# Behavioral Implementation under Incomplete Information*

Mehmet Barlo[†] Nuh Aygün Dalkıran[‡]

March 27, 2022

**Abstract**

We investigate implementation under incomplete information allowing for individuals' choices featuring violations of rationality. Our primitives are individual choices, which 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.

**Keywords:** Behavioral Implementation, Incomplete Information, Bounded Rationality, Interim Implementation, Ex-Post Implementation.

**JEL Classification:** D60, D79, D82, D90

*We would like to 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 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 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*.

*Full implementation* of a predetermined social choice rule requires that the set of equilibrium outcomes of the associated mechanism fully coincides with the given social choice rule. On the other hand, *partial implementation* only requires that the social choice rule be sustained by an equilibrium of the mechanism; hence, it allows for other equilibria associated with outcomes that are not aligned with the social goal at hand. An important appeal of partial implementation involves *the revelation principle*, which implies, in the rational domain and under incomplete information, the following: if there is a mechanism that partially implements a predetermined goal, then there is a *direct mechanism* that truthfully implements it. The undesired equilibria are then often disregarded on the basis of the equilibrium with truthful revelation being the *salient* equilibrium, elegantly pointed out by Postlewaite and Schmeidler (1986) among others.[2] However, in our setting, we

[1] This is why the recent trend involving the use of behavioral insights in policy-making has been growing stronger, implying an increased interest in adapting economic models to allow behavioral biases. In particular, Thaler and Sunstein's book *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.

[2] "[The] problem [of multiple equilibria] is sometimes dismissed with an argument that as long as truthful revelation is an equilibrium, it will somehow be the salient equilibrium even if there are other

reaffirm that the revelation principle (for partial implementation) fails when individuals' choices do not satisfy WARP.[3] Hence, in our environment, one cannot restrict attention to direct mechanisms without a loss of generality. Thus, focusing on full implementation rather than partial implementation becomes crucial in our setup.

In our behavioral environment, a given mechanism induces an incomplete information game in which given a strategy profile, an individual choosing a message provides her an interim Anscombe-Aumann (IAA) act that maps each type profile of others to alternatives. Individuals' choices on IAA acts are given and fail to satisfy WARP. As a result, we obtain a general setup with incomplete information allowing a wide variety of behavioral aspects.

In this context, the 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.

To highlight the new grounds our results cover relative to the complete information analysis of de Clippel (2014), we emphasize that full implementation in BIE is not the same as behavioral (Nash) implementation on every complete information type space: Each individual's strategy is measurable with respect to her type and cannot vary with others' type profiles. This, therefore, results in a requirement akin to the addition of interim incentive compatibility constraints.

Our necessity result, Theorem 1, shows that if a mechanism implements a social choice set (SCS) in BIE, then the opportunity sets sustained in BIE of this mechanism, IAA acts that an individual can obtain by changing her messages while her opponents' strategies remain the same, form a profile of sets of IAA acts with two desirable properties. We refer to such a profile as a profile of sets of IAA acts *interim consistent with the given SCS*. Each set that appears on this profile is associated with an individual, a social choice function (SCF) in the SCS, and a deception profile of the other individuals with the following property novel to the case of incomplete information—causing a significant difference with its complete information counterpart, consistency of de Clippel (2014): Each such set is independent of the type of the individual whom this set is associated with. Moreover, the following hold: ($i$) For all individuals, all her types, and all SCFs in the SCS, this individual's choices at the interim stage from the set of corresponding IAA acts associated with the truthtelling profile of the other individuals contain the IAA act associated with

equilibria as well" (Postlewaite & Schmeidler, 1986).

[3] To the best of our knowledge, the failure of the revelation principle in behavioral environments is first documented by Saran (2011). Meanwhile, Bierbrauer and Netzer (2016) notes that the revelation principle fails with intention-based social preferences.

the SCF given her type; ($ii$) whenever there is a deception that leads to an outcome that is incompatible with the SCS, there is an informant individual along with her informant type such that she does not (interim) choose at her informant type the IAA act given her informant type generated by this deception from her set of IAA acts associated with others' types identified via their deception. We show that the first of these, ($i$), implies quasi incentive compatibility (Proposition 1) and establish that this condition is necessary and sufficient for the partial implementation of an SCF in BIE (Proposition 2). Another implication (Proposition 3) is that the revelation principle holds if individuals' interim choices satisfy the independence of irrelevant alternatives (IIA).

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

To showcase the applicability of our findings in economically relevant domains, *first*, we analyze the implementability of efficiency in BIE. Due to the well-known conflict between efficiency and incentive compatibility, we restrict feasibility based on quasi incentive compatibility in our behavioral incomplete information environment. Consequently, we introduce the behavioral counterpart of interim incentive efficiency of Holmström and Myerson (1983). Our construction parallels that of de Clippel's complete information efficiency and is obtained by entangling its structure with quasi incentive compatibility. Further, we establish that the interim incentive efficient SCS is implementable in BIE under standard conditions (Proposition 5). *Second*, we examine the implementability of the constrained rational expectations equilibrium (CREE) in BIE. To do so, we introduce a behavioral formulation of CREE and show that the corresponding SCS is implementable in BIE under standard assumptions (Proposition 6).[4] *Third*, we display the full implementability in BIE of a selection from the interim incentive efficient SCS on an example using minimax-regret preferences of Savage (1951), while an SCF in this set of SCS 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).

In the implementation literature, positive sufficiency results often rely on "augmented" mechanisms asking individuals to report more than their types. Such mechanisms seem less intuitive and less practical than direct mechanisms to many researchers. That is why we analyze the scope of situations where implementation in BIE is achievable via direct mechanisms: Theorem 3 identifies a necessary and sufficient condition for an SCF to be

[4]See Palfrey and Srivastava (1987) and Bochet (2007) for the implementability of CREE in rational domains.

implementable in BIE via its associated direct mechanism.

The use of the ex-post approach in the analysis of implementation under incomplete information in rational domains provides additional practicality and tractability, especially when individuals' ex-post choices are interdependent. (Bergemann & Morris, 2008) To capture these attractive aspects in our environment, we propose a condition, Property B, that links individuals' ex-post choices with their choices on IAA acts. This condition is in the spirit of Savage's sure-thing principle as well as Property P of de Clippel (2020). It demands that an IAA act be chosen from a set of IAA acts whenever for any state, the corresponding realization of this IAA act is ex-post chosen from the set of alternatives sustained by IAA acts in that set of IAA acts at that state. Thanks to this condition, the notion of (behavioral) ex-post equilibrium (EPE)—a strategy profile in the incomplete information game induced by the mechanism such that individual's plans of action result in (behavioral) Nash equilibrium play at every state—carries over the desirable robustness properties of its counterpart in rational domains. Under Property B, every EPE is a BIE. Hence, EPE induces *robust* behavior thanks to the following ex-post no-regret property: no individual has any incentive to go back to the interim stage and find out others' private information. Moreover, EPE makes no use of any probabilistic information. It is belief-free, does not involve any belief updating or expectation considerations, and does not require any common prior assumption. We obtain a necessity result for implementation in EPE, Theorem 4, and show that it implies a pseudo ex-post incentive compatibility (Proposition 8) along with an ex-post choice monotonicity condition (Proposition 7). Furthermore, we establish that under rationality, our pseudo ex-post choice incentive compatibility and ex-post choice monotonicity are equivalent to the ex-post incentive compatibility and ex-post choice monotonicity of Bergemann and Morris (2008). We also establish a sufficiency result for implementation in EPE, Theorem 5, using the ex-post version of our choice incompatibility condition in addition to the condition obtained in our necessity result.

While Property B instigates appealing properties for implementation in EPE, it comes with a severe warning in environments with individuals' ex-post choices failing WARP. In such environments, the analysis of de Clippel (2020) exhorts us to be wary of the use of EPE as Property B and the failure of WARP may generate a contradiction. We demonstrate situations in which such a contradiction may appear and hence provide a scope of safety for the justifiability of EPE.

Our paper is mostly related to de Clippel (2014), which provides necessary as well as sufficient conditions for behavioral implementation under complete information. Besides de Clippel (2014), another closely related paper is Saran (2011), which considers behavioral partial implementation under incomplete information formalizing behavioral aspects with menu-dependent preferences over IAA acts. It establishes that weak contraction consistency, a condition implied by WARP, is sufficient for the revelation principle. Another related paper is Jackson (1991), which analyzes Bayesian implementation in the rational domain. It generalizes the analysis of Maskin (1999) (on Nash implementation under complete information) to the case of incomplete information. In this sense, what Jackson (1991) is to the seminal work in Maskin (1999), our paper is to de Clippel (2014). Alternatively, our paper can also 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 individual choices' need not satisfy WARP. Another significant and related paper is Bergemann and Morris (2008), analyzing ex-post implementation in the rational domain under incomplete information.

Hurwicz (1986), Eliaz (2002), Korpela (2012), and Ray (2018) have also investigated the problem of behavioral implementation under complete information. 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. Then, the mechanism has to deal with the complications that emerge due to each individual "optimally respond[ing] to the non-faulty players regardless of the identity and actions of the faulty players." On the other hand, Korpela (2012) shows that when individual choices fail rationality axioms, the independence of irrelevant alternatives, also known as Chernoff's $\alpha$, is key to obtaining the necessary and sufficient condition synonymous to that of Moore and Repullo (1990).[5]

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 result. In Section 5, we analyze the implementation of interim incentive efficiency in BIE, while Section 6 provides an example with minimax-regret preferences. Section 7 presents our results analyzing when implementation in BIE is possible via direct mechanisms. Section 8 contains an application, the implementation in BIE of the constrained rational

[5] There have been other papers investigating implementation under complete information that allow for "non-rational" behavior of individuals. An earlier paper of ours, Barlo and Dalkiran (2009), provides an analysis of implementation for the case of epsilon-Nash equilibrium, i.e., when individuals are satisfied by getting close to (but not necessarily achieving) their best responses. 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. Some of the other related work include Benoit and Ok (2006), Cabrales and Serrano (2011), Kucuksenel (2012), Saran (2016), Kunimoto and Saran (2020), Kunimoto and Serrano (2020), Bochet and Tumennasan (2021), Kunimoto et al. (2021), and Barlo and Dalkıran (2022). For more on full implementation, we refer the reader to surveys such as Moore (1992), Jackson (2001), Maskin and Sjöström (2002), Palfrey (2002), and Serrano (2004).

expectations equilibrium. Section 9 contains our analysis with the ex-post approach. Finally, Section 10 concludes.

## 2 Notation and Definitions

Consider a set of individuals, denoted by $N = \{1, \ldots, n\}$, who have to select an alternative from a non-empty set of alternatives $X$. $\mathcal{X}$ stands for the set of all non-empty subsets of $X$. Let $\Theta$ denote the set of all relevant states of the world regarding the choices of the individuals. We assume that there is incomplete information among the individuals regarding the true state of the world and that the true state of the world is distributed knowledge. That is, $\Theta$ 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 $a_i : \Theta_{-i} \to X$, a function mapping $\Theta_{-i}$ into $X$. We denote the set of $i$'s IAA acts sustained by $\Theta_{-i}$ on $X$ by $\mathcal{A}_i$ and let $\mathcal{A}_i^c := \cup_{x \in X} \{a_i^x \in \mathcal{A}_i\}$ where $a_i^x$ is the *constant* IAA act with $a_i^x(\theta_{-i}) = x$ for all $\theta_{-i} \in \Theta_{-i}$. For any $\tilde{\mathcal{A}}_i \subset \mathcal{A}_i$ and any $\theta_{-i}$, $\tilde{\mathcal{A}}_i(\theta_{-i}) := \{x \in X \mid a_i(\theta_{-i}) = x \text{ for some } a_i \in \tilde{\mathcal{A}}_i\}$. Given $i \in N$, her type $\theta_i \in \Theta_i$, and a non-empty subset of IAA acts $\mathcal{S} \subset \mathcal{A}_i$, the *choice of individual $i$ of type $\theta_i$ from the set of IAA acts $\mathcal{S}$* is given by $\mathcal{C}_i^{\theta_i}(\mathcal{S}) \subset \mathcal{S}$ with $\mathcal{C}_i^{\theta_i}(\mathcal{S}) \neq \emptyset$. We impose no further restrictions on $\mathcal{C}_i^{\theta_i} : \mathscr{A}_i \to \mathscr{A}_i$, where $\mathscr{A}_i$ denotes the set of non-empty subsets of $\mathcal{A}_i$.[6]

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

A *state-contingent allocation* (SCA) is a function mapping $\Theta$ into $X$. The SCA $h : \Theta \to X$ induces the following associated IAA act that individual $i$ of type $\theta_i$ faces: $h_{i,\theta_i} \in \mathcal{A}_i$ defined by $h_{i,\theta_i}(\theta_{-i}) = h(\theta_i, \theta_{-i})$ for all $\theta_{-i} \in \Theta_{-i}$. We denote the set of all SCAs by $H := \{h \mid h : \Theta \to X\}$.

An SCF $f : \Theta \to X$ is an SCA that specifies a socially optimal alternative—as evaluated by the planner—for each state. As there may be many socially optimal SCAs 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., $F \subset \{f \mid f : \Theta \to X\}$ and $F \neq \emptyset$.[7] $\mathscr{F}$ denotes the set of all SCSs.

[6] Sen (1971) shows that a choice correspondence satisfies WARP (and be 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 $\alpha$ (Chernoff, 1954)) and an expansion consistency axiom (known as Sen's $\beta$). Letting $\mathcal{X}$ be the set of all non-empty subsets of alternatives, we say that the individual choice correspondence $C : \mathcal{X} \to \mathcal{X}$ 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 $\beta$ 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 $y \in C(T)$.

[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).

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.

An *(interim) strategy* of individual $i$ under mechanism $\mu$ 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 $\mu$ for $\sigma_{-i}$* 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

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

Given individuals' choices on IAA acts, a natural equilibrium concept that parallels the interim equilibrium of Saran (2011) is 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 $\theta_i \in \Theta_i$, $\boldsymbol{a}_{i,\theta_i}^* \in \mathcal{C}_i^{\theta_i}(\mathcal{O}_i^{\mu}(\sigma_{-i}^*))$, where $\boldsymbol{a}_{i,\theta_i}^*$ is the IAA act defined by $\boldsymbol{a}_{i,\theta_i}^*(\theta_{-i}) = g(\sigma_i^*(\theta_i), \sigma_{-i}^*(\theta_{-i}))$ for all $\theta_{-i} \in \Theta_{-i}$.*

In words, the strategy profile $\sigma^*$ is a BIE of $\mu$ if for any player $i$ and any type $\theta_i$, the IAA act generated in the mechanism via the prescribed action $\sigma_i^*(\theta_i)$ is chosen by $i$ of type $\theta_i$ from the opportunity set of IAA acts of $i$ given others' strategy $\sigma_{-i}^*$. 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 $\mu$ such that*

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

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

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 $\alpha^{\text{id}}$ 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 $\Lambda_i$ and $\Lambda := \times_{i \in N} \Lambda_i$. Finally, $\alpha_{-i}(\theta_{-i}) := (\alpha_j(\theta_j))_{j \neq i}$ and $\Lambda_{-i} := \times_{j \neq i} \Lambda_j$.

# 3 Necessity

Below, we introduce the notion of *interim consistency* and show that it is necessary for implementation in BIE.

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

(*i*) *for all* $i \in N$ *and all* $\theta_i \in \Theta_i$, $f_{i,\theta_i} \in C_i^{\theta_i}(\mathcal{S}_i(f, \alpha_{-i}^{\mathrm{id}}))$, *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* $f_{i^*,\theta_{i^*}^*}^{\alpha} \notin C_{i^*}^{\theta_{i^*}^*}(\mathcal{S}_{i^*}(f, \alpha_{-i^*}))$, *where* $f_{i^*,\theta_{i^*}^*}^{\alpha} : \Theta_{-i^*} \to X$ *is given by* $f_{i^*,\theta_{i^*}^*}^{\alpha}(\theta_{-i^*}) = f(\alpha_{i^*}(\theta_{i^*}^*), \alpha_{-i^*}(\theta_{-i^*}))$ *for all* $\theta_{-i^*} \in \Theta_{-i^*}$.

A profile of sets of IAA acts $\mathbb{S} := (\mathcal{S}_i(f, \alpha_{-i}))_{i \in N,\ f \in F,\ \alpha_{-i} \in \Lambda_{-i}}$ satisfies interim consistency with an SCS $F$ if for each SCF $f$ in SCS $F$, the following hold: ($i$) Given any $i \in N$ and any $\theta_i \in \Theta_i$, it must be that $i$'s choices when she is of type $\theta_i$ from $\mathcal{S}_i(f, \alpha_{-i}^{\mathrm{id}})$ (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 $f_{i,\theta_i}$; and ($ii$) if there is a deception profile $\alpha$ that leads to an outcome not compatible with the SCS, i.e., $f \circ \alpha \notin F$, then there exist an informant individual $i^*$ of type $\theta_{i^*}^*$ who does not choose $f_{i^*,\theta_{i^*}^*}^{\alpha}$ (the IAA act associated with $f \circ \alpha$ that $i^*$ of type $\theta_{i^*}^*$ faces) from $\mathcal{S}_{i^*}(f, \alpha_{-i^*})$ (the set of IAA acts in $\mathbb{S}$ associated with $i^*$, $f$, and $\alpha_{-i^*}$, the deception profile of individuals other than $i^*$).[8]

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 $\sigma^f$ of $\mu$ such that $f = g \circ \sigma^f$. Thus, for each $i \in N$ and $\theta_i \in \Theta_i$, the IAA act associated with $f$ that $i$ of type $\theta_i$ faces, $f_{i,\theta_i}$, is in $C_i^{\theta_i}(O_i^{\mu}(\sigma_{-i}^f))$. Defining $\mathbb{S}$ by $\mathcal{S}_i(f, \alpha_{-i}) := O_i^{\mu}(\sigma_{-i}^f \circ \alpha_{-i})$ with $i \in N$, $f \in F$, and $\alpha_{-i} \in \Lambda_{-i}$, we observe that ($i$) of interim consistency of $\mathbb{S}$ with $F$ holds as $\sigma_{-i}^f \circ \alpha_{-i}^{\mathrm{id}} = \sigma_{-i}^f$.

On the other hand, if a deception profile $\alpha$ is such that $f \circ \alpha \notin F$, then $\sigma^f \circ \alpha$ cannot be a BIE of $\mu$. Otherwise, by ($ii$) of implementability in BIE (Definition 2), 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 $f_{i^*,\theta_{i^*}^*}^{\alpha}$, the IAA act associated with $f \circ \alpha$ that $i^*$ of type $\theta_{i^*}^*$ faces, from $O_{i^*}^{\mu}(\sigma_{-i^*}^f(\alpha_{-i^*}))$, which equals $\mathcal{S}_{i^*}(f, \alpha_{-i^*})$. This delivers ($ii$) of interim consistency of $\mathbb{S}$ with $F$.

[8] 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 $x \in \Phi(\theta)$, $x \in C_i^{\theta}(S_i(x, \theta))$; ($ii$) $x \in \Phi(\theta) \setminus \Phi(\theta')$ with $\theta, \theta' \in \Theta$ implies there is $i^* \in N$ such that $x \notin C_{i^*}^{\theta'}(S_{i^*}(x, \theta))$. The critical difference between de Clippel's consistency and ours is that, with incomplete information, each choice set must be independent of the type of the individual whom this set is associated with.

This 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 with 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.

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 $\mathcal{T}_i \in \mathscr{A}_i$ with $\{f_{i,\tilde{\theta}_i} \mid \tilde{\theta}_i \in \Theta_i\} \subset \mathcal{T}_i$ and $f_{i,\theta_i} \in \mathcal{C}_i^{\theta_i}(\mathcal{T}_i)$ 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 $\mathbb{S} := (\mathcal{S}_i(f, \alpha_{-i}))_{i \in N,\ f \in F,\ \alpha_{-i} \in \Lambda_{-i}}$. To see this, for any given $f \in F$ and any $i \in N$, let $\mathcal{T}_i = \mathcal{S}_i(f, \alpha_{-i}^{\mathrm{id}})$. Then, for all $i \in N$, $\{f_{i,\tilde{\theta}_i} \mid \tilde{\theta}_i \in \Theta_i\} \subset \mathcal{T}_i$, and $(i)$ of interim consistency implies $f_{i,\theta_i} \in \mathcal{C}_i^{\theta_i}(\mathcal{T}_i)$ 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.*

As full implementability implies partial implementability, we obtain the following result, which is a restatement of Proposition 4.9 of Saran (2011) in our setting.

**Proposition 2.** *Let $f : \Theta \to X$ be an SCF. Then, $f$ is partially implementable in BIE if and only if the SCS $F = \{f\}$ is quasi incentive compatible.*

**Proof of Proposition 2.** The necessity direction of this proposition follows from Theorem 1 and Proposition 1. On the other hand, the sufficiency direction demands the construction of an indirect mechanism that sustains partial implementability of the given quasi incentive compatible SCF. To see this, suppose that SCF $f$ is quasi incentive compatible (i.e., $F = \{f\}$ is a quasi incentive compatible SCS). The associated profile of sets of IAA acts $(\mathcal{T}_i)_{i \in N}$ are such that for all $i$, $\{f_{i,\tilde{\theta}_i} \mid \tilde{\theta}_i \in \Theta_i\} \subset \mathcal{T}_i$ and $f_{i,\theta_i} \in \mathcal{C}_i^{\theta_i}(\mathcal{T}_i)$ for all $\theta_i \in \Theta_i$. The mechanism we use is $\mu = (M, g)$ where for each $i \in N$, $M_i = \Theta_i \times \mathcal{A}_i \times \{0, 1, \ldots, n\}$, and a generic message is denoted by $m_i = (\theta_i^{(i)}, a_i^{(i)}, k^{(i)})$. We let

$j^* := \sum_{i \in N} k^{(i)}$ (mod n), and specify $g : M \to X$ via the following rules:

$$\textbf{Rule 1}: \quad g(m) = f(\theta) \qquad \text{if } m_i = (\theta_i, f_{i,\theta_i}, 0) \text{ for all } i \in N,$$

$$\textbf{Rule 2}: \quad g(m) = \begin{cases} a_j^{(j)}(\theta_{-j}) & \text{if } j = j^* \text{ and } a_j^{(j)} \in \mathcal{T}_j, \\ f_{j,\tilde{\theta}_j}(\theta_{-j}) & \text{otherwise.} \end{cases} \qquad \text{otherwise,}$$

Now, we show that there is a BIE of $\mu$, $\sigma^f$, such that $f = g \circ \sigma^f$. Let $\sigma_i^f(\theta_i) = (\theta_i, f_{i,\theta_i}, 0)$ for each $i \in N$. Then, Rule 1 applies and we have $g(\sigma^f(\theta)) = f(\theta)$ for each $\theta \in \Theta$, i.e., $f = g \circ \sigma^f$. Observe that for any unilateral deviation of individual $i$ from $\sigma^f$, by construction, $\mathcal{O}_i^\mu(\sigma_{-i}^f) = \mathcal{T}_i$ for all $i \in N$. Then, by quasi incentive compatibility of $F = \{f\}$, $f_{i,\theta_i} \in \mathcal{C}_i^{\theta_i}(\mathcal{T}_i)$ for each $i \in N$ and each $\theta_i \in \Theta_i$. Ergo, for all $i \in N$ and all $\theta_i \in \Theta_i$, $a_{i,\theta_i}^* \in \mathcal{C}_i^{\theta_i}(\mathcal{O}_i^\mu(\sigma_{-i}^f))$ where $a_{i,\theta_i}^*(\theta_{-i}) = g(\sigma_i^f(\theta_i), \sigma_{-i}^f(\theta_{-i}))$ for all $\theta_{-i} \in \Theta_{-i}$. That is, $\sigma^f$ is a BIE of $\mu$ such that $f = g \circ \sigma^f$. ■

Proposition 2 establishes that quasi incentive compatibility of an SCF $f$ is equivalent to partial implementability of $f$ in BIE. In general, quasi incentive compatibility of an SCF $f$ does not imply that truthtelling is a BIE in the *direct mechanism associated with* $f$.[9] We note that if $f$ is partially truthfully implemented in BIE by the direct mechanism $\mu^f$, then the opportunity set of $i$ under truthtelling is $\mathcal{O}_i^{\mu^f}(\alpha_{-i}^{\text{id}}) = \{f_{i,\tilde{\theta}_i} \mid \tilde{\theta}_i \in \Theta_i\}$. Indeed, an SCF $f$ is said to be *incentive compatible* if for all $i \in N$, $f_{i,\theta_i} \in \mathcal{C}_i^{\theta_i}(\{f_{i,\tilde{\theta}_i} \mid \tilde{\theta}_i \in \Theta_i\})$ for all $\theta_i \in \Theta_i$. Consequently, if attention is restricted to choices satisfying the IIA, then implementability of $f$ in BIE (thanks to Theorem 1 and Proposition 1) implies that for all $i \in N$, the existence of a set of IAA acts $\mathcal{T}_i$ such that $f_{i,\theta_i} \in \mathcal{C}_i^{\theta_i}(\mathcal{T}_i)$ for all $\theta_i \in \Theta_i$. As for all $i \in N$, $\{f_{i,\tilde{\theta}_i} \mid \tilde{\theta}_i \in \Theta_i\} \subset \mathcal{T}_i$, thanks to the IIA, for all $i \in N$ and all $\theta_i \in \Theta_i$, $f_{i,\theta_i} \in \mathcal{C}_i^{\theta_i}(\{f_{i,\tilde{\theta}_i} \mid \tilde{\theta}_i \in \Theta_i\})$. This observation establishes that under the IIA, truthtelling is a BIE in the direct mechanism associated with $f$ whenever $f$ is partially implementable in BIE. Indeed, the revelation principle in our setting demands that if $\sigma^*$ is a BIE of a mechanism $\mu = (M, g)$, then truthtelling, $\alpha^{\text{id}}$, is a BIE of the direct mechanism associated with the SCA $g \circ \sigma^*$. Hence, we attain the following result:

**Proposition 3.** *If individual choices satisfy the IIA, then the revelation principle holds.*

In summary, if a mechanism $\mu$ partially implements an SCF $f$ in BIE and individuals' choices satisfy the IIA, then there is a direct mechanism that partially implements $f$ in truthful BIE. That is, the IIA is sufficient for the revelation principle. However, when the IIA does not hold, the revelation principle fails as we display in Section 6. Indeed, Saran

[9]The direct mechanism associated with an SCF $f : \Theta \to X$ is $\mu^f = (\Theta, g^f)$ where for all $i \in N$, the set of messages is given by $\Theta_i$, and the outcome function $g^f = f$.

(2011) is the first study to observe the failure of the revelation principle in behavioral domains, and it identifies a condition that is weaker than the IIA and sufficient for the revelation principle.

## 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. Below, we present such conditions.[10]

**Definition 5.** *The* **choice incompatible pair property** *holds in a given environment $\mathcal{E}$ if the following holds: For each profile of sets of IAA acts $(\tilde{\mathcal{A}}_i)_{i\in N}$ and each SCA $h \in H$ such that for some $\bar{\theta} \in \Theta$ and for some $\bar{j} \in N$,*

$$\tilde{\mathcal{A}}_i(\bar{\theta}_{-i}) = X \text{ for all } i \in N \setminus \{\bar{j}\},$$

*and*

$$\hbar_{i,\bar{\theta}_i} \in \tilde{\mathcal{A}}_i \text{ for all } i \in N,$$

*there are $i^*, j^* \in N$ with $i^* \neq j^*$ such that $\hbar_{i^*,\bar{\theta}_{i^*}} \notin \mathcal{C}_{i^*}^{\bar{\theta}_{i^*}}(\tilde{\mathcal{A}}_{i^*})$ and $\hbar_{j^*,\bar{\theta}_{j^*}} \notin \mathcal{C}_{j^*}^{\bar{\theta}_{j^*}}(\tilde{\mathcal{A}}_{j^*})$.*

In words, the choice incompatible pair property demands that for any profile of sets of IAA acts, $(\tilde{\mathcal{A}}_i)_{i\in N}$, and any SCA $h$ such that for some state $\bar{\theta}$ and individual $\bar{j} \in N$, the alternatives sustained in projection by an IAA act in $\tilde{\mathcal{A}}_i$, namely $\tilde{\mathcal{A}}_i(\bar{\theta}_{-i})$, equals $X$ for all $i$ other than $\bar{j}$, and the IAA act induced by $h$ for $i$'s type at $\bar{\theta}$, $\hbar_{i,\bar{\theta}_i}$, is in $\tilde{\mathcal{A}}_i$ for all individuals $i$ (including $\bar{j}$), there are two distinct individuals $i^*$ and $j^*$ who do not choose $\hbar_{i^*,\bar{\theta}_{i^*}}$ and $\hbar_{j^*,\bar{\theta}_{j^*}}$ at their types $\bar{\theta}_{i^*}$ and $\bar{\theta}_{j^*}$ from $\tilde{\mathcal{A}}_{i^*}$ and $\tilde{\mathcal{A}}_{j^*}$, respectively.

Vaguely put, this condition implies some level of disagreement among individuals regarding their evaluation of SCAs and requires that the choices of at most $n-2$ individuals are aligned with each other.

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

**Theorem 2.** *Suppose that the environment $\mathcal{E}$ is such that $n \geq 3$ and the choice incompatible pair property holds. Then, if there exists a profile of sets of IAA acts interim consistent with the 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 incompatible pair property holds, and $F$ is an SCS for which

[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.

the profile $\mathbb{S} := (\mathcal{S}_i(f, \alpha_{-i}))_{i\in N,\ f\in F,\ \alpha_{-i}\in\Lambda_{-i}}$ is interim consistent. The mechanism we employ makes use of the following observations: $(i)$ the outcome should be $f(\theta)$ when there is unanimous agreement between the individuals over $f \in F$ and the realized state is $\theta$; $(ii)$ under such a unanimous agreement each individual $j$ should be able to generate unilaterally the set $\mathcal{S}_j(f, \alpha^{\mathrm{id}}_{-j})$ (i.e., when all other individuals $i \neq j$ have unanimously decided on SCF $f \in F$ and are reporting their types truthfully, then $j$ should be able to generate $\mathcal{S}_j(f, \alpha^{\mathrm{id}}_{-j})$); $(iii)$ whenever there is an attempt to deceive the designer so that an outcome not compatible with the SCS emerges, a whistleblower should be able to alert the designer; $(iv)$ undesirable BIE should be eliminated (e.g., by using a modulo or an integer game).[11]

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

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

**Table 1:** The outcome function of the mechanism.

In words, each individual $i$ is required to send a message that specifies an SCF $f^{(i)} \in F$, a type for himself $\theta_i^{(i)} \in \Theta_i$, an IAA act $\boldsymbol{a}_i^{(i)} \in \mathcal{A}_i$, an alternative $x^{(i)} \in X$, and 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 the 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 $\mathcal{S}_j(f, \alpha^{\mathrm{id}}_{-j})$, otherwise the outcome is in line with the SCF $f$ (while the IAA act associated with $f$ that $j$ faces is also in $\mathcal{S}_j(f, \alpha^{\mathrm{id}}_{-j})$). Finally, Rule 3 applies when both Rules 1 and 2 fail, then the outcome is determined by the winner of the modulo game.

[11] Our mechanism resembles those used for sufficiency in the implementation literature. See for example, Repullo (1987), Saijo (1988), Moore and Repullo (1990), Jackson (1991), Danilov (1992), Maskin (1999), Bergemann and Morris (2008), de Clippel (2014), Koray and Yildiz (2018).

If Rule 1 holds at a state $\theta \in \Theta$, the opportunity sets of IAA acts under the associated (possibly deceptive) strategy profile are as given by the interim consistent profile $\mathbb{S}$. Thus, the opportunity sets of IAA acts under truthtelling satisfy $(i)$ of interim consistency and hence 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 incompatible pair property. This property is similar to the *economic environment* assumption in the rational domain (Jackson, 1991).[12] As a result, every BIE must be aligned with the SCS $F$; because otherwise, by $(ii)$ of interim consistency, there is a whistleblower 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 that the choice incompatible pair property holds. Let $F \in \mathscr{F}$ be an SCS for which the profile $\mathbb{S} := (\mathcal{S}_i(f, \alpha_{-i}))_{i \in N,\ f \in F,\ \alpha_{-i} \in \Lambda_{-i}}$ is interim consistent. The mechanism we use is as defined on page 12.

First, we show that for any $f \in F$, there exists $\sigma^f$ 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., $f = g \circ \sigma^f$. Observe that for any unilateral deviation of individual $i$ from $\sigma^f$, either Rule 1 or Rule 2 applies, while Rule 3 is not attainable. So, by construction $\mathcal{O}_i^\mu(\sigma_{-i}^f) = \mathcal{S}_i(f, \alpha_{-i}^{\mathrm{id}})$ for all $i \in N$. Recall that, by $(i)$ of interim consistency, $\mathbf{f}_{i,\theta_i} \in \mathcal{C}_i^{\theta_i}(\mathcal{S}_i(f, \alpha_{-i}^{\mathrm{id}}))$ for each $i \in N$ and each $\theta_i \in \Theta_i$. Ergo, for all $i \in N$ and all $\theta_i \in \Theta_i$, $\mathbf{a}_{i,\theta_i}^* \in \mathcal{C}_i^{\theta_i}(\mathcal{O}_i^\mu(\sigma_{-i}^f))$ where $\mathbf{a}_{i,\theta_i}^*(\theta_{-i}) = g(\sigma_i^f(\theta_i), \sigma_{-i}^f(\theta_{-i}))$ for all $\theta_{-i} \in \Theta_{-i}$. That is, $\sigma^f$ is a BIE of $\mu$ such that $f = g \circ \sigma^f$.

Consider now any BIE $\sigma^*$ of $\mu$ denoted as $\sigma_i^*(\theta_i) = (f_i(\theta_i), \alpha_i(\theta_i), \mathbf{a}_i(\theta_i), x_i(\theta_i), k_i(\theta_i))$ for each $i \in N$. That is, $f_i(\theta_i)$ denotes the proposed SCF, $\alpha_i(\theta_i)$ 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 $\theta_i$.

Next, we show that, under any BIE $\sigma^*$ of $\mu$, Rule 1 must apply at every state $\theta \in \Theta$: Suppose for a contradiction that either Rule 2 or Rule 3 applies under $\sigma^*$ at $\theta$. Then, by construction, $\mathcal{O}_i^\mu(\sigma_{-i}^*)$ is given by some $\tilde{\mathcal{A}}_i \subset \mathcal{A}_i$ such that $\tilde{\mathcal{A}}_i(\theta_{-i}) = X$ for all $i \neq j$ for some $j \in N$ (as any $i \neq j$ can change their proposed integer and win the modulo game at $\theta$). As $\sigma^*$ is a BIE of $\mu$, the SCA $h^* := g \circ \sigma^*$ is such that $\mathbf{h}_{i,\theta_i}^* \in \mathcal{C}_i^{\theta_i}(\tilde{\mathcal{A}}_i)$ for at least $n-1$ individuals where by construction $h^*$ is such that $\mathbf{h}_{\ell,\theta_\ell}^* \in \tilde{\mathcal{A}}_\ell$ for all $\ell \in N$ as $\tilde{\mathcal{A}}_\ell = \mathcal{O}_j^\mu(\sigma_{-\ell}^*)$. This contradicts the choice incompatible pair property. Thus, under any BIE of $\mu$, Rule

[12] For other versions of this assumption, see Bergemann and Morris (2008) and Barlo and Dalkıran (2021).

1 must apply at every $\theta \in \Theta$.

Moreover, the product structure of the state space guarantees that under any BIE $\sigma^*$ of $\mu$, there is a unique $f \in F$ such that $f_i(\theta_i) = f$ for all $i \in N$ and for all $\theta_i \in \Theta_i$.

Finally, we show that it must be 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 $\sigma^*$, by construction, we have, at each $\theta \in \Theta$, $\mathcal{O}_i^\mu(\sigma_{-i}^*) = \mathcal{S}_i(f, \alpha_{-i})$ for all $i \in N$. 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 $\mathcal{f}_{i^*,\theta_{i^*}^*}(\alpha_{-i^*}) \notin \mathcal{C}_{i^*}^{\theta_{i^*}^*}(\mathcal{S}_{i^*}(f, \alpha_{-i^*}))$. But this implies $\mathcal{a}_{i^*,\theta_{i^*}^*}^* \notin \mathcal{C}_{i^*}^{\theta_{i^*}^*}(\mathcal{O}_{i^*}^\mu(\sigma_{-i^*}^*))$, where $\mathcal{a}_{i^*,\theta_{i^*}^*}^*(\theta_{-i^*}) = g(\sigma_{i^*}^*(\theta_{i^*}^*), \sigma_{-i^*}^*(\theta_{-i^*}))$ for all $\theta_{-i^*} \in \Theta_{-i^*}$. This is a contradiction to $\sigma^*$ being a BIE of $\mu$. Therefore, $g \circ \sigma^* = f \circ \alpha \in F$, which implies that condition $(ii)$ of implementability in BIE holds as well. ■

## 5 Interim Incentive Efficiency

"[T]he proper object for welfare analysis with incomplete information is the [SCA], rather than the actual decision or allocation ultimately chosen. Furthermore, any efficiency criterion for evaluating [SCAs] must be defined independently of [individuals' private information]." (Holmström & Myerson, 1983). In what follows, we introduce the behavioral counterpart of interim incentive 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 $\theta$ if each individual has an implicit opportunity set from which she chooses $x$ at $\theta$, and each alternative is in at least one of the implicit opportunity sets of an individual.[13]

Below, we propose the notion of *interim incentive efficiency* following Holmström and Myerson (1983). That study introduces this notion in rational domains under incomplete information by restricting feasibility based on incentive compatibility. In our setting, the relevant restriction takes the form of quasi incentive compatibility, as is displayed by our necessity results. Letting the set of all quasi incentive compatible SCAs be denoted by $H^*$, we obtain the following definition:

**Definition 6.** *Given an environment $\mathcal{E}$, the interim incentive efficient SCS $E \subset H^*$ consists of all SCAs $e : \Theta \to X$ for which there is a profile of sets of IAA acts $(\mathcal{Y}_i)_{i \in N}$ such that*

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

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

[13] Attaining individuals' choices on alternatives at a given state requires further structure in our setup, as individuals' choices on IAA acts depend only on their own types. We obtain such a structure with ex-post considerations. See Section 9 for further details.

This welfare criterion internalizes quasi incentive compatibility into efficiency: An SCA $e$ is 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 $e$ (and hence $e \in H^*$) with the additional property that for every possible quasi incentive compatible SCA, there is an individual and her type for which the IAA act associated with this SCA is in her implicit opportunity set. Therefore, interim incentive efficient SCS, $E$, can be considered as an interim counterpart of de Clippel's efficiency entangled with quasi incentive compatibility.[14]

The next result paves the way to the implementability in BIE of our efficiency notion:

**Proposition 4.** *If the interim incentive efficient SCS $E$ is non-empty, then it has an interim consistent profile of sets of IAA acts.*

**Proof of Proposition 4.** Consider $\mathbb{Y} := (\mathcal{Y}_i(e, \alpha_{-i}))_{i \in N,\ e \in E,\ \alpha_{-i} \in \Lambda_{-i}}$ such that for all $i \in N$, all $e \in E$, and all $\alpha_{-i} \in \Lambda_{-i}$, $\mathcal{Y}_i(e, \alpha_{-i})$ equals $\mathcal{Y}_i(e)$ associated with $e$ as in Definition 6. Below, we show that $\mathbb{Y}$ is interim consistent with $E$.

Let $e \in E$. Then, $(i)$ of Definition 6 implies $(i)$ of interim consistency of $\mathbb{Y}$. Now, suppose there is a deception $\alpha$ such that $e \circ \alpha \notin E$. Then, by letting $e \circ \alpha = e^\alpha$ and noting that $e^\alpha$ is an SCA, we observe that for all $(\mathcal{Y}_j)_{j \in N}$ satisfying $(ii)$ of Definition 6, there is $(i, \theta_i)$ such that $\boldsymbol{e}^\alpha_{i,\theta_i} \notin \mathcal{C}_i^{\theta_i}(\mathcal{Y}_i)$; because otherwise, $e^\alpha \in E$. Thus, in particular for the profile $(\mathcal{Y}_i(e))_{i \in N}$, there exists $(i^*, \theta^*_{i^*})$ such that $\boldsymbol{e}^\alpha_{i^*,\theta^*_{i^*}} \notin \mathcal{C}_{i^*}^{\theta^*_{i^*}}(\mathcal{Y}_{i^*}(e))$; ergo, as $\mathcal{Y}_{i^*}(e) = \mathcal{Y}_{i^*}(e, \alpha_{-i^*})$, $(ii)$ of interim consistency follows. ∎

Theorem 2 and Proposition 4 deliver the following result presented without proof: $E$ is implementable in BIE when there is a weak form of disagreement in the society.

**Proposition 5.** *Suppose that the environment $\mathcal{E}$ is such that $n \geq 3$, the choice incompatible pair property holds, and the interim incentive efficient SCS $E$ is non-empty. Then, $E$ is implementable in BIE.*

## 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. Nonetheless, we attain full implementability in BIE of an SCS that constitutes a selection from the interim incentive efficient SCS. Meanwhile, we also attain partial implementation of an SCF in BIE even though partial direct implementation of that same

[14]The SCS of the example we analyze in the next section is a selection from the interim efficient SCS of that setting and demonstrates interim efficiency with minimax-regret preferences.

SCF in truthful BIE is not possible. In other words, the revelation principle fails to hold in our setting. This parallels 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 payoff possible in that state. Then, the maximum regret of individual $i$ of type $\theta_i$ 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 $\theta_i$. Formally, for any two IAA acts $\boldsymbol{a}_i$ and $\tilde{\boldsymbol{a}}_i$ in a given set of IAA acts $\mathcal{S}_i \subset \mathcal{A}_i$, $\boldsymbol{a}_i$ is weakly preferred to $\tilde{\boldsymbol{a}}_i$ in the minimax-regret setting if

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

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 $\theta$.

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 2. An SCA $h : \Theta \to X$ is denoted by $h = \langle abcd\rangle$ where $h(\alpha, \alpha) = a$,

| $\Theta$ | $(\alpha,\alpha)$ | $(\alpha,\beta)$ | $(\beta,\alpha)$ | $(\beta,\beta)$ |
|---|---|---|---|---|
| $(u_1(x \mid \theta), u_2(x \mid \theta))$ | $(0,1)$ | $(0,1/2)$ | $(1/2,1)$ | $(1/2,1/2)$ |
| $(u_1(y \mid \theta), u_2(y \mid \theta))$ | $(1,0)$ | $(1,1)$ | $(1,0)$ | $(1,1)$ |
| $(u_1(z \mid \theta), u_2(z \mid \theta))$ | $(2,0)$ | $(2,0)$ | $(0,0)$ | $(0,0)$ |

**Table 2:** State-contingent utilities of our example.

$h(\alpha, \beta) = b$, $h(\beta, \alpha) = c$, and $h(\beta, \beta) = d$ with $a, b, c, d \in \{x, y, z\}$. Similarly, an IAA act that $i$ faces, $\boldsymbol{a}_i : \Theta_{-i} \to X$ is denoted by $\boldsymbol{a}_i = \langle ab\rangle$ where $\boldsymbol{a}_i(\alpha) = a$ and $\boldsymbol{a}_i(\beta) = b$ with $a, b \in \{x, y, z\}$.

The SCS we consider in this example for full implementation is $F = \{\langle xyyy\rangle, \langle yyyy\rangle\}$. We note that $F \subset E$, i.e., both of these SCAs are interim incentive efficient. It is tedious but straightforward to show that the following profile of sets of IAA acts sustain both SCAs in $F$ in interim incentive efficiency as formalized in Definition 6: $\mathcal{Y}_1 = \{\langle xx\rangle, \langle xy\rangle, \langle xz\rangle, \langle yy\rangle\}$ and $\mathcal{Y}_2 = \{\langle xy\rangle, \langle yx\rangle, \langle yy\rangle, \langle yz\rangle, \langle zx\rangle, \langle zy\rangle, \langle zz\rangle\}$.

The mechanism given in Table 3 implements the 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 the SCA $\langle xyyy\rangle$; $\sigma_1^{(2)}(\alpha) = M$, $\sigma_1^{(2)}(\beta) = M$, $\sigma_2^{(2)}(\alpha) = L$, and $\sigma_2^{(2)}(\beta) = L$ inducing $\langle yyyy\rangle$; $\sigma_1^{(3)}(\alpha) = M$, $\sigma_1^{(3)}(\beta) = M$, $\sigma_2^{(3)}(\alpha) = L$, and $\sigma_2^{(3)}(\beta) = R$ inducing $\langle yyyy\rangle$; $\sigma_1^{(4)}(\alpha) = M$, $\sigma_1^{(4)}(\beta) = M$, $\sigma_2^{(4)}(\alpha) = R$, and $\sigma_2^{(4)}(\beta) = L$ inducing $\langle yyyy\rangle$.

| | | Individual 2 | |
|---|---|---|---|
| | | $L$ | $R$ |
| Individual 1 | $U$ | $x$ | $y$ |
| | $M$ | $y$ | $y$ |
| | $D$ | $x$ | $z$ |

**Table 3:** The mechanism fully implementing SCS $F$ in BIE.

The SCA $\langle xyyy \rangle$ is partially implementable in BIE by the mechanism defined in Table 3 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 4. This follows from type $\alpha$ of individual 1 choosing the IAA

| | | Individual 2 | |
|---|---|---|---|
| | | $\alpha$ | $\beta$ |
| Individual 1 | $\alpha$ | $x$ | $y$ |
| | $\beta$ | $y$ | $y$ |

**Table 4:** The direct mechanism $\mu^d$ 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 Direct Mechanisms

We now evaluate the significance of direct mechanisms pertinent to full implementation in BIE in general environments. Thereby, we portray settings where implementation in BIE is attainable using intuitive mechanisms. We focus on SCFs instead of SCSs since direct mechanisms cannot coordinate selections of SCFs from an SCS.

Given an SCF $f : \Theta \to X$, the relevant profile of sets of IAA acts is given by $\mathbb{F} := (\mathcal{F}_i(\alpha_{-i}))_{i \in N,\ \alpha_{-i} \in \Lambda_{-i}}$ where for any $i \in N$ and any $\alpha_{-i} \in \Lambda_{-i}$

$$\mathcal{F}_i(\alpha_{-i}) := \{\boldsymbol{a}_i \in \mathcal{A}_i \mid \exists \tilde{\theta}_i \in \Theta_i \text{ s.t. } \boldsymbol{a}_i(\theta_{-i}) = f(\tilde{\theta}_i, \alpha_{-i}(\theta_{-i}))\ \forall \theta_{-i} \in \Theta_{-i}\}.$$

In words, $\mathcal{F}_i(\alpha_{-i})$ is the set of IAA acts that SCF $f$ induces for all possible types of individual $i$ given that individuals other than $i$ are using the deception profile $\alpha_{-i}$. The resulting profile of sets of IAA acts, $\mathbb{F}$, corresponds to the opportunity sets of IAA acts in the direct mechanism associated with the SCF $f$. As a result, there is an intertwined link between the interim consistency of $\mathbb{F}$ with the SCF $f$ and the BIE implementability of $f$ via its direct mechanism. This connection results in Theorem 3, a characterization of situations in which implementability in BIE is equivalent to the BIE implementability via the direct mechanism possessing a truthful BIE.

**Theorem 3.** *An SCF $f : \Theta \to X$ is fully implementable in BIE by its associated direct mechanism possessing a truthful BIE if and only if the profile of sets of IAA acts $\mathbb{F} := (\mathcal{F}_i(\alpha_{-i}))_{i\in N,\ \alpha_{-i}\in\Lambda_{-i}}$ is interim consistent with $f$.*[15]

**Proof of Theorem 3.** For the *necessity* direction, suppose $f$ is implementable in BIE by its direct mechanism $\mu^f = (\Theta, g^f)$ with $g^f = f$. Due to full BIE implementation, let the truthful BIE be $\sigma^f$ with $\sigma_i^f : \Theta_i \to \Theta_i$ for all $i \in N$ and $f = g^f \circ \sigma^f$. Clearly, $\sigma^f = \alpha^{\mathrm{id}}$.

Then, $\mathcal{O}_i^{\mu^f}(\alpha_{-i}^{\mathrm{id}}) = \mathcal{F}_i(\alpha_{-i}^{\mathrm{id}})$ for all $i \in N$ implies (as $\alpha^{\mathrm{id}}$ is a BIE of $\mu^f$) $f_{i,\theta_i} \in \mathcal{C}_i^{\theta_i}(\mathcal{F}_i(\alpha_{-i}^{\mathrm{id}}))$ for all $i \in N$ and $\theta_i \in \Theta_i$ establishing $(i)$ of interim consistency of $\mathbb{F}$.

For any deception $\alpha$ with $f \circ \alpha \neq f$, $\sigma^f \circ \alpha$ cannot be a BIE of $\mu^f$ because otherwise $g^f \circ \sigma^f \circ \alpha = f \circ \alpha$ and hence by $(ii)$ of BIE implementation $f \circ \alpha$ equals $f$, a contradiction. Let $f^\alpha := f \circ \alpha$. Because that $\sigma^f \circ \alpha$ is not a BIE, there is $i^* \in N$ and $\theta_{i^*}^* \in \Theta_{i^*}$ such that $f^\alpha_{i^*,\theta^*_{i^*}} \notin \mathcal{C}_{i^*}^{\theta^*_{i^*}}(\mathcal{F}_{i^*}(\alpha_{-i^*}))$ since $\mathcal{O}_{i^*}^{\mu^f}((\sigma_j^f(\alpha_j)_{j\neq i^*}) = \mathcal{F}_{i^*}(\alpha_{-i^*})$. This delivers $(ii)$ of interim consistency of $\mathbb{F}$.

Next, we handle the *sufficiency* of Theorem 3. Suppose that $\mathbb{F}$ is interim consistent with $f$. Consider the truthtelling strategy $\alpha^{\mathrm{id}}$. Since $g^f(\alpha^{\mathrm{id}}(\theta)) = f(\theta)$ for all $\theta \in \Theta$; for all $i \in N$, $\mathcal{O}_i^{\mu^f}(\alpha_{-i}^{\mathrm{id}}) = \mathcal{F}_i(\alpha_i^{\mathrm{id}})$ and $f_{i,\theta_i} \in \mathcal{C}_i^{\theta_i}(\mathcal{F}_i(\alpha_{-i}^{\mathrm{id}}))$ for all $\theta_i \in \Theta_i$ (by $(i)$ of interim consistency), $\alpha^{\mathrm{id}}$ is a truthful BIE strategy with $g^f \circ \alpha^{\mathrm{id}} = f$.

Now, we show that if $\sigma^*$ is a BIE of $\mu^f$, then $g^f \circ \sigma^*$ equals $f$. Suppose $\sigma^*$ is a BIE and $g^f(\sigma^*(\theta)) \neq f(\theta)$ for some $\theta \in \Theta$. Let $\alpha$ be such that $\alpha(\theta) = \sigma^*(\theta) \neq \theta$. Then, $f(\alpha(\theta)) \neq f(\theta)$; ergo, $f \circ \alpha \neq f$. Hence, by $(ii)$ interim consistency of $\mathbb{F}$ with SCF $f$, there is $i^* \in N$ and $\theta^*_{i^*} \in \Theta_{i^*}$ with $g^f(\sigma^*) = f \circ \alpha =: f^\alpha$ we have $f^\alpha_{i^*,\theta^*_{i^*}} \notin \mathcal{C}_{i^*}^{\theta^*_{i^*}}(\mathcal{F}_{i^*}(\alpha_{-i^*}))$, contradicting to $\sigma^*$ being a BIE as $\mathcal{O}_{i^*}^{\mu^f}(\alpha_{-i^*}) = \mathcal{F}_{i^*}(\alpha_{-i^*})$. ■

## 8 Constrained Rational Expectations Equilibrium

In this section, we analyze behavioral implementation under incomplete information of the *constrained rational expectations equilibrium* (CREE), "the most frequently studied notion of market equilibrium [with incomplete information]" (Palfrey & Srivastava, 1987).[16] First, we present the formulation of CREE in the behavioral domain. Similar to its rational counterpart, behavioral CREE consists of rational expectations equilibrium with agents choosing from budgetarily feasible bundles that do not exceed the aggregate

[15]The direct mechanism associated with $f$ may also have an untruthful BIE. But then, its outcome must coincide with $f$ whenever $f$ is fully implementable in BIE by its direct mechanism.

[16]It is appropriate to point out that "the extension of the concept of Walrasian equilibrium to the case of incomplete information is not yet fully understood." (Bochet, 2007)

endowment. Second, using a mild restriction on individuals' choices, we prove the full implementability of behavioral CREE for economies that possess a behavioral CREE.

Each individual's consumption set is $\mathbb{R}^K_+$ capturing the situation with $K \in \mathbb{N}$ different perfectly divisible goods. The endowment of individual $i \in N$ is denoted by $\mathbf{w}_i \in \mathbb{R}^K_{++} := \{\mathbf{w} \in \mathbb{R}^K_+ \mid \mathbf{w} \gg \mathbf{0}\}$. An SCA is a map $x : \Theta \to \mathbb{R}^{N\times K}_+$, and $x$ is a feasible SCA if it maps each state $\theta \in \Theta$ into $\{\mathbf{y} \in \mathbb{R}^{N\times K}_+ \mid \sum_{i\in N}(\mathbf{y}_i - \mathbf{w}_i) \leq \mathbf{0}\}$. We define the aggregate endowment by $\bar{\mathbf{w}} := \sum_{i\in N} \mathbf{w}_i$, and let $[\mathbf{0}, \bar{\mathbf{w}}] := \{\mathbf{w} \in \mathbb{R}^K_+ \mid w_k \in [0, \bar{w}_k]\ \forall k = 1, \ldots, K\}$. The choices of individual $i \in N$ of type $\theta_i \in \Theta_i$ from a non-empty set of IAA acts $\tilde{\mathcal{X}}_i \subset \mathcal{X}_i := \{\chi_i : \Theta_{-i} \to \mathbb{R}^K_+\}$ are given by $\mathcal{C}^{\theta_i}_i(\tilde{\mathcal{X}}_i)$ where $\mathcal{C}^{\theta_i}_i(\tilde{\mathcal{X}}_i) \subset \tilde{\mathcal{X}}_i$ with $\mathcal{C}^{\theta_i}_i(\tilde{\mathcal{X}}_i) \neq \emptyset$.

An economy $\mathcal{E}$ is summarized by $\langle(\mathcal{C}^{\theta_i}_i)_{\theta_i\in\Theta_i}, \mathbf{w}_i\rangle_{i\in N}$.

Following Palfrey and Srivastava (1987) and Bochet (2007), we consider information structures where the set of admissible states $\Theta^* \subset \Theta$ is such that the following hold:

$(i)$ *Non-exclusive Information* (NEI): For each $\theta \in \Theta^*$, there does not exist $\theta'_i \in \Theta_i \backslash \{\theta_i\}$ with $(\theta'_i, \theta_{-i}) \in \Theta^*$; and

$(ii)$ *No redundant type* (NRT): For all $i \in N$ and all $\theta_i \in \Theta_i$, there is $\theta_{-i} \in \Theta_{-i}$ such that $(\theta_i, \theta_{-i}) \in \Theta^*$.

For all individuals $i \in N$ and all of $i$'s types $\theta_i \in \Theta_i$, we let $\Theta^*_{-i}(\theta_i) := \{\theta_{-i} \in \Theta_{-i} \mid (\theta_i, \theta_{-i}) \in \Theta^*\}$ and note that $\Theta^*_{-i}(\theta_{-i})$ is non-empty (thanks to NRT) and does not have a product structure (due to NEI).

Recall that given SCA $x : \Theta^* \to \mathbb{R}^{N\times K}_+$, the IAA act that individual $i$ of type $\theta_i$ faces is given by $\chi_{i,\theta_i} : \Theta^*_{-i}(\theta_i) \to \mathbb{R}^K_+$ with $\chi_{i,\theta_i}(\theta_{-i}) = x_i(\theta_i, \theta_{-i})$ for all $\theta_{-i} \in \Theta^*_{-i}(\theta_i)$.

For any state-contingent strictly positive prices $p : \Theta^* \to \mathbb{R}^K_{++}$, the budgetarily feasible IAA acts individual $i$ of type $\theta_i$ faces that are constrained by $\bar{\mathbf{w}}$ are given as follows:[17]

$$\mathcal{B}^{\bar{\mathbf{w}}}_{i,\theta_i}(p) := \left\{ y_{i,\theta_i} : \Theta^*_{-i}(\theta_i) \to [\mathbf{0}, \bar{\mathbf{w}}] \;\middle|\; \begin{array}{ll} (i) & y_{i,\theta_i} \text{is measurable,} \\ (ii) & p(\theta_i, \theta_{-i}) \cdot (y_{i,\theta_i}(\theta_{-i}) - \mathbf{w}_i) \leq 0 \quad \forall \theta_{-i} \in \Theta^*_{-i}(\theta_i) \end{array} \right\}.$$

Now, we are ready to define a behavioral CREE:

**Definition 7.** *Prices $p^* : \Theta^* \to \mathbb{R}^K_{++}$ and SCA $x^* : \Theta^* \to \mathbb{R}^{N\times K}_+$ constitute a (behavioral)* **constrained rational expectations equilibrium** *of an economy $\mathcal{E}$ if*

$(i)$ *For all $i \in N$ and all $\theta_i \in \Theta_i$, $p^*_{i,\theta_i} : \Theta^*_{-i}(\theta_i) \to \mathbb{R}^K_{++}$ and $\chi^*_{i,\theta_i} : \Theta^*_{-i}(\theta_i) \to \mathbb{R}^K_+$ are measurable where $p^*_{i,\theta_i}(\theta_{-i}) = p^*(\theta_i, \theta_{-i})$ and $\chi^*_{i,\theta_i}(\theta_{-i}) = x^*_i(\theta_i, \theta_{-i})$ for all $\theta_{-i} \in \Theta^*_{-i}(\theta_i)$; and*

$(ii)$ *For all $i \in N$ and all $\theta_i \in \Theta_i$, we have that*

$(a)$ $\chi^*_{i,\theta_i} \in \mathcal{C}^{\theta_i}_i(\mathcal{B}^{\bar{\mathbf{w}}}_{i,\theta_i}(p^*))$, *and*

[17] If $\Theta$ is finite, then we can dispense with the measurability requirements.

(b) $p^*(\theta_i, \theta_{-i}) = p^*(\theta_i, \theta'_{-i})$ *with* $\theta_{-i}, \theta'_{-i} \in \Theta^*_{-i}(\theta_i)$ *implies* $\mathcal{X}^*_{i,\theta_i}(\theta_{-i}) = \mathcal{X}^*_{i,\theta_i}(\theta'_{-i})$; *and*

(iii) *For all* $\theta \in \Theta^*$, $\sum_{i \in N} \mathcal{X}^*_{i,\theta_i}(\theta_{-i}) \leq \bar{\mathbf{w}}$.

*We refer to the set* $\{x^* : \Theta^* \to \mathbb{R}^{N\times K}_+ \mid \exists p^* : \Theta^* \to \mathbb{R}^K_{++}$ *s.t.* $(p^*, x^*)$ *is a CREE of* $\mathcal{E}\}$ *as the CREE SCS on* $\mathcal{E}$.

The notion of (behavioral) CREE requires a strictly positive price rule $p^* : \Theta^* \to \mathbb{R}^K_{++}$ and SCA $x^* : \Theta^* \to \mathbb{R}^{N\times K}_+$ to satisfy three requirements. The first is that $p^*$ and $x^*$ are measurable with respect to the private information of each of the individuals. That is, for every individual $i$ who observes any one of her types $\theta_i$, the price IAA act that she faces, $\mathcal{p}^*_{i,\theta_i} : \Theta^*_{-i}(\theta_i) \to \mathbb{R}^K_{++}$, and the consumption IAA act she encounters, $\mathcal{X}^*_{i,\theta_i} : \Theta^*_{-i}(\theta_i) \to \mathbb{R}^K_+$, are both measurable. The item $(a)$ of the second requirement demands the optimality of individuals' choices from budgetarily feasible IAA acts while $(b)$ is a requirement associated with rational expectations setting: If the private information of an individual is the same in two different states for which the price rule induces the same price vector, then her optimal choice must deliver the same bundle at these two states. Finally, the third condition involves feasibility in every state.

We employ the following condition when establishing full implementability of (behavioral) CREE in BIE:

**Definition 8.** *We say that an economy* $\mathcal{E}$ *satisfies* **Condition M** *if it satisfies the following for all* $i \in N$ *and all* $\theta_i \in \Theta_i$*: If* $\tilde{\Theta}_{-i} \subset \Theta^*_{-i}(\theta_i)$ *with* $\tilde{\Theta}_{-i} \neq \emptyset$ *and* $\mathcal{X}_{i,\theta_i}, \mathcal{X}'_{i,\theta_i} \in \mathcal{X}_i$ *are such that* $\mathcal{X}_{i,\theta_i}(\tilde{\theta}_{-i}) = \mathbf{w}_i$ *and* $\mathcal{X}'_{i,\theta_i}(\tilde{\theta}_{-i}) = \mathbf{0}$ *for all* $\tilde{\theta}_{-i} \in \tilde{\Theta}_{-i}$ *while* $\mathcal{X}_{i,\theta_i}(\theta'_{-i}) = \mathcal{X}'_{i,\theta_i}(\theta'_{-i})$ *for all* $\theta'_{-i} \in \Theta^*_{-i}(\theta_i) \setminus \tilde{\Theta}_{-i}$*, then* $\mathcal{X}'_{i,\theta_i}(\theta'_{-i}) \notin \mathcal{C}^{\theta_i}_i(\tilde{\mathcal{X}}_i)$ *for all* $\tilde{\mathcal{X}}_i \subset \mathcal{X}_i$ *with* $\mathcal{X}_{i,\theta_i}, \mathcal{X}'_{i,\theta_i} \in \tilde{\mathcal{X}}_i$.

Intuitively, this condition demands that no consumption is not desirable when endowments are available. It requires that for each individual $i$ and any one of her types $\theta_i$ the following holds: If two IAA acts agree on a subset of others' types compatible with $\theta_i$ and differ on the remaining type profiles of others so that in such profiles, the first provides the endowment of that individual, $\mathbf{w}_i$, while the second involves no consumption, $\mathbf{0}$, then the first IAA act prevents the second to be chosen by $i$ of type $\theta_i$ from any set of IAA acts containing these IAA acts. We wish to emphasize that Condition M is in the same spirit as Savage's **P3** (*Monotonicity*) and involves only a comparison of $\mathbf{w}_i$ with $\mathbf{0}$. It, thereby, allows a wide class of behavioral aspects.[18]

We are now ready to present our result about the full implementability of (behavioral) CREE in BIE.

[18]For example, Condition M holds in economies where individuals' choices originate from minimax-regret preferences (which we describe in Section 6) that possess the following structure: For all $i \in N$ and all $\theta_i \in \Theta_i$, $u_i(\mathbf{0} \mid (\theta_i, \theta'_{-i})) < u_i(\mathbf{w}_i \mid (\theta_i, \theta'_{-i}))$ for all $\theta'_{-i} \in \Theta^*_{-i}(\theta_i)$.

**Proposition 6.** *If there exists a CREE of an economy $\mathcal{E}$ satisfying Condition M, then the CREE SCS on $\mathcal{E}$ is implementable in BIE.*

**Proof of Proposition 6.** Suppose that the economy $\mathcal{E}$ satisfies Condition M, there is a CREE of $\mathcal{E}$, and the information structure satisfies NEI and NRT.

We employ a variant of the mechanism of Bochet (2007) obtained by extending the elementary mechanisms of Dutta et al. (1995) to incomplete and non-exclusive information. The mechanism $\mu = (M, g)$ is such that $M_i := Y \times P \times \Theta_i \times \mathbb{N}$ where $Y := \{y : \Theta^* \to \mathbb{R}_+^{N\times K}\}$, $P := \{p : \Theta^* \to \mathbb{R}_{++}^K\}$, and a generic message of $i$ is denoted by $m_i = (y^{(i)}, p^{(i)}, \theta_i^{(i)}, k^{(i)})$. The outcome function $g : M \to Y$ is defined via the following rules: For any given $m \in M$, let $i^* := \min\{i \in N \mid k^{(i)} \geq k^{(j)}\ \forall j \in N\}$ and $j^* := \max\{j \in N \mid k^{(j)} \leq k^{(i)}\ \forall i \in N\}$.

**Rule 1:** If $m_i = (\bar{x}, \bar{p}, \theta_i^{(i)}, k^{(i)})$ for all $i \in N$ is such that $(\theta_i^{(i)})_{i\in N} = \bar{\theta} \in \Theta^*$ and $(\bar{p}, \bar{x})$ satisfies the following requirements

- $(i)$ for all $i \in N$ and all $\theta_i \in \Theta_i$, $\bar{p}_{i,\theta_i} : \Theta_{-i}^*(\theta_i) \to \mathbb{R}_{++}^K$ is measurable, and $\bar{x}_{i,\theta_i} : \Theta_{-i}^*(\theta_i) \to \mathbb{R}_+^K$ is in $\mathcal{B}_{i,\theta_i}^{\bar{\mathbf{w}}}(\bar{p})$, and
- $(ii)$ for all $i \in N$ and all $\theta_i \in \Theta_i$, $\bar{p}(\theta_i, \theta_{-i}) = \bar{p}(\theta_i, \theta'_{-i})$ for some $\theta_{-i}, \theta'_{-i} \in \Theta_{-i}^*(\theta_i)$ implies $\bar{x}_{i,\theta_i}(\theta_{-i}) = \bar{x}_{i,\theta_i}(\theta'_{-i})$, and
- $(iii)$ for all $\theta \in \Theta^*$, $\sum_{i\in N} \bar{x}_{i,\theta_i}(\theta_{-i}) \leq \bar{\mathbf{w}}$,

then $g(m) = \bar{x}(\bar{\theta})$.

**Rule 2:** If $m_j = (\bar{x}, \bar{p}, \theta_j^{(j)}, k^{(j)})$ for all $j \in N \setminus \{i^*\}$, $(\theta_i^{(i)})_{i\in N} = \bar{\theta} \in \Theta^*$, $(\bar{p}, \bar{x})$ satisfies $(i), (ii), (iii)$ of Rule 1, $p^{(i^*)} = \bar{p}$, and $x^{(i^*)} \neq \bar{x}$ but

- $(iv)$ $x_{i^*,\bar{\theta}_{i^*}}^{(i^*)} \in \mathcal{B}_{i^*,\bar{\theta}_{i^*}}^{\bar{\mathbf{w}}}(\bar{p})$, and
- $(v)$ for all $\theta_{i^*} \in \Theta_{i^*}$, $\bar{p}(\theta_{i^*}, \theta_{-i^*}) = \bar{p}(\theta_{i^*}, \theta'_{-i^*})$ for some $\theta_{-i^*}, \theta'_{-i^*} \in \Theta_{-i^*}^*(\theta_{i^*})$ implies $x_{i^*,\theta_{i^*}}^{(i^*)}(\theta_{-i^*}) = x_{i^*,\theta_{i^*}}^{(i^*)}(\theta'_{-i^*})$,

then $g(m) = (\mathbf{y}_i)_{i\in N}$ with $\mathbf{y}_i = \mathbf{0}$ for all $i \neq i^*$ and $\mathbf{y}_{i^*} = x_{i^*,\bar{\theta}_{i^*}}^{(i^*)}(\bar{\theta}_{-i^*})$.

**Rule 2':** If $m_j = (\bar{x}, \bar{p}, \theta_j^{(j)}, k^{(j)})$ for all $j \in N \setminus \{i^\diamond\}$ for some $i^\diamond \in N \setminus \{i^*\}$, $(\theta_\ell^{(\ell)})_{\ell\in N} = \bar{\theta} \in \Theta^*$, $(\bar{p}, \bar{x})$ satisfies $(i), (ii), (iii)$ of Rule 1, $p^{(i^\diamond)} = \bar{p}$, and $x^{(i^\diamond)} \neq \bar{x}$ but

- $(iv)$ $x_{i^\diamond,\bar{\theta}_{i^\diamond}}^{(i^\diamond)} \in \mathcal{B}_{i^\diamond,\bar{\theta}_{i^\diamond}}^{\bar{\mathbf{w}}}(\bar{p})$, and
- $(v)$ for all $\theta_{i^\diamond} \in \Theta_{i^\diamond}$, $\bar{p}(\theta_{i^\diamond}, \theta_{-i^\diamond}) = \bar{p}(\theta_{i^\diamond}, \theta'_{-i^\diamond})$ for some $\theta_{-i^\diamond}, \theta'_{-i^\diamond} \in \Theta_{-i^\diamond}^*(\theta_{i^\diamond})$ implies $x_{i^\diamond,\theta_{i^\diamond}}^{(i^\diamond)}(\theta_{-i^\diamond}) = x_{i^\diamond,\theta_{i^\diamond}}^{(i^\diamond)}(\theta'_{-i^\diamond})$,

then $g(m) = (\mathbf{y}_j)_{j\in N}$ with $\mathbf{y}_j = \mathbf{0}$ for all $j \neq i^\diamond$ and $\mathbf{y}_{i^\diamond} = x_{i^\diamond,\bar{\theta}_{i^\diamond}}^{(i^\diamond)}(\bar{\theta}_{-i^\diamond})$.

**Rule 3:** For all other cases, each individual $i \in N \setminus \{j^*\}$ receives her endowment $\mathbf{w}_i$ while individual $j^*$ gets $\mathbf{0}$.

**Claim 1.** *For any CREE $(p^*, x^*)$ of $\mathcal{E}$, $\sigma^*$ is a BIE of $\mu = (M, g)$ with $g \circ \sigma^* = x^*$ where $\sigma_i^*(\theta_i) = (x^*, p^*, \theta_i, 1)$ for all $i \in N$ and all $\theta_i \in \Theta_i$.*

**Proof of Claim 1.** Under $\sigma^*$, Rule 1 applies at every $\theta \in \Theta^*$ since $(p^*, x^*)$ is a CREE of $\mathcal{E}$. Hence, $g \circ \sigma^* = x^*$ on $\Theta^*$. Thanks to NEI, for each $\theta' \in \Theta^*$, $(\theta_i'', \theta_{-i}') \notin \Theta^*$ for all $\theta_i'' \neq \theta_i'$. Thus, any unilateral deviation from $\sigma^*$ leads to either of Rules 2, 2' or 3 applying at some $\tilde{\theta} \in \Theta^*$. As a result, for any $\theta \in \Theta^*$, $\mathcal{O}_i^{\mu}(\sigma_{-i}^*) = \mathcal{B}_{i,\theta_i}^{\bar{\mathbf{w}}}(p^*)$ for all $i \in N$. As $(p^*, x^*)$ is a CREE of $\mathcal{E}$, for all $i \in N$ and all $\theta_i \in \Theta_i$, $\chi_{i,\theta_i}^* \in \mathcal{C}_i^{\theta_i}(\mathcal{O}_i^{\mu}(\sigma_{-i}^*))$. Therefore, $\sigma^*$ is a BIE of $\mu$ sustaining $x^*$. ■

**Claim 2.** *For any $\sigma^*$ BIE of $\mu = (M, g)$, Rule 1 applies at every $\theta \in \Theta^*$.*

**Proof of Claim 2.** We let $\sigma_i^*(\theta_i) = (x^{*(i)}(\theta_i), p^{*(i)}(\theta_i), \alpha_i^{*(i)}(\theta_i), k_i^{*(i)}(\theta_i))$ for all $i \in N$ all $\theta_i \in \Theta_i$.

For any strategy profile $\sigma$ and any agent $i$ of type $\theta_i$, we let $\mathcal{g}_{i,\theta_i}^{\sigma} : \Theta_{-i}^*(\theta_i) \to \mathbb{R}_+^K$ (where $\mathcal{g}_{i,\theta_i}^{\sigma}(\theta_{-i}) := g_i(\sigma_i(\theta_i), \sigma_{-i}(\theta_{-i}))$ for any $\theta_{-i} \in \Theta_{-i}^*(\theta_i)$) denote the IAA act that $i$ of $\theta_i$ faces under $\sigma$ in mechanism $\mu = (M, g)$; where $g_i(m) = \mathbf{y}_i$ whenever $g(m) = (\mathbf{y}_j)_{j \in N}$.

Now, consider any $\tilde{\theta} \in \Theta^*$. Then, the following cases emerge:

**Case 1.** *Rule 2 applies at $\tilde{\theta}$ under $\sigma^*$.*

Consider $i \in N \setminus \{i^*\}$ and note that $g_i(\sigma_i^*(\tilde{\theta}_i), \sigma_{-i}^*(\theta_{-i})) = \mathbf{0}$. Next, we define a unilateral deviation of $i$ of type $\tilde{\theta}_i$, $\tilde{\sigma}_i$, as follows: $\tilde{\sigma}_i(\theta_i) = \sigma_i^*(\theta_i)$ for all $\theta_i \neq \tilde{\theta}_i$ while $\tilde{\sigma}_i(\tilde{\theta}_i) = (x^{*(i)}(\tilde{\theta}_i), p^{*(i)}(\tilde{\theta}_i), \alpha_i^{*(i)}(\tilde{\theta}_i), \tilde{k})$ where $\tilde{k} > \max_{j \in N \setminus \{i\}} k_j^{*(j)}(\tilde{\theta}_j)$. That is, $i$ of type $\tilde{\theta}_i$ (who is not $i^*$ at $\tilde{\theta}$ under $\sigma^*$) deviates unilaterally only in her integer choice so that following this deviation $i$ becomes $i^*$ at $\tilde{\theta}$ under $(\tilde{\sigma}_i, \sigma_{-i}^*)$. Rule 3 applies at $\tilde{\theta}$ under $(\tilde{\sigma}_i, \sigma_{-i}^*)$ because $(i)$ Rule 2 applies at $\tilde{\theta}$ under $\sigma^*$; $(ii)$ $i$ is not $i^*$ at $\tilde{\theta}$ under $\sigma^*$; $(iii)$ $i^*$ at $\tilde{\theta}$ under $\sigma^*$ is choosing an SCA different from all the others (including $i$); and $(iv)$ $\tilde{\sigma}_i$ involves a unilateral deviation only concerning $i$'s integer choice at $\tilde{\theta}_i$. Therefore, $g_i(\tilde{\sigma}_i(\tilde{\theta}_i), \sigma_{-i}^*(\tilde{\theta}_{-i})) = \mathbf{w}_i$ as clearly $i$ is not $j^*$ at $\tilde{\theta}$ under $(\tilde{\sigma}_i, \sigma_{-i}^*)$.

At states $(\tilde{\theta}_i, \hat{\theta}_{-i})$ with $\hat{\theta}_{-i} \in \Theta_{-i}^*(\tilde{\theta}_i) \setminus \{\tilde{\theta}_{-i}\}$, Rules 1, 2, 2', or 3 may apply under $\sigma^*$.

**Subcase 1.** *Rule 1 applies at $(\tilde{\theta}_i, \hat{\theta}_{-i})$ under $\sigma^*$.*

Then (as a deviation in the integer choice does not trigger any other rules) Rule 1 still applies at $(\tilde{\theta}_i, \hat{\theta}_{-i})$ under $(\tilde{\sigma}_i, \sigma_{-i}^*)$. Thus, $g_i(\sigma_i^*(\tilde{\theta}_i), \sigma_{-i}^*(\hat{\theta}_{-i})) = g_i(\tilde{\sigma}_i(\tilde{\theta}_i), \sigma_{-i}^*(\hat{\theta}_{-i}))$.

**Subcase 2.** *Rule 2 applies at $(\tilde{\theta}_i, \hat{\theta}_{-i})$ under $\sigma^*$.*

Whether or not $i$ is $i^*$ at $(\tilde{\theta}_i, \hat{\theta}_{-i})$ under $\sigma^*$ determines $i$'s consumption bundle at $(\tilde{\theta}_i, \hat{\theta}_{-i})$.

If $i$ is $i^*$ at $(\tilde{\theta}_i, \hat{\theta}_{-i})$ under $\sigma^*$, then $i$ is also $i^*$ at $(\tilde{\theta}_i, \hat{\theta}_{-i})$ under $(\tilde{\sigma}_i, \sigma^*_{-i})$ as $i$'s integer choice for her type $\tilde{\theta}_i$ has increased, and Rule 2 also applies at $(\tilde{\theta}_i, \hat{\theta}_{-i})$ under $(\tilde{\sigma}_i, \sigma^*_{-i})$; therefore, $g_i(\sigma^*_i(\tilde{\theta}_i), \sigma^*_{-i}(\hat{\theta}_{-i})) = g_i(\tilde{\sigma}_i(\tilde{\theta}_i), \sigma^*_{-i}(\hat{\theta}_{-i}))$.

If $i$ is not $i^*$ at $(\tilde{\theta}_i, \hat{\theta}_{-i})$ under $\sigma^*$, then $i$ may become $i^*$ or not at $(\tilde{\theta}_i, \hat{\theta}_{-i})$ under $(\tilde{\sigma}_i, \sigma^*_{-i})$ since $i$'s integer for her type $\tilde{\theta}_i$ goes up; however, $i$ cannot become $i^\diamond$ at $(\tilde{\theta}_i, \hat{\theta}_{-i})$ under $(\tilde{\sigma}_i, \sigma^*_{-i})$ as $i$ chooses the same SCA with all but $i^*$ at $(\tilde{\theta}_i, \hat{\theta}_{-i})$ under $\sigma^*$. Therefore, if $i$ does not become $i^*$ at $(\tilde{\theta}_i, \hat{\theta}_{-i})$ under $(\tilde{\sigma}_i, \sigma^*_{-i})$, then Rule 2 still applies at $(\tilde{\theta}_i, \hat{\theta}_{-i})$ under $(\tilde{\sigma}_i, \sigma^*_{-i})$, and $g_i(\sigma^*_i(\tilde{\theta}_i), \sigma^*_{-i}(\hat{\theta}_{-i})) = g_i(\tilde{\sigma}_i(\tilde{\theta}_i), \sigma^*_{-i}(\hat{\theta}_{-i})) = \mathbf{0}$. On the other hand, if $i$ becomes $i^*$ at $(\tilde{\theta}_i, \hat{\theta}_{-i})$ under $(\tilde{\sigma}_i, \sigma^*_{-i})$, then Rule 2' is triggered at $(\tilde{\theta}_i, \hat{\theta}_{-i})$ under $(\tilde{\sigma}_i, \sigma^*_{-i})$. This follows from $(i)$ Rule 2 applies at $(\tilde{\theta}_i, \hat{\theta}_{-i})$ under $\sigma^*$; $(ii)$ $i$ is not $i^*$ at $(\tilde{\theta}_i, \hat{\theta}_{-i})$ under $\sigma^*$; $(iii)$ $i^*$ at $(\tilde{\theta}_i, \hat{\theta}_{-i})$ under $\sigma^*$ is choosing an SCA different from all the others (including $i$); and $(iv)$ $\tilde{\sigma}_i$ involves a unilateral deviation only concerning $i$'s integer choice at $\tilde{\theta}_i$. Therefore, after this change $i^*$ at $(\tilde{\theta}_i, \hat{\theta}_{-i})$ under $\sigma^*$ becomes $i^\diamond$ at $(\tilde{\theta}_i, \hat{\theta}_{-i})$ under $(\tilde{\sigma}_i, \sigma^*_{-i})$. Consequently, $g_i(\tilde{\sigma}_i(\tilde{\theta}_i), \sigma^*_{-i}(\hat{\theta}_{-i})) = \mathbf{0} = g_i(\sigma^*_i(\tilde{\theta}_i), \sigma^*_{-i}(\hat{\theta}_{-i}))$.

**Subcase 3.** *Rule 2' applies at $(\tilde{\theta}_i, \hat{\theta}_{-i})$ under $\sigma^*$.*

At $(\tilde{\theta}_i, \hat{\theta}_{-i})$ under $\sigma^*$, $i$ may be either $(i)$ $i^*$ (and hence neither $i^\diamond$ nor $j^*$); or $(ii)$ $i^\diamond$ (and hence not $i^*$); or $(iii)$ neither $i^*$ nor $i^\diamond$.

If $i$ is $i^*$ at $(\tilde{\theta}_i, \hat{\theta}_{-i})$ under $\sigma^*$, then $i$ is still $i^*$ at $(\tilde{\theta}_i, \hat{\theta}_{-i})$ and Rule 2' continues to hold under $(\tilde{\sigma}_i, \sigma^*_{-i})$ (as the only change involves an increase of $i$'s proposed integer); so, $g_i(\tilde{\sigma}_i(\tilde{\theta}_i), \sigma^*_{-i}(\hat{\theta}_{-i})) = g_i(\sigma^*_i(\tilde{\theta}_i), \sigma^*_{-i}(\hat{\theta}_{-i}))$.

If $i$ is $i^\diamond$ at $(\tilde{\theta}_i, \hat{\theta}_{-i})$ under $\sigma^*$, then as a result of $i$'s deviation to $\tilde{\sigma}_i$, $i$ may or may not become $i^*$ at $(\tilde{\theta}_i, \hat{\theta}_{-i})$ under $(\tilde{\sigma}_i, \sigma^*_{-i})$. If $i$ becomes $i^*$ at $(\tilde{\theta}_i, \hat{\theta}_{-i})$ under $(\tilde{\sigma}_i, \sigma^*_{-i})$, then Rule 2 is triggered (as $i^*$ at $(\tilde{\theta}_i, \hat{\theta}_{-i})$ under $\sigma^*$ chooses the same SCA with all others apart from $i$) and hence $g_i(\tilde{\sigma}_i(\tilde{\theta}_i), \sigma^*_{-i}(\hat{\theta}_{-i})) = g_i(\sigma^*_i(\tilde{\theta}_i), \sigma^*_{-i}(\hat{\theta}_{-i}))$. If $i$ does not become $i^*$ at $(\tilde{\theta}_i, \hat{\theta}_{-i})$ under $(\tilde{\sigma}_i, \sigma^*_{-i})$, then still Rule 2' holds (as $i^*$ at $(\tilde{\theta}_i, \hat{\theta}_{-i})$ under $\sigma^*$ chooses the same SCA with all others apart from $i^\diamond$, and $i$ is $i^\diamond$ at $(\tilde{\theta}_i, \hat{\theta}_{-i})$ under $\sigma^*$) $i$ continues to be $i^\diamond$ at $(\tilde{\theta}_i, \hat{\theta}_{-i})$ under $(\tilde{\sigma}_i, \sigma^*_{-i})$; ergo, $g_i(\tilde{\sigma}_i(\tilde{\theta}_i), \sigma^*_{-i}(\hat{\theta}_{-i})) = g_i(\sigma^*_i(\tilde{\theta}_i), \sigma^*_{-i}(\hat{\theta}_{-i}))$.

If $i$ is neither $i^*$ nor $i^\diamond$ at $(\tilde{\theta}_i, \hat{\theta}_{-i})$ under $\sigma^*$, then $i$ may or may not become $i^*$ (but $i$ cannot turn into $i^\diamond$) at $(\tilde{\theta}_i, \hat{\theta}_{-i})$ under $(\tilde{\sigma}_i, \sigma^*_{-i})$. As only $i^\diamond$ chooses an SCA different from all others and $i^\diamond \neq i^*$ at $(\tilde{\theta}_i, \hat{\theta}_{-i})$ under $\sigma^*$, Rule 2' still applies at $(\tilde{\theta}_i, \hat{\theta}_{-i})$ under $(\tilde{\sigma}_i, \sigma^*_{-i})$; so, $g_i(\tilde{\sigma}_i(\tilde{\theta}_i), \sigma^*_{-i}(\hat{\theta}_{-i})) = \mathbf{0} = g_i(\sigma^*_i(\tilde{\theta}_i), \sigma^*_{-i}(\hat{\theta}_{-i}))$.

**Subcase 4.** *Rule 3 applies at $(\tilde{\theta}_i, \hat{\theta}_{-i})$ under $\sigma^*$.*

Then there are three possibilities: $i$ is $i^*$ at $(\tilde{\theta}_i, \hat{\theta}_{-i})$ under $\sigma^*$; $i$ is neither $i^*$ nor $j^*$ at $(\tilde{\theta}_i, \hat{\theta}_{-i})$ under $\sigma^*$; and $i$ is $j^*$ at $(\tilde{\theta}_i, \hat{\theta}_{-i})$ under $\sigma^*$.

If $i$ is $i^*$ at $(\tilde{\theta}_i, \hat{\theta}_{-i})$ under $\sigma^*$, then $i$ increasing her integer choice at $\tilde{\theta}_i$ does not trigger any other rules, and Rule 3 still applies at $(\tilde{\theta}_i, \hat{\theta}_{-i})$ under $(\tilde{\sigma}_i, \sigma^*_{-i})$; so, $g_i(\sigma^*_i(\tilde{\theta}_i), \sigma^*_{-i}(\hat{\theta}_{-i})) = g_i(\tilde{\sigma}_i(\tilde{\theta}_i), \sigma^*_{-i}(\hat{\theta}_{-i}))$.

If $i$ is neither $i^*$ nor $j^*$ at $(\tilde{\theta}_i, \hat{\theta}_{-i})$ under $\sigma^*$, then $g_i(\sigma^*_i(\tilde{\theta}_i), \sigma^*_{-i}(\hat{\theta}_{-i})) = \mathbf{w}_i$. Individual $i$ choosing a higher integer at $\tilde{\theta}_i$ under $\tilde{\sigma}_i$ implies that $i$ cannot be $j^*$ at $(\tilde{\theta}_i, \hat{\theta}_{-i})$ under $(\tilde{\sigma}_i, \sigma^*_{-i})$. However, $i$ may or may not become $i^*$ at $(\tilde{\theta}_i, \hat{\theta}_{-i})$ under $(\tilde{\sigma}_i, \sigma^*_{-i})$. If $i$ is not $i^*$ at $(\tilde{\theta}_i, \hat{\theta}_{-i})$ under $(\tilde{\sigma}_i, \sigma^*_{-i})$, Rule 3 continues to hold and $g_i(\sigma^*_i(\tilde{\theta}_i), \sigma^*_{-i}(\hat{\theta}_{-i})) = \mathbf{w}_i$. On the other hand, if $i$ becomes $i^*$ at $(\tilde{\theta}_i, \hat{\theta}_{-i})$ under $(\tilde{\sigma}_i, \sigma^*_{-i})$, then still Rule 3 holds at $(\tilde{\theta}_i, \hat{\theta}_{-i})$ under $(\tilde{\sigma}_i, \sigma^*_{-i})$.[19] So, Rule 3 applies at $(\tilde{\theta}_i, \hat{\theta}_{-i})$ under $(\tilde{\sigma}_i, \sigma^*_{-i})$ implies $g_i(\tilde{\sigma}_i(\tilde{\theta}_i), \sigma^*_{-i}(\hat{\theta}_{-i})) = \mathbf{w}_i = g_i(\sigma^*_i(\tilde{\theta}_i), \sigma^*_{-i}(\hat{\theta}_{-i}))$.

If $i$ is $j^*$ at $(\tilde{\theta}_i, \hat{\theta}_{-i})$ under $\sigma^*$, then $g_i(\sigma^*_i(\tilde{\theta}_i), \sigma^*_{-i}(\hat{\theta}_{-i})) = \mathbf{0}$. As individual $i$ of type $\tilde{\theta}_i$ chooses a higher integer under $\tilde{\sigma}_i$, $i$ may be either still $j^*$, or be neither $j^*$ nor $i^*$, or become $i^*$ at $(\tilde{\theta}_i, \hat{\theta}_{-i})$ under $(\tilde{\sigma}_i, \sigma^*_{-i})$. If $i$ remains to be $j^*$ at $(\tilde{\theta}_i, \hat{\theta}_{-i})$ under $(\tilde{\sigma}_i, \sigma^*_{-i})$, then $g_i(\tilde{\sigma}_i(\tilde{\theta}_i), \sigma^*_{-i}(\hat{\theta}_{-i})) = \mathbf{0} = g_i(\sigma^*_i(\tilde{\theta}_i), \sigma^*_{-i}(\hat{\theta}_{-i}))$. If $i$ becomes neither $i^*$ nor $j^*$ at $(\tilde{\theta}_i, \hat{\theta}_{-i})$ under $(\tilde{\sigma}_i, \sigma^*_{-i})$, then (as Rule 3 applies at $(\tilde{\theta}_i, \hat{\theta}_{-i})$ under $\sigma^*$ by hypothesis) Rule 3 holds at at $(\tilde{\theta}_i, \hat{\theta}_{-i})$ under $(\tilde{\sigma}_i, \sigma^*_{-i})$; hence, $g_i(\tilde{\sigma}_i(\tilde{\theta}_i), \sigma^*_{-i}(\hat{\theta}_{-i})) = \mathbf{w}_i$ while $g_i(\sigma^*_i(\tilde{\theta}_i), \sigma^*_{-i}(\hat{\theta}_{-i})) = \mathbf{0}$. If $i$ becomes $i^*$ at $(\tilde{\theta}_i, \hat{\theta}_{-i})$ under $(\tilde{\sigma}_i, \sigma^*_{-i})$, then Rule 3 continues to hold at $(\tilde{\theta}_i, \hat{\theta}_{-i})$ under $(\tilde{\sigma}_i, \sigma^*_{-i})$; as an $i^*$ cannot be a $j^*$, $g_i(\tilde{\sigma}_i(\tilde{\theta}_i), \sigma^*_{-i}(\hat{\theta}_{-i})) = \mathbf{w}_i$ while $g_i(\sigma^*_i(\tilde{\theta}_i), \sigma^*_{-i}(\hat{\theta}_{-i})) = \mathbf{0}$.

By hypothesis, $\sigma^*$ is a BIE of $\mu$. Following the feasible unilateral deviation of $i$ from $\sigma^*_i$ to $\tilde{\sigma}_i$, we observe that the resulting IAA acts $\mathcal{g}^{\sigma^*}_{i,\tilde{\theta}_i}$ and $\mathcal{g}^{(\tilde{\sigma}_i, \sigma^*_{-i})}_{i,\tilde{\theta}_i}$ in $\mathcal{X}_i$ are such that there is a non-emtpy $\Theta'_{-i} \subset \Theta^*_{-i}(\tilde{\theta}_i)$ with $\mathcal{g}^{\sigma^*}_{i,\tilde{\theta}_i}(\theta'_{-i}) = \mathbf{0}$ whereas $\mathcal{g}^{(\tilde{\sigma}_i, \sigma^*_{-i})}_{i,\tilde{\theta}_i}(\theta'_{-i}) = \mathbf{w}_i$ for all $\theta'_{-i} \in \Theta'_{-i}$ while $\mathcal{g}^{\sigma^*}_{i,\tilde{\theta}_i}(\theta''_{-i}) = \mathcal{g}^{(\tilde{\sigma}_i, \sigma^*_{-i})}_{i,\tilde{\theta}_i}(\theta''_{-i})$ for all $\theta''_{-i} \in \Theta^*_{-i}(\tilde{\theta}_i) \setminus \Theta'_{-i}$.[20] Moreover, $\mathcal{g}^{\sigma^*}_{i,\tilde{\theta}_i}$ and $\mathcal{g}^{(\tilde{\sigma}_i, \sigma^*_{-i})}_{i,\tilde{\theta}_i}$ in $\mathcal{O}^{\mu}_i(\sigma^*_{-i})$. Therefore, by Condition M, $\mathcal{g}^{\sigma^*}_{i,\tilde{\theta}_i} \notin \mathcal{O}^{\mu}_i(\sigma^*_{-i})$. This contradicts $\sigma^*$ being a BIE of $\mu$, finishing the analysis of Case 1.

**Case 2.** *Rule 2' applies at $\tilde{\theta}$ under $\sigma^*$.*

Take $i \in N \setminus \{i^{\diamond}\}$ and consider the following unilateral deviation of $i$ of type $\tilde{\theta}_i$ that we denote by $\tilde{\sigma}_i$: $\tilde{\sigma}_i(\theta_i) = \sigma^*_i(\theta_i)$ for all $\theta_i \neq \tilde{\theta}_i$, while $\tilde{\sigma}_i(\tilde{\theta}_i) = (x^{*(i)}(\tilde{\theta}_i), p^{*(i)}(\tilde{\theta}_i), \alpha^{*(i)}_i(\tilde{\theta}_i), \tilde{k})$ where $\tilde{k} > \max_{j \in N \setminus \{i\}} k^{*(j)}_j(\tilde{\theta}_j)$. That is, $i$ of type $\tilde{\theta}_i$ (who is not $i^{\diamond}$ at $\tilde{\theta}$ under $\sigma^*$) deviates unilaterally only in her integer choice so that following this deviation $i$ becomes $i^*$ at $\tilde{\theta}$

[19]Observe that if $m_j = (\bar{x}, \bar{p}, \theta^{(j)}_j, k^{(j)})$ for all $j \in N \setminus \{i\}$, $(\theta^{(j)}_j)_{j \in N} \in \Theta^*$, $(\bar{p}, \bar{x})$ satisfies $(i), (ii), (iii)$ of Rule 1, $p^{(i)} = \bar{p}$, $x^{(i)} \neq \bar{x}$, $\mathcal{X}^{(i)}_{i,\bar{\theta}_i} \in \mathcal{B}^{\bar{\mathbf{w}}}_{i,\bar{\theta}_i}(\bar{p})$, for all $\theta_i \in \Theta_i$, $\bar{p}(\theta_i, \theta_{-i}) = \bar{p}(\theta_i, \theta'_{-i})$ for some $\theta_{-i}, \theta'_{-i} \in \Theta^*_{-i}(\theta_i)$ implies $\mathcal{X}^{(i)}_{i,\theta_i}(\theta_{-i}) = \mathcal{X}^{(i)}_{i,\theta_i}(\theta'_{-i})$, but $i$ is not $i^*$ (according to individuals' integer choices), then Rule 2' applies. So, if $i$ increases her integer and becomes $i^*$, then Rule 2 applies. This displays that Rule 2 are 2' are not reachable from Rule 3 via a unilateral deviation involving only the integer choice.

[20]The non-emptiness of $\Theta'_{-i}$ follows from the arguments made for $\tilde{\theta}$ at the start of the proof of Case 1: Note that $g_i(\tilde{\sigma}_i(\tilde{\theta}_i), \sigma^*_{-i}(\tilde{\theta}_{-i})) = \mathbf{w}_i$ while $g_i(\sigma^*_i(\tilde{\theta}_i), \sigma^*_{-i}(\tilde{\theta}_{-i})) = \mathbf{0}$; ergo, $\tilde{\theta}_{-i} \in \Theta'_{-i}$.

under $(\tilde{\sigma}_i, \sigma^*_{-i})$. Rule 3 applies at $\tilde{\theta}$ under $(\tilde{\sigma}_i, \sigma^*_{-i})$ because $(i)$ Rule 2' applies at $\tilde{\theta}$ under $\sigma^*$; $(ii)$ $i$ is not $i^\diamond$ at $\tilde{\theta}$ under $\sigma^*$; $(iii)$ $i^\diamond$ at $\tilde{\theta}$ under $\sigma^*$ chooses an SCA different from all the others (including $i$); and $(iv)$ $\tilde{\sigma}_i$ involves a unilateral deviation only concerning $i$'s integer choice at $\tilde{\theta}_i$. Thus, $g_i(\tilde{\sigma}_i(\tilde{\theta}_i), \sigma^*_{-i}(\tilde{\theta}_{-i})) = \mathbf{w}_i$ as clearly $i$ is not $j^*$ at $\tilde{\theta}$ under $(\tilde{\sigma}_i, \sigma^*_{-i})$. Meanwhile, $g_i(\sigma^*_i(\tilde{\theta}_i), \sigma^*_{-i}(\theta_{-i})) = \mathbf{0}$ by Rule 2' as $i \neq i^\diamond$ at $\tilde{\theta}$ under $\sigma^*$.

Rules 1, 2, 2', or 3 may apply at states $(\tilde{\theta}_i, \hat{\theta}_{-i})$ with $\hat{\theta}_{-i} \in \Theta^*_{-i}(\tilde{\theta}_i) \setminus \{\tilde{\theta}_{-i}\}$ under $\sigma^*$. The arguments in Subcases 1 – 4 of Case 1 continue to hold verbatim. Similar to Case 1, Condition M contradicts to $\sigma^*$ being a BIE of $\mu$, concluding the analysis of Case 2.

**Case 3.** *Rule 3 applies at $\tilde{\theta}$ under $\sigma^*$.*

Consider $i$ who is $j^*$ at $\tilde{\theta}$ under $\sigma^*$ and notice that $g_i(\sigma^*_i(\tilde{\theta}_i), \sigma^*_{-i}(\tilde{\theta}_{-i})) = \mathbf{0}$. The unilateral deviation of $i$ of type $\tilde{\theta}_i$, $\tilde{\sigma}_i$ is defined by $\tilde{\sigma}_i(\theta_i) = \sigma^*_i(\theta_i)$ for all $\theta_i \neq \tilde{\theta}_i$ while $\tilde{\sigma}_i(\tilde{\theta}_i) = (x^{*(i)}(\tilde{\theta}_i), p^{*(i)}(\tilde{\theta}_i), \alpha_i^{*(i)}(\tilde{\theta}_i), \tilde{k})$ where $\tilde{k} > \max_{j \in N \setminus \{i\}} k_j^{*(j)}(\tilde{\theta}_j)$. So, $i$ of type $\tilde{\theta}_i$ deviates unilaterally only in her integer choice, and following this deviation $i$ becomes $i^*$ at $\tilde{\theta}$ under $(\tilde{\sigma}_i, \sigma^*_{-i})$. Therefore, Rule 3 continues to hold at $\tilde{\theta}$ under $(\tilde{\sigma}_i, \sigma^*_{-i})$ (because an $i^*$ cannot be a $j^*$) and hence $g_i(\tilde{\sigma}_i(\tilde{\theta}_i), \sigma^*_{-i}(\tilde{\theta}_{-i})) = \mathbf{w}_i$ while $g_i(\sigma^*_i(\tilde{\theta}_i), \sigma^*_{-i}(\tilde{\theta}_{-i})) = \mathbf{0}$.

At any $(\tilde{\theta}_i, \hat{\theta}_{-i})$ with $\hat{\theta}_{-i} \in \Theta^*_{-i}(\tilde{\theta}_i) \setminus \{\tilde{\theta}_{-i}\}$, Rules 1, 2, 2', or 3 may apply under $\sigma^*$. As before, the arguments of Subcases 1 – 4 in Case 1 hold verbatim. Similar to Cases 1 and 2, Condition M is at odds with $\sigma^*$ being a BIE of $\mu$, concluding the analysis of Case 3.

This finishes the proof of Claim 2. ■

**Claim 3.** *For any $\sigma^*$ BIE of $\mu = (M, g)$, $g \circ \sigma^*$ is a CREE SCA of $\mathcal{E}$.*

**Proof of Claim 3.** Thanks to Claim 2, let $\sigma^*$ be a BIE such that Rule 1 applies at every $\theta \in \Theta^*$. We define $\sigma^*_i(\theta_i) = (x^{*(i)}(\theta_i), p^{*(i)}(\theta_i), \alpha_i^{*(i)}(\theta_i), k_i^{*(i)}(\theta_i))$ for all $i \in N$ all $\theta_i \in \Theta_i$. Notice that for any $\theta \in \Theta^*$ and any $i \in N$, $\mathcal{O}_i^\mu(\sigma^*_{-i}) = \mathcal{B}^{\bar{\mathbf{w}}}_{i,\theta_i}(\bar{p})$ where $(x^{*(i)}(\theta_i), p^{*(i)}(\theta_i)) = (\bar{x}, \bar{p})$ for some $(\bar{x}, \bar{p})$ satisfying the requirements featured in Rule 1. Then, by the defining properties of Rule 1, $\sigma^*$ being a BIE of $\mu$ under Rule 1 implies $(\bar{p}, \bar{x})$ is a CREE of $\mathcal{E}$. ■

This finishes the proof of the theorem. ■

## 9 Ex–Post Approach

In this section, we analyze behavioral implementation under incomplete information using an ex-post approach.

As highlighted in the robust mechanism design literature (see, e.g., Bergemann and Morris (2005, 2008, 2009, 2011), and Jehiel et al. (2006, 2008)), the appeal of ex-post equilibrium relies on the fact that in the rational domain, an ex-post equilibrium of an incomplete information game is a Bayesian Nash Equilibrium of that game for every belief profile. To reach a similar conclusion in behavioral domains, we need to relate individuals' ex-post choices with their interim choices.

We define individuals' ex-post choices as follows: Individual $i$'s *ex-post choice* at state $\theta$ is described by $C_i^\theta : \mathcal{X} \to \mathcal{X}$, such that $C_i^\theta(S) \subseteq S$ for all non-empty sets of alternatives $S \in \mathcal{X}$. The ex-post environment is summarized by $\mathcal{E}^{\text{ep}} := \langle N, X, \Theta, (C_i^\theta)_{i\in N,\ \theta\in\Theta}\rangle$. We assume that $\mathcal{E}^{\text{ep}}$ is common knowledge among the individuals and note that our setup allows (but does not depend on) individual ex-post choices to be interdependent.

**Definition 9.** *Given an ex-post environment $\mathcal{E}^{\text{ep}} = \langle N, X, \Theta, (C_i^\theta)_{i\in N,\theta\in\Theta}\rangle$, the associated interim environment $\mathcal{E} = \langle N, X, (\Theta_i)_{i\in N}, (\mathcal{C}_i^{\theta_i})_{i\in N,\ \theta_i\in\Theta_i}\rangle$ satisfies* **Property B** *if the following holds for each individual $i \in N$ and each of her type $\theta_i \in \Theta_i$: if for all non-empty $\tilde{\mathcal{A}}_i \subset \mathcal{A}_i$ and all $\boldsymbol{a}_i \in \tilde{\mathcal{A}}_i$, $\boldsymbol{a}_i(\theta'_{-i}) \in C_i^{(\theta_i,\theta'_{-i})}(\tilde{\mathcal{A}}_i(\theta'_{-i}))$ for all $\theta'_{-i} \in \Theta_{-i}$, then $\boldsymbol{a}_i \in \mathcal{C}_i^{\theta_i}(\tilde{\mathcal{A}}_i)$.*

Given an ex-post environment $\mathcal{E}^{\text{ep}}$, its associated interim counterpart $\mathcal{E}$ satisfies Property B if individuals' ex-post choices imply the following on individuals' interim choices: For any individual $i$ of type $\theta_i$ and any subset of her IAA acts, $\tilde{\mathcal{A}}_i$, if an IAA act $\boldsymbol{a}_i \in \tilde{\mathcal{A}}_i$ is such that for any one of others' type profile $\theta_{-i}$, alternative $\boldsymbol{a}_i(\theta_{-i})$ (the image of $\boldsymbol{a}_i$ at $\theta_{-i}$) is in $i$'s ex-post choice from the set of alternatives $\tilde{\mathcal{A}}_i(\theta_{-i})$ (alternatives sustained in projection by an IAA act in $\tilde{\mathcal{A}}_i$), then $\boldsymbol{a}_i$ is in $i$'s interim choice from $\tilde{\mathcal{A}}_i$.[21]

Mechanism $\mu = (M, g)$ induces a game of incomplete information in a given ex-post environment. In this environment, the relevant concept of opportunity sets involves alternatives rather than IAA acts: *Individual $i$'s opportunity set of alternatives under mechanism $\mu$ for a given message profile of other individuals $m_{-i} \in M_{-i}$* equals $O_i^\mu(m_{-i}) := \{g(m_i, m_{-i}) \in X \mid m_i \in M_i\}$.

The following presents the notion of EPE in ex-post environments:

**Definition 10.** *A strategy profile $\sigma^* : \Theta \to M$ is an* **ex-post equilibrium** *of $\mu$ if for each $\theta \in \Theta$, we have $g(\sigma^*(\theta)) \in C_i^\theta(O_i^\mu(\sigma^*_{-i}(\theta_{-i})))$ for all $i \in N$.*

In words, an EPE requires that the outcomes generated by the mechanism be an NE at every state of the world, while individuals' strategies have to be measurable with respect to only their own types.[22]

Under Property B, we obtain arguments similar to those of Bergemann and Morris (2008, 2011) and justify the use of EPE in behavioral domains. Every EPE of mechanism $\mu$ is a BIE of $\mu$. Therefore, no individual has any incentives to find out others' private

[21] We are grateful to an anonymous referee for pointing us towards Property B and suggesting its useful implications for our construction with ex-post choices.

[22] A message profile $m^* \in M$ is an NE of mechanism $\mu = (M, g)$ at $\theta$ if $g(m^*) \in \cap_{i\in N} C_i^\theta(O_i^\mu(m^*_{-i}))$.

| | $\theta_{-i}$ | $\tilde{\theta}_{-i}$ |
|---|---|---|
| $\boldsymbol{a}_i$ | $C_i^\theta(T)$ | $C_i^{(\theta_i,\tilde{\theta}_{-i})}(T \setminus \{x\})$ |
| $\boldsymbol{a}'_i$ | $C_i^\theta(T \setminus \{C_i^\theta(T)\})$ | $C_i^{(\theta_i,\tilde{\theta}_{-i})}(T \setminus \{x\})$ |
| $\boldsymbol{a}''_i$ | $C_i^{(\theta_i,\tilde{\theta}_{-i})}(T \setminus \{x\})$ | $C_i^\theta(T)$ |
| $\boldsymbol{a}'''_i$ | $x$ | $C_i^\theta(T)$ |
| $\boldsymbol{a}_i^{y,\tilde{y}}$ | $y$ | $\tilde{y}$ |

**Table 5:** An example in conjunction with Property B.

information at the interim stage. In other words, "no agent would like to change his message even if he were to know the true type profile of the remaining agents" (Bergemann & Morris, 2008), and hence EPE induces *robust* behavior on account of this ex-post no-regret property.

Notwithstanding, de Clippel (2020) presents a serious *warning* for the use of behavioral ex-post/dominant equilibrium in environments with probabilistically sophisticated individuals having singleton valued choices over alternatives: The failure of the IIA may be at odds with the plausibility of the ex-post/dominant equilibrium notion. The condition he analyzes is in the spirit of Savage's sure-thing principle and is systematically violated by choices that do not satisfy the IIA (and hence WARP). As Property B is related to this condition, in what follows, we discuss situations in which a contradiction along the lines of de Clippel (2020) may emerge in our behavioral setting.

To that regard, we construct an example mimicking the construction in the proof of de Clippel (2020, Theorem 1): Suppose that individuals' ex-post choices are singleton valued while the IIA does not hold for some individual's ex-post choices. Hence, there is an individual $i$, a state $\theta \in \Theta$, a non-empty set of alternatives $T \in \mathcal{X}$, and an alternative $x \in T \setminus C_i^\theta(T)$ such that $C_i^\theta(T) \neq C_i^\theta(T \setminus \{x\})$. Given $i$ and her type $\theta_i$, if there are two distinct type profiles of others, $\theta_{-i}, \tilde{\theta}_{-i} \in \Theta_{-i}$, then we can construct the following set of IAA acts: $\tilde{\boldsymbol{\mathcal{A}}}_i = \{\boldsymbol{a}_i, \boldsymbol{a}'_i, \boldsymbol{a}''_i, \boldsymbol{a}'''_i\} \bigcup \left(\cup_{y \in Y,\ \tilde{y} \in \tilde{Y}} \{\boldsymbol{a}_i^{y,\tilde{y}}\}\right)$ where these IAA acts are as specified in Table 5 and $Y, \tilde{Y} \in \mathcal{X}$ are as follows:

$$Y = T \setminus \left\{x, C_i^\theta(T), C_i^{(\theta_i,\tilde{\theta}_{-i})}(T \setminus \{x\}), C_i^\theta(T \setminus \{C_i^\theta(T)\})\right\},$$
$$\tilde{Y} = T \setminus \left\{x, C_i^\theta(T), C_i^{(\theta_i,\tilde{\theta}_{-i})}(T \setminus \{x\})\right\},$$

Meanwhile, we let $\hat{\boldsymbol{\mathcal{A}}}_i = \tilde{\boldsymbol{\mathcal{A}}}_i \setminus \{\boldsymbol{a}_i\}$. Then, we observe that $\tilde{\boldsymbol{\mathcal{A}}}_i(\theta_{-i}) = T$, $\tilde{\boldsymbol{\mathcal{A}}}_i(\tilde{\theta}_{-i}) = T \setminus \{x\}$, $\hat{\boldsymbol{\mathcal{A}}}_i(\theta_{-i}) = T \setminus \{C_i^\theta(T)\}$, and $\hat{\boldsymbol{\mathcal{A}}}_i(\tilde{\theta}_{-i}) = T \setminus \{x\}$. Thus, by Property B, $\boldsymbol{a}_i \in \mathcal{C}_i^{\theta_i}(\tilde{\boldsymbol{\mathcal{A}}}_i)$ as $\boldsymbol{a}_i(\theta_{-i}) = C_i^\theta(\tilde{\boldsymbol{\mathcal{A}}}_i(\theta_{-i}))$ and $\boldsymbol{a}_i(\tilde{\theta}_{-i}) = C_i^{(\theta_i,\tilde{\theta}_{-i})}(\tilde{\boldsymbol{\mathcal{A}}}_i(\tilde{\theta}_{-i}))$. Similarly, $\boldsymbol{a}'_i \in \mathcal{C}_i^{\theta_i}(\hat{\boldsymbol{\mathcal{A}}}_i)$ since $\boldsymbol{a}'_i(\theta_{-i}) = C_i^\theta(\hat{\boldsymbol{\mathcal{A}}}_i(\theta_{-i}))$ and $\boldsymbol{a}'_i(\tilde{\theta}_{-i}) = C_i^{(\theta_i,\tilde{\theta}_{-i})}(\hat{\boldsymbol{\mathcal{A}}}_i(\tilde{\theta}_{-i}))$.

We need the following additional requirements to reach a contradiction as in de Clippel (2020): Individual $i$ should perceive IAA acts $\boldsymbol{a}_i$ and $\boldsymbol{a}_i''$ to be equivalent to each other on grounds of $\boldsymbol{a}_i(\theta_{-i}) = \boldsymbol{a}_i''(\tilde{\theta}_{-i})$ and $\boldsymbol{a}_i(\tilde{\theta}_{-i}) = \boldsymbol{a}_i''(\theta_{-i})$. That is, when considering $\theta_{-i}$ and $\tilde{\theta}_{-i}$, only the underlying alternatives associated with these IAA acts matter for her. As a result, she perceives the IAA act that delivers $x'$ at $\theta_{-i}$ and $y'$ at $\tilde{\theta}_{-i}$ to be equivalent to another that provides $y'$ at $\theta_{-i}$ and $x'$ at $\tilde{\theta}_{-i}$ where $x', y' \in X$. We model $i$'s equivalence perception via an equivalence relation $\doteqdot$ defined on $\mathcal{A}_i$ and let $[\bar{\boldsymbol{a}}_i]_{\doteqdot} := \{\tilde{\boldsymbol{a}}_i \in \mathcal{A}_i \mid \bar{\boldsymbol{a}}_i \doteqdot \tilde{\boldsymbol{a}}_i\}$ be the equivalence class of $i$'s perception associated with $\bar{\boldsymbol{a}}_i \in \mathcal{A}_i$. For any set of IAA acts $\mathcal{A}_i' \subset \mathcal{A}_i$, the relation $\doteqdot$ partitions $\mathcal{A}_i'$ into equivalence classes. We assume that $i$ perceives two sets of IAA acts $\mathcal{A}_i'$ and $\mathcal{A}_i''$ as equivalent if the collection of equivalence classes in $\mathcal{A}_i$ that the IAA acts in $\mathcal{A}_i'$ and $\mathcal{A}_i''$ belong to are equal to one another. We denote such a case by $\mathcal{A}_i' \doteqdot \mathcal{A}_i''$. Formally, $\mathcal{A}_i^{(1)} \doteqdot \mathcal{A}_i^{(2)}$ if for all $\bar{\boldsymbol{a}}_i \in \mathcal{A}_i^{(k)}$, $[\bar{\boldsymbol{a}}_i]_{\doteqdot} \cap \mathcal{A}_i^{(\ell)} \neq \emptyset$ for all $k, \ell = 1, 2$. In furtherance, we assume that $i$'s interim choices from a set of IAA acts respect the resulting equivalence classes: For any set of IAA acts $\mathcal{A}_i'$, $\bar{\boldsymbol{a}}_i \in \mathcal{C}_i^{\theta_i}(\mathcal{A}_i')$ if and only if $\tilde{\boldsymbol{a}}_i \in \mathcal{C}_i^{\theta_i}(\mathcal{A}_i')$ for all $\tilde{\boldsymbol{a}}_i \in [\bar{\boldsymbol{a}}_i]_{\doteqdot} \cap \mathcal{A}_i'$. For example, such equivalence classes emerge under probabilistic sophistication with the use of lotteries when $i$'s belief is such that $\theta_{-i}$ and $\tilde{\theta}_{-i}$ are equally likely. Then, going back to our example, we see that $\boldsymbol{a}_i \doteqdot \boldsymbol{a}_i''$ and $\tilde{\mathcal{A}}_i \doteqdot \hat{\mathcal{A}}_i$. Recall that by Property B, $\boldsymbol{a}_i \in \mathcal{C}_i^{\theta_i}(\tilde{\mathcal{A}}_i)$ and $\boldsymbol{a}_i' \in \mathcal{C}_i^{\theta_i}(\hat{\mathcal{A}}_i)$. If, in addition, we have that for any set of IAA acts $\mathcal{A}_i'$, if $\bar{\boldsymbol{a}}_i \in \mathcal{C}_i^{\theta_i}(\mathcal{A}_i')$, then $\tilde{\boldsymbol{a}}_i \in \mathcal{C}_i^{\theta_i}(\mathcal{A}_i')$ implies $\tilde{\boldsymbol{a}}_i \in [\bar{\boldsymbol{a}}_i]_{\doteqdot} \cap \mathcal{A}_i'$ (i.e., the interim choices are singleton valued upto the equivalence classes), then the desired contradiction emerges: $\boldsymbol{a}_i \in \mathcal{C}_i^{\theta_i}(\tilde{\mathcal{A}}_i)$ implies $\boldsymbol{a}_i'' \in \mathcal{C}_i^{\theta_i}(\tilde{\mathcal{A}}_i)$ as $\boldsymbol{a}_i \doteqdot \boldsymbol{a}_i''$; further, $\tilde{\mathcal{A}}_i \doteqdot \hat{\mathcal{A}}_i$ implies $\boldsymbol{a}_i'' \in \mathcal{C}_i^{\theta_i}(\hat{\mathcal{A}}_i)$; but $\boldsymbol{a}_i'' \notin [\boldsymbol{a}_i']_{\doteqdot}$ while $\boldsymbol{a}_i' \in \mathcal{C}_i^{\theta_i}(\hat{\mathcal{A}}_i)$.

This discussion establishes that the planner has to take de Clippel's warning seriously when using EPE to design mechanisms with individuals whose ex-post choices over alternatives do not satisfy the IIA even if the environment does not feature probabilistic sophistication.

On the other hand, interesting behavioral settings often involve ex-post choices that fail the IIA.[23] In such environments and with incomplete information, as EPE is particularly suited to handle interdependencies, Property B may have a practical appeal as a starting point of analysis, especially when the aforementioned restrictions are not admissible. For example, if there are no two states that are equally likely in the sense of the perception equivalence elaborated above, or the interim choices are not unique up to the resulting

[23]See for example the rational shortlist method of Manzini and Mariotti (2007); the choice under status-quo bias analyzed in Samuelson and Zeckhauser (1988), Masatlioglu and Ok (2014), and Dean et al. (2017); the choice with attraction effect studied in Huber et al. (1982), Ok et al. (2015), and de Clippel and Eliaz (2012); choices of committees involving Condorcet cycles as in Hurwicz (1986); among other such behavioral settings.

equivalence classes, then implementation in EPE may be plausible.[24]

**Definition 11.** *We say that an SCS $F \in \mathcal{F}$ is* **ex-post implementable** *if there exists a mechanism $\mu$ such that*

*(i) for every $f \in F$, there exists an EPE $\sigma^*$ of $\mu$ that satisfies $f = g \circ \sigma^*$, and*

*(ii) for every EPE $\sigma^*$ of $\mu$, there exists $f \in F$ such that $g \circ \sigma^* = f$.*

We refer to an SCF $f$ as being *partially ex-post implementable* whenever condition $(i)$ in Definition 11 holds for $F = \{f\}$.

We show that the notion of *ex-post consistency* is necessary for ex-post implementation.

**Definition 12.** *A profile of sets of alternatives $\mathbf{S} := (S_i(f, \theta_{-i}))_{i \in N,\ f \in F,\ \theta_{-i} \in \Theta_{-i}}$ is* **ex-post consistent with the SCS** *$F \in \mathcal{F}$ if for every SCF $f \in F$,*

*(i) for all $i \in N$ and all $\theta'_i \in \Theta_i$, $f(\theta'_i, \theta_{-i}) \in C_i^{(\theta'_i, \theta_{-i})}(S_i(f, \theta_{-i}))$ for all $\theta_{-i} \in \Theta_{-i}$, and*

*(ii) for any deception profile $\alpha$ with $f \circ \alpha \notin F$, there exists $\theta^* \in \Theta$ and $i^* \in N$ such that $f(\alpha(\theta^*)) \notin C_{i^*}^{\theta^*}(S_{i^*}(f, \alpha_{-i^*}(\theta^*_{-i^*})))$.*

A profile of sets of alternatives $\mathbf{S}$ is ex-post consistent with an SCS $F$ if the following hold: $(i)$ Given any $i \in N$ and any $f \in F$ and any $\theta_{-i} \in \Theta_{-i}$, it must be that $i$'s ex-post choices when she is of type $\theta'_i$ at state $(\theta'_i, \theta_{-i})$ contains $f(\theta'_i, \theta_{-i})$ for all $\theta'_i \in \Theta_i$; $(ii)$ given any $f \in F$, whenever there is a deception profile $\alpha$ that leads to an outcome not compatible with the SCS, there exist an informant state $\theta^*$ and an informant individual $i^*$ such that $f(\alpha(\theta^*))$ is not in the ex-post choice of $i^*$ at $\theta^*$ from $S_{i^*}(f, \alpha_{-i^*}(\theta^*_{-i^*}))$.

If mechanism $\mu$ ex-post implements a given SCS $F \in \mathcal{F}$, then for any SCF $f \in F$, there is an EPE $\sigma^f$ of $\mu$ such that $f = g \circ \sigma^f$. Thus, for all $\theta \in \Theta$, $g(\sigma^f(\theta)) = f(\theta) \in \cap_{i \in N} C_i^{\theta}(O_i^{\mu}(\sigma^f_{-i}(\theta_{-i})))$. Defining $\mathbf{S}$ by $S_i(f, \theta_{-i}) := O_i^{\mu}(\sigma^f_{-i}(\theta_{-i}))$ with $i \in N$, $f \in F$, and $\theta_{-i} \in \Theta_{-i}$ implies $(i)$ of ex-post consistency of $\mathbf{S}$ with $F$. Meanwhile, if a deception profile $\alpha$ is such that $f \circ \alpha \notin F$, then $\sigma^f \circ \alpha$ cannot be an EPE of $\mu$; because otherwise, by $(ii)$ of ex-post implementability, there is $\tilde{f} \in F$ with $\tilde{f} = g \circ \sigma^f \circ \alpha$. But, since $f = g \circ \sigma^f$, $\tilde{f} = f \circ \alpha \in F$, a contradiction. So, there is a state $\theta^*$ and an individual $i^*$ whose ex-post choice at $\theta^*$ from $O_{i^*}^{\mu}(\sigma^f_{-i^*}(\alpha_{-i^*}(\theta_{-i^*})))$ (which equals $S_{i^*}(f, \alpha_{-i^*}(\theta_{-i^*})))$ does not include $f(\alpha(\theta^*))$. This delivers $(ii)$ of ex-post consistency of $\mathbf{S}$ with $F$. This discussion proves the following necessity result for implementation in EPE:

**Theorem 4.** *If an SCS $F \in \mathcal{F}$ is ex-post implementable, then there is a profile of sets of alternatives ex-post consistent with $F$.*

[24] We wish to note that implementation in EPE neither implies nor is implied by Nash implementation even in the rational domain. (Bergemann & Morris, 2008)

Next, to establish that our ex-post necessity result extends the analysis of Bergemann and Morris (2008) to behavioral domains, we show that our necessary condition implies *analogs* of theirs: ex-post choice monotonicity and quasi-ex-post choice incentive compatibility. Then, we display that under WARP, these conditions are equivalent to ex-post monotonicity and ex-post incentive compatibility of Bergemann and Morris (2008).

**Proposition 7.** *If there exists a profile of sets of alternatives ex-post consistent with an SCS $F \in \mathcal{F}$, then $F$ is* **ex-post choice monotonic***; i.e., for every SCF $f \in F$ and deception profile $\alpha$ with $f \circ \alpha \notin F$, there is a state $\theta^* \in \Theta$ and an individual $i^* \in N$ and a set of alternatives $S^* \in \mathcal{X}$ such that*

*(i)* $f(\alpha(\theta^*)) \notin C_{i^*}^{\theta^*}(S^*)$, *and*

*(ii)* $f(\theta'_{i^*}, \alpha_{-i^*}(\theta^*_{-i^*})) \in C_{i^*}^{(\theta'_{i^*}, \alpha_{-i^*}(\theta^*_{-i^*}))}(S^*)$ *for all* $\theta'_{i^*} \in \Theta_{i^*}$.

Proposition 7 directly follows from the existence of a profile of sets of alternatives that are ex-post consistent with the given SCS $F$: Given a profile of sets of alternatives $\mathbf{S} := (S_i(f, \theta_{-i}))_{i \in N,\ f \in F,\ \theta_{-i} \in \Theta_{-i}}$ ex-post consistent with $F$, let $S^* := S_{i^*}(f, \alpha_{-i^*}(\theta_{-i^*}))$. Then, $(i)$ of ex-post choice monotonicity follows from $(ii)$ of ex-post consistency while $(ii)$ of ex-post choice monotonicity follows from $(i)$ of ex-post consistency.

**Proposition 8.** *If there exists a profile of sets of alternatives ex-post consistent with an SCS $F \in \mathcal{F}$, then $F$ is* **quasi-ex-post choice incentive compatible***; i.e., for every SCF $f \in F$ and state $\theta \in \Theta$ and individual $i \in N$, there exists a set of alternatives $S \in \mathcal{X}$ such that $f(\theta) \in C_i^{\theta}(S)$ and $f(\Theta_i, \theta_{-i}) := \{f(\theta'_i, \theta_{-i}) \in X \mid \theta'_i \in \Theta_i\} \subseteq S$.*

To see the arguments needed to establish this result, let $\mathbf{S} := (S_i(f, \theta_{-i}))_{i \in N,\ f \in F,\ \theta_{-i} \in \Theta_{-i}}$ be a profile of sets of alternatives ex-post consistent with $F$ and set $S := S_i(f, \theta_{-i})$. By $(i)$ of ex-post consistency, $f(\theta) \in C_i^{\theta}(S)$ establishing the first condition of quasi-ex-post choice incentive compatibility. Since $f(\theta'_i, \theta_{-i}) \in C_i^{(\theta'_i, \theta_{-i})}(S_i(f, \theta_{-i}))$ for each $\theta'_i \in \Theta_i$ due to $(i)$ ex-post consistency, $f(\theta'_i, \theta_{-i}) \in S$ for each $\theta'_i \in \Theta_i$, establishing $f(\Theta_i, \theta_{-i}) \subseteq S$.

To analyze ex-post implementation in the rational domain, we denote the utility of individual $i \in N$ at state $\theta \in \Theta$ of alternative $x \in X$ by $u_i(x, \theta)$, and let $C_i^{\theta}(S) := \{y \in S : u_i(y, \theta) \geq u_i(x, \theta)\ \forall x \in S\}$ for any $S \in \mathcal{X}$. Then, the necessary conditions of Bergemann and Morris (2008) are defined as follows: An SCS $F$ is ***ex-post incentive compatible*** if for every $f \in F$, $u_i(f(\theta), \theta) \geq u_i(f(\theta'_i, \theta_{-i}), \theta)$ for all $i \in N$, all $\theta \in \Theta$, and all $\theta'_i \in \Theta_i$. Meanwhile, an SCS $F$ is ***ex-post monotonic*** if for every $f \in F$ and $\alpha$ with $f \circ \alpha \notin F$ there exist $i \in N$, $\theta \in \Theta$, and $y \in X$ such that

*(i)* $u_i(y, \theta) > u_i(f(\alpha(\theta)), \theta)$, and

*(ii)* $u_i(f(\theta'_i, \alpha_{-i}(\theta_{-i})), (\theta'_i, \alpha_{-i}(\theta_{-i}))) \geq u_i(y, (\theta'_i, \alpha_{-i}(\theta_{-i})))$ for all $\theta'_i \in \Theta_i$.

We now establish that under WARP, the necessary conditions of Bergemann and Morris (2008) are equivalent to our ex-post choice monotonicity coupled with quasi-ex-post choice incentive compatibility.

**Proposition 9.** *Under WARP, ex-post choice monotonicity coupled with quasi-ex-post choice incentive compatibility is equivalent to ex-post monotonicity coupled with ex-post incentive compatibility.*

**Proof of Proposition 9.** The result follows from Lemmas 1 and 2.

**Lemma 1.** *Under WARP, quasi-ex-post choice incentive compatibility is equivalent to ex-post incentive compatibility.*

**Proof.** Suppose that individuals' choices satisfy WARP. If $F$ is quasi-ex-post choice incentive compatible, then for all $f \in F$, all $\theta \in \Theta$, and all $i \in N$, there exists $S \in \mathcal{X}$ such that $f(\Theta_i, \theta_{-i}) \subset S$ and $f(\theta) \in C_i^{\theta}(S)$. Hence, the definition of $C_i^{\theta}$ under WARP implies $u_i(f(\theta), \theta) \geq u_i(f(\theta_i', \theta_{-i}), \theta)$ for all $\theta_i' \in \Theta_i$, i.e., $F$ is ex-post incentive compatible. Conversely, if $F$ is ex-post incentive compatible, then for all $f \in F$, all $\theta \in \Theta$, and all $i \in N$, $u_i(f(\theta), \theta) \geq u_i(f(\theta_i', \theta_{-i}), \theta)$ for all $\theta_i' \in \Theta_i$. Letting $S = f(\Theta_i, \theta_{-i})$ delivers the desired conclusion. ■

**Lemma 2.** *Under WARP, the following hold:*

*($i$) if an SCS $F$ is ex-post choice monotonic, then it is ex-post monotonic, and*

*($ii$) if an SCS $F$ is ex-post monotonic and ex-post incentive compatible, then it is ex-post choice monotonic.*

**Proof.** Suppose that individuals' choices satisfy WARP.

For ($i$), suppose that for all $f \in F$ and $\alpha$ with $f \circ \alpha \notin F$, there exist $i \in N$, $\theta \in \Theta$, and $S \in \mathcal{X}$ such that $f(\alpha(\theta)) \notin C_i^{\theta}(S)$ while $f(\theta_i', \alpha_{-i}(\theta_{-i})) \in C_i^{\theta_i', \alpha_{-i}(\theta_{-i})}(S)$. Then, let $y \in C_i^{\theta}(S)$. Then, by the definition of $C_i^{\theta}$ under WARP, we have $u_i(y, \theta) > u_i(f(\alpha(\theta)), \theta)$ and $u_i(f(\theta_i', \alpha_{-i}(\theta_{-i})), (\theta_i', \alpha_{-i}(\theta_{-i}))) \geq u_i(y, (\theta_i', \alpha_{-i}(\theta_{-i})))$ for all $\theta_i' \in \Theta_i$.

For ($ii$), suppose that for all $f \in F$ and $\alpha$ with $f \circ \alpha \notin F$, there exist $i \in N$, $\theta \in \Theta$, and $y \in X$ such that $u_i(y, \theta) > u_i(f(\alpha(\theta)), \theta)$ and $u_i(f(\theta_i', \alpha_{-i}(\theta_{-i})), (\theta_i', \alpha_{-i}(\theta_{-i}))) \geq u_i(y, (\theta_i', \alpha_{-i}(\theta_{-i})))$ for all $\theta_i' \in \Theta_i$. Let $S = f(\Theta_i, \alpha_{-i}(\theta_{-i})) \cup \{y\}$. Note that $f(\alpha(\theta)) \in S$ and (by the definition of $C_i^{\theta}$ under WARP) $f(\alpha(\theta)) \notin C_i^{\theta}(S)$. Since $F$ is ex-post incentive compatible by hypothesis, for all $\theta_i' \in \Theta_i$ we have that $u_i(f(\theta_i', \alpha_{-i}(\theta_{-i})), (\theta_i', \alpha_{-i}(\theta_{-i}))) \geq u_i(f(\tilde{\theta}_i, \alpha_{-i}(\theta_{-i})), (\theta_i', \alpha_{-i}(\theta_{-i})))$ for all $\tilde{\theta}_i \in \Theta_i$ and $u_i(f(\theta_i', \alpha_{-i}(\theta_{-i})), (\theta_i', \alpha_{-i}(\theta_{-i}))) \geq u_i(y, (\theta_i', \alpha_{-i}(\theta_{-i})))$. Hence, $f(\theta_i', \alpha_{-i}(\theta_{-i})) \in C_i^{\theta_i', \alpha_{-i}(\theta_{-i})}(S)$ for all $\theta_i' \in \Theta_i$. ■ ■

We need the following to establish our sufficiency result for implementation in EPE.

**Definition 13.** *The* **ex-post choice incompatible pair property** *holds in a given ex-post environment $\mathcal{E}^{\text{ep}}$ if for each state $\theta$ and each alternative $x \in X$, there exist individuals $i, j \in N$ with $i \neq j$ such that $x \notin C_i^\theta(X)$ and $x \notin C_j^\theta(X)$.*

Similar to its interim counterpart, this condition implies some level of disagreement among individuals regarding their ex-post choices at every state. It ensures that any alternative can be (ex-post) chosen by at most $n-2$ individuals at any state.

The ex-post choice incompatible pair property coupled with ex-post consistency is sufficient for implementation in EPE:

**Theorem 5.** *Suppose that the ex-post environment $\mathcal{E}^{\text{ep}}$ is such that $n \geq 3$ and ex-post choice incompatible pair property holds. Then, if there is a profile of sets of alternatives ex-post consistent with the SCS $F \in \mathcal{F}$, then $F$ is implementable in EPE.*

**Proof of Theorem 5.** Consider an SCS $F \in \mathcal{F}$ with which the profile of sets of alternatives $\mathbf{S} := (S_i(f, \theta_{-i}))_{i \in N,\ f \in F,\ \theta_{-i} \in \Theta_{-i}}$ is ex-post consistent. For any $i \in N$, $f \in F$, $\theta_{-i} \in \Theta_{-i}$, let $\bar{x}(i, f, \theta_{-i})$ be an arbitrary alternative in $S_i(f, \theta_{-i})$.

We use the following mechanism $\mu = (M, g)$: For each $i \in N$, her set of messages is $M_i = F \times \Theta_i \times X \times N$, while a generic message is denoted by $m_i = (f, \theta_i, x_i, k_i)$, and the outcome function $g : M \to X$ is as specified in Table 6.

| | | |
|---|---|---|
| **Rule 1** : | $g(m) = f(\theta)$ | if $m_i = (f, \theta_i, \cdot, \cdot)$ for all $i \in N$, |
| **Rule 2** : | $g(m) = \begin{cases} x_j & \text{if } x_j \in S_j(f, \theta_{-j}), \\ \bar{x}(j, f, \theta_{-j}) & \text{otherwise.} \end{cases}$ | if $m_i = (f, \theta_i, \cdot, \cdot)$ for all $i \in N \setminus \{j\}$ and $m_j = (\tilde{f}, \tilde{\theta}_j, x_j, \cdot)$ with $\tilde{f} \neq f$, |
| **Rule 3** : | $g(m) = x_j$ where $j = \sum_i k_i \pmod{\text{n}}$ | otherwise. |

**Table 6:** The outcome function of the mechanism.

First, we show that for any $f \in F$, there exists an EPE, $\sigma^f$, of $\mu = (M, g)$ such that $f = g \circ \sigma^f$. Take any $f \in F$, let $\sigma_i^f(\theta_i) = (f, \theta_i, x, 1)$ for each $i \in N$ and for some arbitrary $x \in \bar{X}$. By Rule 1, we have $g(\sigma^f(\theta)) = f(\theta)$ for each $\theta \in \Theta$, i.e., $f = g \circ \sigma^f$. Observe that for any unilateral deviation by individual $i$ from $\sigma^f$, either Rule 1 or Rule 2 applies, i.e., Rule 3 is not attainable by any unilateral deviation from $\sigma^f$. By construction, $O_i^\mu(\sigma_{-i}^f(\theta_{-i})) = S_i(f, \theta_{-i})$ for each $\theta \in \Theta$, $i \in N$. Since, by (i) of ex-post consistency, $f(\theta) \in C_i^\theta(S_i(f, \theta_{-i}))$ for each $i \in N$, we have for each $\theta \in \Theta$, $g(\sigma^f(\theta)) \in C_i^\theta(O_i^\mu(\sigma_{-i}^f(\theta_{-i})))$ for all $i \in N$, i.e., $\sigma^f$ is an EPE of $\mu$ such that $f = g \circ \sigma^f$.

Consider now any EPE $\sigma^*$ of $\mu$ denoted as $\sigma_i^*(\theta_i) = (f_i(\theta_i), \alpha_i(\theta_i), x_i(\theta_i), k_i(\theta_i))$ for each $i \in N$. That is, $f_i(\theta_i)$ denotes the SCF proposed by $i$ when her type is $\theta_i$; $\alpha_i(\theta_i)$, the reported type of $i$ when her type is $\theta_i$; $x_i(\theta_i)$, the alternative proposed by $i$ when her type is $\theta_i$; and $k_i(\theta_i)$, the number proposed by $i$ when her type is $\theta_i$.

Next, we show that, under any EPE $\sigma^*$ of $\mu$, Rule 1 must apply at each $\theta \in \Theta$: Suppose, for contradiction, that either Rule 2 or Rule 3 applies at some $\tilde{\theta} \in \Theta$ under $\sigma^*$. If Rule 2 applies at $\tilde{\theta}$, by construction, we have $O_j^{\mu}(\sigma^*_{-j}(\tilde{\theta}_{-j})) = S_j(f, \alpha_j(\tilde{\theta}_{-j}))$ for the odd-man-out $j \in N$ and $O_i^{\mu}(\sigma^*_{-i}(\tilde{\theta}_{-i})) = X$ for all $i \neq j$, i.e., for all the other $n-1$ individuals. On the other hand, if Rule 3 applies at $\tilde{\theta}$, we have, by construction, $O_i^{\mu}(\sigma^*_{-i}(\tilde{\theta}_{-i})) = X$ for all $i \in N$. Therefore, under both Rule 2 and Rule 3, at least $n-1$ individuals have the opportunity set $X$. Since $\sigma^*$ is an EPE of $\mu$, it follows that $g(\sigma^*(\tilde{\theta})) \in C_i^{\theta}(X)$ for at least $n-1$ individuals. This contradicts the ex-post choice incompatible pair property.

Moreover, under any EPE $\sigma^*$ of $\mu$, the product structure of $\Theta$, there is a unique $f \in F$ such that $f_i(\theta_i) = f$ for all $i \in N$ and for all $\theta_i \in \Theta_i$. Hence, by Rule 1, $g(\sigma^*(\theta)) = f(\alpha(\theta))$ for 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 $\sigma^*$, by construction, at each $\theta \in \Theta$, $O_i^{\mu}(\sigma^*_{-i}(\theta_{-i})) = S_i(f, \alpha_{-i}(\theta_{-i}))$ for all $i \in N$. If $f \circ \alpha \notin F$, then by $(ii)$ of ex-post consistency, there are $\theta^* \in \Theta$ and $i^* \in N$ such that $f(\alpha(\theta^*)) \notin C_{i^*}^{\theta^*}(S_{i^*}(f, \alpha_{-i^*}(\theta_{-i^*})))$. But this implies $g(\sigma^*(\theta^*)) \notin C_{i^*}^{\theta^*}(O_{i^*}^{\mu}(\sigma^*_{-i^*}(\theta^*_{-i^*})))$, a contradiction to $\sigma^*$ being an EPE of $\mu$. Thus, $f \circ \alpha \in F$. So, $g \circ \sigma^* = f \circ \alpha \in F$, which implies that condition $(ii)$ of ex-post implementability holds as well. ■

## 10 Concluding Remarks

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

Our focus is on full implementation in BIE because the revelation principle for partial implementation fails, and hence, one cannot restrict attention to direct mechanisms without a loss of generality. Consequently, we provide necessary as well as sufficient conditions for the implementation of SCSs in BIE. These help us analyze the implementability in BIE of behavioral versions of interim incentive efficiency and CREE. Further, we identify conditions characterizing instances when an SCF is implementable in BIE via its associated direct mechanism. Finally, we study the ex-post approach and inspect the full implementation of SCSs in EPE.

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


Ariely, D. (2009). *Predictably irrational: The hidden forces that shape our decisions.* Harper Perennial.

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

Barlo, M., & Dalkıran, N. A. (2022). Computational implementation. *Review of Economic Design*, 1–29.

Barlo, M., & Dalkıran, N. A. (2021). Implementation with missing data. *Mimeo*.

Benoit, J.-P., & Ok, E. A. (2006). Maskin's theorem with limited veto power. *Games and Economic Behavior*, *55*(2), 331–339.

Bergemann, D., & Morris, S. (2005). Robust mechanism design. *Econometrica*, *73*(6), 1771–1813.

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

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

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

Bierbrauer, F., & Netzer, N. (2016). Mechanism design and intentions. *Journal of Economic Theory*, *163*, 557–603.

Bochet, O. (2007). Switching from complete to incomplete information. *Journal of Mathematical Economics*, *43*(6), 735–748.

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.

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

Danilov, V. (1992). Implementation via Nash equilibria. *Econometrica*, 43–56.

Dean, M., Kıbrıs, Ö., & Masatlioglu, Y. (2017). Limited attention and status quo bias. *Journal of Economic Theory*, *169*, 93–127.

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

de Clippel, G. (2020). Incorporating risk in choice theory: Some observations. *Mimeo*.

de Clippel, G., & Eliaz, K. (2012). Reason-based choice: A bargaining rationale for the attraction and compromise effects. *Theoretical Economics*, *7*(1), 125–162.

Dutta, B., Sen, A., & Vohra, R. (1995). Nash implementation through elementary mechanisms in economic environments. *Economic Design*, *1*(1), 173–203.

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.

Huber, J., Payne, J. W., & Puto, C. (1982). Adding asymmetrically dominated alternatives: Violations of regularity and the similarity hypothesis. *Journal of Consumer Research*, *9*(1), 90–98.

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.

Jehiel, P., Meyer-ter Vehn, M., & Moldovanu, B. (2008). Ex-post implementation and preference aggregation via potentials. *Economic Theory*, *37*(3), 469–490.

Jehiel, P., Meyer-ter Vehn, M., Moldovanu, B., & Zame, W. R. (2006). The limits of ex post implementation. *Econometrica*, *74*(3), 585–610.

Kahneman, D. (2011). *Thinking, fast and slow.* Penguin Books.

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.

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

Kunimoto, T., & Saran, R. (2020). Robust implementation in rationalizable strategies in general mechanisms. *Mimeo*.

Kunimoto, T., Saran, R., & Serrano, R. (2021). Interim rationalizable implementation of functions. *Mimeo*.

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

Manzini, P., & Mariotti, M. (2007). Sequentially rationalizable choice. *American Economic Review*, *97*(5), 1824–1839.

Masatlioglu, Y., & Ok, E. A. (2014). A Canonical Model of Choice with Initial Endowments. *The Review of Economic Studies*, *81*(2), 851-883.

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.

Ok, E. A., Ortoleva, P., & Riella, G. (2015). Revealed (p)reference theory. *American Economic Review*, *105*(1), 299–321.

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.

Ray, K. T. (2018). Nash implementation under choice functions. *Mimeo*.

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.

Samuelson, W., & Zeckhauser, R. (1988). Status quo bias in decision making. *Journal of Risk and Uncertainty*, *1*(1), 7–59.

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.

Savage, L. J. (1972). *The foundations of statistics.* Courier Corporation.

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.

Thaler, R., & Sunstein, C. (2008). *Nudge: Improving decisions about health, wealth, and happiness.* Yale University Press.