Behavioral Implementation under Incomplete Information*

Mehmet Barlo[†] Nuh Aygün Dalkıran[‡]

July 23, 2023

Abstract

We investigate implementation under incomplete information allowing for individuals' choices featuring violations of rationality. Our primitives are individuals' interim choices that do not have to satisfy the weak axiom of revealed preferences. In this setting, we provide necessary as well as sufficient conditions for behavioral implementation under incomplete information. We also introduce behavioral interim incentive Pareto efficiency and investigate its implementability under incomplete information.

Keywords: Behavioral Implementation, Incomplete Information, Interim Implementation, Behavioral Interim Incentive Efficiency.

JEL Classification: D60, D79, D82, D90

*We thank the editor, the associate editor, and three anonymous referees for helpful suggestions. We would like to also thank Fuad Aleskerov, Ahmet Alkan, Dirk Bergemann, Eddie Dekel, Mehmet Ekmekci, Jeff Ely, Ehud Kalai, Semih Koray, Stephen Morris, Wojciech Olszewski, Alessandro Pavan, Luciano Pomatto, Yuval Salant, Marciano Siniscalchi, Asher Wolinsky, Kemal Yıldız, and the seminar participants at Northwestern University, The North American Summer Meeting of the Econometric Society 2019, The European Meeting of the Econometric Society 2019, Bosphorus Workshop on Economic Design 2019, Social Choice and Welfare 2018 for helpful comments. Muhammed Dönmez and Ece Yegane provided excellent research assistance. The authors gratefully acknowledge the financial support by TÜBİTAK under grant no. 119K957. Part of this work was completed while Nuh Aygün Dalkıran was visiting the Managerial Economics & Decision Sciences Department at Kellogg School of Management. He is thankful to The Fulbright Program for making this visit possible and to Northwestern University | Kellogg School of Management for their hospitality. Any remaining errors are ours.

[†]Faculty of Arts and Social Sciences, Sabancı University; barlo@sabanciuniv.edu

[‡]Corresponding author, Department of Economics, Bilkent University; dalkiran@bilkent.edu.tr

Table of Contents

1 Introduction

People have limited cognitive abilities and are prone to various behavioral biases; this is documented by ample evidence in marketing, psychology, and behavioral economics literature. Thus, it is not surprising that the behavior of individuals may not be consistent with the *standard axioms of rationality*.[1] What shall a planner do if she wants to implement a goal when the relevant information is distributed among "predictably irrational" individuals?

The present paper provides an analysis of the theory of implementation under incomplete information when individuals' choices do not necessarily comply with the weak axiom of revealed preferences (WARP), the condition corresponding to the standard axioms of rationality. Our results provide useful insights into behavioral mechanism design as information asymmetries are inescapable in many economic settings.

In particular, we analyze full implementation under incomplete information when individuals' interim choices do not necessarily satisfy WARP. Using the concept of behavioral interim equilibrium (BIE) that parallels the equilibrium at the interim stage of Saran (2011), we provide necessary as well as sufficient conditions for the full implementation of social choice rules in BIE. Therefore, our paper can be viewed as the *incomplete information* counterpart of de Clippel (2014), which is one of the pioneering papers on behavioral implementation under *complete information*.

In behavioral environments of incomplete information, each mechanism induces an incomplete information game where an individual's message (action) generates an interim Anscombe-Aumann (IAA) act, mapping each type profile of other individuals to alternatives. We assume that the planner knows how each type of each individual makes their choices on IAA acts, while these choices do not necessarily satisfy WARP. By following Saran (2011), who introduced this general setup and the corresponding equilibrium concept, we analyze full implementation under incomplete information allowing for a wide variety of behavioral biases.

A BIE of a mechanism is a strategy profile such that each individual's plan of action depends only on her type (her private information) and is one of her 'best' responses at the interim stage. It is well suited to our environment with interim choices on IAA acts, not necessarily derived from preference maximization. Indeed, BIE reduces to Bayesian Nash equilibrium in a rational setting under probabilistic sophistication.

[1] This is why the trend involving behavioral insights in policy-making has been growing, implying an increased interest in adopting economic models to allow behavioral biases. Thaler and Sunstein's *Nudge*, Ariely's *Predictably Irrational*, and Kahneman's *Thinking, Fast and Slow* have been influential in guiding real-life policies. There is such a trend in the academic literature as well, e.g., Spiegler (2011) provides a synthesis of efforts to adapt models in industrial organization to bounded rationality.

Our necessity result, Theorem 1, shows that if a mechanism implements a social choice set (SCS) in BIE, then the resulting opportunity sets, IAA acts individuals obtain via unilateral deviations, form a profile of sets with some desirable properties. In this profile, each set corresponds to an individual, a social choice function (SCF) in the SCS, and a deception profile of the other individuals.[2] First, this profile is *closed under deception*: IAA acts generated via any given deception from any set in the profile constitute another set in this profile. Second, each individual chooses the IAA act induced by the given SCF from the set corresponding to others' truthful type reports, implying a quasi-incentive compatibility condition (Proposition 1). Finally, a Maskin monotonicity-like requirement holds: For any deception turning a desirable SCF into a state-contingent allocation not aligned with the SCS, there is a whistle-blower who objects to this deception. A profile of sets satisfying these properties for a given SCS, we call *interim consistent* with the SCS.

Our sufficiency result for implementation in BIE, Theorem 2, uses a mild condition, *choice incompatibility*, that requires some level of disagreement in the society in addition to interim consistency.

To showcase the applicability of our findings, we analyze the implementability of efficiency in BIE. First, we introduce behavioral interim efficiency by modifying de Clippel's (complete information) efficiency and show that it exists when individuals' choices are non-empty valued. However, it is not implementable in BIE due to the well-known conflict between efficiency and incentive compatibility. This is why we entangle behavioral interim efficiency with quasi-incentive compatibility and obtain the behavioral counterpart of interim incentive Pareto efficiency of Holmström and Myerson (1983). We also provide a sufficient condition for its existence. Moreover, we establish that under rationality, behavioral interim efficiency and behavioral interim incentive efficiency coincide with interim Pareto efficiency and interim incentive Pareto efficiency, respectively.

Second, we show that in behavioral environments, blending efficiency with quasi-incentive compatibility does not suffice for its implementability in BIE, even under choice incompatibility. This is because the profile of implicit opportunity sets associated with behavioral interim incentive efficiency fails to ensure the existence of a profile closed under deception, a necessary condition for BIE implementation. Demanding a monotonicity-like condition from the profile of sets generated by deceptions from the implicit opportunity sets sustaining behavioral interim incentive efficiency equips us with a profile closed under deception. As a result, we obtain the BIE implementability of this SCS under choice incompatibility (Proposition 5). We display the full implementability in BIE of a selection

[2]An individual's deception maps her type space into itself and hence contains the identity function, an individual's truthful type report.

from the behavioral interim incentive efficient SCS on an example using the minimax-regret preferences of Savage (1951), while an SCF in this selection is not partially implementable in BIE by its associated direct mechanism. This, therefore, reiterates the failure of the revelation principle in behavioral domains, first observed by Saran (2011).

Our paper is mostly related to de Clippel (2014) and Saran (2011). de Clippel (2014) identifies the key condition —existence of opportunity sets that are consistent with the social goal—that is necessary and, with some additional restrictions, sufficient for behavioral (full) implementation under complete information. On the other hand, Saran (2011) develops a general framework and an equilibrium concept to study mechanism design in behavioral domains and shows that quasi-incentive compatibility characterizes behavioral partial implementation under incomplete information. Our paper fills the gap by providing interim consistency and showing that it is necessary and, with some additional restrictions, sufficient for behavioral (full) implementation under incomplete information. Another related paper is Jackson (1991), which analyzes Bayesian implementation in the rational domain and generalizes the analysis of Maskin (1999) (on Nash implementation under complete information) to the case of incomplete information. Our paper can be thought of as an envelope of de Clippel (2014) and Jackson (1991). We extend de Clippel (2014)'s analysis to the case of incomplete information and Jackson (1991)'s analysis to the case where individuals' interim choices need not satisfy WARP.

Hurwicz (1986), Eliaz (2002), Barlo and Dalkiran (2009), Ray (2010), and Korpela (2012) have also investigated the problem of implementation under complete information in behavioral/non-rational domains. Hurwicz (1986) considers choices that can be represented by a well-defined preference relation that does not have to be acyclic. Eliaz (2002), a seminal paper containing pioneering research on behavioral implementation, provides an analysis of full implementation when some of the individuals might be "faulty" and hence fail to act optimally. Barlo and Dalkiran (2009) provides an analysis of implementation in epsilon-Nash equilibrium, i.e., when individuals are satisficing so that they get close to (but not necessarily achieve) their best responses. Ray (2010) is among the first papers that extend the standard implementation problem to include behavior not representable by a preference relation. Korpela (2012) shows that when individual choices fail rationality axioms, the independence of irrelevant alternatives is key to obtaining the necessary and sufficient condition synonymous to that of Moore and Repullo (1990).[3]

[3]There have been other papers investigating implementation under complete information that allow for 'non-rational' behavior of individuals. Glazer and Rubinstein (2012) provides a mechanism design approach where the content and the framing of the mechanism affect individuals' ability to manipulate their information. Meanwhile, some of the other related works include Cabrales and Serrano (2011), Kucuksenel (2012), Saran (2016), Kunimoto and Serrano (2020), Bochet and Tumennasan (2021), Chen et al. (2021), Kunimoto and Saran (2022), Barlo and Dalkıran (2022a), Kunimoto et al. (2023), and

Another significant and related paper is Bergemann and Morris (2008), which analyzes ex-post implementation in the rational domain under incomplete information. We analyze behavioral ex-post implementation in Barlo and Dalkıran (2023) and extend the analysis of Bergemann and Morris (2008) to behavioral domains. Meanwhile, we adhere to the warning of de Clippel (2022) involving the use of behavioral ex-post equilibrium in environments where ex-post choices fail WARP.[4]

The organization of the paper is as follows: In Section 2, we provide the notation and the definitions. Section 3 contains our necessity results; Section 4 our sufficiency results. In Section 5, we analyze behavioral interim (incentive) efficiency, while Section 6 provides an example with minimax-regret preferences. Section 7 concludes.

2 Notation and Definitions

Consider a set of individuals $N = \{1, \ldots, n\}$ and a non-empty set of alternatives X. Let Θ denote the set of all relevant states of the world regarding individuals' choices. We assume that there is incomplete information among the individuals regarding the true state of the world and the private information of each individual is exclusive. Hence, the true state of the world is distributed knowledge.[5] Therefore, Θ has a product structure, i.e., $\Theta = \times_{i \in N} \Theta_i$ where $\theta_i \in \Theta_i$ denotes the private information (type) of individual $i \in N$ at state $\theta = (\theta_1, \ldots, \theta_n) \in \Theta$.

For any individual $i \in N$, an *interim Anscombe-Aumann act* (IAA) sustained by $\Theta_{-i} := \times_{j \neq i} \Theta_j$ on X is $\mathbf{a}_i : \Theta_{-i} \to X$, a function mapping Θ_{-i} into X. We denote the set of i's IAA acts sustained by Θ_{-i} on X by $\mathbf{A}_i$ and let $\mathbf{A}_i^c := \cup_{x \in X} \{\mathbf{a}_i^x \in \mathbf{A}_i\}$ where $\mathbf{a}_i^x$ is the *constant* IAA act with $\mathbf{a}_i^x(\theta_{-i}) = x$ for all $\theta_{-i} \in \Theta_{-i}$. For any $\tilde{\mathbf{A}}_i \subset \mathbf{A}_i$ and any θ_{-i}, $\tilde{\mathbf{A}}_i(\theta_{-i}) := \{x \in X \mid \mathbf{a}_i(\theta_{-i}) = x \text{ for some } \mathbf{a}_i \in \tilde{\mathbf{A}}_i\}$. Given $i \in N$, her type $\theta_i \in \Theta_i$, and a non-empty subset of IAA acts $\mathbf{S} \subset \mathbf{A}_i$, the *choice of individual i of type θ_i from the set of IAA acts* $\mathbf{S}$ is given by $\mathbf{C}_i^{\theta_i}(\mathbf{S}) \subset \mathbf{S}$. Individuals' choices are not necessarily non-empty valued unless otherwise stated explicitly, and we impose no further restrictions.[6]

Xiong (2023). For more on full implementation, we refer the reader to Moore (1992), Jackson (2001), Maskin and Sjöström (2002), Palfrey (2002), and Serrano (2004).

[4] Previous versions of the current paper contain some of this analysis.

[5] Unlike Jackson (1991), our formulation does not encompass the case of complete information because individuals' private information is exclusive. Consequently, the only common knowledge event among the individuals is the whole state space.

[6] Sen (1971) shows that a choice correspondence satisfies WARP (and is represented by a complete and transitive preference relation) if and only if it satisfies independence of irrelevant alternatives (referred to as IIA or Chernoff's α (Chernoff, 1954)) and an expansion consistency axiom (known as Sen's β). Letting $\mathcal{X}$ be the set of all non-empty subsets of X, we say that the choice correspondence c on X satisfies (i) the IIA if $x \in S \cap c(T)$ for some $S, T \in \mathcal{X}$ with $S \subset T$ implies $x \in c(S)$; (ii) Sen's β if $x, y \in S \subset T$ for some $S, T \in \mathcal{X}$, and $x, y \in c(S)$ implies $x \in c(T)$ if and only if $y \in c(T)$.

We summarize our environment by $\mathcal{E} = \langle N, X, (\Theta_i)_{i\in N}, (\mathbf{C}_i^{\theta_i})_{i\in N,\ \theta_i\in\Theta_i}\rangle$, which is common knowledge among the individuals.

An SCF $h : \Theta \to X$ is a state-contingent allocation mapping Θ into X. SCF $h : \Theta \to X$ induces an associated IAA act that individual i of type θ_i faces: $\mathbf{h}_{i,\theta_i} \in \mathbf{A}_i$ defined by $\mathbf{h}_{i,\theta_i}(\theta_{-i}) = h(\theta_i, \theta_{-i})$ for all $\theta_{-i} \in \Theta_{-i}$. The set of all SCFs is $H := \{h \mid h : \Theta \to X\}$.

As there may be many socially optimal SCFs that a designer may wish to consider simultaneously, we focus on SCSs: An SCS, denoted by F, is a non-empty set of SCFs, i.e., $F \subset H$ and $F \neq \emptyset$. An SCF $f \in F$ specifies a socially optimal alternative—as evaluated by the planner—for each state.[7] $\mathscr{F}$ denotes the set of all SCSs.

We denote a *mechanism* by $\mu = (M, g)$ where M_i denotes the non-empty set of *messages* available to individual i with $M = \times_{i\in N} M_i$, and $g : M \to X$ describes the *outcome function* that specifies the alternative to be selected for each message profile. A mechanism induces an incomplete information game form in our environment. A *strategy* of individual i under mechanism μ specifies a message for each possible type of i and is denoted by $\sigma_i : \Theta_i \to M_i$. *Individual i's opportunity set of IAA acts under μ for σ_{-i}* consists of IAA acts that i can unilaterally generate when the other individuals use $\sigma_{-i} := (\sigma_j)_{j\neq i}$, and it is given by

$$\mathbf{O}_i^{\mu}(\sigma_{-i}) := \bigcup_{m_i\in M_i} \{\mathbf{a}_i \in \mathbf{A}_i \mid \mathbf{a}_i(\theta_{-i}) = g(m_i, \sigma_{-i}(\theta_{-i})) \text{ for all } \theta_{-i} \in \Theta_{-i}\}\,.$$

Given individuals' choices on IAA acts, a natural equilibrium concept that Saran (2011) introduces is the behavioral interim equilibrium, which is defined as follows:

Definition 1. *A strategy profile $\sigma^* = (\sigma_i^*)_{i\in N}$ is a* **behavioral interim equilibrium** *(BIE) of mechanism $\mu = (M, g)$ if for all $i \in N$ and all $\theta_i \in \Theta_i$, $\mathbf{a}^*_{i,\theta_i} \in \mathbf{C}_i^{\theta_i}(\mathbf{O}_i^{\mu}(\sigma^*_{-i}))$, where $\mathbf{a}^*_{i,\theta_i}$ is the IAA act defined by $\mathbf{a}^*_{i,\theta_i}(\theta_{-i}) = g(\sigma_i^*(\theta_i), \sigma^*_{-i}(\theta_{-i}))$ for all $\theta_{-i} \in \Theta_{-i}$.*

In words, the strategy profile σ^* is a BIE of μ if for any player i and any type θ_i, the IAA act generated in the mechanism via the prescribed action $\sigma_i^*(\theta_i)$ is chosen by i of type θ_i from the opportunity set of IAA acts of i given others' strategy σ^*_{-i}. The concept of BIE gives rise to the following notion of implementation:

Definition 2. *An SCS $F \in \mathscr{F}$ is* **implementable in BIE** *if there is a mechanism μ such that*

(i) for all $f \in F$, there exists a BIE σ^f of μ such that $g \circ \sigma^f = f$, and

(ii) if σ^* is a BIE of μ, then $g \circ \sigma^* \in F$.

[7] We note that it is customary to denote a social choice rule as an SCS rather than a social choice correspondence under incomplete information. We refer to Holmström and Myerson (1983), Postlewaite and Schmeidler (1986), Palfrey and Srivastava (1987), Jackson (1991), and Bergemann and Morris (2008).

Given an SCS, implementability in BIE demands the existence of a mechanism such that (i) every SCF in the SCS must be sustained by a BIE of that mechanism, and (ii) every BIE of the mechanism must correspond to an SCF in the SCS. Hence, our focus is on full implementation. We refer to an SCF f as being *partially implementable in BIE* whenever condition (i) of Definition 2 holds for $F = \{f\}$.

Any mechanism that implements an SCS in BIE should take into consideration individuals' private information. However, individuals may misreport their private information. We denote a *deception* by individual i as $\alpha_i : \Theta_i \to \Theta_i$. The interpretation is that $\alpha_i(\theta_i)$ is individual i's reported type. Therefore, $\alpha(\theta) := (\alpha_1(\theta_1), \alpha_2(\theta_2), \ldots, \alpha_n(\theta_n))$ is a profile of (possibly deceptive) reported types while α^{id} denotes the *truthtelling profile*, i.e., $\alpha_i^{\text{id}}(\theta_i) = \theta_i$ for all $i \in N$ and all $\theta_i \in \Theta_i$. We denote the set of all possible deceptions of individual i by Λ_i. We let $\Lambda := \times_{i \in N} \Lambda_i$, $\Lambda_{-i} := \times_{j \neq i} \Lambda_j$, and $\alpha_{-i}(\theta_{-i}) := (\alpha_j(\theta_j))_{j \neq i}$. Moreover, a *garbling of an IAA act* $\mathbf{a}_i$ that individual i of type θ_i faces when the other individuals are using deception $\alpha_{-i} \in \Lambda_{-i}$ is the IAA act $\mathbf{a}_i^{\alpha} := \mathbf{a}_i \circ \alpha_{-i}$. Similarly, a *garbling of an SCF* $h \in H$ that individual i of type θ_i faces when the others are using deception $\alpha_{-i} \in \Lambda_{-i}$ is the IAA act $\mathbf{h}_{i,\theta_i}^{\alpha} := \mathbf{h}_{i,\theta_i} \circ \alpha_{-i}$.

3 Necessity

If an SCS F is implementable in BIE, then there is a mechanism μ such that for every SCF f in F, there is a BIE σ^f of μ that generates f. Thus, σ_{-i}^f induces an opportunity set for individual i from which each type θ_i of i chooses $\mathbf{f}_{i,\theta_i}$, the IAA act induced by f that i of type θ_i faces. Meanwhile, any individual $j \in N$ may employ any deception $\alpha_j : \Theta_j \to \Theta_j$ whereby j of type θ_j acts as if she is of type $\alpha_j(\theta_j)$. So, individual i of type θ_i faces the IAA act induced by $f \circ \alpha$, i.e., $\mathbf{f}_{i,\theta_i}^{\alpha}$, the garbling of $\mathbf{f}_{i,\theta_i}$ when others are using deception α_{-i}. Thus, i's opportunity set induced by $\sigma_{-i}^f \circ \alpha_{-i}$ must include this garbling.

Consequently, the notion of closedness under deception emerges naturally thanks to implementation under incomplete information:[8] Given an SCS $F \in \mathscr{F}$, a profile of sets of IAA acts $\mathbb{S} := (\mathbf{S}_i(f, \alpha_{-i}))_{i \in N,\ f \in F,\ \alpha_{-i} \in \Lambda_{-i}}$ is **closed under deception** if $\mathbf{a}_i \in \mathbf{S}_i(f, \alpha_{-i})$ implies $\mathbf{a}_i^{\tilde{\alpha}} \in \mathbf{S}_i(f, \tilde{\alpha}_{-i} \circ \alpha_{-i})$ for all $i \in N$, all $f \in F$, and all $\alpha, \tilde{\alpha} \in \Lambda_{-i}$.

We now introduce our necessary condition *interim consistency*:

Definition 3. *A profile of sets of IAA acts* $\mathbb{S} := (\mathbf{S}_i(f, \alpha_{-i}))_{i \in N,\ f \in F,\ \alpha_{-i} \in \Lambda_{-i}}$ *is* **interim consistent with SCS** $F \in \mathscr{F}$ *if it is closed under deception and for every SCF* $f \in F$*,*

(i) *for all* $i \in N$ *and all* $\theta_i \in \Theta_i$, $\mathbf{f}_{i,\theta_i} \in \mathbf{C}_i^{\theta_i}(\mathbf{S}_i(f, \alpha_{-i}^{\text{id}}))$*, and*

(ii) *for any deception profile* $\alpha \in \Lambda$ *with* $f \circ \alpha \notin F$*, there exists* $i^* \in N$ *and* $\theta_{i^*}^* \in \Theta_{i^*}$ *such that* $\mathbf{f}_{i^*,\theta_{i^*}^*}^{\alpha} \notin \mathbf{C}_{i^*}^{\theta_{i^*}^*}(\mathbf{S}_{i^*}(f, \alpha_{-i^*}))$.

[8] We thank an anonymous referee for pointing us toward the notion of closedness under deception.

A profile of sets of IAA acts $\mathbb{S} := (\mathbf{S}_i(f, \alpha_{-i}))_{i\in N,\ f\in F,\ \alpha_{-i}\in\Lambda_{-i}}$ satisfies interim consistency with an SCS F if $\mathbb{S}$ is closed under deception and for each $f \in F$, the following hold: (i) Given any $i \in N$ and any $\theta_i \in \Theta_i$, i's choices when she is of type θ_i from $\mathbf{S}_i(f, \alpha^{\mathrm{id}}_{-i})$ (the set of IAA acts in $\mathbb{S}$ associated with i, f, and the truthtelling profile of individuals other than i) contains the IAA act associated with f that she faces, namely $\mathbf{f}_{i,\theta_i}$; and (ii) if there is a deception profile α that leads to an outcome not compatible with the SCS, i.e., $f\circ\alpha \notin F$, then there exists an informant individual i^* of type $\theta^*_{i^*}$ who does not choose $\mathbf{f}^{\alpha}_{i^*,\theta^*_{i^*}}$ (the garbling of $\mathbf{f}_{i^*,\theta^*_{i^*}}$ that i^* of type $\theta^*_{i^*}$ faces when the others use deception α_{-i^*}) from $\mathbf{S}_{i^*}(f, \alpha_{-i^*})$ (the set of IAA acts in $\mathbb{S}$ associated with i^*, f, and α_{-i^*}).[9]

If a mechanism $\mu = (M, g)$ implements a given SCS $F \in \mathscr{F}$ in BIE, then for any SCF $f \in F$, there exists a BIE σ^f of μ such that $f = g \circ \sigma^f$. Let us define $\mathbb{S}$ by $\mathbf{S}_i(f, \alpha_{-i}) := \mathbf{O}^{\mu}_i(\sigma^f_{-i} \circ \alpha_{-i})$ for each $i \in N$, $f \in F$, and $\alpha_{-i} \in \Lambda_{-i}$.

First, we observe that $\mathbb{S}$ is closed under deception: If for any $i \in N$, $\mathbf{a}_i \in \mathbf{S}_i(f, \alpha_{-i})$, then $\mathbf{a}_i(\theta_{-i}) = g(m_i, \sigma^f_{-i}(\alpha_{-i}(\theta_{-i})))$ for some $m_i \in M_i$, for all $\theta_{-i} \in \Theta_{-i}$; hence, for any other deception profile $\tilde{\alpha} \in \Lambda$, $\mathbf{a}_i(\tilde{\alpha}_{-i}(\theta_{-i})) = g(m_i, \sigma^f_{-i}(\tilde{\alpha}_{-i}(\alpha_{-i}(\theta_{-i}))))$ for all $\theta_{-i} \in \Theta_{-i}$. Therefore, $\mathbf{a}^{\tilde{\alpha}}_i \in \mathbf{O}^{\mu}_i(\sigma^f_{-i} \circ \tilde{\alpha}_{-i} \circ \alpha_{-i})$ and hence $\mathbf{a}^{\tilde{\alpha}}_i \in \mathbf{S}_i(f, \tilde{\alpha}_{-i} \circ \alpha_{-i})$.

As for each $i \in N$ and $\theta_i \in \Theta_i$, the IAA act associated with f that i of type θ_i faces, $\mathbf{f}_{i,\theta_i}$, is in $\mathbf{C}^{\theta_i}_i(\mathbf{O}^{\mu}_i(\sigma^f_{-i}))$, (i) of interim consistency of $\mathbb{S}$ with F holds since σ^f is a BIE of μ such that $f = g \circ \sigma^f$ while $\sigma^f_{-i} \circ \alpha^{\mathrm{id}}_{-i} = \sigma^f_{-i}$ implies that $\mathbf{O}^{\mu}_i(\sigma^f_{-i}) = \mathbf{S}_i(f, \alpha^{\mathrm{id}}_{-i})$.

On the other hand, if a deception profile α is such that $f \circ \alpha \notin F$, then $\sigma^f \circ \alpha$ cannot be a BIE of μ. Otherwise, by (ii) of implementability in BIE, there exists $\tilde{f} \in F$ with $\tilde{f} = g \circ \sigma^f \circ \alpha$. But, since $f = g \circ \sigma^f$, we have $\tilde{f} = f \circ \alpha \notin F$, a contradiction. So, there is an individual i^* of type $\theta^*_{i^*}$ who does not choose $\mathbf{f}^{\alpha}_{i^*,\theta^*_{i^*}}$, the IAA act associated with $f \circ \alpha$ that i^* of type $\theta^*_{i^*}$ faces, from $\mathbf{O}^{\mu}_{i^*}(\sigma^f_{-i^*} \circ \alpha_{-i^*})$, which equals $\mathbf{S}_{i^*}(f, \alpha_{-i^*})$. This delivers (ii) of interim consistency of $\mathbb{S}$ with F.

The above discussion proves that the existence of a profile interim consistent with an SCS is a necessary condition for this SCS to be implementable in BIE:

Theorem 1. *If an SCS $F \in \mathscr{F}$ is implementable in BIE, then there is a profile of sets of IAA acts interim consistent with F.*

Theorem 1 affirms the following intuition along the same lines as de Clippel (2014): If the designer cannot identify sets from which individuals make choices compatible with the social goal, then she cannot succeed in the corresponding implementation attempt. Moreover, these sets should be constructed such that 'bad equilibria' do not emerge.

[9] Consistency of de Clippel (2014), a necessary condition for behavioral implementation under complete information, requires that, given a social choice correspondence $\Phi : \Theta \to \mathcal{X}$, there exists a collection $\{S_i(x, \theta) \in \mathcal{X} \mid i \in N, \theta \in \Theta, x \in \Phi(\theta)\}$, such that (i) for all $i \in N$, all $\theta \in \Theta$, and all $x \in \Phi(\theta)$, $x \in c^{\theta}_i(S_i(x, \theta))$; (ii) $x \in \Phi(\theta) \setminus \Phi(\theta')$ with $\theta, \theta' \in \Theta$ implies there is $i^* \in N$ such that $x \notin c^{\theta'}_{i^*}(S_{i^*}(x, \theta))$.

The identification of sets of IAA acts from which agents' choices are aligned with the social goal brings about the following incentive compatibility condition:

Definition 4. *An SCS $F \in \mathscr{F}$ is* **quasi-incentive compatible** *if for all $f \in F$ and all $i \in N$, there exists a set of IAA acts $\mathbf{T}_i \subset \mathbf{A}_i$ with $\{\mathbf{f}_{i,\tilde{\theta}_i} \mid \tilde{\theta}_i \in \Theta_i\} \subset \mathbf{T}_i$ and $\mathbf{f}_{i,\theta_i} \in \mathbf{C}_i^{\theta_i}(\mathbf{T}_i)$ for all $\theta_i \in \Theta_i$.*

We note that quasi-incentive compatibility of an SCS F follows from the existence of an interim consistent profile of sets of IAA acts $\mathbb{S} := (\mathbf{S}_i(f, \alpha_{-i}))_{i\in N,\ f\in F,\ \alpha_{-i}\in\Lambda_{-i}}$. To see this, for any given $f \in F$ and any $i \in N$, let $\mathbf{T}_i = \mathbf{S}_i(f, \alpha_{-i}^{\mathrm{id}})$. Then, for all $i \in N$, $\{\mathbf{f}_{i,\tilde{\theta}_i} \mid \tilde{\theta}_i \in \Theta_i\} \subset \mathbf{T}_i$, and (i) of interim consistency implies $\mathbf{f}_{i,\theta_i} \in \mathbf{C}_i^{\theta_i}(\mathbf{T}_i)$ for all $\theta_i \in \Theta_i$. This proves the following result:

Proposition 1. *If there exists a profile of sets of IAA acts interim consistent with an SCS $F \in \mathscr{F}$, then F satisfies quasi-incentive compatibility.*

The quasi-incentive compatibility of SCS $F = \{f\}$ is equivalent to the partial implementability of SCF f in BIE as established in Proposition 4.9 of Saran (2011). That study identifies a condition called *weak contraction consistency* (implied by the IIA), proves that this condition is necessary and sufficient for the revelation principle, and shows that the revelation principle fails under non-rational choices.

4 Sufficiency

Implementation of an SCS F in BIE is not feasible when there is no profile of sets of IAA acts that is interim consistent with F. Therefore, the planner should start the design by identifying such profiles and then explore additional requirements to be imposed on these for sufficiency. In this section, we present such conditions.[10]

Definition 5. *The* **choice incompatibility** *condition holds in environment $\mathcal{E}$ whenever the following holds: If for any SCF $h \in H$ and any $\bar{\theta} \in \Theta$, a profile of sets of IAA acts $(\tilde{\mathbf{A}}_i)_{i\in N}$ is such that*

(i) for all $i \in N$, $\mathbf{h}_{i,\bar{\theta}_i} \in \tilde{\mathbf{A}}_i$, and

(ii) there is $\bar{j} \in N$ such that for all $i \in N \setminus \{\bar{j}\}$, $\tilde{\mathbf{A}}_i(\bar{\theta}_{-i}) = X$,

then there is $i^* \in N \setminus \{\bar{j}\}$ such that $\mathbf{h}_{i^*,\bar{\theta}_{i^*}} \notin \mathbf{C}_{i^*}^{\bar{\theta}_{i^*}}(\tilde{\mathbf{A}}_{i^*})$.

[10] There is room for other sufficient conditions since we do not restrict choices using universal axioms. But, it seems neither easy nor practical to close the gap between the necessary and sufficient conditions.

In words, the choice incompatibility condition demands the following: Suppose SCF h, state $\bar{\theta}$, and a profile of sets of IAA acts $(\tilde{\mathbf{A}}_i)_{i\in N}$ are such that (i) for each individual $i \in N$, the IAA act induced by h for i's type $\bar{\theta}_i$, $\mathbf{h}_{i,\bar{\theta}_i}$, is in $\tilde{\mathbf{A}}_i$, and (ii) apart from an odd-man-out $\bar{j} \in N$, for all $i \neq \bar{j}$, the alternatives sustained in projection by an IAA act in $\tilde{\mathbf{A}}_i$, namely $\tilde{\mathbf{A}}_i(\bar{\theta}_{-i})$, equals X. Then, there is an individual i^*, different than the odd-man-out, who does not choose $\mathbf{h}_{i^*,\bar{\theta}_{i^*}}$ at her type $\bar{\theta}_{i^*}$ from $\tilde{\mathbf{A}}_{i^*}$. This condition implies some level of disagreement among individuals regarding their evaluation of SCFs in some circumstances. To see why note that if SCF h, state $\bar{\theta}$, and the profile of IAA acts $(\tilde{\mathbf{A}}_i)_{i\in N}$ are such that $\mathbf{h}_{i,\bar{\theta}_i} \in \mathbf{C}_i^{\bar{\theta}_i}(\tilde{\mathbf{A}}_i)$ for all $i \in N$, then (i) of choice incompatibility follows trivially and hence existence of $\bar{j} \in N$ with $\tilde{\mathbf{A}}_i(\bar{\theta}_{-i}) = X$ for all $i \neq \bar{j}$ implies choice incompatibility cannot hold.

The choice incompatibility condition coupled with interim consistency is sufficient for implementation in BIE whenever there are at least three individuals in the society.

Theorem 2. *Suppose that environment $\mathcal{E}$ is such that $n \geq 3$ and the choice incompatibility condition holds. Then, if there exists a profile of sets of IAA acts interim consistent with SCS $F \in \mathscr{F}$, then F is implementable in BIE.*

To see the details of the proof of our sufficiency result, assume that there are at least three individuals, the choice incompatibility condition holds, and F is an SCS for which the profile $\mathbb{S} := (\mathbf{S}_i(f, \alpha_{-i}))_{i\in N,\ f\in F,\ \alpha_{-i}\in\Lambda_{-i}}$ is interim consistent. The mechanism we employ makes use of the following observations: (i) the outcome should be $f(\alpha(\theta))$ when there is unanimous agreement between the individuals over $f \in F$, the realized state is θ, and all individuals are reporting their types following some deception $\alpha \in \Lambda$;[11] (ii) then, under such a unanimous agreement, each individual j should be able to generate unilaterally the set $\mathbf{S}_j(f, \alpha_{-j})$; (iii) whenever there is an attempt to deceive the designer with a deceptive unanimous agreement that results in an outcome not compatible with the SCS, a whistle-blower should be able to alert the designer; (iv) the remaining undesirable BIE should be eliminated (e.g., by using a modulo or an integer game).[12]

The mechanism $\mu = (M, g)$ we define below satisfies the properties discussed above: For each $i \in N$, $M_i = F \cup \{\emptyset\} \times \Theta_i \times \mathbf{A}_i \times X \times N$, while a generic message is denoted by $m_i = (m_i^1, \theta_i^{(i)}, \mathbf{a}_i^{(i)}, x^{(i)}, k^{(i)})$, and the outcome function $g : M \to X$ is as in Table 1.

In words, each individual i is required to send a message that consists of five components. The first component specifies either an SCF $f^{(i)} \in F$ or a flag denoted by $\emptyset$, the

[11]We note that the outcome equals $f(\theta)$ if there is unanimous agreement among the individuals over $f \in F$, the realized state is θ, and all individuals are reporting their types truthfully as $\alpha^{\mathrm{id}}(\theta) = \theta$.

[12]Our mechanism resembles those used for sufficiency in the implementation literature. See for example, Repullo (1987), Saijo (1988), Moore and Repullo (1990), Maskin (1999), Bergemann and Morris (2008), de Clippel (2014), Koray and Yildiz (2018), and Altun et al. (2023).

Rule 1 :	$g(m) = f(\theta)$	if $m_i = (f, \theta_i, \cdot, \cdot, \cdot)$ for all $i \in N$,
Rule 2 :	$g(m) = \begin{cases} \tilde{\mathbf{a}}_j(\theta_{-j}) & \text{if } \tilde{\mathbf{a}}_j \in \mathbf{S}_j(f, \alpha^{\text{id}}_{-j}), \\ \mathbf{f}_{j,\tilde{\theta}_j}(\theta_{-j}) & \text{otherwise.} \end{cases}$	if $m_i = (f, \theta_i, \cdot, \cdot, \cdot)$ for all $i \in N \setminus \{j\}$ and $m_j = (m^1_j, \tilde{\theta}_j, \tilde{\mathbf{a}}_j, \cdot, \cdot)$ with $m^1_j \neq f$,
Rule 3 :	$g(m) = x^{(j)}$ where $j = \sum_{i \in N} k^{(i)}$ (mod n)	otherwise.

Table 1: The outcome function of the mechanism for Theorem 2.

second a type of herself $\theta_i^{(i)} \in \Theta_i$, the third an IAA act $\mathbf{a}_i^{(i)} \in \mathbf{A}_i$, the fourth an alternative $x^{(i)} \in X$, and the fifth a number $k^{(i)} \in N = \{1, 2, \ldots, n\}$.

Rule 1 indicates that if there is unanimity among the individuals' messages regarding the SCF to be implemented, then the outcome is determined according to this SCF and the reported type profile in the messages. Rule 2 indicates that if there is an agreement between all the individuals but one regarding SCF $f \in F$ in their messages, then the outcome is in line with the IAA act proposed by the odd-man-out, j, whenever this act is in $\mathbf{S}_j(f, \alpha^{\text{id}}_{-j})$, otherwise the outcome is in line with SCF f (while the IAA act associated with f that j faces is also in $\mathbf{S}_j(f, \alpha^{\text{id}}_{-j})$). Finally, Rule 3 applies when both Rules 1 and 2 fail, then the outcome is determined by the winner of the modulo game.

If there is unanimity among the individuals' messages regarding the SCF to be implemented and all individuals are reporting their types truthfully, then Rule 1 applies at every state, and hence the opportunity set of IAA acts of any individual i is $\mathbf{S}_i(f, \alpha^{\text{id}}_{-i})$. As the opportunity sets of IAA acts under truthtelling satisfy (i) of interim consistency, the unanimous agreement on an SCF $f \in F$ along with the truthful revelation of types is a BIE of this mechanism sustaining f. Further, under any BIE of our mechanism, Rule 1 applies at every state $\theta \in \Theta$ thanks to the choice incompatibility condition. Consequently, the opportunity sets of IAA acts under any (possibly deceptive) BIE are induced by the interim consistent profile $\mathbb{S}$ thanks to this profile being closed under deception. That is why every BIE must be aligned with SCS F; otherwise, by (ii) of interim consistency, there is a whistle-blower who does not choose to go along with the others' deception. Below, we provide the proof that is sketched above.

Proof of Theorem 2. Suppose that $n \geq 3$ and the choice incompatibility condition holds. Let $F \in \mathscr{F}$ be an SCS for which the profile $\mathbb{S} := (\mathbf{S}_i(f, \alpha_{-i}))_{i \in N,\ f \in F,\ \alpha_{-i} \in \Lambda_{-i}}$ is interim consistent. The mechanism we use is as defined on page 9.

First, we show that for any $f \in F$, there exists σ^f a BIE of $\mu = (M, g)$ such that $f = g \circ \sigma^f$. That is, condition (i) of implementability in BIE holds.

Take any $f \in F$, let $\sigma_i^f(\theta_i) = (f, \theta_i, \mathbf{f}_{i,\theta_i}, \bar{x}, 1)$ for each $i \in N$ and some $\bar{x} \in X$. Then, Rule 1 applies and we have $g(\sigma^f(\theta)) = f(\theta)$ for each $\theta \in \Theta$, i.e., $f = g \circ \sigma^f$. For any unilateral deviation of individual i from σ^f, either Rule 1 or Rule 2 applies, while Rule 3 is not attainable. Hence, by construction, $\mathbf{O}_i^\mu(\sigma_{-i}^f) = \mathbf{S}_i(f, \alpha_{-i}^{\text{id}})$ for all $i \in N$. Recall that, by (i) of interim consistency, $\mathbf{f}_{i,\theta_i} \in \mathbf{C}_i^{\theta_i}(\mathbf{S}_i(f, \alpha_{-i}^{\text{id}}))$ for each $i \in N$ and each $\theta_i \in \Theta_i$. Ergo, for all $i \in N$ and all $\theta_i \in \Theta_i$, $\mathbf{a}_{i,\theta_i}^* \in \mathbf{C}_i^{\theta_i}(\mathbf{O}_i^\mu(\sigma_{-i}^f))$ where $\mathbf{a}_{i,\theta_i}^*(\theta_{-i}) = g(\sigma_i^f(\theta_i), \sigma_{-i}^f(\theta_{-i}))$ for all $\theta_{-i} \in \Theta_{-i}$. So, σ^f is a BIE of μ such that $f = g \circ \sigma^f$.

Consider now any BIE σ^* of μ denoted as $\sigma_i^*(\theta_i) = (m_i^1(\theta_i), \alpha_i(\theta_i), \mathbf{a}_i(\theta_i), x_i(\theta_i), k_i(\theta_i))$ for each $i \in N$. That is, $m_i^1(\theta_i)$ denotes the first component (either a proposed SCF or a flag), $\alpha_i(\theta_i)$ the reported type, $\mathbf{a}_i(\theta_i)$ the proposed act, $x_i(\theta_i)$ the proposed alternative, and $k_i(\theta_i)$ the proposed number by i when her realized type is θ_i.

Next, we show that, under any BIE σ^* of μ, Rule 1 must apply at every state $\theta \in \Theta$: Suppose for a contradiction that either Rule 2 or Rule 3 applies under σ^* at $\bar{\theta}$ and consider $(\tilde{\mathbf{A}}_i)_{i\in N}$ with $\tilde{\mathbf{A}}_i := \mathbf{O}_i^\mu(\sigma_{-i}^*)$ for all $i \in N$. Let SCF $h^* := g \circ \sigma^*$. Then, $\mathbf{h}_{i,\bar{\theta}_i}^* \in \tilde{\mathbf{A}}_i$ for all $i \in N$. Further, $\tilde{\mathbf{A}}_i(\bar{\theta}_{-i}) = X$ for all $i \neq \bar{j}$ for some $\bar{j} \in N$. To see why consider the following: If Rule 2 applies under σ^* at $\bar{\theta}$ with $\bar{j}$ as the odd-man-out, then, for any $x \in X$ and $i \neq \bar{j}$, $(\tilde{\sigma}_i, \sigma_{-i}^*)$ triggers Rule 3 at $\bar{\theta}$ and delivers x where $\tilde{\sigma}_i$ is such that $\tilde{\sigma}_i(\theta_i) = \sigma_i^*(\theta_i)$ for all $\theta_i \neq \bar{\theta}_i$ and $\tilde{\sigma}_i(\bar{\theta}_i) = (\emptyset, \alpha_i(\bar{\theta}_i), x, k^*)$ where k^* is the number that makes i the winner of the modulo game at $\bar{\theta}$ given σ_{-i}^*. If Rule 3 applies under σ^* at $\bar{\theta}$, one can simply take $\bar{j} = 1$ and repeat the steps above. Thus, for SCF h^* and state $\bar{\theta}$, the profile of sets of IAA acts $(\tilde{\mathbf{A}}_i)_{i\in N}$ satisfies both (i) and (ii) of the choice incompatibility condition with the odd-man-out given by $\bar{j}$. So, there is $i^* \neq \bar{j}$ with $\mathbf{h}_{i^*,\bar{\theta}_{i^*}} \notin \mathbf{C}_{i^*}^{\bar{\theta}_{i^*}}(\tilde{\mathbf{A}}_{i^*})$. This delivers the desired contradiction as σ^* is a BIE of μ and hence $\mathbf{h}_{i,\bar{\theta}_i}^* \in \mathbf{C}_i^{\bar{\theta}_i}(\tilde{\mathbf{A}}_i)$ for all $i \in N$.

Moreover, due to the product structure of the state space, under any BIE σ^* of μ, there is a unique $f \in F$ such that $m_i^1(\theta_i) = f$ for all $i \in N$ and all $\theta_i \in \Theta_i$. To see why, suppose there are i, j with $i \neq j$, who propose different SCFs under σ^*, say $f, f' \in F$ with $f \neq f'$ for their types θ_i and θ_j, respectively. Then, Rule 1 cannot apply at $(\theta_i, \theta_j, \theta_{-\{i,j\}})$, a contradiction to the conclusion that under any BIE σ^* of μ, Rule 1 holds at each $\theta \in \Theta$.

Finally, we show that $f \circ \alpha \in F$: Since Rule 1 applies at each $\theta \in \Theta$, and each $i \in N$ reports the type $\alpha_i(\theta_i) \in \Theta_i$ as the second entry of their messages at $\theta \in \Theta$ under σ^*, by construction and $\mathbb{S}$ being closed under deception, we have, at each $\theta \in \Theta$, $\mathbf{O}_i^\mu(\sigma_{-i}^*) = \cup_{\mathbf{a}_i \in \mathbf{S}_i(f, \alpha_{-i}^{\text{id}})}\{\mathbf{a}_i \circ \alpha_{-i}\} = \mathbf{S}_i(f, \alpha_{-i}^{\text{id}} \circ \alpha_{-i}) = \mathbf{S}_i(f, \alpha_{-i})$ for all $i \in N$.[13] If $f \circ \alpha \notin F$, then by (ii) of interim consistency, there exists $i^* \in N$ and $\theta_{i^*}^* \in \Theta_{i^*}$ such that $\mathbf{f}_{i^*,\theta_{i^*}^*}^\alpha \notin \mathbf{C}_{i^*}^{\theta_{i^*}^*}(\mathbf{S}_{i^*}(f, \alpha_{-i^*}))$. But this implies $\mathbf{f}_{i^*,\theta_{i^*}^*}^\alpha = \mathbf{h}_{i^*,\theta_{i^*}^*}^* \notin \mathbf{C}_{i^*}^{\theta_{i^*}^*}(\mathbf{O}_{i^*}^\mu(\sigma_{-i^*}^*))$. This contradicts σ^* being a BIE of μ. Therefore, $h^* = g \circ \sigma^* = f \circ \alpha \in F$, which implies that

[13] We note that $\cup_{\tilde{\theta}_i \in \Theta_i}\{\mathbf{f}_{i,\tilde{\theta}_i}^\alpha\} \subset \cup_{\mathbf{a}_i \in \mathbf{S}_i(f,\alpha_{-i}^{\text{id}})}\{\mathbf{a}_i \circ \alpha_{-i}\}$ as $\mathbf{f}_{i,\tilde{\theta}_i} \in \mathbf{S}_i(f, \alpha_{-i}^{\text{id}})$ for all $\tilde{\theta}_i$ thanks to (i) of interim consistency.

condition (ii) of implementability in BIE holds as well. $\blacksquare$

We wish to emphasize that the choice incompatibility condition allows us to construct a canonical mechanism under which BIE emerges only under Rule 1. Moreover, thanks to this condition, we do not need finite state spaces or the existence of special (state-contingent) alternatives such as the 0 alternative in Corollary 2 of Jackson (1991).

On the other hand, we can weaken choice incompatibility due to the following observation: This condition specified for h, $\bar{\theta}$, $(\tilde{\mathbf{A}}_i)_{i\in N}$, and the odd-man-out $\bar{j}$ does not require the entanglement of h with any alternative $x \in X$ at $\bar{\theta}$ to induce an IAA act individual $i \neq \bar{j}$ of type $\bar{\theta}_i$ faces to be in $\tilde{\mathbf{A}}_i$. Nevertheless, such entanglements may be achievable via unilateral deviations under Rule 2 and Rule 3 of our canonical mechanism. To clarify, let us consider the following example:[14]

Let $X = \{1, 2\}$, $N = \{1, 2, 3\}$, $\Theta_i = \{\theta_i, \theta_i'\}$ for all $i \in N$, and SCF h be as follows:

	$(\theta_1,\theta_2,\theta_3)$	$(\theta_1',\theta_2,\theta_3)$	$(\theta_1,\theta_2',\theta_3)$	$(\theta_1,\theta_2,\theta_3')$	$(\theta_1',\theta_2',\theta_3)$	$(\theta_1',\theta_2,\theta_3')$	$(\theta_1,\theta_2',\theta_3')$	$(\theta_1',\theta_2',\theta_3')$
$h(\theta)$	x	y	y	y	x	x	x	y

SCF h induces the following IAA acts for each player i:

	(θ_j,θ_k)	(θ_j',θ_k)	(θ_j,θ_k')	(θ_j',θ_k')
$\mathbf{h}_{i,\theta_i}(\theta_{-i})$	x	y	y	x
$\mathbf{h}_{i,\theta_i'}(\theta_{-i})$	y	x	x	y

Consider the profile of sets of IAA acts $(\tilde{\mathbf{A}}_i)_{i\in N}$ with $\tilde{\mathbf{A}}_i := \{\langle xyyx\rangle, \langle yxxy\rangle\}$ for each $i \in N$ where $\mathbf{h}_{i,\theta_i}$ and $\mathbf{h}_{i,\theta_i'}$ are denoted by $\langle xyyx\rangle$ and $\langle yxxy\rangle$, respectively. For SCF h and any $\bar{\theta} \in \Theta$, the profile of sets of IAA acts $(\tilde{\mathbf{A}}_i)_{i\in N}$ with any odd-man-out $\bar{j}$ satisfy (i) and (ii) of the hypothesis of choice incompatibility. Thus, there is $i^* \neq \bar{j}$ such that $\mathbf{h}_{i^*,\bar{\theta}_{i^*}} \notin \mathbf{C}_{i^*}^{\bar{\theta}_{i^*}}(\tilde{\mathbf{A}}_{i^*})$ whenever the choice incompatibility condition holds.

In general, the choice incompatibility condition guarantees that Rule 2 or Rule 3 cannot arise at any state under any BIE of our canonical mechanism specified in Table 1. The key observation is that the profile of opportunity sets of IAA acts in this mechanism satisfies both (i) and (ii) of the hypothesis of the choice incompatibility condition whenever Rule 2 or Rule 3 holds. However, the sets in the profile $(\tilde{\mathbf{A}}_i)_{i\in N}$ in the above example are not guaranteed to arise as opportunity sets in our canonical mechanism under Rule 2 or Rule 3. To see why, let $\bar{\theta} = (\theta_1, \theta_2, \theta_3)$, $\bar{j} = 3$, suppose that Rule 2 holds at $\bar{\theta}$ under σ, and Individual 1 of type θ_1 deviates to $m_1 = (\emptyset, ., ., y, k^*)$. Then, Rule 3 applies at $\bar{\theta}$. Additionally, if Rule 3 applies at $(\theta_1, \tilde{\theta}_{-1})$ for all $\tilde{\theta}_{-1} \in \Theta_{-1}$ and k^* makes her the

[14] We thank an anonymous referee for pointing out this example.

winner of the modulo game in all such states, given m_1, σ_2, and σ_3, then as a result of this deviation, Individual 1 of type θ_1 faces the IAA act $\langle yyyy\rangle$ which is not in $\tilde{\mathbf{A}}_1$.

The basis of the weaker choice incompatibility condition we provide below involves the observation that unilateral deviations in the canonical mechanism under Rule 2 or Rule 3 induce IAA acts achievable via the entanglements exemplified above. To formalize such entanglements, we need the following: For any pair of IAA acts $\mathbf{a}_i, \tilde{\mathbf{a}}_i \in \mathbf{A}_i$, we define the *splicing* of $\mathbf{a}_i$ with $\tilde{\mathbf{a}}_i$ along a set $\Theta' \subset \Theta$ as follows:

$$[\mathbf{a}_i/_{\Theta'}\tilde{\mathbf{a}}_i](\theta_{-i}) = \begin{cases} \mathbf{a}_i(\theta_{-i}) & \text{if } \theta_{-i} \in \Theta'_{-i}, \\ \tilde{\mathbf{a}}_i(\theta_{-i}) & \text{otherwise.} \end{cases}$$

Recall that for any $x \in X$ and any individual $i \in N$, $\mathbf{a}_i^x \in \mathbf{A}_i^c$ is i's constant IAA act resulting in x, i.e., $\mathbf{a}_i^x(\theta_{-i}) = x$ for all $\theta_{-i} \in \Theta_{-i}$.

Definition 6. *The* **weak choice incompatibility** *condition holds in environment $\mathcal{E}$ whenever the following holds: If for any SCF $h \in H$ and any $\bar{\theta} \in \Theta$, a profile of sets of IAA acts $(\tilde{\mathbf{A}}_i)_{i\in N}$ is such that*

- *(i) for all $i \in N$, $\mathbf{h}_{i,\bar{\theta}_i} \in \tilde{\mathbf{A}}_i$, and*
- *(ii) there is $\bar{j} \in N$ such that for all $i \in N \setminus \{\bar{j}\}$, for any $x \in X$, $[\mathbf{a}_i^x/_{\Theta'}\mathbf{h}_{i,\bar{\theta}_i}] \in \tilde{\mathbf{A}}_i$ for some $\Theta' \subset \Theta$ with $\bar{\theta} \in \Theta'$,*

then there is $i^* \in N \setminus \{\bar{j}\}$ such that $\mathbf{h}_{i^*,\bar{\theta}_{i^*}} \notin \mathbf{C}_{i^*}^{\bar{\theta}_{i^*}}(\tilde{\mathbf{A}}_{i^*})$.

We note that (ii) of the hypothesis of weak choice incompatibility implies (ii) of the hypothesis of choice incompatibility. That is, for any SCF h and state $\bar{\theta}$, if the profile of sets of IAA acts $(\tilde{\mathbf{A}}_i)_{i\in N}$ satisfies (ii) of weak choice incompatibility, then $\tilde{\mathbf{A}}_i(\bar{\theta}_{-i}) = X$ for all $i \neq \bar{j}$. Thus, choice incompatibility implies weak choice incompatibility.

We would like to also highlight that the weak choice incompatibility condition implies the following: If for any SCF $h \in H$ and any $\bar{\theta} \in \Theta$, a profile of sets of IAA acts $(\tilde{\mathbf{A}}_i)_{i\in N}$ is such that for all $i \in N$, $\mathbf{h}_{i,\bar{\theta}_i} \in \tilde{\mathbf{A}}_i$, and for any $x \in X$, $[\mathbf{a}_i^x/_{\Theta'}\mathbf{h}_{i,\bar{\theta}_i}] \in \tilde{\mathbf{A}}_i$ for some $\Theta' \subset \Theta$ with $\bar{\theta} \in \Theta'$, then there is i^* such that $\mathbf{h}_{i^*,\bar{\theta}_{i^*}} \notin \mathbf{C}_{i^*}^{\bar{\theta}_{i^*}}(\tilde{\mathbf{A}}_{i^*})$.[15]

In what follows, we strengthen our sufficiency result in environments with *finite state spaces* and *null alternatives* by employing the weak choice incompatibility condition.

Definition 7. *An alternative $z \in X$ is a* ***null alternative of individiual*** *$i \in N$ if for all $\theta_i \in \Theta_i$, $\mathbf{C}_i^{\theta_i}(\tilde{\mathbf{A}}_i) = \mathbf{C}_i^{\theta_i}(\tilde{\mathbf{A}}_i \cup \{\mathbf{a}_i^z\})$ for any non-empty $\tilde{\mathbf{A}}_i \subset \mathbf{A}_i$.*

[15] To see why, let $h \in H$, $\bar{\theta} \in \Theta$, and consider a profile of sets of IAA acts $(\tilde{\mathbf{A}}_i)_{i\in N}$ such that for all $i \in N$, $\mathbf{h}_{i,\bar{\theta}_i} \in \tilde{\mathbf{A}}_i$, and for any $x \in X$, $[\mathbf{a}_i^x/_{\Theta'}\mathbf{h}_{i,\bar{\theta}_i}] \in \tilde{\mathbf{A}}_i$ for some $\Theta' \subset \Theta$ with $\bar{\theta} \in \Theta'$. Then, h, $\bar{\theta}$, and $(\tilde{\mathbf{A}}_i)_{i\in N}$ satisfies (ii) of weak choice incompatibility by letting $\bar{j} = 1$ and hence there is $i^* \neq 1$ such that $\mathbf{h}_{i^*,\bar{\theta}_{i^*}} \notin \mathbf{C}_{i^*}^{\bar{\theta}_{i^*}}(\tilde{\mathbf{A}}_{i^*})$.

Intuitively, z is a null alternative of player i if $\mathbf{a}_i^z$, the constant IAA act that results in z, does not affect the choices of any type of individual i whenever $\mathbf{a}_i^z$ is added to the set of IAA acts under consideration. We refer to $\mathbf{a}_i^z$ as a null IAA act of individual i.

We are now ready to present our second sufficiency result:

Theorem 3. *Suppose that environment $\mathcal{E}$ is such that $|\Theta| < \infty$, $n \geq 3$, the weak choice incompatibility condition holds, and for each individual, there is a null alternative. Then, if there exists a profile of sets of IAA acts interim consistent with SCS $F \in \mathscr{F}$, then F is implementable in BIE.*

The mechanism we construct for the proof of Theorem 3 is as follows: For each individual $i \in N$, $M_i = (F \cup \{\emptyset\}) \times \Theta_i \times (\mathbf{A}_i \cup \{\emptyset\}) \times (H \cup \{\emptyset\}) \times \mathbb{N}$, and a generic message is denoted by $m_i = (m_i^1, \theta_i^{(i)}, \mathbf{a}_i^{(i)}, m_i^4, k^{(i)})$, and $g : M \to X$ is as specified in Table 2 for a given profile of null alternatives $(z^i)_{i \in N}$ with $j^* := \min\{i \in N \mid k^{(i)} \geq k^{(j)} \text{ for all } j \in N\}$:

Rule 1 :	$g(m) = f(\theta)$	if $m_i = (f, \theta_i, \emptyset, \emptyset, \cdot)$ for all $i \in N$,
Rule 1′ :	$g(m) = f(\theta)$	if $m_i = (f, \theta_i, \emptyset, \emptyset, \cdot)$ for all $i \in N \setminus \{j\}$ and $m_j = (\emptyset, \theta_j, \emptyset, h, \cdot)$,
Rule 2 :	$g(m) = \begin{cases} \mathbf{a}_j(\theta_{-j}) & \text{if } \mathbf{a}_j \in \mathbf{S}_j(f, \alpha_{-j}^{\text{id}}), \\ \mathbf{f}_{j,\theta_j}(\theta_{-j}) & \text{otherwise.} \end{cases}$	if $m_i = (f, \theta_i, \emptyset, \emptyset, \cdot)$ for all $i \in N \setminus \{j\}$ and $m_j = (m_j^1, \theta_j, \mathbf{a}_j, \cdot, \cdot)$,
Rule 2′ :	$g(m) = z^j$	if $m_i = (f, \theta_i, \emptyset, \emptyset, \cdot)$ for all $i \in N \setminus \{j\}$ and $m_j = (m_j^1, \cdot, \emptyset, \emptyset, \cdot)$ with $m_j^1 \neq f$,
Rule 3 :	$g(m) = h^{(j^*)}(\theta)$ where $\theta = (\theta_i^{(i)})_{i \in N}$	otherwise.

Table 2: The outcome function of the mechanism for Theorem 3.

In words, each individual i is required to send a message that consists of five components. The first component specifies either an SCF $f^{(i)} \in F$ or a flag, the second a type of herself $\theta_i^{(i)} \in \Theta_i$, the third either an IAA act $\mathbf{a}_i^{(i)} \in \mathbf{A}_i$ or a flag, the fourth either an SCF $h^{(i)} \in H$ or a flag, and the fifth an integer $k^{(i)} \in \mathbb{N} = \{1, 2, \ldots\}$. The flags are denoted by $\emptyset$. Rule 1 demands that if all individuals agree upon the same SCF and raise flags as their third and fourth components, then the outcome is determined according to the agreed upon SCF and the reported type profile in their messages. Rule 1′ indicates that if $n - 1$ individuals agree upon the same SCF and raise flags as their third and fourth components while the odd-man-out proposes a flag as her first and third components but not as her fourth component, then the outcome is again determined according to the SCF agreed upon by $n - 1$ individuals and the reported type profile in the messages. Rule 2 decrees that if there is an agreement between all the individuals but one regarding SCF $f \in F$ and these $n-1$ individuals raise flags as their third and fourth components and the odd-man-out j proposes an IAA act as her third component, this IAA act determines the outcome according to the reported types of the individuals other than j whenever that IAA act is in $\mathbf{S}_j(f, \alpha^{\text{id}}_{-j})$, otherwise the outcome is determined according to the IAA act induced by SCF f for individual j of her reported type θ_j, which is also in $\mathbf{S}_j(f, \alpha^{\text{id}}_{-j})$ due to (i) of interim consistency. Rule 2′ applies when the SCF proposed by the odd-man-out j in Rule 2 is different than the others, and she proposes flags as her third and fourth components. Then, the outcome is the odd-man-out's null alternative z^j. Finally, Rule 3 applies when other rules fail; the outcome is determined according to the SCF proposed by the winner of the integer game (the individual who has the minimum index among those who propose the highest integer) and the reported types in the messages.

Proof of Theorem 3. Suppose $|\Theta| < \infty$, $n \geq 3$, the choice incompatibility condition holds, and $(z^i)_{i \in N}$ is a profile of null alternatives. Let F be an SCS for which the profile $\mathbb{S} := (\mathbf{S}_i(f, \alpha_{-i}))_{i \in N,\ f \in F,\ \alpha_{-i} \in \Lambda_{-i}}$ is interim consistent. We use the mechanism on page 14.

First, we show that for any $f \in F$, there is σ^f a BIE of $\mu = (M, g)$ such that $f = g \circ \sigma^f$. That is, condition (i) of implementability in BIE holds. Take any $f \in F$, let $\sigma^f_i(\theta_i) = (f, \theta_i, \emptyset, \emptyset, 1)$ for each $i \in N$. Then, at every $\theta \in \Theta$, Rule 1 applies and $g(\sigma^f(\theta)) = f(\theta)$, i.e., $f = g \circ \sigma^f$. For any unilateral deviation of individual i from σ^f, either Rule 1 or Rule 1′ or Rule 2 or Rule 2′ applies, while Rule 3 is not attainable. So, $\mathbf{O}^\mu_i(\sigma^f_{-i}) = \mathbf{S}_i(f, \alpha^{\text{id}}_{-i}) \cup \{\mathbf{a}^{z^i}\}$ for all $i \in N$. Recall that, by (i) of interim consistency, $\mathbf{f}_{i,\theta_i} \in \mathbf{C}^{\theta_i}_i(\mathbf{S}_i(f, \alpha^{\text{id}}_{-i}))$ for each $i \in N$ and each $\theta_i \in \Theta_i$. Because $\mathbf{a}^{z^i}$ is a null IAA act of individual i, $\mathbf{f}_{i,\theta_i} \in \mathbf{C}^{\theta_i}_i(\mathbf{S}_i(f, \alpha^{\text{id}}_{-i}) \cup \{\mathbf{a}^{z^i}\})$. Hence, for all $i \in N$ and all $\theta_i \in \Theta_i$, $\mathbf{a}^*_{i,\theta_i} \in \mathbf{C}^{\theta_i}_i(\mathbf{O}^\mu_i(\sigma^f_{-i}))$ where $\mathbf{a}^*_{i,\theta_i}(\theta_{-i}) = g(\sigma^f_i(\theta_i), \sigma^f_{-i}(\theta_{-i}))$ for all $\theta_{-i} \in \Theta_{-i}$. Thus, σ^f is a BIE of μ such that $f = g \circ \sigma^f$.

Consider next any BIE σ^* of μ denoted as $\sigma^*_i(\theta_i) = (m^1_i(\theta_i), \alpha_i(\theta_i), m^3_i(\theta_i), m^4_i(\theta_i), k_i(\theta_i))$ for each $i \in N$. That is, $m^1_i(\theta_i)$ denotes the first component, i.e., either a proposed SCF or a flag, $\alpha_i(\theta_i)$ the reported type, $m^3_i(\theta_i)$ denotes the third component, i.e., either a proposed act or a flag, $m^4_i(\theta_i)$ the fourth component, i.e., either a proposed SCF or a flag, and $k_i(\theta_i)$ the proposed integer by i when her realized type is θ_i.

We now show that, under any BIE σ^* of μ, Rule 1 must apply at every state $\theta \in \Theta$. Let us denote the SCF that arises when individuals follow σ^* by h^*, i.e., $h^* := g \circ \sigma^*$. Hence, individual i of type θ_i faces IAA act $\mathbf{h}^*_{i,\theta_i}$ under σ^* and $\mathbf{h}^*_{i,\theta_i} \in \mathbf{C}^{\theta_i}_i(\mathbf{O}^\mu_i(\sigma^*_{-i}))$ for all $i \in N$ as σ^* is a BIE of μ.

Suppose Rule 1′ or Rule 2 or Rule 2′ or Rule 3 holds at $\bar{\theta}$ under σ^*. If Rules 1′, 2 or 2′ holds, let us denote the odd-man-out by $\bar{j}$. If Rule 3 applies, let $\bar{j} = 1$. Let $\ell \neq \bar{j}$

and $\bar{\Theta} := \{\theta \in \Theta \mid \theta_\ell = \bar{\theta}_\ell\}$. Let α_*^{-1} be a deception such that $h^* \circ \alpha_*^{-1}(\alpha(\theta)) = h^*(\theta)$ for all $\theta \in \Theta$.[16] For any $x \in X$, consider the deviation by ℓ to $\tilde{\sigma}_\ell$ such that $\tilde{\sigma}_\ell(\theta_\ell) = \sigma_\ell^*(\theta_\ell)$ for all $\theta_\ell \neq \bar{\theta}_\ell$ and $\tilde{\sigma}_\ell(\bar{\theta}_\ell) = (m_\ell^1(\bar{\theta}_\ell), \alpha_\ell(\bar{\theta}_\ell), \mathbf{h}^*_{\ell,\bar{\theta}_\ell} \circ \alpha^{-1}_{*-\ell}, [x/_{\bar{\Theta}}h^*] \circ \alpha_*^{-1}, k^*)$ where $k^* = \max_{i \neq \ell,\ \tilde{\theta}_i \in \Theta_i} k_i(\tilde{\theta}_i) + 1$, and $[x/_{\bar{\Theta}}h^*]$ is the entanglement of alternative x with SCF h^* along $\bar{\Theta}$ that is defined by $[x/_{\bar{\Theta}}h^*](\theta) = x$ for all $\theta \in \bar{\Theta}$ and $[x/_{\bar{\Theta}}h^*](\theta) = h^*(\theta)$ otherwise. That is, individual ℓ deviates from σ_ℓ^* only when her type is $\bar{\theta}_\ell$ so that the first and the second component of her deviation is the same as that of σ_ℓ^*, the third component is the IAA act that ℓ faces under σ^* when her true type is $\bar{\theta}_\ell$ and the others are reporting their types truthfully (as $\alpha_*^{-1} \circ \alpha = \alpha^{\mathrm{id}}$), the fourth component is the SCF obtained by composing the entanglement of alternative x with SCF h^* along $\bar{\Theta}$ with α_*^{-1}, and the last component is an integer that is higher than all proposed integers of the other individuals at all states, which is well-defined as $|\Theta| < \infty$.

Below, we analyze the effect of ℓ's deviation to the outcome at $(\bar{\theta}_i, \theta'_{-\ell})$ for any $\theta'_{-\ell} \in \Theta_{-\ell}$. Note that the outcome at $(\bar{\theta}_\ell, \theta'_{-\ell})$ under σ^* is $g(\sigma^*_\ell(\bar{\theta}_\ell), \sigma^*_{-\ell}(\theta'_{-\ell})) = h^*(\bar{\theta}_\ell, \theta'_{-\ell})$. Observe that, after ℓ's deviation, either Rule 2 or Rule 3 applies under $(\tilde{\sigma}_\ell, \sigma^*_{-\ell})$ at $(\bar{\theta}_\ell, \theta'_{-\ell})$. This is because ℓ does not raise any flag in $\tilde{\sigma}_\ell$ when her type is $\bar{\theta}_\ell$ and hence Rules 1, 1′, and 2′ cannot apply under $(\tilde{\sigma}_\ell, \sigma^*_{-\ell})$ at $(\bar{\theta}_\ell, \theta'_{-\ell})$.

For Rule 2 to apply under $(\tilde{\sigma}_\ell, \sigma^*_{-\ell})$ at $(\bar{\theta}_\ell, \theta'_{-\ell})$, ℓ has to be the odd-man-out after the deviation as she does not propose any flag in $\tilde{\sigma}_\ell$ when her type is $\bar{\theta}_\ell$. Then, independent of whether or not $\mathbf{h}^*_{\ell,\bar{\theta}_\ell} \circ \alpha^{-1}_{*-\ell}$ is in $\mathbf{S}_i(\bar{f}, \alpha^{\mathrm{id}}_{-\ell})$, by Rule 2, the outcome under $(\tilde{\sigma}_\ell, \sigma^*_{-\ell})$ at $(\bar{\theta}_\ell, \theta'_{-\ell})$ is $g(\tilde{\sigma}_\ell(\bar{\theta}_\ell), \sigma^*_{-\ell}(\theta'_{-\ell})) = h^*(\bar{\theta}_\ell, \theta'_{-\ell})$ because $\mathbf{h}^*_{\ell,\bar{\theta}_\ell}(\alpha^{-1}_{*-\ell}(\alpha_{-\ell}(\theta'_{-\ell}))) = h^*(\bar{\theta}_\ell, \theta'_{-\ell})$, which equals the outcome under σ^* at $(\bar{\theta}_\ell, \theta'_{-\ell})$. If Rule 3 applies under $(\tilde{\sigma}_\ell, \sigma^*_{-\ell})$ at $(\bar{\theta}_\ell, \theta'_{-\ell})$, then ℓ becomes the winner of the integer game under $(\tilde{\sigma}_\ell, \sigma^*_{-\ell})$ at $(\bar{\theta}_\ell, \theta'_{-\ell})$ and the outcome equals x as $[x/_{\bar{\Theta}}h^*](\alpha_*^{-1}(\alpha_\ell(\bar{\theta}_\ell), \alpha_{-\ell}(\theta'_{-\ell}))) = [x/_{\bar{\Theta}}h^*](\bar{\theta}_\ell, \theta'_{-\ell}) = x$ since $(\bar{\theta}_\ell, \theta'_{-\ell}) \in \bar{\Theta}$.

Thus, as a result of $\ell's$ deviation, the outcome at any state either stays the same or becomes x. That is, for all $\theta \in \Theta$, the outcome under $(\tilde{\sigma}_\ell, \sigma^*_{-\ell})$ at θ equals either that under σ^* at θ or x. Therefore, for any $x \in X$, at least $n-1$ individuals (each $\ell \in N$ other than the odd-man-out $\bar{j}$ if Rule 2 holds at $\bar{\theta}$ or all individuals if Rule 3 holds at $\bar{\theta}$) can deviate so that the following holds: the outcome stays the same in any state θ with $\theta_\ell \neq \bar{\theta}_\ell$; for all states with $\theta_\ell = \bar{\theta}_\ell$, the outcome either stays the same or it changes to x.

Now, we employ weak choice incompatibility: Consider SCF $h^* = g \circ \sigma^* \in H$, $\bar{\theta} \in \Theta$, and the profile $(\tilde{\mathbf{A}}_i)_{i \in N}$ defined by $\tilde{\mathbf{A}}_i := \mathbf{O}_i^\mu(\sigma^*_{-i})$ for all $i \in N$. Then, trivially, $\mathbf{h}^*_{i,\bar{\theta}_i} \in \tilde{\mathbf{A}}_i$ for all $i \in N$, and hence, (i) of hypothesis of weak choice incompatibility holds. For (ii) of the hypothesis of weak choice incompatibility, consider each individual $\ell \neq \bar{j}$ and let

[16] Even if α is not invertible, one can always identify such a function: Let $\alpha^{-1}(\theta) \subset \Theta$ be the set of states that are mapped to θ under deception α. For any $\theta \in \Theta$, if $\alpha^{-1}(\theta) = \emptyset$, then set $\alpha_*^{-1}(\theta) = \theta$, else pick an arbitrary $\bar{\theta} \in \alpha^{-1}(\theta)$ and set $\alpha_*^{-1}(\theta') = \bar{\theta}$ whenever $\alpha(\theta') = \theta$ for some $\theta' \in \Theta$.

Θ' be the set of states such that the outcome changes to x after ℓ deviates unilaterally to $\tilde{\sigma}_\ell$. Observe that $\bar{\theta} \in \Theta'$ and for any $x \in X$, $[\mathbf{a}_\ell^x/_{\Theta'}\mathbf{h}^*_{\ell,\bar{\theta}_\ell}] \in \tilde{\mathbf{A}}_\ell$ for all $\ell \neq \bar{j}$ because, by definition, $[\mathbf{a}_\ell^x/_{\Theta'}\mathbf{h}^*_{\ell,\bar{\theta}_\ell}]$ is the IAA act individual ℓ of type $\bar{\theta}_\ell$ faces after deviating to $\tilde{\sigma}_\ell$. Hence, by weak choice incompatibility, there is $i^* \in N$ with $i^* \neq \bar{j}$ such that $\mathbf{h}^*_{i^*,\bar{\theta}_{i^*}} \notin \mathbf{C}_{i^*}^{\bar{\theta}_{i^*}}(\tilde{\mathbf{A}}_{i^*})$, which means $\mathbf{h}^*_{i^*,\bar{\theta}_{i^*}} \notin \mathbf{C}_{i^*}^{\bar{\theta}_{i^*}}(\mathbf{O}^\mu_{i^*}(\sigma^*_{-i^*}))$, which contradicts σ^* being a BIE of μ as $h^* = g \circ \sigma^*$.

This establishes that Rule 1 applies at every $\theta \in \Theta$ under any BIE of μ.

Due to the product structure of the state space, under any BIE σ^* of μ, there is a unique $f \in F$ such that $m_i^1(\theta_i) = f$ for all $i \in N$ and all $\theta_i \in \Theta_i$. Otherwise, if there were i, j with $i \neq j$, who propose $f, f' \in F$ with $f \neq f'$ for their types θ_i and θ_j under σ^*, respectively, then Rule 1 cannot apply at $(\theta_i, \theta_j, \theta_{-\{i,j\}})$, a contradiction.

Finally, we show that $f \circ \alpha \in F$: Since Rule 1 applies at each $\theta \in \Theta$, and each $i \in N$ reports the type $\alpha_i(\theta_i) \in \Theta_i$ as the second entry of their messages at $\theta \in \Theta$ under σ^*, $g \circ \sigma^* = h^* = f \circ \alpha$. As $\mathbb{S}$ is closed under deception, at each $\theta \in \Theta$, $\mathbf{O}_i^\mu(\sigma^*_{-i}) = \cup_{\mathbf{a}_i \in \mathbf{S}_i(f,\alpha^{\mathrm{id}}_{-i})}\{\mathbf{a}_i \circ \alpha_{-i}\} \cup \{\mathbf{a}^{z^i}\} = \mathbf{S}_i(f, \alpha^{\mathrm{id}}_{-i} \circ \alpha_{-i}) \cup \{\mathbf{a}^{z^i}\} = \mathbf{S}_i(f, \alpha_{-i}) \cup \{\mathbf{a}^{z^i}\}$ for all $i \in N$. If $f \circ \alpha \notin F$, then by (ii) of interim consistency, there is $i^* \in N$ and $\theta^*_{i^*} \in \Theta_{i^*}$ such that $\mathbf{f}^\alpha_{i^*,\theta^*_{i^*}} \notin \mathbf{C}_{i^*}^{\theta^*_{i^*}}(\mathbf{S}_{i^*}(f, \alpha_{-i^*}))$. Since $(z^i)_{i\in N}$ is a profile of null alternatives, $\mathbf{a}^{z^{i^*}}$ is a null IAA act of individual i^* and this implies $\mathbf{f}^\alpha_{i^*,\theta^*_{i^*}} \notin \mathbf{C}_{i^*}^{\theta^*_{i^*}}(\mathbf{S}_{i^*}(f, \alpha_{-i^*}) \cup \{\mathbf{a}^{z^{i^*}}\})$. Then, $\mathbf{f}^\alpha_{i^*,\theta^*_{i^*}} = \mathbf{h}^*_{i^*,\theta^*_{i^*}} \notin \mathbf{C}_{i^*}^{\theta^*_{i^*}}(\mathbf{O}^\mu_{i^*}(\sigma^*_{-i^*}))$, which contradicts σ^* being a BIE of μ. Thus, $g \circ \sigma^* = f \circ \alpha \in F$, which implies that condition (ii) of implementability in BIE also holds. ■

Now, we discuss the relation between the weak choice incompatibility condition and the economic environment assumption introduced by Jackson (1991).[17] To formalize Jackson's economic environment condition, we need to consider rational choices: For all $i \in N$ and all $\theta_i \in \Theta_i$, let $\mathbf{R}_i^{\theta_i}$ defined over the set of all IAA acts $\mathbf{A}_i$ be a complete and transitive preference relation such that for any non-empty $\mathbf{S} \subset \mathbf{A}_i$, $\mathbf{a}_i \in \mathbf{C}_i^{\theta_i}(\mathbf{S})$ if and only if $\mathbf{a}_i \mathbf{R}_i^{\theta_i} \mathbf{a}'_i$ for all $\mathbf{a}'_i \in \mathbf{S}$. Let $\mathbf{P}_i^{\theta_i}$ denote the strict counterpart of $\mathbf{R}_i^{\theta_i}$.

Definition 8 (Jackson (1991))**.** *A rational environment $\mathcal{E}$ is* **economic** *if for any SCF $h \in H$, and any $\bar{\theta} \in \Theta$, there are $i^*, j^* \in N$ with $i^* \neq j^*$ and $x^*, y^* \in X$ so that $[\mathbf{a}_{i^*}^{x^*}/_{\Theta'}\mathbf{h}_{i^*,\bar{\theta}_{i^*}}]\ \mathbf{P}_{i^*}^{\bar{\theta}^*_i}\ \mathbf{h}_{i^*,\bar{\theta}_{i^*}}$ and $[\mathbf{a}_{j^*}^{y^*}/_{\Theta'}\mathbf{h}_{j^*,\bar{\theta}_{j^*}}]\ \mathbf{P}_{j^*}^{\bar{\theta}^*_j}\ \mathbf{h}_{j^*,\bar{\theta}_{j^*}}$ for all $\Theta' \subset \Theta$ with $\bar{\theta} \in \Theta'$.*

We transform Jackson's economic environment assumption to our behavioral setting:

Definition 9. *Environment $\mathcal{E}$ is* **behaviorally economic** *if for any SCF $h \in H$, and any $\bar{\theta} \in \Theta$, there are $i^*, j^* \in N$ and $x^*, y^* \in X$ with $i^* \neq j^*$ such that*

[17] For other versions, see Bergemann and Morris (2008) and Barlo and Dalkıran (2022b, 2023).

$\mathbf{h}_{i^*,\bar{\theta}_{i^*}} \notin \mathbf{C}_{i^*}^{\bar{\theta}_{i^*}}(\tilde{\mathbf{A}}_{i^*})$ *whenever* $[\mathbf{a}_{i^*}^{x^*}/_{\Theta'}\mathbf{h}_{i^*,\bar{\theta}_{i^*}}] \in \tilde{\mathbf{A}}_{i^*}$ *for some* $\Theta' \subset \Theta$ *with* $\bar{\theta} \in \Theta'$, *and* $\mathbf{h}_{j^*,\bar{\theta}_{j^*}} \notin \mathbf{C}_{j^*}^{\bar{\theta}_{j^*}}(\tilde{\mathbf{A}}_{j^*})$ *whenever* $[\mathbf{a}_{j^*}^{y^*}/_{\Theta'}\mathbf{h}_{j^*,\bar{\theta}_{j^*}}] \in \tilde{\mathbf{A}}_{j^*}$ *for some* $\Theta' \subset \Theta$ *with* $\bar{\theta} \in \Theta'$.

It is clear that under rationality, if an environment is economic, then it is behaviorally economic. In what follows, we show that the latter implies weak choice incompatibility in both rational and behavioral domains.

Proposition 2. *If environment $\mathcal{E}$ is behaviorally economic, then the weak choice incompatibility condition holds in $\mathcal{E}$.*

Proof. As in the hypothesis of the weak choice incompatibility condition, let $h \in H$, $\bar{\theta} \in \Theta$, and $(\tilde{\mathbf{A}}_i)_{i\in N}$ be a profile of sets of IAA acts such that (i) for all $i \in N$, $\mathbf{h}_{i,\bar{\theta}_i} \in \tilde{\mathbf{A}}_i$, and (ii) there is $\bar{j} \in N$ such that for all $i \in N \setminus \{\bar{j}\}$, for any $x \in X$, $[\mathbf{a}_i^x/_{\Theta'}\mathbf{h}_{i,\bar{\theta}_i}] \in \tilde{\mathbf{A}}_i$ for some $\Theta' \subset \Theta$ with $\bar{\theta} \in \Theta'$. $\mathcal{E}$ being behaviorally economic implies that there are $i^*, j^* \in N$ with $i^* \neq j^*$ and $x^*, y^* \in X$ be such that $\mathbf{h}_{i^*,\bar{\theta}_{i^*}} \notin \mathbf{C}_{i^*}^{\bar{\theta}_{i^*}}(\hat{\mathbf{A}}_{i^*})$ whenever $[\mathbf{a}_{i^*}^{x^*}/_{\Theta'}\mathbf{h}_{i^*,\bar{\theta}_{i^*}}] \in \hat{\mathbf{A}}_{i^*}$ and $\mathbf{h}_{j^*,\bar{\theta}_{j^*}} \notin \mathbf{C}_{j^*}^{\bar{\theta}_{j^*}}(\hat{\mathbf{A}}_{j^*})$ whenever $[\mathbf{a}_{j^*}^{y^*}/_{\Theta'}\mathbf{h}_{j^*,\bar{\theta}_{j^*}}] \in \hat{\mathbf{A}}_{j^*}$. Thus, either $i^* \neq \bar{j}$ or $j^* \neq \bar{j}$. If $i^* \neq \bar{j}$, then $[\mathbf{a}_{i^*}^{x^*}/_{\Theta'}\mathbf{h}_{i^*,\bar{\theta}_{i^*}}] \in \tilde{\mathbf{A}}_{i^*}$ and hence $\mathbf{h}_{i^*,\bar{\theta}_{i^*}} \notin \mathbf{C}_{i^*}^{\bar{\theta}_{i^*}}(\tilde{\mathbf{A}}_{i^*})$ because $\mathcal{E}$ is behaviorally economic. If $j^* \neq \bar{j}$, then $[\mathbf{a}_{j^*}^{y^*}/_{\Theta'}\mathbf{h}_{j^*,\bar{\theta}_{j^*}}] \in \tilde{\mathbf{A}}_{j^*}$ and so $\mathbf{h}_{j^*,\bar{\theta}_{j^*}} \notin \mathbf{C}_{j^*}^{\bar{\theta}_{j^*}}(\tilde{\mathbf{A}}_{j^*})$ as $\mathcal{E}$ is behaviorally economic. In either case, weak choice incompatibility holds in $\mathcal{E}$. ■

Finally, we compare our second sufficiency result with Theorem 1 of Jackson (1991): These results share the assumption of finite state spaces. However, Jackson's does not need a null alternative: In the rational domain, Jackson eliminates the undesirable equilibria by making use of rationality on top of Bayesian monotonicity and the economic environment assumption. On the other hand, in ours, we employ null alternatives on top of interim consistency and the weak choice incompatibility condition to get rid of undesirable equilibria.

5 Behavioral Interim Efficiency

"[T]he proper object for welfare analysis with incomplete information is the [SCF], rather than the actual decision or allocation ultimately chosen. Furthermore, any efficiency criterion for evaluating [SCFs] must be defined independently of [individuals' private information]" (Holmström & Myerson, 1983). In what follows, we introduce the behavioral counterpart of interim incentive Pareto efficiency of Holmström and Myerson (1983). Our construction is in line with de Clippel (2014), introducing the following efficiency notion in behavioral domains of complete information: An alternative x is efficient at state θ if each individual has an implicit opportunity set from which she chooses x at θ, and each alternative is in at least one of these implicit opportunity sets.

Extending the behavioral efficiency notion of de Clippel (2014) to incomplete information environments, we define the notion of *behavioral interim efficiency* as follows:

Definition 10. *An SCF $f : \Theta \to X$ is* ***behaviorally interim efficient*** *if there is a profile of sets of IAA acts $(\mathbf{Y}_{i,\theta_i})_{i \in N,\ \theta_i \in \Theta_i}$ such that*

(i) for all $i \in N$ and all $\theta_i \in \Theta_i$, $\mathbf{f}_{i,\theta_i} \in \mathbf{C}_i^{\theta_i}(\mathbf{Y}_{i,\theta_i})$, and

(ii) for all $h \in H$, there is $i \in N$ and $\theta_i \in \Theta_i$ with $\mathbf{h}_{i,\theta_i} \in \mathbf{Y}_{i,\theta_i}$.

We refer to the set of all behaviorally efficient SCFs as the ***behavioral interim efficient SCS*** *and denote it as $IE \in \mathscr{F}$.*

In words, an SCF f is behaviorally interim efficient if there is a profile of implicit opportunity sets of IAA acts, one for each type of each individual, such that type θ_i of individual i chooses the IAA act induced by f from her implicit opportunity set and all possible SCFs are accounted for (meaning that for any SCF, there are an individual and her type such that the corresponding IAA act induced by this SCF belongs to the associated implicit opportunity set).

Behavioral interim efficiency is an extension of interim Pareto efficiency to behavioral domains. In the rational domain, an SCF is interim Pareto efficient if there is no other SCF that makes every type of every individual strictly better off.[18] In particular, individuals' choices are rational when for all $i \in N$ and all $\theta_i \in \Theta_i$, there is a complete and transitive preference relation $\mathbf{R}_{i,\theta_i} \subset \mathbf{A}_i \times \mathbf{A}_i$ such that for any non-empty $\mathbf{S} \subset \mathbf{A}_i$, $\mathbf{a}_i \in \mathbf{C}_i^{\theta_i}(\mathbf{S})$ if and only if $\mathbf{a}_i \mathbf{R}_{i,\theta_i} \mathbf{a}_i'$ for all $\mathbf{a}_i' \in \mathbf{S}$. Then, an SCF f is **interim Pareto efficient** (in the rational domain) if there is no $h \in H$ such that $\mathbf{h}_{i,\theta_i} \mathbf{P}_{i,\theta_i} \mathbf{f}_{i,\theta_i}$ for all $i \in N$ and all $\theta_i \in \Theta_i$.

Behavioral interim efficiency coincides with interim Pareto efficiency under rationality: To see that interim Pareto efficiency implies behavioral interim efficiency, let f be interim Pareto efficient and set $\mathbf{Y}_{i,\theta_i} = LCS_{i,\theta_i}(\mathbf{f}_{i,\theta_i}) := \{\mathbf{h}_{i,\theta_i} \in \mathbf{A}_i \mid \mathbf{f}_{i,\theta_i} \mathbf{R}_{i,\theta_i} \mathbf{h}_{i,\theta_i}\}$ for all $i \in N$ and all $\theta_i \in \Theta_i$. Then, for all $i \in N$ and all $\theta_i \in \Theta_i$, $\mathbf{f}_{i,\theta_i} \in \mathbf{C}_i^{\theta_i}(\mathbf{Y}_{i,\theta_i})$, delivering (i) of behavioral interim efficiency of f. For (ii) of behavioral interim efficiency of f, notice that if there were some $h \in H$ such that for all $i \in N$ and all $\theta_i \in \Theta_i$, $\mathbf{h}_{i,\theta_i} \notin \mathbf{Y}_{i,\theta_i}$, then, by construction, $\mathbf{h}_{i,\theta_i} \mathbf{P}_{i,\theta_i} \mathbf{f}_{i,\theta_i}$ for all $i \in N$ and all $\theta_i \in \Theta_i$; this leads to a contradiction of f being interim Pareto efficient. For the converse, let f be behaviorally interim efficient but not interim Pareto efficient, i.e., there is $h \in H$ such that $\mathbf{h}_{i,\theta_i} \mathbf{P}_{i,\theta_i} \mathbf{f}_{i,\theta_i}$ for all $i \in N$ and all $\theta_i \in \Theta_i$. Then, by (ii) of behavioral interim efficiency of f, $\mathbf{h}_{j,\theta_j} \in \mathbf{Y}_{j,\theta_j}$ for some $j \in N$ and $\theta_j \in \Theta_j$. But this implies that $\mathbf{f}_{j,\theta_j} \notin \mathbf{C}_j^{\theta_j}(\mathbf{Y}_{j,\theta_j})$, a contradiction to (i) of behavioral interim efficiency of f.

[18] This notion is the weak version of interim Pareto efficiency in Holmström and Myerson (1983).

Whether or not the behaviorally interim efficient SCS is non-empty is a relevant question. Below, we show that the answer is affirmative if individuals' choices are non-empty valued. This is because any SCF inducing an IAA act that is chosen by a type of an individual from the set of all IAA acts of that individual is behaviorally interim efficient:

Proposition 3. *Suppose individuals' choices over IAA acts are non-empty valued. Then, the behavioral interim efficient SCS, IE, is non-empty.*

Proof. Suppose individuals' choices over IAA acts are non-empty valued. Let us fix individual j of type $\bar{\theta}_j$ and an alternative $\bar{x} \in X$. Consider an IAA act $\mathbf{a}_j \in \mathbf{A}_j$ such that $\mathbf{a}_j \in \mathbf{C}_j^{\bar{\theta}_j}(\mathbf{A}_j)$. Let SCF f be defined as follows: for all $\theta \in \Theta$, $f(\theta) = \mathbf{a}_j(\theta_{-j})$ if $\theta_j = \bar{\theta}_j$ and $f(\theta) = \bar{x}$ if $\theta_j \neq \bar{\theta}_j$. Then, the IAA act induced by f for type $\bar{\theta}_j$ of individual j, $\mathbf{f}_{j,\bar{\theta}_j} = \mathbf{a}_j$. Let the implicit opportunity set profile $(\mathbf{Y}_{i,\theta_i})_{i\in N,\theta_i\in\Theta_i}$ be such that $\mathbf{Y}_{i,\theta_i} = \{\mathbf{f}_{i,\theta_i}\}$ for all $i \neq j$, and $\mathbf{Y}_{j,\theta_j}$ equals $\mathbf{A}_j$ if $\theta_j = \bar{\theta}_j$ and $\{\mathbf{f}_{j,\theta_j}\}$ otherwise. Then, (i) of behavioral interim efficiency holds as $\mathbf{f}_{j,\bar{\theta}_j} \in \mathbf{C}_j^{\bar{\theta}_j}(\mathbf{A}_j)$ and all other implicit opportunity sets are singletons. Furthermore, for all $h \in H$, we have $\mathbf{h}_{j,\bar{\theta}_j} \in \mathbf{Y}_{j,\bar{\theta}_j} = \mathbf{A}_j$, implying (ii) of behavioral interim efficiency. Therefore, $f \in IE$. ■

The notion of behavioral interim efficiency above depends on the private information of the individuals. Holmström and Myerson (1983) argues that any efficiency criterion for evaluating SCFs under incomplete information must be defined independently of individuals' private information. By restricting feasibility based on incentive compatibility, they come up with the following notion of *interim incentive Pareto efficiency* in the rational domain: An incentive compatible SCF is interim incentive efficient if no "individual would surely prefer another [incentive compatible SCF] over [the given SCF] when he knows his own type, whatever his type might be" (Holmström & Myerson, 1983).

In our behavioral setting, the relevant restriction takes the form of quasi-incentive compatibility. Let us denote the set of all quasi-incentive compatible SCFs by $H^* \subset H$. Consequently, we define *behavioral interim incentive efficiency* as follows:

Definition 11. *An SCF $f : \Theta \to X$ is* ***behaviorally interim incentive efficient*** *if there is a profile of sets of IAA acts $(\mathbf{Y}_i)_{i\in N}$ such that*

(i) for all $i \in N$ and all $\theta_i \in \Theta_i$, $\mathbf{f}_{i,\theta_i} \in \mathbf{C}_i^{\theta_i}(\mathbf{Y}_i)$, and

(ii) for all $h \in H^*$, there is $i \in N$ and $\theta_i \in \Theta_i$ with $\mathbf{h}_{i,\theta_i} \in \mathbf{Y}_i$.

We refer to the set of all behaviorally incentive efficient SCFs as the ***behavioral interim incentive efficient SCS*** *and denote it as $IIE \in \mathscr{F}$.*

This welfare criterion internalizes quasi-incentive compatibility into behavioral interim efficiency: An SCF is behaviorally interim incentive efficient if, for any individual, there exists an implicit opportunity set of IAA acts such that her choices from this set for each of her types are aligned with the SCF with the additional property that for every possible quasi-incentive compatible SCF, there is an individual and her type for which the IAA act associated with this SCF is in her implicit opportunity set. The key difference between behavioral interim incentive efficiency and behavioral interim efficiency is that the implicit opportunity sets in the former do not depend on individuals' private information.

Behavioral interim incentive efficiency extends interim incentive Pareto efficiency to behavioral domains as quasi-incentive compatibility boils down to incentive compatibility under rationality. However, quasi-incentive compatibility embedded in the definition of behavioral interim incentive efficiency brings about a more stringent requirement for the existence of the IIE SCS. Below, we provide a sufficient condition for its existence.

Proposition 4. *Suppose individuals' choices over IAA acts are non-empty valued, and there exist $j \in N$ and $h^* \in H^*$ with $\mathbf{h}^*_{j,\theta_j} \in \mathbf{C}_j^{\theta_j}(\tilde{\mathbf{A}}_j)$ for all $\theta_j \in \Theta_j$ for some $\tilde{\mathbf{A}}_j \subset \mathbf{A}_j$ with $\mathbf{A}^*_j := \{\mathbf{h}_{j,\theta_j} \in \mathbf{A}_j \mid h \in H^* \text{ and } \theta_j \in \Theta_j\} \subset \tilde{\mathbf{A}}_j$. Then, the IIE SCS is non-empty.*

Proof. Suppose the hypothesis holds. Let the profile of sets of IAA acts associated with SCF h^* be $(\mathbf{T}_i)_{i\in N}$ as in Definition 4 so that $\mathbf{h}^*_{i,\theta_i} \in \mathbf{C}_i^{\theta_i}(\mathbf{T}_i)$ for all $i \in N$ and all $\theta_i \in \Theta_i$. Recall that $\{\mathbf{h}^*_{i,\tilde{\theta}_i} \mid \tilde{\theta}_i \in \Theta_i\} \subset \mathbf{T}_i$ for all $i \in N$. Let the implicit opportunity set profile $(\mathbf{Y}_i)_{i\in N}$ be such that $\mathbf{Y}_j = \tilde{\mathbf{A}}_j$ and $\mathbf{Y}_i = \mathbf{T}_i$ for all $i \neq j$. Then, (i) of behavioral interim incentive efficiency holds. Also, for all $h \in H^*$ and for all $\theta_j \in \Theta_j$, $\mathbf{h}_{j,\theta_j} \in \mathbf{A}^*_j \subset \tilde{\mathbf{A}}_j = \mathbf{Y}_j$, implying (ii) of behavioral interim incentive efficiency. Thus, $f \in IIE$. ■

On the other hand, when analyzing the implementability of behavioral efficiency, we note that there is no hope for BIE implementability of the IE SCS without its quasi-incentive compatibility thanks to Theorem 1 and Proposition 1. Meanwhile, we observe that the implicit opportunity sets in the definition of the IIE SCS are not necessarily aligned with a profile of IAA acts closed under deception. This leads to the failure of implementability of the IIE SCS in BIE, as we display below:

Example. Let $N = \{1, 2, 3\}$, $X = \{x, y\}$, $\Theta_1 = \{t_1^1, t_1^2\}$, $\Theta_2 = \{t_2^1, t_2^2\}$, and $\Theta_3 = \{t_3\}$. Then, the set of all possible SCFs is given by $H = \{\langle abcd \rangle \mid a, b, c, d \in \{x, y\}\}$ where $\langle abcd \rangle$ denotes the SCF h such that $h(t_1^1, t_2^1, t_3) = a$, $h(t_1^2, t_2^1, t_3) = b$, $h(t_1^1, t_2^2, t_3) = c$, and $h(t_1^2, t_2^2, t_3) = d$. Because there are two types of individuals 1 and 2, the set of all IAA acts of individual 3 is given by $\mathbf{A}_3 = H$. Since there is only one type of individual 3, the set of all IAA acts each type of individual 1 and 2 faces is $\mathbf{A}_1 = \mathbf{A}_2 = \{\langle xx \rangle, \langle xy \rangle, \langle yx \rangle, \langle yy \rangle\}$ where $\mathbf{a}_i = \langle ab \rangle$ represents the IAA act such that $\mathbf{a}_i(t_j^1, t_3) = a$ and $\mathbf{a}_i(t_j^2, t_3) = b$ where $a, b \in \{x, y\}$, $i, j \in \{1, 2\}$, and $i \neq j$.

In what follows, we require that the choices are non-empty valued and satisfy the following: The choices of each type of individuals 1 and 2 over sets of IAA acts are as in

Table 3; individual 3's choices over the IAA acts are such that $\mathbf{C}_3^{t_3}(\{\langle xxxx\rangle, \langle yyyy\rangle\}) = \{\langle yyyy\rangle\}$ and for any set of IAA acts $\mathbf{S} \subset \mathbf{A}_3$ with $\langle xyyx\rangle, \langle yyyy\rangle \in \mathbf{S}$, $\langle yyyy\rangle \notin \mathbf{C}_3^{t_3}(\mathbf{S})$.

	$\mathbf{C}_1^{t_1^1}$	$\mathbf{C}_1^{t_1^2}$	$\mathbf{C}_2^{t_2^1}$	$\mathbf{C}_2^{t_2^2}$
$\{\langle xx\rangle, \langle xy\rangle, \langle yx\rangle, \langle yy\rangle\}$	$\{\langle xy\rangle\}$	$\{\langle yx\rangle\}$	$\{\langle xy\rangle\}$	$\{\langle yx\rangle\}$
$\{\langle xx\rangle, \langle xy\rangle, \langle yy\rangle\}$	$\{\langle xy\rangle\}$	$\{\langle \cdot\rangle\}$	$\{\langle xy\rangle\}$	$\{\langle \cdot\rangle\}$
$\{\langle xx\rangle, \langle yx\rangle, \langle yy\rangle\}$	$\{\langle \cdot\rangle\}$	$\{\langle yx\rangle\}$	$\{\langle \cdot\rangle\}$	$\{\langle yx\rangle\}$
$\{\langle xy\rangle, \langle yx\rangle, \langle yy\rangle\}$	$\{\langle xy\rangle\}$	$\{\langle yx\rangle\}$	$\{\langle xy\rangle\}$	$\{\langle yx\rangle\}$
$\{\langle xx\rangle, \langle yy\rangle\}$	$\{\langle yy\rangle\}$	$\{\langle yy\rangle\}$	$\{\langle yy\rangle\}$	$\{\langle yy\rangle\}$
$\{\langle xy\rangle, \langle yy\rangle\}$	$\{\langle xy\rangle\}$	$\{\langle \cdot\rangle\}$	$\{\langle xy\rangle\}$	$\{\langle \cdot\rangle\}$
$\{\langle yx\rangle, \langle yy\rangle\}$	$\{\langle \cdot\rangle\}$	$\{\langle yx\rangle\}$	$\{\langle \cdot\rangle\}$	$\{\langle yx\rangle\}$

Table 3: The choices of each type of individuals 1 and 2.

Given the choices described above, SCF $f = \langle xyyx\rangle$ is a behaviorally interim incentive efficient SCF. Letting $\mathbf{Y}_1 = \mathbf{Y}_2 = \{\langle xx\rangle, \langle xy\rangle, \langle yx\rangle, \langle yy\rangle\}$ and $\mathbf{Y}_3 = \{\langle xyyx\rangle\}$, we observe that (i) and (ii) of behavioral interim incentive efficiency hold for $f = \langle xyyx\rangle$.[19]

Consider deception profile $\tilde{\alpha} \in \Lambda$ such that $\tilde{\alpha}_1(t_1^1) = \tilde{\alpha}_1(t_1^2) = t_1^1$ and $\tilde{\alpha}_2(t_2^1) = \tilde{\alpha}_1(t_2^2) = t_2^2$. Then, $f \circ \tilde{\alpha} = \langle yyyy\rangle$. Hence, $\mathbf{f}^{\tilde{\alpha}}_{i,t_i^j} = \langle yy\rangle$ for both $i, j = 1, 2$ and $\mathbf{f}^{\tilde{\alpha}}_{3,t_3} = \langle yyyy\rangle$. Any mechanism that implements the IIE SCS must possess a BIE that corresponds to SCF $f = \langle xyyx\rangle$. Therefore, it follows from our necessity result (Theorem 1) that for deception $\tilde{\alpha}$, we have $\mathbf{S}_i(f, \tilde{\alpha}_{-i}) = \{\langle xx\rangle, \langle yy\rangle\}$ for both $i = 1, 2$ and either $\mathbf{S}_3(f, \tilde{\alpha}_{-3}) = \{\langle xxxx\rangle, \langle yyyy\rangle\}$ or $\mathbf{S}_3(f, \tilde{\alpha}_{-3}) = \{\langle yyyy\rangle\}$.[20]

Furthermore, it follows from (i) and (ii) of behavioral interim incentive efficiency that for $\langle yyyy\rangle$ to be a behaviorally interim incentive efficient SCF, the IAA act it induces must be chosen by a type of an individual from a set of IAA acts including the IAA act induced by $\langle xyyx\rangle$ for the same type of the same individual. However, this is not the case because when $\langle xy\rangle$ and $\langle yy\rangle$ are in the same set of IAA acts type t_i^1 of individual $i = 1, 2$ does not choose $\langle yy\rangle$ from this set; when $\langle yx\rangle$ and $\langle yy\rangle$ are in the same set of IAA acts type t_i^2 of individual $i = 1, 2$ does not choose $\langle yy\rangle$ from this set; and for any set of IAA acts $\mathbf{S} \subset \mathbf{A}_3$ with $\langle xyyx\rangle, \langle yyyy\rangle \in \mathbf{S}$, we have $\langle yyyy\rangle \notin \mathbf{C}_3^{t_3}(\mathbf{S})$. Therefore, $\langle yyyy\rangle$ is not behaviorally interim incentive efficient.

Finally, from the choices described above, we observe that $\mathbf{C}_i^{t_i^j}(\{\langle xx\rangle, \langle yy\rangle\}) = \{\langle yy\rangle\}$

[19]We note that $\mathbf{h}_{i,\theta_i} \in \mathbf{Y}_i = \{\langle xx\rangle, \langle xy\rangle, \langle yx\rangle, \langle yy\rangle\}$ for any $h \in H$ for both $i = 1, 2$. Indeed, letting $\mathbf{Y}_{1,t_1^1} = \mathbf{Y}_{1,t_1^2} = \mathbf{Y}_{2,t_2^1} = \mathbf{Y}_{2,t_2^2} = \{\langle xx\rangle, \langle xy\rangle, \langle yx\rangle, \langle yy\rangle\}$ and $\mathbf{Y}_{3,t_3} = \{\langle xyyx\rangle\}$, we also observe that $f = \langle xyyx\rangle$ is behaviorally interim efficient.

[20]For SCF $f = \langle xyyx\rangle$ to arise as a BIE outcome in any mechanism, the corresponding opportunity set of individuals 1 and 2 must include the IAA acts $\langle xy\rangle$ and $\langle yx\rangle$. Therefore, it must be that $\langle xy\rangle, \langle yx\rangle \in \mathbf{S}_i(f, \alpha^{\mathrm{id}}_{-i})$ for any interim consistent profile of the IIE SCS. As $\langle xy\rangle^{\tilde{\alpha}} = \langle xx\rangle$ and $\langle yx\rangle^{\tilde{\alpha}} = \langle yy\rangle$. It follows from any interim consistent profile being closed under deception that $\mathbf{S}_i(f, \tilde{\alpha}_{-i}) = \{\langle xx\rangle, \langle yy\rangle\}$ for both $i = 1, 2$. Similarly, we have $\langle xyyx\rangle \in \mathbf{S}_3(f, \alpha^{\mathrm{id}}_{-3})$ for any interim consistent profile of the IIE SCS. This implies that $\langle xyyx\rangle^{\tilde{\alpha}} = \langle yyyy\rangle \in \mathbf{S}_3(f, \tilde{\alpha}_{-3})$. As individuals 1 and 2 always report the same types under deception $\tilde{\alpha}$, $\mathbf{S}_3(f, \tilde{\alpha}_{-3})$ cannot include any other IAA act apart from $\langle xxxx\rangle$ and $\langle yyyy\rangle$.

for all $i, j = 1, 2$ and $\mathbf{C}_3^{t_3}(\{\langle xxxx\rangle, \langle yyyy\rangle\}) = \{\langle yyyy\rangle\}$. But, this means for deception profile $\tilde{\alpha}$, we have $f \circ \tilde{\alpha} \notin IIE$ and for all $i \in N$ and $\theta_i \in \Theta_i$, $\mathbf{f}_{i,\theta_i}^{\tilde{\alpha}} \in \mathbf{C}_i^{\theta_i}(\mathbf{S}_i(f, \tilde{\alpha}_{-i}))$, which implies (ii) of interim consistency cannot hold. Thus, IIE SCS does not have an interim consistent profile; hence IIE SCS cannot be implemented in BIE in this example.[21]

The failure of BIE implementability of the IIE SCS is due to the existence of a garbling appearing as a bad equilibrium in any mechanism that has the potential to implement the IIE SCS in BIE. Indeed, the key is that the implicit opportunity sets in the definition of behavioral interim incentive efficiency do not necessarily provide us with a profile of IAA acts closed under deception. This leads us to Proposition 5, which we present after providing the following necessary formalities: For any SCF $f \in IIE$, let us denote the profile of associated implicit opportunity sets of IAA acts (as in Definition 11) by $(\mathbf{Y}_i^f)_{i\in N}$. For any $f \in IIE$, any deception profile α, and any individual i, we refer to the set of IAA acts obtained by garbling those in $\mathbf{Y}_i^f$ as $\mathbf{Y}_i^{f\circ\alpha}$.

Proposition 5. *Suppose that environment $\mathcal{E}$ is such that $n \geq 3$, the choice incompatibility condition holds, and the IIE SCS is non-empty. Then, the IIE SCS is implementable in BIE if $f \in IIE$ and $f \circ \alpha \notin IIE$ implies that there are $i \in N$ and $\theta_i \in \Theta_i$ such that $\mathbf{f}_{i,\theta_i}^{\alpha} \notin \mathbf{C}_i^{\theta_i}(\mathbf{Y}_i^{f\circ\alpha})$.*

Proof. Let $f \in IIE$, $\alpha \in \Lambda$, and define $\mathbb{S} := (\mathbf{S}_i(f, \alpha_{-i}))_{i\in N,\ f\in F,\ \alpha_{-i}\in\Lambda_{-i}}$ by $\mathbf{S}_i(f, \alpha_{-i}) := \mathbf{Y}_i^{f\circ\alpha}$ for all $i \in N$. $\mathbb{S}$ is closed under deception because $\mathbf{a}_i \in \mathbf{S}_i(f, \alpha_{-i}) = \mathbf{Y}_i^{f\circ\alpha}$ implies that for any $\tilde{\alpha} \in \Lambda$, $\mathbf{a}_i^{\tilde{\alpha}} \in \mathbf{Y}_i^{f\circ\alpha\circ\tilde{\alpha}} = \mathbf{S}_i(f, \alpha_{-i} \circ \tilde{\alpha}_{-i})$ since $\alpha \circ \tilde{\alpha} \in \Lambda$. Further, as $\mathbf{S}_i(f, \alpha_{-i}^{\mathrm{id}}) = \mathbf{Y}_i^f$ for all $i \in N$, (i) of behavioral interim incentive efficiency implies (i) of interim consistency. Finally, (ii) of interim consistency follows directly from the hypothesis of the proposition as $\mathbf{Y}_i^{f\circ\alpha} = \mathbf{S}_i(f, \alpha_{-i})$. Hence, $\mathbb{S}$ is interim consistent with the IIE SCS. The result follows immediately from Theorem 2. ■

Proposition 5 delivers BIE implementability of the IIE SCS by embedding the standard monotonicity condition into the profile of implicit opportunity sets (sustaining behavioral interim incentive efficiency) by requiring this profile to be closed under deception. This parallels considering incentive compatibility in conjunction with interim efficiency in rational domains when analyzing its implementability under incomplete information.[22]

[21] We wish to emphasize that the failure of the IIE SCS in this example is not due to the failure of the (weak) choice incompatibility condition but rather due to a garbling appearing as a bad BIE in any mechanism that has the potential to implement IIE SCS in BIE. Indeed, by adding more individuals and alternatives, one can extend this example to a choice structure (with a considerable tedious effort) where (weak) choice incompatibility holds while IIE SCS is not implementable in BIE.

[22] To the best of our knowledge, there are only a couple of papers that study Bayesian implementability of interim incentive Pareto efficiency, both with private values and independent types: Korpela and Lombardi (2020) presents a general property, called closure under interim utility equivalence, and shows

In our analysis, closedness under deception emerges as a necessary condition for BIE implementation; hence, it is natural to incorporate this condition into behavioral interim incentive efficiency to obtain implementability in BIE. Finally, we note that, thanks to Theorem 3, we can replace choice incompatibility in Proposition 5 with weak choice incompatibility when the state space is finite and every individual has a null alternative.

6 An Example with Minimax-Regret Preferences

The following example involves minimax-regret preferences of Savage (1951) and is modified from Example 4.8 of Saran (2011). In our example, the IIA and hence WARP does not hold for interim choices. Nonetheless, we attain full implementability in BIE of an SCS that constitutes a selection from the behavioral interim incentive efficient SCS, IIE. Meanwhile, we also attain partial implementation of an SCF in BIE even though partial direct implementation of that same SCF in truthful BIE is not possible. In other words, the revelation principle fails, which echoes Saran's findings about the failure of the revelation principle in behavioral domains.

Under minimax-regret preferences, each type of every individual chooses the IAA act that minimizes her maximum regret. The regret of choosing an IAA act at a state is given by the difference between the payoff obtained and the maximum possible payoff in that state. Then, the maximum regret of individual i of type θ_i as a result of an IAA act is the highest regret i suffers because of this IAA act across all states in which he is of type θ_i. Formally, for any two IAA acts $\mathbf{a}_i$ and $\tilde{\mathbf{a}}_i$ in a given set of IAA acts $\mathbf{S}_i \subset \mathbf{A}_i$, individual i of type θ_i weakly prefers $\mathbf{a}_i$ to $\tilde{\mathbf{a}}_i$ in the minimax-regret setting if

$$\max_{\theta_{-i}\in\Theta_{-i}} \left[\max_{\mathbf{a}'_i\in\mathbf{S}_i} \left(u_i(\mathbf{a}'_i(\theta_{-i}) \mid (\theta_i,\theta_{-i})) - u_i(\mathbf{a}_i(\theta_{-i}) \mid (\theta_i,\theta_{-i}))\right)\right]$$
$$\leq \max_{\theta_{-i}\in\Theta_{-i}} \left[\max_{\mathbf{a}''_i\in\mathbf{S}_i} \left(u_i(\mathbf{a}''_i(\theta_{-i}) \mid (\theta_i,\theta_{-i})) - u_i(\tilde{\mathbf{a}}_i(\theta_{-i}) \mid (\theta_i,\theta_{-i}))\right)\right],$$

where for any $x \in X$ and state $\theta = (\theta_i, \theta_{-i})$, $u_i(x \mid \theta)$ denotes i's payoffs from x at θ.

We let $N = \{1, 2\}$, $\Theta_1 = \Theta_2 = \{\alpha, \beta\}$, $X = \{x, y, z\}$, and the state-contingent utilities are as given in Table 4. An SCF $h : \Theta \to X$ is denoted by $h = \langle abcd\rangle$ where $h(\alpha, \alpha) = a$, $h(\alpha, \beta) = b$, $h(\beta, \alpha) = c$, and $h(\beta, \beta) = d$ with $a, b, c, d \in \{x, y, z\}$. Similarly, an IAA act that i faces, $\mathbf{a}_i : \Theta_{-i} \to X$ is denoted by $\mathbf{a}_i = \langle ab\rangle$ where $\mathbf{a}_i(\alpha) = a$ and $\mathbf{a}_i(\beta) = b$ with $a, b \in \{x, y, z\}$.

The SCS we consider in this example for full implementation is $F = \{\langle xyyy\rangle, \langle yyyy\rangle\}$.

that it is sufficient for Bayesian implementation of any interim incentive efficient SCF when there are two individuals. Meanwhile, Pram (2020) shows that interim incentive Pareto efficiency is fully Bayesian implementable when there are at least four individuals, and a no-total-indifference condition holds.

Θ	(α, α)	(α, β)	(β, α)	(β, β)
$(u_1(x \mid \theta), u_2(x \mid \theta))$	$(0, 1)$	$(0, 1/2)$	$(1/2, 1)$	$(1/2, 1/2)$
$(u_1(y \mid \theta), u_2(y \mid \theta))$	$(1, 0)$	$(1, 1)$	$(1, 0)$	$(1, 1)$
$(u_1(z \mid \theta), u_2(z \mid \theta))$	$(2, 0)$	$(2, 0)$	$(0, 0)$	$(0, 0)$

Table 4: State-contingent utilities of our example.

We note that $F \subset IIE$, i.e., both of these SCFs are behaviorally interim incentive efficient. It is tedious but straightforward to show that the following profile of sets of IAA acts sustains both SCFs in F in behavioral interim incentive efficiency as formalized in Definition 11: $\mathbf{Y}_1 = \{\langle xx \rangle, \langle xy \rangle, \langle xz \rangle, \langle yy \rangle\}$ and $\mathbf{Y}_2 = \{\langle xy \rangle, \langle yx \rangle, \langle yy \rangle, \langle yz \rangle, \langle zx \rangle, \langle zy \rangle, \langle zz \rangle\}$.

The mechanism given in Table 5 implements SCS F in BIE:

		Individual 2	
		L	R
Individual 1	U	x	y
	M	y	y
	D	x	z

Table 5: The mechanism fully implementing SCS F in BIE.

The set of BIE equals $\{\sigma^{(1)}, \ldots, \sigma^{(4)}\}$ where $\sigma_1^{(1)}(\alpha) = U$, $\sigma_1^{(1)}(\beta) = M$, $\sigma_2^{(1)}(\alpha) = L$, and $\sigma_2^{(1)}(\beta) = R$ inducing SCF $\langle xyyy \rangle$; $\sigma_1^{(2)}(\alpha) = M$, $\sigma_1^{(2)}(\beta) = M$, $\sigma_2^{(2)}(\alpha) = L$, and $\sigma_2^{(2)}(\beta) = L$; $\sigma_1^{(3)}(\alpha) = M$, $\sigma_1^{(3)}(\beta) = M$, $\sigma_2^{(3)}(\alpha) = L$, and $\sigma_2^{(3)}(\beta) = R$; $\sigma_1^{(4)}(\alpha) = M$, $\sigma_1^{(4)}(\beta) = M$, $\sigma_2^{(4)}(\alpha) = R$, and $\sigma_2^{(4)}(\beta) = L$ with $\sigma^{(2)}$, $\sigma^{(3)}$, and $\sigma^{(4)}$ all inducing $\langle yyyy \rangle$.

The SCF $\langle xyyy \rangle$ is partially implementable in BIE by the mechanism defined in Table 5 as full implementation in BIE of SCS $F = \{\langle xyyy \rangle, \langle yyyy \rangle\}$ implies partial implementation in BIE of SCF $\langle xyyy \rangle$. However, $\langle xyyy \rangle$ is not a BIE outcome in the corresponding direct mechanism given in Table 6. This follows from type α of individual 1 choosing the IAA

		Individual 2	
		α	β
Individual 1	α	x	y
	β	y	y

Table 6: The direct mechanism μ^d for Saran's example augmented.

act $\langle xy \rangle$ from the set of IAA acts $\{\langle xy \rangle, \langle xz \rangle, \langle yy \rangle\}$, but not from $\{\langle xy \rangle, \langle yy \rangle\}$. Besides displaying the failure of the revelation principle, this observation also exhibits that the IIA, and hence WARP, does not hold in this example.

7 Concluding Remarks

We investigate the problem of full implementation under incomplete information when individuals' interim choices need not satisfy the standard axioms of rationality.

We provide necessary as well as sufficient conditions for the implementation of SCSs in BIE. These help us analyze the implementability of behavioral interim incentive Pareto efficiency in BIE.

An interesting direction for future research is to analyze whether practical and simple mechanisms are available for specific types of behavioral biases under incomplete information. We hope that our results pave the way for contributions in this direction.

References

Altun, O. A., Barlo, M., & Dalkıran, N. A. (2023). Implementation with a sympathizer. *Mathematical Social Sciences*, *121*, 36–49.

Barlo, M., & Dalkiran, N. A. (2009). Epsilon-Nash implementation. *Economics Letters*, *102*(1), 36–38.

Barlo, M., & Dalkıran, N. A. (2022a). Computational implementation. *Review of Economic Design*, *26*(4), 605–633.

Barlo, M., & Dalkıran, N. A. (2022b). Implementation with missing data. *Mimeo*.

Barlo, M., & Dalkıran, N. A. (2023). Behavioral ex-post implementation. *Mimeo*.

Bergemann, D., & Morris, S. (2008). Ex post implementation. *Games and Economic Behavior*, *63*(2), 527–566.

Bochet, O., & Tumennasan, N. (2021). One truth and a thousand lies: Defaults and benchmarks in mechanism design. *Mimeo*.

Cabrales, A., & Serrano, R. (2011). Implementation in adaptive better-response dynamics: Towards a general theory of bounded rationality in mechanisms. *Games and Economic Behavior*, *73*(2), 360–374.

Chen, Y.-C., Kunimoto, T., Sun, Y., & Xiong, S. (2021). Rationalizable implementation in finite mechanisms. *Games and Economic Behavior*, *129*, 181–197.

Chernoff, H. (1954). Rational selection of decision functions. *Econometrica*, 422–443.

de Clippel, G. (2014). Behavioral implementation. *American Economic Review*, *104*(10), 2975–3002.

de Clippel, G. (2022). Departures from preference maximization, violations of the surething principle, and relevant implications. *Mimeo*.

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

Glazer, J., & Rubinstein, A. (2012). A model of persuasion with boundedly rational agents. *Journal of Political Economy*, *120*(6), 1057–1082.

Holmström, B., & Myerson, R. B. (1983). Efficient and durable decision rules with incomplete information. *Econometrica*, 1799–1819.

Hurwicz, L. (1986). On the implementation of social choice rules in irrational societies. *Social Choice and Public Decision Making: Essays in Honor of Kenneth J. Arrow*.

Jackson, M. O. (1991). Bayesian implementation. *Econometrica*, 461–477.

Jackson, M. O. (2001). A crash course in implementation theory. *Social Choice and Welfare*, *18*(4), 655–708.

Koray, S., & Yildiz, K. (2018). Implementation via rights structures. *Journal of Economic Theory*, *176*, 479–502.

Korpela, V. (2012). Implementation without rationality assumptions. *Theory and Decision*, *72*(2), 189–203.

Korpela, V., & Lombardi, M. (2020). Closure under interim utility equivalence implies two-agent bayesian implementation. *Games and Economic Behavior*, *121*, 108–116.

Kucuksenel, S. (2012). Behavioral mechanism design. *Journal of Public Economic Theory*, *14*(5), 767–789.

Kunimoto, T., & Saran, R. (2022). Robust implementation in rationalizable strategies in general mechanisms. *Mimeo*.

Kunimoto, T., Saran, R., & Serrano, R. (2023). Interim rationalizable implementation of functions. *Mimeo*.

Kunimoto, T., & Serrano, R. (2020). Rationalizable incentives: Interim implementation of sets in rationalizable strategies. *Mimeo*.

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

Maskin, E., & Sjöström, T. (2002). Implementation theory. *Handbook of Social Choice and Welfare*, *1*, 237–288.

Moore, J. (1992). Implementation, contracts, and renegotiation in environments with complete information. *Advances in Economic Theory*, *1*, 182–281.

Moore, J., & Repullo, R. (1990). Nash implementation: a full characterization. *Econometrica*, 1083–1099.

Palfrey, T. R. (2002). Implementation theory. *Handbook of Game Theory with Economic Applications*, *3*, 2271–2326.

Palfrey, T. R., & Srivastava, S. (1987). On Bayesian implementable allocations. *The Review of Economic Studies*, *54*(2), 193–208.

Postlewaite, A., & Schmeidler, D. (1986). Implementation in differential information economies. *Journal of Economic Theory*, *39*(1), 14–33.

Pram, K. (2020). Weak implementation. *Economic Theory*, *69*(3), 569–594.

Ray, K. T. (2010). Nash implementation under irrational preferences. *Unpublished manuscript*.

Repullo, R. (1987). A simple proof of Maskin's theorem on Nash implementation. *Social Choice and Welfare*, *4*(1), 39–41.

Saijo, T. (1988). Strategy space reduction in Maskin's theorem: sufficient conditions for Nash implementation. *Econometrica*, 693–700.

Saran, R. (2011). Menu-dependent preferences and revelation principle. *Journal of Economic Theory*, *146*(4), 1712–1720.

Saran, R. (2016). Bounded depths of rationality and implementation with complete information. *Journal of Economic Theory*, *165*, 517–564.

Savage, L. J. (1951). The theory of statistical decision. *Journal of the American Statistical Association*, *46*(253), 55–67.

Sen, A. K. (1971). Choice functions and revealed preference. *The Review of Economic Studies*, *38*(3), 307–317.

Serrano, R. (2004). The theory of implementation of social choice rules. *SIAM Review*, *46*(3), 377–414.

Spiegler, R. (2011). *Bounded eationality and industrial organization.* Oxford University Press.

Xiong, S. (2023). Rationalizable implementation of social choice functions: complete characterization. *Theoretical Economics*, *18*, 197–230.