in I \times T_i$. Then:

$$\begin{aligned} U_i(m_i, \mu^i, \hat{\theta}_i(t_i)) &= \sum_{(\theta_0, \theta_{-i}, m_{-i}) \in \Theta_0 \times \Theta_{-i} \times M_{-i}} \mu^i(t_i, \sigma_{-i}) \, u_i\big(g(m_i, m_{-i}), \big(\theta_0, \hat{\theta}_i(t_i), \theta_{-i}\big)\big) \\ &= \sum_{(\theta_0, t_{-i}) \in \Theta_0 \times T_{-i}} \kappa(t_i)[\theta_0, t_{-i}] \, u_i\left(f\left(t_i, \tilde{\beta}_{-i}(t_{-i}) \right), \left(\theta_0, \hat{\theta}_i(t_i), \hat{\theta}_{-i}(t_{-i}) \right) \right) \\ &= \sum_{(\theta_0, t_{-i}) \in \Theta_0 \times T_{-i}} \kappa(t_i)[\theta_0, t_{-i}] \, u_i\left(f(t_i, t_{-i}), \left(\theta_0, \hat{\theta}_i(t_i), \hat{\theta}_{-i}(t_{-i}) \right) \right). \end{aligned}$$

Moreover, for all $\hat{m}_i \in M_i$ s.t. $\hat{m}_i^2 = 1$, it holds that

$$\begin{aligned} U_i(\hat{m}_i, \mu^i, \hat{\theta}_i(t_i)) &= \sum_{(\theta_0, \theta_{-i}, m_{-i}) \in \Theta_0 \times \Theta_{-i} \times M_{-i}} \mu^i(t_i, \sigma_{-i}) \, u_i\big(g(\hat{m}_i, m_{-i}), \big(\theta_0, \hat{\theta}_i(t_i), \theta_{-i}\big)\big) \\ &= \sum_{(\theta_0, t_{-i}) \in \Theta_0 \times T_{-i}} \kappa(t_i)[\theta_0, t_{-i}] \, u_i\left(f\left(\hat{m}_i^1, \tilde{\beta}_{-i}(t_{-i}) \right), \left(\theta_0, \hat{\theta}_i(t_i), \hat{\theta}_{-i}(t_{-i}) \right) \right). \\ &= \sum_{(\theta_0, t_{-i}) \in \Theta_0 \times T_{-i}} \kappa(t_i)[\theta_0, t_{-i}] \, u_i\left(f(\hat{m}_i^1, t_{-i}), \left(\theta_0, \hat{\theta}_i(t_i), \hat{\theta}_{-i}(t_{-i}) \right) \right). \end{aligned}$$

Since f satisfies β^{id}-IIC, it follows that $U_i(m_i, \mu^i, \hat{\theta}_i(t_i)) \geq U_i(\hat{m}_i, \mu^i, \hat{\theta}_i(t_i))$.

Finally, for all $\hat{m}_i \in M_i$ s.t. $\hat{m}_i^2 \neq 1$, it holds for all $y_i \in Y_i(\beta^{lim})$,

$$\begin{aligned} U_i(\hat{m}_i,\mu^i,\hat{\theta}_i(t_i)) &= \sum_{(\theta_0,\theta_{-i},m_{-i})\in\Theta_0\times\Theta_{-i}\times M_{-i}} \mu^i\left(t_i,\sigma_{-i}\right) u_i\big(g(\hat{m}_i,m_{-i}),\big(\theta_0,\hat{\theta}_i(t_i),\theta_{-i}\big)\big) \\ &= \sum_{(\theta_0,t_{-i})\in\Theta_0\times T_{-i}} \kappa\left(t_i\right)\left[\theta_0,t_{-i}\right] u_i\left(y_i(\tilde{\beta}_i(t_{-i})),\big(\theta_0,\hat{\theta}_i(t_i),\hat{\theta}_{-i}(t_{-i})\big)\right). \end{aligned}$$

It follows that $m_i=(t_i,1,\cdot,\cdot)\in r_i(\mu^i\left(t_i,\sigma_{-i}\right),\hat{\theta}_i(t_i))$.

Since the choice of $(i,t_i)\in I\times T_i$ was arbitrary, it follows that $m_i=(t_i,1,\cdot,\cdot)\in r_i(\mu^i_i\left(t_i,\sigma_{-i}\right),\hat{\theta}_i(t_i))$ for all $(i,t_i)\in I\times T_i$.

Fix any $(i,t_i)\in I\times T_i$. Since β^{lim} is acceptable for f on $\mathcal{T}$, it follows that $f(t_i,\cdot)=f(\tilde{\beta}_i(t_i),\cdot)=f(\tilde{t}_i,\cdot)$ for all $\tilde{t}_i\in\beta_i(t_i)$. Thus, there exists $\mu^i\left(t_i,\sigma_{-i}\right)\in C_i(t_i,\tilde{\Sigma}^{\beta}_{-i})$ s.t. $\tilde{\Sigma}^{\beta_i}_i(t_i)\subseteq r_i(\mu^i\left(t_i,\sigma_{-i}\right),\hat{\theta}_i(t_i))$. Since the choice of (i,t_i) was arbitrary, it follows that $\tilde{\Sigma}^{\beta}$ is a best-reply set in $(\mathcal{M},\mathcal{T})$. ■

Lemma 9. For all $(\alpha,i,t_i)\in\Omega\times I\times T^*_i(W^{\alpha,\beta^{lim}})$ and all $\mu^i\in C_i\left(t_i,\Sigma_{-i}(W^{\alpha,\beta^{lim}}_{-i})\right)$, if $m_i\in r_i(\mu^i,\hat{\theta}_i(t_i))$, then $m^2_i=1$, $\mu^i\in C_i\left(t_i,\tilde{\Sigma}_{-i}(W^{\alpha,\beta^{lim}}_{-i})\right)$ and $m^1_i\in W^{\alpha+1,\beta^{lim}}_i(t_i)$.

Proof. Fix any $(\alpha,i,t_i)\in\Omega\times I\times T^*_i(W^{\alpha,\beta^{lim}})$. Since f satisfies $\beta^{lim}-RNWA$ and $t_i\in T^*_i(W^{\alpha,\beta^{lim}})$, it follows that $\alpha\geq\hat{\alpha}(t_i)$. Take any $\mu^i\in C_i(t_i,\Sigma_{-i}(W^{\alpha,\beta^{lim}}_{-i}))$ and let $m_i\in r_i(\mu^i,\hat{\theta}_i(t_i))$. We proceed by contradiction. Suppose that $m^2_i>1$.

We need to distinguish whether *Rule 2* applies or *Rule 3* applies. To this end, let us note that $\mu^i\in C_i(t_i,\Sigma_{-i}(W^{\alpha,\beta^{lim}}_{-i}))$ can be decomposed as follows. To save writing, let $A=\Theta_0\times\Theta_{-i}\times\tilde{\Sigma}_{-i}(W^{\alpha,\beta^{lim}}_{-i})$ and $B=\Theta_0\times\Theta_{-i}\times\left(\Sigma_{-i}(W^{\alpha,\beta^{lim}}_{-i})\setminus\tilde{\Sigma}_{-i}(W^{\alpha,\beta^{lim}}_{-i})\right)$. Let $Prob(A)$ be denoted by ν, which is defined by

$$\nu=\sum_{(\theta_0,\theta_{-i},m_{-i})\in A}\mu^i[\theta_0,\theta_{-i},m_{-i}]. \tag{27}$$

and $Prob(B) = 1 - \nu$. Let $\bar{\mu}^i$ be defined over $\Theta_0 \times \Theta_{-i} \times \hat{T}_{-i}$ by

$$\bar{\mu}^i[\theta_0, \theta_{-i}, m^1_{-i}] = \frac{\sum\limits_{\bar{m}_{-i} \in \tilde{\Sigma}_{-i}(W^{\alpha,\beta^{lim}}_{-i})[m^1_{-i}]} \mu^i[\theta_0, \theta_{-i}, \bar{m}_{-i}]}{\nu}. \tag{28}$$

for all $(\theta_0, \theta_{-i}, m^1_{-i}) \in \Theta_0 \times \Theta_{-i} \times \hat{T}_{-i}$, where $\tilde{\Sigma}_{-i}(W^{\alpha,\beta^{lim}}_{-i})[m^1_{-i}] = \{\hat{m}_{-i} \in \tilde{\Sigma}_{-i}(W^{\alpha,\beta^{lim}}_{-i}) : m^1_{-i} = \hat{m}^1_{-i}\}$. It can be checked that $\sum\limits_{(\theta_0,\theta_{-i},m_{-i}) \in A} \bar{\mu}^i[\theta_0, \theta_{-i}, m_{-i}] = 1$ and that $\bar{\mu}^i \in C_i(t_i, W^{\alpha,\beta^{lim}}_{-i})$.

Moreover, for all $(\theta_0, \theta_{-i}) \in \Theta_0 \times \Theta_{-i}$, let $\phi_i(\theta_0, \theta_{-i})$ be defined by

$$\phi_i(\theta_0, \theta_{-i}) = \frac{\sum\limits_{m_{-i} \in \Sigma_{-i}(W^{\alpha,\beta^{lim}}_{-i}) \backslash \tilde{\Sigma}_{-i}(W^{\alpha,\beta^{lim}}_{-i})} \mu^i[\theta_0, \theta_{-i}, m_{-i}]}{1 - \nu}. \tag{29}$$

Thus, the expected utility of t_i of playing m_i with $m^2_i > 1$ is given by

$$\begin{aligned} U_i(m_i, \mu^i, \hat{\theta}_i(t_i)) = \nu &\sum_{(\theta_0,\theta_{-i},m_{-i}) \in A} \bar{\mu}^i[\theta_0, \theta_{-i}, m_{-i}] u_i\Big(g(m_i, m_{-i}), \big(\theta_0, \hat{\theta}_i(t_i), \theta_{-i}\big)\Big) \\ &+ (1 - \nu) \sum_{(\theta_0,\theta_{-i},m_{-i}) \in B} \phi_i(\theta_0, \theta_{-i}) u_i\Big(g(m_i, m_{-i}), \big(\theta_0, \hat{\theta}_i(t_i), \theta_{-i}\big)\Big), \end{aligned} \tag{30}$$

which simplifies to

$$\begin{aligned} &\nu\Big[\left(1 - \frac{1}{1 + m^2_i}\right) U_i(m^3_i(\cdot), \bar{\mu}^i, \hat{\theta}_i(t_i)) + \left(\frac{1}{1 + m^2_i}\right) U_i(\overline{y}_i[\hat{\alpha}(i)], \bar{\mu}^i, \hat{\theta}_i(t_i))\Big] \\ &+ (1 - \nu) \sum_{(\theta_0,\theta_{-i},m_{-i}) \in B} \phi_i(\theta_0, \theta_{-i}) u_i\Big[\Big(m^4_i\left(1 - \frac{1}{1 + m^2_i}\right) \oplus \hat{y}_i\left(\frac{1}{1 + m^2_i}\right)\Big), \big(\theta_0, \hat{\theta}_i(t_i), \theta_{-i}\big)\Big]. \end{aligned} \tag{31}$$

Since $\bar{\mu}^i \in C_i(t_i, W^{\alpha,\beta^{lim}}_{-i})$ and $t_i \in T^*_i(W^{\alpha,\beta^{lim}})$ and since $\alpha \geq \hat{\alpha}(t_i)$, Lemma 7 implies that there exists $y_i(t_i, \cdot) : C_i(t_i, W_{-i}^{\hat{\alpha}(t_i),\beta^*}) \to Y_i(\beta^*)$ s.t.

$$U_i(y_i(t_i, \bar{\mu}^i), \bar{\mu}^i, \hat{\theta}_i(t_i)) > U_i(\bar{y}_i[\hat{\alpha}(i)], \bar{\mu}^i, \hat{\theta}_i(t_i)). \tag{32}$$

Since Lemma 2 implies that $W^{\alpha,\beta^{lim}} \subseteq W^{\hat{\alpha}(t_i),\beta^{lim}}$ for all $\alpha \in \Omega$ s.t. $\alpha \geq \alpha(t_i)$, we can see that the inequality in (32) holds for all $\bar{\mu}^i \in C_i(t_i, W_{-i}^{\alpha,\beta^{lim}})$ with $\alpha \geq \alpha(t_i)$.

Furthermore, Lemma 6 implies that for all $t_i \in T_i$, there exists $y_i \in X^*$ s.t.

$$\sum_{(\theta_0,\theta_{-i})\in\Theta_0\times\Theta_{-i}} \phi_i(\theta_0, \theta_{-i}) u_i\big(y_i, \big(\theta_0, \hat{\theta}_i(t_i), \theta_{-i}\big)\big) > \sum_{(\theta_0,\theta_{-i})\in\Theta_0\times\Theta_{-i}} \phi_i(\theta) u_i\big(\hat{y}_i, \big(\theta_0, \hat{\theta}_i(t_i), \theta_{-i}\big)\big). \tag{33}$$

Since $m_i \in r_i(\mu^i, \hat{\theta}_i(t_i))$, it follows that

$$U_i(m_i^3(\cdot), \bar{\mu}^i, \hat{\theta}_i(t_i)) \geq U_i(y_i(t_i, \bar{\mu}^i), \bar{\mu}^i, \hat{\theta}_i(t_i)) \tag{34}$$

and that

$$\sum_{(\theta_0,\theta_{-i})\in\Theta_0\times\Theta_{-i}} \phi_i(\theta_0, \theta_{-i}) u_i\big(m_i^4, \big(\theta_0, \hat{\theta}_i(t_i), \theta_{-i}\big)\big) \geq \sum_{(\theta_0,\theta_{-i})\in\Theta_0\times\Theta_{-i}} \phi_i(\theta_0, \theta_{-i}) u_i\big(y_i, \big(\theta_0, \hat{\theta}_i(t_i), \theta_{-i}\big)\big). \tag{35}$$

Inequalities in (32)-(35) imply that $U_i(m_i, \mu^i, \hat{\theta}_i(t_i))$ is strictly increasing in m_i^2, which contradicts our supposition that $m_i \in r_i(\mu^i, \hat{\theta}_i(t_i))$. Thus, $m_i^2 = 1$.

Suppose that $\mu^i \notin C_i\left(t_i, \tilde{\Sigma}_{-i}(W_{-i}^{\alpha,\beta^{lim}})\right)$. Then, since $m_i^2 = 1$, either *Rule 2* applies where $m_j^2 > 1$ for some $j \in I\backslash\{i\}$ or *Rule 3* applies. In what follows, we focus only on the case that *Rule 2* applies.[22]

By the definition of g, for all $(\theta_0, \theta_i, m_{-i}) \in \text{Supp}(\mu^i(t_i))$, it holds that

$$g(m_i, m_{-i}) = \left(1 - \frac{1}{m_j^2 + 1}\right) m_j^3(m_{-j}^1) + \left(\frac{1}{m_j^2 + 1}\right) \bar{y}_j[\hat{\alpha}(j)](m_{-j}^1). \tag{36}$$

[22] When *Rule 3* applies, we can see, by the arguments provided above, that player i can find a profitable deviation.

To show that player i can gain by triggering *Rule 3*, we need to define a lottery $\hat{m}_i^4 \in X^* = M_i^4$ that can be used by player i. To this end, we first define the allocation h over M as follows: For all (m_i, m_{-i}) s.t. $(\theta_0, \theta_i, m_{-i}) \in$Supp$(\mu^i(t_i))$,

$$h(m_i, m_{-i}) = \left(1 - \frac{1}{m_j^2 + 1}\right) m_j^3(m_{-j}^1) + \left(\frac{1}{m_j^2 + 1}\right) \tilde{y}_j[\hat{\alpha}(j), \varepsilon](m_{-j}^1) \tag{37}$$

where $\tilde{y}_j[\hat{\alpha}(j), \varepsilon](m_{-j}^1) = (1 - \varepsilon)\, \bar{y}_j[\hat{\alpha}(j)](m_{-j}^1) + \varepsilon[\sum\limits_{j \neq i} \frac{1}{I}\hat{y}_j + \frac{1}{I} y_i]$ and where y_i is s.t. (13) is satisfied. Finally, let us define $\hat{m}_i^4$ by

$$\hat{m}_i^4 = \sum_{(\theta_0, \theta_{-i}, m_{-i})} \mu^i\left(t_i\right)\left[\theta_0, \theta_{-i}, m_{-i}\right] h\left(\cdot, m_{-i}\right). \tag{38}$$

Since player i's utility is strictly higher under $h(m_i, m_{-i})$ than under $g(m_i, m_{-i})$ for each $(\theta_0, \theta_{-i} a, m_{-i}) \in$Supp$(\mu^i(t_i))$ and since, moreover, player i's utility function is continuous, we can assume without loss of generality that $\hat{m}_i^4 \in \Delta^*(A) = M_i^4$.

Since player i's utility is strictly higher under $h(m_i, m_{-i})$ than under $g(m_i, m_{-i})$, for all $(\theta_0, \theta_{-i}, m_{-i}) \in$Supp$\left(\mu^i(t_i)\right)$, player i can change m_i with $m_i' \in M_i$, where its fourth component is $\hat{m}_i^4$ and its second component is $\hat{m}_i^2 > 1$, so that he can trigger *Rule 3*. Since the utility gain of player i is obtained point-wise in the Supp$(\mu^i(t_i))$, we obtain the desired contradiction. Thus, $\mu^i \in C_i\left(t_i, \tilde{\Sigma}_{-i}(W_{-i}^{\alpha, \beta^{lim}})\right)$.

Therefore, it must be the case that $m_i^1 \notin W_i^{\alpha+1, \beta^{lim}}(t_i)$. Since $m_i \in r_i(\mu^i, \hat{\theta}_i(t_i))$ and $m_i^2 = 1$ and since $\mu^i \in C_i\left(t_i, \tilde{\Sigma}_{-i}(W_{-i}^{\alpha, \beta^{lim}})\right)$, it follows that Rule 1 applies with probability 1, and so

$$U_i(m_i, \mu^i, \hat{\theta}_i(t_i)) = U_i(f(m_i^1, \cdot), \bar{\mu}^i, \hat{\theta}_i(t_i)) \tag{39}$$

Since $m_i \in r_i(\mu^i, \hat{\theta}_i(t_i))$ and since player i can never induce *Rule 3*, it follows from the definition of g that

$$U_i(f(m_i^1, \cdot), \bar{\mu}^i, \hat{\theta}_i(t_i)) \geq U_i(m_i^3, \bar{\mu}^i, \hat{\theta}_i(t_i)) \tag{40}$$

for all $m_i^3 \in Y_i^*(\beta^{lim})$. Since $Y_i^*(\beta^{lim})$ is a countable, dense subset of $Y_i(\beta^{lim})$ and since u_i is continuous, we have that the inequality in (49) holds for all $m_i^3 \in Y_i(\beta^{lim})$. Thus, $m_i^1 \in \rho_i^{\beta^{lim}}(\bar{\mu}^i, t_i)$, where $\rho_i^{\beta^{lim}}$ is defined in (7). Since $\bar{\mu}^i \in C_i\left(t_i, W_{-i}^{\alpha,\beta^{lim}}\right)$, it follows from (9) that $m_i^1 \in W_i^{\alpha+1,\beta^{lim}}(t_i)$, which is a contradiction. ■

Lemma 10. For all $\alpha \in \Omega$, $R_i^\alpha \subseteq \Sigma_i(W_i^{\alpha,\beta^{lim}})$ for all $i \in I$.

Proof. Let us proceed by transfinite induction over Ω. It is clear that $R_i^0 \subseteq \Sigma_i(W_i^{0,\beta^{lim}}) = M_i$. Fix any $\alpha \in \Omega \setminus \{0\}$. Suppose that for all $\gamma < \alpha$, $R_i^\gamma \subseteq \Sigma_i(W_i^{\gamma,\beta^{lim}})$ for all $i \in I$. Fix any $i \in I$. We show that $R_i^\alpha \subseteq \Sigma_i(W_i^{\alpha,\beta^{lim}})$. We proceed according to whether α is a successor ordinal or not.

Suppose that α is a limit ordinal. Since $\bigcap_{\gamma<\alpha} R_i^\gamma = R_i^\alpha$ (by Definition 4), it follows that $R_i^\alpha \subseteq \bigcap_{\gamma<\alpha} \Sigma_i(W_i^{\gamma,\beta^{lim}})$. Since $\bigcap_{\gamma<\alpha} \Sigma_i(W_i^{\gamma,\beta^{lim}}) \subseteq \Sigma_i(W_i^{\alpha,\beta^{lim}})$, it follows that $R_i^\alpha \subseteq \Sigma_i(W_i^{\alpha,\beta^{lim}})$.

Suppose that α is a successor ordinal. Fix any $t_i \in T_i$. We proceed according to whether $t_i \in T_i^*(W^{\alpha-1,\beta^{lim}})$ or not.

Suppose that $t_i \in T_i^*(W^{\alpha-1,\beta^{lim}})$. Fix any $m_i \in R_i^\alpha(t_i)$. The inductive hypothesis implies that $R_i^{\alpha-1} \subseteq \Sigma_i(W_i^{\alpha-1,\beta^{lim}})$. Since $m_i \in R_i^\alpha(t_i)$, Definition 4 implies that that there exists $\mu^i \in C_i(t_i, R_{-i}^{\alpha-1})$ s.t. $m_i \in r_i(\mu^i, \hat{\theta}_i(t_i))$. Since $R_i^{\alpha-1} \subseteq \Sigma_i(W_i^{\alpha-1,\beta^{lim}})$, it follows that $\mu^i \in C_i(t_i, \Sigma_i(W_i^{\alpha-1,\beta^{lim}}))$. Since $t_i \in T_i^*(W^{\alpha-1,\beta^{lim}})$, $\mu^i \in C_i(t_i, \Sigma_i(W_i^{\alpha-1,\beta^{lim}}))$ and $m_i \in r_i(\mu^i, \hat{\theta}_i(t_i))$, Lemma 9 implies that $m_i^2 = 1$ and $m_i^1 \in W_i^{\alpha,\beta^{lim}}(t_i)$. Thus, $m_i \in \Sigma_i(W_i^{\alpha,\beta^{lim}})(t_i)$.

Suppose that $t_i \in T_i^{*c}\left(W^{\alpha-1,\beta^{lim}}\right)$. Part 2 of Lemma 5 implies that $W_i^{\alpha,\beta^{lim}}(t_i) = \bar{\beta}(t_i)$. It follows from (25) that $R_i^\alpha(t_i) \subseteq \Sigma_i(W_i^{\alpha,\beta^{lim}})(t_i)$.

Since the choice of player i and of player i's type t_i were arbitrary, we conclude that for all $i \in I$, $R_i^\alpha \subseteq \Sigma_i(W_i^{\alpha,\beta^{lim}})$. By the principle of transfinite induction, the statement follows. ■

Lemma 11. For all $\alpha \in \Omega$, all $i \in I$, and all $t_i \in T_i^*(W^{\alpha,\beta^{lim}})$, if $m_i \in R_i^{\alpha+1}(t_i)$, then $m_i^2 = 1$ and $m_i^1 \in W_i^{\alpha+1,\beta^{lim}}(t_i)$.

Proof. Fix $(\alpha, i, t_i) \in \Omega \times I \times T_i^*(W^{\alpha,\beta^{lim}})$. Suppose that $m_i \in R_i^{\alpha+1}(t_i)$. Def. 4 implies that there exists $\mu^i \in C_i(t_i, R_{-i}^{\alpha})$ s.t. $m_i \in r_i(\mu^i, \hat{\theta}_i(t_i))$. Lemma 10 implies that $R_{-i}^{\alpha} \subseteq \Sigma_{-i}(W_{-i}^{\alpha,\beta^{lim}})$, and so $\mu^i \in C_i(t_i, \Sigma_{-i}(W_{-i}^{\alpha,\beta^{lim}}))$. Lemma 9 implies that $m_i^2 = 1$ and that $m_i^1 \in W_i^{\alpha+1,\beta^{lim}}(t_i)$. Since the choice of $(\alpha, i, t_i) \in \Omega \times I \times T_i(W^{\alpha,\beta^{lim}})$ was arbitrary, the proof is complete. ■

Let us show that $\mathcal{M}$ ICRR-implements f on $\mathcal{T}$. Lemma 8 and Lemma 1 imply that for all $i \in I$ and $t_i \in T_i$, $R_i(t_i) \neq \emptyset$. Thus, part (i) of Def. 6 is satisfied. Recall that Lemma 2 implies that there exists an α s.t. $W^{\alpha,\beta^{lim}} = W^{\alpha+1,\beta^{lim}} = \beta^{lim}$. Since f is non-trivial and satisfies β-RNWA, it holds that $T^*\left(\beta^{lim}\right) = T$. Thus, $T^*\left(W^{\alpha,\beta^{lim}}\right) = T$. Fix any $t \in T$ and any $m \in R(t)$. Since $R(t) \subseteq R^{\alpha+1}(t)$, then $m \in R^{\alpha+1}(t)$. Lemma 11 implies that $m_i^2 = 1$ and $m_i^1 \in W_i^{\alpha+1,\beta^{lim}}(t_i) = W_i^{\beta^{lim}}(t_i) = \beta_i^{lim}(t_i)$ for all $(i, t_i) \in I \times T_i$. *Rule 1* implies that $g(m) = f(m^1)$. Since β^{lim} is acceptable for f on $\mathcal{T}$, it follows that $f(m^1) = f(t)$. Since the choice of $(t, m) \in T \times R(t)$ was arbitrary, we conclude that part (ii) of Def. 6 is satisfied. Thus, f is ICR-implementable on $\mathcal{T}$, and so part (i) of Def. 7 is satisfied. Finally, Lemma 8 implies that part (ii) of Def. 7 is satisfied, and so $\mathcal{M}$ ICRR-implements f on $\mathcal{T}$.

B. Proof of Lemma 4

Fix any $\mathcal{T}$.

Proof of Part 1: Fix any $\beta \in \mathcal{B}$. Let us show that $\beta \subseteq W^{\beta}$. Let us proceed by transfinite induction. By definition, it follows that $\beta \subseteq W^{0,\beta}$. Fix any arbitrary $\alpha \in \Omega$ and suppose that $\beta \subseteq W^{\gamma,\beta}$ for all $\gamma < \alpha$. We show that $\beta \subseteq W^{\alpha,\beta}$. We proceed according to whether α is a limit ordinal or a successor ordinal. When α is a limit ordinal, the induction hypothesis and the definition of $W^{\alpha,\beta}$ implies that $\beta \subseteq \bigcap_{\gamma<\alpha} W^{\gamma,\beta} = W^{\alpha,\beta}$.

Suppose that α is a successor ordinal. The induction hypothesis implies that $\beta \subseteq W^{\alpha-1,\beta}$. Fix any $i \in I$ and any $t_i \in T_i$. Take any $\hat{t}_i \in \beta_i(t_i)$. Since $\beta \subseteq W^{\alpha-1,\beta}$, it holds that $\hat{t}_i \in W_i^{\alpha-1,\beta}(t_i)$. Moreover, since $\beta \subseteq W^{\alpha-1,\beta}$, it follows that $C_i\left(t_i,\beta_{-i}\right) \subseteq C_i\left(t_i, W_{-i}^{\alpha-1,\beta}\right)$. Since β is non-refutable w.r.t. itself and since $C_i\left(t_i,\beta_{-i}\right) \subseteq C_i\left(t_i, W_{-i}^{\alpha-1,\beta}\right)$, it follows that $\hat{t}_i \in \rho_i^{\beta}(\mu^i, t_i)$ for some $\mu^i \in C_i\left(t_i, W_{-i}^{\alpha-1,\beta}\right)$. Since $\hat{t}_i \in W_i^{\alpha-1,\beta}(t_i)$, (9) implies that $\hat{t}_i \in W_i^{\alpha,\beta}(t_i)$.

Since the choice of $\left(i,t_i,\hat{t}_i\right) \in I \times T_i \times \beta_i\left(t_i\right)$ was arbitrary, it follows that $\beta \subseteq W^{\alpha,\beta}$. By the principle of transfinite induction, we have that $\beta \subseteq W^{\alpha,\beta}$ for all $\alpha \in \Omega$. Since Lemma 2 implies that $\lim_{\alpha\in\Omega} W^{\alpha,\beta} = W^{\beta}$, it follows that $\beta \subseteq W^{\beta}$.

Proof of Part 2: Let us show that the net $\{\beta_\alpha\}_{\alpha\in\Omega}$ is monotone increasing w.r.t. set inclusion and that $\lim_{\alpha\in\Omega}\beta_\alpha = \beta^{lim} \in \mathcal{E}(W)$. To see that $\{\beta_\alpha\}_{\alpha\in\Omega}$ is monotone increasing w.r.t. set inclusion, fix any $\alpha \in \Omega$. Suppose that $\alpha = 0$. Then, $\beta_1 = W^{0,\beta} \in \mathcal{B}$, and so $\beta_0 \subseteq \beta_1$. Then, let $\alpha \in \Omega\setminus\{0\}$. Since $\alpha+1$ is a successor ordinal, it follows that $\beta_{\alpha+1} = W^{\alpha,\beta}$. Since part 1) of Lemma 4 (proved above) implies that $\beta_\alpha \subseteq W^{\alpha,\beta}$, it follows that $\beta_\alpha \subseteq \beta_{\alpha+1}$. Since the choice of α was arbitrary, it follows that $\{\beta_\alpha\}_{\alpha\in\Omega}$ is monotone increasing w.r.t. set inclusion. Since $\{\beta_\alpha\}_{\alpha\in\Omega}$ is monotone increasing w.r.t. set inclusion, Lemma 2 implies that its limit exists, i.e., $\lim_{\alpha\in\Omega}\beta_\alpha = \beta^{lim}$. It also follows that $\beta^{lim} = W^{\beta^{lim}}$, and so $\beta^{lim} \in \mathcal{E}(W)$.

Proof of Part 3: The proof is based on the following lemma.

Lemma 12. For all $\beta, \beta' \in \mathcal{B}$, if $\beta \subseteq \beta'$, then $W^{\beta} \subseteq W^{\beta'}$.

Proof. Fix any $\beta, \beta' \in \mathcal{B}$ s.t. $\beta \subseteq \beta'$. Let us show that $W^{\beta} \subseteq W^{\beta'}$. Firstly, let us observe that since $\beta \subseteq \beta'$, it follows that $Y(\beta') \subseteq Y(\beta)$. Let us proceed by transfinite induction. It is clear from the definition that $W^{0,\beta} \subseteq W^{0,\beta'}$. Fix any $\alpha \in \Omega\setminus\{0\}$ and suppose that $W^{\gamma,\beta} \subseteq W^{\gamma,\beta'}$ for all $\gamma < \alpha$. We show that $W^{\alpha,\beta} \subseteq W^{\alpha,\beta'}$. We proceed according to whether α is a limit ordinal or a successor ordinal.

When α is a limit ordinal, it follows from (10) and the induction hypothesis that $W^{\alpha,\beta} \subseteq W^{\alpha,\beta'}$. Thus, let us suppose that α is a successor ordinal. Fix any $(i,t_i) \in I \times T$. Let us show that $W_i^{\alpha,\beta}(t_i) \subseteq W_i^{\alpha,\beta'}(t_i)$. To this end, take any $\hat{t}_i \in W_i^{\alpha,\beta}(t_i)$. (9) implies that $\hat{t}_i \in W_i^{\alpha-1,\beta}(t_i)$ and $\hat{t}_i \in \rho_i^{\beta}(\mu^i, t_i)$ for some $\mu^i \in C_i\left(t_i, W_{-i}^{\alpha-1,\beta}\right)$. Since the induction hypothesis implies that $W_i^{\alpha-1,\beta} \subseteq W_i^{\alpha-1,\beta'}$, it follows that $C_i\left(t_i, W_{-i}^{\alpha-1,\beta}\right) \subseteq C_i\left(t_i, W_{-i}^{\alpha-1,\beta'}\right)$. Since $\hat{t}_i \in \rho_i^{\beta}(\mu^i, t_i)$ for some $\mu^i \in C_i\left(t_i, W_{-i}^{\alpha-1,\beta}\right)$ and $Y(\beta') \subseteq Y(\beta)$, it follows that $\hat{t}_i \in \rho_i^{\beta'}(\mu^i, t_i)$ for some $\mu^i \in C_i\left(t_i, W_{-i}^{\alpha-1,\beta'}\right)$. Since $W_i^{\alpha-1,\beta}(t_i) \subseteq W_i^{\alpha-1,\beta'}(t_i)$ and $\hat{t}_i \in W_i^{\alpha-1,\beta}(t_i)$, we have that $\hat{t}_i \in W_i^{\alpha-1,\beta'}(t_i)$. Since $\hat{t}_i \in \rho_i^{\beta'}(\mu^i, t_i)$ for some $\mu^i \in C_i\left(t_i, W_{-i}^{\alpha-1,\beta'}\right)$ and since $\hat{t}_i \in W_i^{\alpha-1,\beta'}(t_i)$, it follows from (9) that $\hat{t}_i \in W_i^{\alpha,\beta'}(t_i)$.

Since the choice of $(i, t_i, \hat{t}_i) \in I \times T_i \times W_i^{\alpha,\beta}(t_i)$ was arbitrary, we have that $W^{\alpha,\beta} \subseteq W^{\alpha,\beta'}$. By the principle of transfinite induction, it follows that for all $\alpha \in \Omega$, $W^{\alpha,\beta} \subseteq W^{\alpha,\beta'}$. Finally, since Lemma 2 implies that the limits of the nets $\{W^{\alpha,\beta}\}_{\alpha\in\Omega}$ and $W^{\alpha,\beta'}\}_{\alpha\in\Omega}$ exist, the statement follows. ■

Since $\mathcal{E}(W) \neq \emptyset$, fix any $\tilde{\beta} \in \mathcal{E}(W)$. Then, $\tilde{\beta} = W^{\tilde{\beta}}$. Since $\tilde{\beta} \in \mathcal{B}$, it follows that $\beta^{id} \subseteq \tilde{\beta}$. Lemma 12 implies that $W^{\beta^{id}} = \beta_1 \subseteq W^{\tilde{\beta}} = \tilde{\beta}$. Take any $\alpha \in \Omega \setminus \{0\}$. Suppose that $W^{\beta_\gamma} \subseteq W^{\tilde{\beta}}$ for all $\gamma < \alpha$. Let us show that $W^{\beta_\alpha} \subseteq W^{\tilde{\beta}} = \tilde{\beta}$. We proceed according to whether α is a limit ordinal or a successor ordinal. Suppose that α is a limit ordinal. Since $\beta_\alpha = \bigcup\limits_{\gamma<\alpha} W^{\beta_\gamma}$, the induction hypothesis implies that $\beta_\alpha \subseteq \tilde{\beta}$. $W^{\beta_\alpha} \subseteq W^{\tilde{\beta}} = \tilde{\beta}$ follows from Lemma 12. Suppose that α is a successor ordinal. Then, $\beta_\alpha = W^{\beta_{\alpha-1}}$. Since the induction hypothesis implies that $\beta_\alpha \subseteq \tilde{\beta}$, it follows from Lemma 12 that $W^{\beta_\alpha} \subseteq W^{\tilde{\beta}} = \tilde{\beta}$. Then, $\beta_\alpha = \beta_{\alpha-1}$. By the principle of transfinite induction, it follows that $W^{\beta_\alpha} \subseteq W^{\tilde{\beta}} = \tilde{\beta}$ for all $\alpha \in \Omega$. Since part 2) of Lemma 4 (proved above) implies that $\lim_{\alpha\in\Omega} \beta_\alpha = \beta^{lim}$ exists, it follows that $\beta^{lim} \subseteq \tilde{\beta}$.

Proof of Part 4. Take any $\beta \in \mathcal{B}$ s.t. β is non-refutable w.r.t. β^{lim}. Let us show that $\beta \subseteq \beta^{lim}$. Assume, to the contrary, that $\beta \nsubseteq \beta^{lim}$. Then, there exists $(i, t_i, t_i') \in I \times T_i \times \beta_i(t_i)$ s.t. $t_i' \in \beta_i(t_i)$ and $t_i' \notin W_i^{\beta^{lim}}(t_i)$. A contradiction is obtained if we show that $\beta \subseteq W^{\alpha,\beta^{lim}}$ for all $\alpha \in \Omega$.

Let us proceed by transfinite induction. By definition, $\beta \subseteq W^{0,\beta^{lim}}$. Fix an arbitrary $\alpha \in \Omega$ and suppose that $\beta \subseteq W^{\gamma,\beta^{lim}}$ for all $\gamma < \alpha$. We show that $\beta \subseteq W^{\alpha,\beta^{lim}}$. We proceed according to whether α is a limit ordinal or a successor ordinal. When

α is a limit ordinal, the induction hypothesis and the definition of $W^{\alpha,\beta^{lim}}$ implies that $\beta \subseteq \bigcap_{\gamma<\alpha} W^{\gamma,\beta^{lim}} = W^{\alpha,\beta^{lim}}$. Otherwise, suppose that α is a successor ordinal. Fix an arbitrary $\left(i, t_i, \hat{t}_i\right) \in I \times T_i \times \beta_i\left(t_i\right)$. The induction hypothesis implies that $C_i(t_i, \beta_{-i}) \subseteq C_i(t_i, W_{-i}^{\alpha-1,\beta^{lim}})$ and $\hat{t}_i \in W_i^{\alpha-1,\beta^{lim}}(t_i)$. Since β is non-refutable w.r.t. β^{lim}, we have that $Y_i(\beta^{lim}) \subseteq \mathcal{L}_i(f(t_i', \cdot), \mu^i, t_i)$ for some $\mu^i \in C_i(t_i, \beta_{-i})$. Since $C_i(t_i, \beta_{-i}) \subseteq C_i(t_i, W_{-i}^{\alpha-1,\beta^{lim}})$ and $\hat{t}_i \in W_i^{\alpha-1,\beta^{lim}}(t_i)$, it follows from (9) that $\hat{t}_i \in W_i^{\alpha,\beta^{lim}}(t_i)$. Since the choice of $\left(i, t_i, \hat{t}_i\right) \in I \times T_i \times \beta_i\left(t_i\right)$ was arbitrary, we conclude that $\beta \subseteq W^{\alpha,\beta^{lim}}$.

By the principle of transfinite induction, it holds that $\beta \subseteq W^{\alpha,\beta^{lim}}$ for all $\alpha \in \Omega$. Since part 2) of Lemma 4 (proved above) implies that $\lim_{\alpha\in\Omega} \beta_\alpha = \beta^{lim}$ exists and that $\beta^{lim} = W^{\beta^{lim}}$, it follows that $\beta \subseteq W^{\beta^{lim}}$, yielding a contradiction.

C. Proof of Lemma 5 and Theorem 4

To prove Lemma 5, we need some additional notation and results.

For all $\mathcal{T}$, all $\beta \in \mathcal{B}$, all $\alpha \in \Omega$ and all $i \in I$, let $T_i\left(W^{\alpha,\beta}\right) \subseteq T_i$ be defined by

$$T_i\left(W^{\alpha,\beta^{lim}}\right) := \left\{ t_i \in T_i \;\middle|\; \begin{array}{l} \text{for all } \mu^i \in C_i(t_i, W_{-i}^{\alpha,\beta^{lim}}) \\ \text{there exist } \bar{y}_i, y_i \in Y_i(\beta^{lim}) \\ \text{s.t. } U_i(y_i, \mu^i, t_i) > U_i(\bar{y}_i, \mu^i, t_i) \end{array} \right\} \tag{41}$$

Lemma 13. For all $\mathcal{T}$ and all $(i, \alpha) \in I \times \Omega$, $T_i^*\left(W^{\alpha,\beta^{lim}}\right) = T_i\left(W^{\alpha,\beta^{lim}}\right)$.

Proof. Let $\mathcal{T}$ be any model. Fix any $(i, \alpha) \in I\times\Omega$. Since it is clear that $T_i^*\left(W^{\alpha,\beta^{lim}}\right) \subseteq T_i\left(W^{\alpha,\beta^{lim}}\right)$, let us show that $T_i\left(W^{\alpha,\beta^{lim}}\right) \subseteq T_i^*\left(W^{\alpha,\beta^{lim}}\right)$. Assume that $t_i \in T_i\left(W^{\alpha,\beta^{lim}}\right)$. (41) implies that for all $\mu^i \in C_i(t_i, W_{-i}^{\alpha,\beta^{lim}})$, there exist $y_i^{\mu^i}, \bar{y}_i^{\mu^i} \in Y_i(\beta^{lim})$ s.t. the inequality in (41) is satisfied. Since $C_i(t_i, W_{-i}^{\alpha,\beta^{lim}})$ is a separable metric space, let $\hat{C}_i(t_i, W_{-i}^{\alpha,\beta^{lim}}) = \cup_{k\in\mathbb{N}} \left\{\mu^{i,k}\right\}$ be a countable, dense subset of

$C_i(t_i, W_{-i}^{\alpha,\beta^{lim}})$. Let $\tilde{y}_i \in Y_i(\beta^{lim})$ be a mapping defined by

$$\tilde{y}_i = \sum_{k=1}^{\infty} \frac{1}{2^k} \bar{y}_i^{\mu^{i,k}}.$$

For any $\bar{k} \in \mathbb{N}$, let $y_i^{\bar{k}} \in Y_i(\beta^{lim})$ be a mapping defined by

$$y_i^{\bar{k}} = \sum_{k \neq \bar{k}} \frac{1}{2^k} \bar{y}_i^{\mu^{i,k}} + \frac{1}{2^{\bar{k}}} y_i^{\mu^{i,\bar{k}}}.$$

Thus, for all $k \in \mathbb{N}$, we have that

$$U_i\left(y_i^k, \mu^{i,k}, t_i\right) - U_i\left(\tilde{y}_i, \mu^{i,k}, t_i\right) = \frac{1}{2^k}\left[U_i\left(y_i^{\mu^{i,k}}, \mu^{i,k}, t_i\right) - U_i\left(\bar{y}_i^{\mu^{i,k}}, \mu^{i,k}, t_i\right)\right] > 0,$$

where the strict inequality is guaranteed by (41). Since t_i's preference over lotteries are continuous and since, moreover, $\hat{C}_i(t_i, W_{-i}^{\alpha,\beta^{lim}})$ is a countable, dense subset of $C_i(t_i, W_{-i}^{\alpha,\beta^{lim}})$, it follows that $t_i \in T_i^*(W^{\alpha,\beta^{lim}})$. Since the choice of $t_i \in T_i(W^{\alpha,\beta^{lim}})$ was arbitrary, it follows that $T_i\left(W^{\alpha,\beta^{lim}}\right) \subseteq T_i^*\left(W^{\alpha,\beta^{lim}}\right)$. ■

Lemma 14. For all $\mathcal{T}$ and all $i \in I$, $\{T_i^*(W^{\alpha,\beta^{lim}})\}_{\alpha\in\Omega}$ is a monotone increasing net w.r.t. set inclusion. Moreover, there exists $\alpha(i) \in \Omega$ s.t. for all $\alpha \geq \alpha(i)$, $T_i^*(W^{\alpha,\beta^{lim}}) = T_i^*(W^{\alpha+1,\beta^{lim}}) = T_i^*(W^{\beta^{lim}})$.

Proof. Fix any $(i, \alpha) \in I \times \Omega$. In light of Lemma 13, it suffices to show that $T_i(W^{\alpha,\beta^{lim}}) \subseteq T_i(W^{\alpha+1,\beta^{lim}})$. Take any $t_i \in T_i(W^{\alpha,\beta^{lim}})$. (41) implies that for all $\mu^i \in C_i(t_i, W_{-i}^{\alpha,\beta^{lim}})$, there are mappings $y_i, \bar{y}_i \in Y_i(\beta^{lim})$ s.t. $U_i(t_i, \mu^i, y_i) > U_i(t_i, \mu^i, \bar{y}_i)$. Since Lemma 2 implies that $\{W^{\alpha,\beta^{lim}}\}_{\alpha\in\Omega}$ is a monotone decreasing net, it holds that $C_i(t_i, W_{-i}^{\alpha+1,\beta^{lim}}) \subseteq C_i(t_i, W_{-i}^{\alpha,\beta^{lim}})$. Therefore, (41) implies that $t_i \in T_i(W^{\alpha+1,\beta^{lim}})$. Thus, $\{T_i^*(W^{\alpha,\beta^{lim}})\}_{\alpha\in\Omega}$ is a monotone increasing net w.r.t. set inclusion. Finally, Lemma 2 implies that there exists $\alpha(i) \in \Omega$ s.t. for all $\alpha \geq \alpha(i)$, $T_i^*(W^{\alpha,\beta^{lim}}) = T_i^*(W^{\alpha+1,\beta^{lim}}) = T_i^*(W^{\beta^{lim}})$. ■

Proof of Lemma 5.

Suppose that $f : T \to X$ satisfies IIM on $\mathcal{T}$.

Proof of Part 1. Fix any $(i, t_i, \alpha) \in I \times T_i \times \Omega$ s.t. $W_i^{\alpha+1,\beta^{lim}}(t_i) \neq W_i^{\alpha,\beta^{lim}}(t_i)$. Let us show that $t_i \in T_i^*(W^{\alpha,\beta^{lim}})$. Since $W_i^{\alpha+1,\beta^{lim}}(t_i) \neq W_i^{\alpha,\beta^{lim}}(t_i)$, there exists $\hat{t}_i \in T_i$ s.t. $\hat{t}_i \in W_i^{\alpha,\beta^{lim}}(t_i)$ and $\hat{t}_i \notin W_i^{\alpha+1,\beta^{lim}}(t_i)$. It follows from (9) that $\hat{t}_i \notin \rho_i^{\beta^{lim}}(\mu^i, t_i)$ for all $\mu^i \in C_i(t_i, W_{-i}^{\alpha,\beta^{lim}})$. (7) implies that for all $\mu^i \in C_i(t_i, W_{-i}^{\alpha,\beta^{lim}})$, there exists $\bar{y}_i \in Y_i(\beta^{lim})$ s.t. $\bar{y}_i \notin \mathcal{L}_i(f(\hat{t}_i, \cdot), \mu^i, t_i)$. Since f satisfies IIM on $\mathcal{T}$, f satisfies β^{lim}-IIC on $\mathcal{T}$ and β^{lim} is acceptable for f on $\mathcal{T}$. Thus, $f\left(\hat{t}_i, \cdot\right) \in Y_i(\beta^{lim})$. Since the inequality in (41) holds for all $\mu^i \in C_i(t_i, W_{-i}^{\alpha,\beta^{lim}})$, we have that $t_i \in T_i\left(W^{\alpha,\beta^{lim}}\right)$. Lemma 13 implies that $t_i \in T_i^*(W^{\alpha,\beta^{lim}})$.

Proof of Part 2.
Fix any $(i, t_i, \alpha) \in I \times T_i \times \Omega$ s.t. $t_i \in T_i^{*c}\left(W^{\alpha,\beta^{lim}}\right)$. We show that $W_i^{\alpha,\beta^{lim}}(t_i) = W_i^{\alpha+1,\beta^{lim}}(t_i) = \bar{\beta}_i(t_i)$. Since $t_i \in T_i^{*c}\left(W^{\alpha,\beta^{lim}}\right)$, Lemma 5 implies that $W_i^{\alpha,\beta^{lim}}(t_i) = W_i^{\alpha+1,\beta^{lim}}(t_i)$. Thus, we are left to show that that $W_i^{\alpha,\beta^{lim}}(t_i) = \bar{\beta}_i(t_i)$. Assume, to the contrary, that $W_i^{\alpha,\beta^{lim}}(t_i) \neq \bar{\beta}_i(t_i)$. Then, there exists a successor ordinal $\hat{\alpha}$ with $0 < \hat{\alpha} \leq \alpha$ s.t. $W_i^{\hat{\alpha},\beta^{lim}}(t_i) \neq W_i^{\hat{\alpha}-1,\beta^{lim}}(t_i)$. Part 1 of Lemma 5 (proved above) implies that $t_i \in T_i^*(W^{\hat{\alpha}-1,\beta^{lim}})$. Since Lemma 14 implies that $T_i^*(W^{\hat{\alpha}-1,\beta^{lim}}) \subseteq T_i^*(W^{\alpha,\beta^{lim}})$, we have $t_i \in T_i^*\left(W^{\alpha,\beta^{lim}}\right)$, yielding a contradiction.

Proof of Part 3.
Fix any $i \in I$ s.t. $T_i^*(\beta^{lim}) \neq T_i$. We show that $\beta_i^*(t_i) = T_i$ for some $t_i \in T_i$. Assume, to the contrary, that $\beta_i^*(t_i) \neq T_i$ for all $t_i \in T_i$. Fix any $t_i \in T_i$. Since $\beta_i^*(t_i) \neq T_i$, there exists $\hat{t}_i \in T_i$ s.t. $\hat{t}_i \notin \beta_i^*(t_i)$. Thus, there exists a successor ordinal α s.t. $\hat{t}_i \notin W_i^{\alpha,\beta^{lim}}(t_i)$ and $\hat{t}_i \in W_i^{\alpha-1,\beta^{lim}}(t_i)$. Part 1 of Lemma 5 (proved above) implies that $t_i \in T_i^*(W^{\alpha-1,\beta^{lim}})$. Since Lemma 14 implies that $T_i^*(W^{\alpha-1,\beta^{lim}}) \subseteq T_i^*(\beta^{lim})$, we have $t_i \in T_i^*\left(\beta^{lim}\right)$. Since the choice of $t_i \in T_i$ was arbitrary, we have that $T_i^*(\beta^{lim}) = T_i$, yielding a contradiction.

Proof of Theorem 4.

Suppose that $f : T \to X$ satisfies IIM on $\mathcal{T}$. We proceed according to whether there exists $i \in I$ s.t. $T_i^* (\beta^*) \neq T_i$.

Suppose that there exists $i \in I$ s.t. $T_i^* (\beta^*) \neq T_i$. Part 3) of Lemma 5 implies that $\beta_i^{lim}(t_i) = T_i$ for some $t_i \in T_i$. Since $f : T \to X$ satisfies IIM on $\mathcal{T}$, it follows from Theorem 3 that β^{lim} is acceptable for f on $\mathcal{T}$. Since $\beta_i^{lim}(t_i) = T_i$ for some $t_i \in T_i$, it follows that for all $t_i, t_i' \in T_i$, $f(t_i, t_{-i}) = f(t_i', t_{-i})$ for all $t_{-i} \in T_{-i}$.

Suppose that $T_i^* (\beta^*) = T_i$ for all $i \in I$. Fix any $(i, t_i) \in I \times T_i$. Let us show that f satisfies β^{lim}-RNWA; that is, that either $t_i \in T_i^* \left(W^{0,\beta^{lim}} \right)$ or there exists $\hat{\alpha}(t_i) \in \Omega \setminus \{0\}$ s.t. $t_i \in T_i^* \left(W^{\hat{\alpha}(t_i),\beta^{lim}} \right)$ and $t_i \in T_i^{*c} \left(W^{\gamma,\beta^{lim}} \right)$ for all $\gamma \in \Omega$ s.t. $\gamma < \hat{\alpha}(t_i)$. Since $T_i^* (\beta^*) = T_i$ and since $\lim_{\alpha \in \Omega} \beta_\alpha = \beta^{lim}$, it follows that there exists $\alpha(t_i) \in \Omega$ s.t. $t_i \in T_i^* \left(W^{\alpha(t_i),\beta^{lim}} \right)$. Let $\Omega(\alpha(t_i)) = \{\alpha(t_i) \geq \bar{\alpha} \geq 0 | t_i \in T_i(W^{\bar{\alpha},\beta^{lim}})\} \neq \emptyset$. Let $\hat{\alpha}(t_i) = min\ \Omega(\alpha)$. Since $\Omega(\alpha)$ is a well-ordered set, $\hat{\alpha}(t_i)$ is well-defined. Thus, f satisfies β^{lim}-RNWA.

D. Example 1

In this section, we will show that the SCF f of Example 1 is implementable in ICR by the mechanism defined in that example. Following Kunimoto et al. (2023), SCF f satisfies weak-IRM and hence incentive compatibility. The following observation will be useful: For every $i \in I$, and every $\theta \in \Theta$, it holds that $u_i(z'', \theta) < u_i(f(\theta), \theta)$.

Let us introduce the following definitions. For all $\beta \in \mathcal{B}$ and all $i \in I$, define $\Sigma_i^{\beta_i} : T_i \to 2^{M_i} \setminus \{\emptyset\}$ by

$$\Sigma_i^{\beta_i} (t_i) = \left\{ m_i \in M_i | m_{i,1}^1 \in \beta_i (t_i) \right\}, \tag{42}$$

and $\tilde{\Sigma}_i^{\beta_i} : T_i \to 2^{M_i} \setminus \{\emptyset\}$ by

$$\tilde{\Sigma}_i^{\beta_i} (t_i) = \left\{ m_i \in \Sigma_i^{\beta_i} (t_i) | m_i^2 = 1 \right\}. \tag{43}$$

Lemma 15. $\tilde{\Sigma}^{\beta^f}$ is a best reply set.

Proof. Fix an i, t_i, and $m_i \in \tilde{\Sigma}^{\beta^f}(t_i)$. Consider the following beliefs: $\mu^i \in C_i(t_i, \tilde{\Sigma}^{\beta^f})$ is such that μ_i assigns probability 1 to the event that $m^1_{-i,2} = t_i$. Under these beliefs, a unilateral deviation for player i induces either Rule 1 or Rule 2. Suppose there is a message $\hat{m}_i$ such that Rule 1 is induced. In this case, incentive compatibility ensures that m_i is better than $\hat{m}_i$. Suppose that there is a message $\hat{m}_i$ such that Rule 2 is induced. In this case, agent i obtains an outcome in $\mathcal{L}_i(f(t_i, \cdot), \mu^i, t_i)$.

Since i, t_i, and m_i were arbitrary, it follows that $\tilde{\Sigma}^{\beta^f}$ is a best reply set. ■

Lemma 16. For all $i \in I$ and all $t_i \in T_i$, if $m_i \in R_i(t_i)$, then $m_i^2 = 1$.

Proof. Suppose $m_i \in R_i(t_i)$ and $m_i^2 > 1$. Then, $m_i \in r_i(\mu_i, \hat{\theta}_i(t_i))$ for some $\mu_i \in C_i(t_i, \Sigma^{\bar{\beta}}_{-i})$. Towards a contradiction, we show that the following message

$$\hat{m}_i = ((m^1_{i,1}, m^1_{i,2}), \hat{m}^2_i, f(m^1_{i,1}, \cdot), a)$$

is better than m_i at t_i. Suppose that $m_{-i} \in \mathit{supp}(\mu_i)$ is s.t. Rule 2 applies to (m_i, m_{-i}). Then,

$$g(m_i, m_{-i}) = \frac{m_i^2}{1+m_i^2} f(m^1_{i,1}, m^1_{-i,1}) + \frac{m_i^2}{1+m_i^2} z''. \tag{44}$$

Suppose that $m_{-i} \in \mathit{supp}(\mu_i)$ is s.t. Rule 3 applies to (m_i, m_{-i}). Then,

$$g(m_i, m_{-i}) = \frac{m_i^2}{1+m_i^2} a + \frac{m_i^2}{1+m_i^2} z'' \tag{45}$$

Since $u_i(f(m^1_{i,1}, m^1_{-i,1}), \theta) > u_i(z'', \theta)$ and $u_i(a, \theta) > u_i(z'', \theta)$, it holds for a sufficiently high $\hat{m}^2_i$ that for every m_{-i} and every $\theta \in \Theta$:

$$u_i(g(\hat{m}_i, m_{-i}), \theta) > u_i(g(m_i, m_{-i}), \theta), \tag{46}$$

which is a contradiction to our initial assumption that $m_i \in r_i(\mu_i, t_i)$ for some $\mu_i \in C_i(t_i, \Sigma^{\bar{\beta}}_{-i})$.

■

Lemma 17. For all $i \in I$ and all $t_i \in T_i$, $R_i(t_i) \subseteq \tilde{\Sigma}_i^{\beta_i^f}(t_i)$.

Proof. Lemma 16 implies that for some β^*, it holds that $R = \tilde{\Sigma}^{\beta^*}$. Fix any i, t_i, and $t_i' \in \beta_i^*(t_i)$. Since $R = \tilde{\Sigma}^{\beta^*}$, there exists an $m_i \in R_i(t_i)$ such that $m_{i,1}^1 = t_i'$. Then, there exists $\mu^i \in C_i(t_i, \tilde{\Sigma}_{-i}^{\beta^*})$ such that $m_i \in r_i(\mu^i, \hat{\theta}_i(t_i))$.

Let $\bar{\mu}^i$ be defined over $\Theta_0 \times \Theta_{-i} \times T_{-i}$ by

$$\bar{\mu}^i[\theta_0, \theta_{-i}, m_{-i}^1] = \sum_{\bar{m}_{-i} \in \tilde{\Sigma}_{-i}^{\beta^*}[m_{-i}^1]} \mu^i[\theta_0, \theta_{-i}, \bar{m}_{-i}]. \tag{47}$$

for all $(\theta_0, \theta_{-i}, m_{-i}^1) \in \Theta_0 \times \Theta_{-i} \times T_{-i}$, where $\tilde{\Sigma}_{-i}^{\beta^*}[m_{-i}^1] = \{\hat{m}_i \in \tilde{\Sigma}_{-i}^{\beta^*} : m_{-i}^1 = \hat{m}_{-i}^1\}$. It can be checked that $\bar{\mu}^i \in C_i(t_i, \beta_{-i}^*)$.

Since $m_i \in r_i(\mu^i, \hat{\theta}_i(t_i))$ and $m_i^2 = 1$ and since $\mu^i \in C_i\left(t_i, \tilde{\Sigma}_{-i}^{\beta^*}\right)$, it follows that Rule 1 applies with probability 1, and so

$$U_i(m_i, \mu^i, \hat{\theta}_i(t_i)) = U_i((t_i', \cdot), \bar{\mu}^i, \hat{\theta}_i(t_i)) \tag{48}$$

Since $m_i \in r_i(\mu^i, \hat{\theta}_i(t_i))$ and since player i can never induce *Rule 3*, it follows from the definition of g that

$$U_i(f(t_i', \cdot), \bar{\mu}^i, \hat{\theta}_i(t_i)) \geq U_i(m_i^3, \bar{\mu}^i, \hat{\theta}_i(t_i)) \tag{49}$$

for all $m_i^3 \in Y_i^f$. Thus, $m_i^1 \in \tilde{\rho}_i^{\beta^{id}}(\bar{\mu}^i, t_i)$, where $\tilde{\rho}_i^{\beta^{lim}}$ is defined in (8). Since $\bar{\mu}^i \in C_i\left(t_i, \beta_{-i}^*\right)$ and i,t_i, and t_i' were arbitrarily chosen, it follows that β^* is weakly non-refutable deception with respect to β^{id}. Since our SCF in the example satisfies weak-IRM, it holds that $\beta^* \subseteq \beta^f$ and thus $R = \tilde{\Sigma}^{\beta^*} \subseteq \tilde{\Sigma}^{\beta^f}$.

■