Robust Behavioral Implementation*

Mehmet Barlo† Nuh Aygün Dalkıran‡

December 1, 2023

Abstract

This paper studies full behavioral implementation under incomplete information from a robust mechanism design perspective without requiring that individuals' ex-post and interim choices satisfy the weak axiom of revealed preferences. We employ *behavioral interim equilibrium* (BIE) and *behavioral ex-post equilibrium* (EPE) to derive necessary as well as sufficient conditions for *quasi-robust behavioral* and *ex-post behavioral implementation* of social choice sets. The former requires every optimal social choice function (SCF) be sustained as both a BIE and an EPE of a mechanism and that there be no 'bad' BIE of this mechanism resulting in an SCF not aligned with the social goal at hand. Meanwhile, the latter demands the optimal SCFs be sustained as EPE and that there be no 'bad' EPE in the mechanism. We also introduce and analyze *ex-post behavioral incentive efficiency* and identify sufficient conditions for its implementability.

Keywords: Behavioral Implementation, Incomplete Information, Ex-Post Implementation, Robustness, Ex-post Behavioral Incentive Efficiency.

JEL Classification: C79, D60, D79, D82, D90

---

*We would like to thank Tilman Börgers for his guidance and encouragement. We are also grateful to Alessandro Pavan for his valuable feedback. 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

Individuals are not perfect decision-makers. They often have difficulty processing information and making rational choices, as documented by behavioral sciences. By using the resulting insights, behavioral economics aims to help governments and other organizations design policies and institutions that are more effective in guiding people's behavior. In this context, we focus on how a planner can achieve her goals under incomplete information without relying extensively on individuals' beliefs when individuals are not necessarily making rational choices.

In this paper, we study full behavioral implementation under incomplete information from a robust mechanism design perspective without requiring individuals' ex-post and interim choices to satisfy the weak axiom of revealed preferences (WARP).

We employ *behavioral interim equilibrium* (BIE) and *behavioral ex-post equilibrium* (EPE) to derive necessary as well as sufficient conditions for *quasi-robust behavioral implementation* of social choice rules. This notion requires every optimal state-contingent alternative be sustained as both a BIE and an EPE of a mechanism and that there be no 'bad' BIE of this mechanism. Our paper can thus be regarded as the robust counterpart of Barlo and Dalkıran (2023), which investigates behavioral interim implementation under incomplete information without any ex-post considerations.

In incomplete information environments, each mechanism induces an incomplete information game where, given a strategy profile, each individual's message generates an interim act that maps each type profile of other individuals to alternatives. As a result, we obtain a general setup with incomplete information that allows for a wide variety of behavioral biases.

In this setup, individuals make their message choices in the mechanism at the interim stage (after observing their own private information). That is why BIE is appropriate when analyzing behavioral implementation under incomplete information. A BIE strategy profile requires that each type of each individual choose the act induced by the BIE from the corresponding opportunity set of acts implied by the mechanism and the strategy profile at hand. This notion of equilibrium is pertinent to environments with interim choices, not necessarily derived from preference maximization. In fact, BIE re-

duces to the Bayesian Nash equilibrium in the standard setting with rationality and Savage-Bayesian probabilistic sophistication. This observation also displays why BIE involves individuals' beliefs that are embedded in their interim choices. On the other hand, the notion of EPE requires a strategy profile in the mechanism be such that individuals' plans of actions are measurable with respect to their private information and result in (behavioral) Nash equilibrium play at every state. That is why EPE is robust to individuals' beliefs.[1]

Consequently, our notion of quasi-robust behavioral implementation is robust to individuals' beliefs when sustaining desirable social choice rules. That is, every desirable social choice rule is sustained in *belief-free* equilibrium behavior (an EPE strategy profile that is also BIE) that also features the *ex-post no regret property*: no individual has any incentive to go back to the interim stage and find out others' private information. Notwithstanding, our focus on *full* implementation and hence our need for dismissal of bad BIE implies that our implementation notion is naturally not entirely free of individuals' beliefs. Indeed, the planner may have to beware of equilibrium behavior sustained by some beliefs and make sure that even then, the resulting equilibrium is aligned with her desiderata.[2]

In our necessity result, Theorem 1, we show that if a mechanism quasi-robust behavioral implements a social choice set (SCS), then the resulting opportunity sets, acts achievable via individuals' unilateral deviations, form a profile of sets with some desirable properties, which we refer to as *quasi-robust consistency*. Each set appearing in this profile of sets of acts corresponds to an individual, a social choice function (SCF) in the SCS, and a deception profile of the other individuals. The first property of this profile is closedness under deception: acts generated via any deception from any set in the profile constitute another set in this profile. The second says that for any type of any individual, the act induced by the given SCF is among his interim choices from the

[1] In our paper, each individual's private information is exclusive and identifies his 'payoff type' but not his beliefs about others' payoff types (Penta, 2015; Barlo & Dalkıran, 2023).

[2] In general, full implementation of a collective goal via a mechanism can be seen as a combination of *partial implementation* (making sure that every desirable goal is obtained in equilibrium of that mechanism) and *weak implementation* (ensuring that every equilibrium of that mechanism results in one of the desirable goals). In that sense, quasi-robust behavioral implementation is belief-free in the partial implementation side but not in the weak implementation part.

set corresponding to others' truthful behavior.[3] Meanwhile, an analogous requirement holds for ex-post choices as well. These imply a quasi-incentive compatibility and a quasi-ex-post incentive compatibility condition (Proposition 2). Finally, a property akin to Maskin monotonicity emerges: Any deception turning a desirable SCF into one not aligned with the SCS creates a whistle-blower objecting to this deception. Our sufficiency result for quasi-robust implementation, Theorem 2, uses a mild condition that requires some level of disagreement in the society, *choice incompatibility*, in addition to quasi-robust consistency.

In behavioral environments of incomplete information, combining an individual's optimal ex-post choices across states does not necessarily induce optimality in the interim stage for that individual. We propose a condition, *Property STP**, in the spirit of Savage's sure-thing principle as well as Property STP of de Clippel (2022) in order to relate individuals' ex-post choices on alternatives to their interim choices on acts. It demands that an act be chosen from a set of acts whenever for any state, the realization of this act (an alternative) is ex-post chosen at that state from the set of alternatives sustained by acts in that set of acts. This property provides us with the practicality and tractability of the ex-post approach featured in the rational domain. Under Property STP*, every EPE is a BIE, and hence, quasi-robust behavioral implementation amounts to double implementation in BIE and EPE. Moreover, in such situations, EPE features the desirable robustness properties of its counterpart in the rational domain with Savage-Bayesian probabilistic sophistication where Property STP* holds.

While Property STP* instigates appealing aspects for EPE, it comes with a severe warning in environments with individuals' ex-post choices failing WARP.[4] In such environments, de Clippel (2022) exhorts us to be wary of the use of EPE because Property STP* and the failure of WARP for ex-post choices may generate a contradiction. We demonstrate situations in which such a contradiction may appear in our setup.

Next, we turn our attention to *ex-post behavioral implementation*: Each desirable goal needs to be achievable by an EPE of the mechanism, while every EPE of this mechanism

[3] A deception of an individual is a permutation on his type space and hence contains the identity function, which corresponds to his truthful behavior.

[4] We note that interim choices may satisfy Property STP* but fail WARP even if the associated ex-post choices obey WARP, as we display in the minimax-regret setting (Proposition 1).

has to result in an outcome aligned with the planner's goal. This notion is particularly well-suited for situations in which both of the following hold: (*i*) as in the minimax-regret setting, ex-post choices satisfy the independence of irrelevant alternatives (IIA) while interim choices may involve violations of WARP; (*ii*) the dismissal of bad EPE suffices, and the planner does not need to worry about bad BIE that is not EPE.

We obtain *ex-post consistency* as a necessary condition for ex-post behavioral implementation, Theorem 3, and show that it implies *quasi-ex-post incentive compatibility* (Proposition 5) along with *behavioral ex-post monotonicity* (Proposition 4). Furthermore, (in Proposition 6) we establish that under rationality, our quasi-ex-post incentive compatibility and behavioral ex-post monotonicity are equivalent to the ex-post incentive compatibility and ex-post monotonicity of Bergemann and Morris (2008). We also present a sufficiency result for ex-post behavioral implementation, Theorem 4, using the ex-post version of our choice incompatibility in addition to ex-post consistency.

To display an application of our findings, we analyze the implementability of *ex-post behavioral efficiency* and *ex-post behavioral incentive efficiency*. Under rationality, these notions are equivalent to ex-post Pareto efficiency and ex-post incentive Pareto efficiency of Holmström and Myerson (1983), respectively. First, we observe that in our setting, the well-known conflict between efficiency and incentive compatibility is still at play, and hence, ex-post behavioral efficiency is neither quasi-robust nor ex-post behavioral implementable. Indeed, we know from our necessity results that, at the very least, combining ex-post quasi-incentive compatibility with ex-post behavioral efficiency is necessary for implementation purposes. Doing so delivers our ex-post behavioral incentive efficiency, and we show that the associated opportunity sets constitute an ex-post consistent profile. Consequently, we obtain ex-post behavioral implementability of this SCS under ex-post choice incompatibility (Proposition 9). Moreover, we observe that ex-post behavioral incentive efficiency is not quasi-robust behavioral implementable in general. We identify sufficient conditions for its quasi-robust behavioral implementation (Proposition 10).

Our paper is mostly related to Barlo and Dalkıran (2023) that analyzes interim behavioral implementation under incomplete information *without any ex-post considerations*. In the rational domain, Bergemann and Morris (2009, 2011) analyze robust full implementation both using direct and indirect mechanisms. Unlike our notion of robust-

ness based on ex-post considerations to sustain belief-free evaluations in a behavioral environment, theirs is related to rationalizable implementation.[5] Another significant and closely related paper is Bergemann and Morris (2008), which analyzes ex-post implementation in the rational domain and under incomplete information. Our results on ex-post behavioral implementation extend their analysis to behavioral domains. In the rational domain, Jackson (1991) analyzes Bayesian implementation, which extends the seminal work of Maskin (1999) on Nash Implementation to incomplete information environments. Other seminal papers that analyze Bayesian implementation are Postlewaite and Schmeidler (1986) and Palfrey and Srivastava (1987). The equilibrium concept that we employ, BIE, is first introduced by Saran (2011), which analyzes partial behavioral implementation under incomplete information using menu-dependent preferences.

The necessary and almost-sufficient conditions we identify for quasi-robust and ex-post behavioral implementation are reminiscent of consistency of de Clippel (2014), which provides necessary as well as sufficient conditions for behavioral implementation under complete information. Other work on behavioral implementation under complete information include Hurwicz (1986), Eliaz (2002), Barlo and Dalkiran (2009), Korpela (2012), and Hayashi et al. (2023). Eliaz (2002) provides an analysis of full implementation when some of the individuals might be "faulty" and hence fail to act optimally. 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. On the other hand, Korpela (2012) shows that when individual choices fail rationality axioms, the IIA is key to obtaining the necessary and sufficient condition synonymous to that of Moore and Repullo (1990). Finally, Hayashi et al. (2023) provides an analysis of behavioral strong

[5]We note that iterative elimination of never-best responses may create complications due to the failure of the IIA in behavioral domains. There is a large literature on rationalizability-based robust mechanism design in the rational domain. Such studies include but are not limited to Bergemann and Morris (2005), Penta (2015), Bergemann and Morris (2017), Ollár and Penta (2017), de Clippel et al. (2019), Kunimoto and Serrano (2020), Chen et al. (2021), Chen et al. (2022), Jain and Lombardi (2022), Jain et al. (2022), Kunimoto and Saran (2022), Chen et al. (2023), Jain et al. (2023), Kunimoto et al. (2023), Xiong (2023). Bochet and Tumennasan (2021) studies interim implementation when individuals have benchmark strategies; focusing on direct mechanisms, their implementation amounts to "interim rationalizability with iterative elimination of strategies that are dominated by [the benchmark of] truthtelling."

implementation under complete information.[6]

The organization of the paper is as follows: Section 2 provides the preliminaries; Section 3, our analysis of quasi-robust behavioral implementation; Section 4, analysis of ex-post behavioral implementation. In Section 5, we analyze ex-post behavioral (incentive) efficiency. Section 6 concludes.

2 Preliminaries

Consider a set of individuals $N = \{1, \dots, n\}$ and a non-empty set of alternatives X where 2^X stands for the set of all subsets of X and $\mathcal{X}$ stands for those that are non-empty. Let Θ denote the set of all relevant states of the world regarding individuals' choices. We assume that there is incomplete information among the individuals regarding the true state of the world and their information is exclusive unless stated otherwise explicitly. Thus, Θ has a product structure, i.e., $\Theta = \times_{i \in N} \Theta_i$ where $\theta_i \in \Theta_i$ denotes the private information (type) of individual $i \in N$ at state $\theta = (\theta_1, \dots, \theta_n) \in \Theta$.

For any individual $i \in N$, an *interim act* sustained by $\Theta_{-i} := \times_{j \neq i} \Theta_j$ on X is $\mathbf{a}_i : \Theta_{-i} \to X$, a function mapping Θ_{-i} into X. We denote the set of all interim acts of individual i by $\mathbf{A}_i$. Meanwhile, $\mathbf{A}_i^{\mathrm{c}} := \cup_{\bar{x} \in X} \{\mathbf{a}_i^{\bar{x}} \in \mathbf{A}_i\}$ denotes the set of all *constant* acts where $\mathbf{a}_i^{\bar{x}}(\theta_{-i}) = \bar{x}$ for all $\theta_{-i} \in \Theta_{-i}$. The *image set associated with a set of acts* $\tilde{\mathbf{A}}_i \subset \mathbf{A}_i$ at θ_{-i} equals $\tilde{\mathbf{A}}_i(\theta_{-i}) := \{x \in X \mid \mathbf{a}_i(\theta_{-i}) = x \text{ for some } \mathbf{a}_i \in \tilde{\mathbf{A}}_i\}$. Given $i \in N$, her type $\theta_i \in \Theta_i$, and a non-empty subset of acts $\mathbf{S} \subset \mathbf{A}_i$, the *choice of individual i of type θ_i from the set of acts* $\mathbf{S}$ is given by $\mathbf{C}_i^{\theta_i}(\mathbf{S}) \subset \mathbf{S}$.

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

We define individuals' ex-post choices as follows: Individual i's *ex-post choice* at state θ is described by $c_i^\theta : \mathcal{X} \to 2^X$, such that $c_i^\theta(S) \subseteq S$ for all $S \in \mathcal{X}$. The ex-post environment is summarized by $\mathcal{E}^{\mathrm{ep}} := \langle N, X, \Theta, (c_i^\theta)_{i \in N,\ \theta \in \Theta} \rangle$. We assume that $\mathcal{E}^{\mathrm{ep}}$ is common knowledge among the individuals as well.

We impose *no restrictions* on choices such as WARP.[7] In particular, we allow individ-

[6] Some of the other related works include Kucuksenel (2012), Saran (2016), Barlo and Dalkıran (2022a, 2022b), Barlo et al. (2023), and Rubbini (2023).

[7] Sen (1971) shows that a choice correspondence satisfies WARP (and is represented by a complete and transitive preference relation) if and only if it satisfies independence of irrelevant alternatives (IIA) and an expansion consistency axiom (known as Sen's β). Letting $\mathcal{Z}$ be the set of all non-empty subsets

uals' interim and ex-post choices to be empty valued unless explicitly stated otherwise.

In our model, individuals' beliefs are embedded into their interim choices on acts.[8]

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

We focus on SCSs because a planner may consider many socially optimal SCFs simultaneously. An SCS F is a non-empty set of SCFs, i.e., $F \subset H$ and $F \neq \emptyset$; an SCF $f \in F$ specifies a socially optimal alternative for each state.[9]

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

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

We employ behavioral interim equilibrium, well-suited to environments with individuals' interim choices on acts[10]:

---

of alternatives, we say that the choice correspondence $c : \mathcal{Z} \to \mathcal{Z}$ satisfies (i) the IIA if $x \in S \cap c(T)$ for some $S, T \in \mathcal{Z}$ with $S \subset T$ implies $x \in c(S)$; (ii) Sen's β if $x, y \in S \subset T$ for some $S, T \in \mathcal{Z}$, and $x, y \in c(S)$ implies $x \in c(T)$ if and only if $y \in c(T)$.

[8]To illustrate this under rationality and Savage-Bayesian probabilistic sophistication, consider the following: Given an ex-post environment $\mathcal{E}^{\text{ep}}$, the ex-post choices of individuals can be captured by state-contingent utilities $(u_i(x \mid \theta))_{i \in N, \theta \in \Theta, x \in X}$ under rationality where $u_i(x \mid \theta)$ is the utility that individual i obtains from alternative x when the individuals' type profile is θ. Moreover, under the standard Savage-Bayesian formulation, the interim beliefs $(\pi_i(\theta_{-i} \mid \theta_i))_{i \in N, \theta_i \in \Theta_i, \theta_{-i} \in \Theta_{-i}}$ emerge where $\pi_i(\theta_{-i} \mid \theta_i) \in [0, 1]$ denotes the belief of individual i of type θ_i about the other individuals' type profile being θ_{-i}. Then, we obtain the corresponding interim environment $\mathcal{E}$ as follows: For any non-empty set of acts $\mathbf{S}$, the choice of individual i of type θ_i equals

$$\mathbf{C}_i^{\theta_i}(\mathbf{S}) = \left\{ \mathbf{a} \in \mathbf{S} \;\middle|\; \begin{array}{c} \sum_{\theta_{-i} \in \Theta_{-i}} \pi_i(\theta_{-i} \mid \theta_i) u_i(\mathbf{a}(\theta_{-i}) \mid \theta_i, \theta_{-i}) \geq \\ \sum_{\theta_{-i} \in \Theta_{-i}} \pi_i(\theta_{-i} \mid \theta_i) u_i(\mathbf{b}(\theta_{-i}) \mid \theta_i, \theta_{-i}) \text{ for all } \mathbf{b} \in \mathbf{S} \end{array} \right\}.$$

[9]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 the interested reader to Postlewaite and Schmeidler (1986), Palfrey and Srivastava (1987), Jackson (1991), and Bergemann and Morris (2008).

[10]See Saran (2011) and Barlo and Dalkıran (2023).

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

Intuitively, strategy profile σ^* is a BIE of μ if any individual i of any type θ_i chooses the interim act generated by the prescribed action, $\sigma_i^*(\theta_i)$, from his opportunity set of acts corresponding to others' strategy profile σ^*_{-i}.

In ex-post environment $\mathcal{E}^{\text{ep}}$, the relevant concept of opportunity sets of mechanism μ involves alternatives rather than interim acts: *Individual i's opportunity set of alternatives under mechanism* μ *for a given message profile of other individuals* $m_{-i} \in M_{-i}$ equals $O_i^{\mu}(m_{-i}) := \{g(m_i, m_{-i}) \in X \mid m_i \in M_i\}$. We note that for all $i \in N$, and all strategy profiles σ in mechanism μ, $O_i^{\mu}(\sigma_{-i}(\theta_{-i})) = \mathbf{O}_i^{\mu}(\sigma_{-i})(\theta_{-i})$ for all $\theta_{-i} \in \Theta_{-i}$ where $\mathbf{O}_i^{\mu}(\sigma_{-i})(\theta_{-i}) \in \mathcal{X}$ is the image set associated with the set of acts $\mathbf{O}_i^{\mu}(\sigma_{-i})$ at θ_{-i}.

We now present the definition of EPE of mechanism μ in ex-post environments:

Definition 2. *A strategy profile* $\sigma^* : \Theta \to M$ *is an* **ex-post equilibrium** *of* μ *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 the outcomes generated by the mechanism be a (behavioral) Nash equilibrium at every state of the world, while individuals' strategies have to be measurable with respect to only their own types.[11]

A natural way to relate ex-post choices of individuals to their interim choices is the following property akin to the sure-thing principle of Savage (1972):

Definition 3. *Given 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* $\mathcal{E} = \langle N, X, (\Theta_i)_{i\in N}, (\mathbf{C}_i^{\theta_i})_{i\in N,\ \theta_i\in\Theta_i}\rangle$ *satisfies* **Property STP*** *if the following holds for each individual* $i \in N$ *and each of his type* $\theta_i \in \Theta_i$: *if for all non-empty* $\tilde{\mathbf{A}}_i \subset \mathbf{A}_i$ *and all* $\mathbf{a}_i \in \tilde{\mathbf{A}}_i$, $\mathbf{a}_i(\theta'_{-i}) \in c_i^{(\theta_i,\theta'_{-i})}(\tilde{\mathbf{A}}_i(\theta'_{-i}))$ *for all* $\theta'_{-i} \in \Theta_{-i}$, *then* $\mathbf{a}_i \in \mathbf{C}_i^{\theta_i}(\tilde{\mathbf{A}}_i)$.

Given ex-post environment $\mathcal{E}^{\text{ep}}$, its associated interim counterpart $\mathcal{E}$ satisfies Property STP* if the following holds: For any individual i of any type θ_i and any subset of his acts, $\tilde{\mathbf{A}}_i$, if an act $\mathbf{a}_i \in \tilde{\mathbf{A}}_i$ is such that for any one of others' type profile θ_{-i}, alternative $\mathbf{a}_i(\theta_{-i})$ (the image of $\mathbf{a}_i$ at θ_{-i}) is in i's ex-post choice from the set of alternatives

[11] A message profile $m^* \in M$ is a (behavioral) Nash equilibrium of μ at θ if $g(m^*) \in \cap_{i\in N} c_i^{\theta}(O_i^{\mu}(m^*_{-i}))$.

$\tilde{\mathbf{A}}_i(\theta_{-i})$ (the image set associated with $\tilde{\mathbf{A}}_i$ at θ_{-i}), then $\mathbf{a}_i$ is in i's interim choice from $\tilde{\mathbf{A}}_i$. Property STP* is in the spirit of Savage's sure-thing principle and Property STP introduced by de Clippel (2022).[12]

Under Property STP*, 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 μ is a BIE of μ.[13] That is, Property STP* is *sufficient* for every EPE of a mechanism μ to be one of its BIE. Therefore, no individual has any incentive to find out others' private 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). 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. Hence, EPE induces *robust* behavior on account of these properties.

We note that Property STP* holds in the standard rational framework under Savage-Bayesian probabilistic sophistication. On the other hand, the minimax-regret preferences of Savage (1951) provide a setting in which the interim choices fail the IIA (and hence WARP), while the Property STP* is satisfied. Thus, the minimax-regret setting delivers an interesting behavioral environment where EPE is a plausible equilibrium notion.

In environments with the minimax-regret preferences, each type of each individual chooses the act that minimizes her maximum regret. The regret of individual i of type θ_i from act $\mathbf{a}_i$ at state $\theta = (\theta_i, \theta_{-i})$ equals the difference between the payoff i obtains and his maximum payoff in this state, i.e., $\max_{\mathbf{a}'_i \in \mathbf{S}_i}(u_i(\mathbf{a}'_i(\theta_{-i}) \mid (\theta_i, \theta_{-i})) - u_i(\mathbf{a}_i(\theta_{-i}) \mid (\theta_i, \theta_{-i}))$ where $u_i(x \mid \theta)$ denotes i's ex-post payoff from alternative x at θ. Hence, individual i of type θ_i weakly prefers act $\mathbf{a}_i$ to act $\tilde{\mathbf{a}}_i$ in a given set of acts $\mathbf{S}_i$ if

$$\begin{aligned} &\max_{\theta_{-i} \in \Theta_{-i}} \left[\max_{\mathbf{a}'_i \in \mathbf{S}_i} \left(u_i(\mathbf{a}'_i(\theta_{-i}) \mid (\theta_i, \theta_{-i})) - u_i(\mathbf{a}_i(\theta_{-i}) \mid (\theta_i, \theta_{-i})) \right) \right] \\ &\leq \max_{\theta_{-i} \in \Theta_{-i}} \left[\max_{\mathbf{a}''_i \in \mathbf{S}_i} \left(u_i(\mathbf{a}''_i(\theta_{-i}) \mid (\theta_i, \theta_{-i})) - u_i(\tilde{\mathbf{a}}_i(\theta_{-i}) \mid (\theta_i, \theta_{-i})) \right) \right]. \end{aligned} \tag{1}$$

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

[13] It is easy to see that in general, a BIE of a mechanism need not be one of its EPE regardless of whether or not Property STP* holds.

Proposition 1. *If ex-post environment $\mathcal{E}^{\text{ep}}$ and the associated interim environment $\mathcal{E}$ are related via minimax-regret preferences, then Property STP* holds.*

Proof. Suppose ex-post environment $\mathcal{E}^{\text{ep}} = \langle N, X, \Theta, (c_i^\theta)_{i\in N, \theta\in\Theta}\rangle$ and the associated interim environment $\mathcal{E} = \langle N, X, (\Theta_i)_{i\in N}, (\mathbf{C}_i^{\theta_i})_{i\in N,\ \theta_i\in\Theta_i}\rangle$ are related via minimax-regret preferences. That is, ex-post choices are represented by state-contingent utility functions —they are rational, and hence satisfy the IIA. Moreover, the interim choices can be represented as follows: For any pair of acts $\mathbf{a}_i$ and $\tilde{\mathbf{a}}_i$ in a given set of acts $\mathbf{S}_i$, individual i of type θ_i weakly prefers $\mathbf{a}_i$ to $\tilde{\mathbf{a}}_i$ in $\mathbf{S}_i$ if inequality (1) holds.

If for any individual i of type θ_i, we have $\mathbf{a}_i^*(\theta'_{-i}) \in c_i^{(\theta_i,\theta'_{-i})}(\tilde{\mathbf{A}}_i(\theta'_{-i}))$ for all $\theta'_{-i} \in \Theta_{-i}$ for some non-empty set of acts $\tilde{\mathbf{A}}_i$, then $\mathbf{a}_i^* \in \mathbf{C}_i^{\theta_i}(\tilde{\mathbf{A}}_i)$; i.e., Property STP* holds. This follows from $\mathbf{a}_i^*$ minimizing the maximum regret: $u_i(\mathbf{a}_i^*(\theta'_{-i}) \mid (\theta_i, \theta'_{-i})) = \max_{x\in\tilde{\mathbf{A}}_i(\theta'_{-i})} u_i(x \mid (\theta_i, \theta'_{-i}))$ for all θ'_{-i}. So, for all θ'_{-i}, $\max_{\mathbf{a}'_i\in\tilde{\mathbf{A}}_i}(u_i(\mathbf{a}'_i(\theta'_{-i}) \mid (\theta_i, \theta'_{-i})) - u_i(\mathbf{a}_i^*(\theta_{-i}) \mid (\theta_i, \theta_{-i}))) = 0$, i.e., i of type θ_i's maximum regret from $\mathbf{a}_i^*$ at (θ_i, θ'_{-i}) is 0. ■

Therefore, in the case of minimax-regret preferences, Property STP* holds even though the interim choices are not rational. In such behavioral environments, EPE is a plausible equilibrium notion because every EPE is a BIE.

Meanwhile, whether or not Property STP* is a *necessary* condition for every EPE of a mechanism being one of its BIE emerges as a natural question. Below, we show that the answer is negative by providing an example (inspired by de Clippel (2022)) that involves a mechanism where every EPE is a BIE, but Property STP* fails to hold.

Example 1. Let $N = \{1, 2\}$, $\Theta_1 = \{t_1\}$, $\Theta_2 = \{t_2^1, t_2^2\}$, $X = \{x, y, z\}$. We denote the act, $\mathbf{a}_1 : \Theta_2 \to X$, individual 1 faces by $\langle ab\rangle$ where $\mathbf{a}_1(t_2^1) = a$ and $\mathbf{a}_1(t_2^2) = b$ with $a, b \in \{x, y, z\}$. As there is only one type of individual 1, any act of individual 2 is merely an alternative, i.e., $\mathbf{A}_2 = X$.

Let the state-contingent payoffs and mechanism μ be as in Table 1.

Θ	(t_1, t_2^1)	(t_1, t_2^2)
$(u_1(x \mid \theta), u_2(x \mid \theta))$	$(2, 2)$	$(0, 1)$
$(u_1(y \mid \theta), u_2(y \mid \theta))$	$(0, 1)$	$(2, 2)$
$(u_1(z \mid \theta), u_2(z \mid \theta))$	$(1, 0)$	$(1, 0)$

		Ind. 2	
		L	R
Ind. 1	U	x	y
	D	y	x

Table 1: State-contingent payoffs and mechanism μ.

Suppose that individual 2 is rational whereas individual 1 is ambiguity averse in the sense that he chooses an action that maximizes his minimum possible payoff with respect

to the stage-contingent payoffs given in Table 1. Formally, for any two acts $\mathbf{a}_1$ and $\tilde{\mathbf{a}}_1$ in a given set of acts $\mathbf{S}_1 \subset \mathbf{A}_1$, individual 1 weakly prefers $\mathbf{a}_1$ to $\tilde{\mathbf{a}}_1$ if

$$\max_{\mathbf{a}_1 \in \mathbf{S}_1} \left[\min_{\theta_2 \in \Theta_2} \left(u_1(\mathbf{a}_1(\theta_2) \mid (t_1, \theta_2)) \right) \right] \geq \max_{\tilde{\mathbf{a}}_1 \in \mathbf{S}_1} \left[\min_{\theta_2 \in \Theta_2} \left(u_1(\tilde{\mathbf{a}}_1(\theta_2) \mid (t_1, \theta_2)) \right) \right].$$

Consider first the ex-post choices of individuals 1 and 2 given the stage-contingent payoffs in Table 1. Because $c_1^{(t_1,t_2^1)}(\{x,y\}) = \{x\}$, $c_1^{(t_1,t_2^2)}(\{x,y\}) = \{y\}$, $c_2^{(t_1,t_2^1)}(\{x,y\}) = \{x\}$, and $c_2^{(t_1,t_2^2)}(\{x,y\}) = \{y\}$, there are two EPEs of mechanism μ both inducing SCF $\langle xy \rangle$: $\sigma^{(*)}$ and $\sigma^{(**)}$ where $\sigma_1^{(*)}(t_1) = U$, $\sigma_2^{(*)}(t_2^1) = L$, and $\sigma_2^{(*)}(t_2^2) = R$; $\sigma_1^{(**)}(t_1) = D$, $\sigma_2^{(**)}(t_2^1) = R$, and $\sigma_2^{(**)}(t_2^2) = L$.

Next, considering the interim choices of individual 1, we see that $\mathbf{C}_1^{t_1}(\{\langle xy \rangle, \langle yx \rangle\}) = \{\langle xy \rangle, \langle yx \rangle\}$ as individual 1's minimum payoffs under $\langle xy \rangle$ and $\langle yx \rangle$ are both 0. Therefore, individual 1 is indifferent between choosing U or D in mechanism μ. On the other hand, because individual 2's interim choices over acts are such that $\mathbf{C}_2^{t_2^1}(\{\langle x \rangle, \langle y \rangle\}) = \{\langle x \rangle\}$ and $\mathbf{C}_2^{t_2^2}(\{\langle x \rangle, \langle y \rangle\}) = \{\langle y \rangle\}$, $\sigma^{(*)}$ and $\sigma^{(**)}$ are the only BIEs of mechanism μ.

Therefore, every EPE of μ is a BIE of μ.

Finally, we show that Property STP* fails given individual 1's interim and ex-post choices. To see why, consider interim choices of individual 1 over $\tilde{\mathbf{A}}_1 = \{\langle xy \rangle, \langle zz \rangle, \langle yx \rangle\}$. We have $\mathbf{C}_1^{t_1}(\tilde{\mathbf{A}}_1) = \{\langle zz \rangle\}$ as $\langle zz \rangle$ guarantees individual 1 a payoff of 1 whereas her minimum payoff under $\langle xy \rangle$ and $\langle yx \rangle$ is 0. Further, $\tilde{\mathbf{A}}_1(t_2^1) = \tilde{\mathbf{A}}_1(t_2^2) = \{x, y, z\}$ and ex-post choices of individual 1 is such that $c_1^{(t_1,t_2^1)}(\{x,y,z\}) = \{x\}$, $c_1^{(t_1,t_2^2)}(\{x,y,z\}) = \{y\}$, but $\langle xy \rangle \notin \mathbf{C}_1^{t_1}(\tilde{\mathbf{A}}_1) = \{\langle zz \rangle\}$. Hence, Property STP* fails to hold.

de Clippel (2022) presents a serious *warning* for the use of behavioral ex-post/dominant equilibrium in environments that involve ex-post choices failing rationality (in particular, with probabilistically sophisticated individuals having singleton valued choices over alternatives): The failure of the IIA is at odds with the plausibility of the ex-post/dominant equilibrium notion.[14]

The condition de Clippel (2022) analyzes, namely Property STP, is closely related to our Property STP* but restricted to probabilistic sophistication. Property STP is systematically violated when ex-post choices do not satisfy the IIA (and hence WARP). In Appendix A, we discuss situations in which a contradiction along the lines of de

[14]There are many interesting behavioral settings that involve ex-post choices that fail rationality in the sense that they fail the IIA. 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), de Clippel and Eliaz (2012), and Ok et al. (2015); choices of committees involving Condorcet cycles as in Hurwicz (1986); among other such behavioral settings.

Clippel (2022) may emerge in our behavioral setting: To justify the use of EPE, one needs to dismiss two states that are perceived to be equivalent or the interim choices being unique up to the resulting equivalence classes (see Appendix A for further details).

We note when ex-post choices satisfy the IIA, a contradiction à la de Clippel cannot arise (even if interim choices fail WARP). Indeed, minimax-regret preferences provide such a setting: the ex-post choices satisfy the IIA but the interim choices do not.

Any mechanism implementing an SCS under incomplete information must consider individuals' private information. However, individuals may be deceitful. We denote a *deception* by individual i as $\alpha_i : \Theta_i \to \Theta_i$. Intuitively, $\alpha_i(\theta_i)$ can be thought of as individual i's reported type. So, $\alpha(\theta) := (\alpha_1(\theta_1), \alpha_2(\theta_2), \ldots, \alpha_n(\theta_n))$ is a profile of possibly deceitful reported types while α^{id} denotes the *truthtelling profile*, i.e., $\alpha_i^{\text{id}}(\theta_i) = \theta_i$ for all $i \in N$ and all $\theta_i \in \Theta_i$. We denote individual i's set of all possible deceptions by Λ_i and let $\Lambda := \times_{i \in N} \Lambda_i$, $\Lambda_{-i} := \times_{j \neq i} \Lambda_j$, and $\alpha_{-i}(\theta_{-i}) := (\alpha_j(\theta_j))_{j \neq i}$. Moreover, a *garbling of an act* $\mathbf{a}_i$ that i of type θ_i faces when the other individuals use deception $\alpha_{-i} \in \Lambda_{-i}$ is the act $\mathbf{a}_i^\alpha := \mathbf{a}_i \circ \alpha_{-i}$. Similarly, a *garbling of an SCF* $h \in H$ that i of type θ_i faces when the others use deception $\alpha_{-i} \in \Lambda_{-i}$ is the act $\mathbf{h}_{i,\theta_i}^\alpha := \mathbf{h}_{i,\theta_i} \circ \alpha_{-i}$.

3 Quasi-robust Behavioral Implementation

In this paper, our main notion of implementation for behavioral environments under incomplete information is as follows:

Definition 4. *Given a pair of associated interim and ex-post environments, an SCS F is* **quasi-robust behavioral implementable** *if there is a mechanism μ such that*

(i) for all $f \in F$, there is σ^f that is both a BIE and an EPE of μ with $g \circ \sigma^f = f$, and

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

Part (i) of quasi-robust behavioral implementation demands that every desirable SCF is sustained by a belief-free interim equilibrium (that is both a BIE and an EPE), regardless of whether or not Property STP* holds. Indeed, in implementation frameworks, sustaining desirable SCFs independently of individuals beliefs has been commended by the robust mechanism design literature. Further, quasi-robust behavioral implementation features the ex-post no-regret property: as the strategy profile is both a BIE and an EPE, no individual has any incentive to go back to the interim stage and find out others'

private information. On the other hand, (ii) of quasi-robust behavioral implementation requires that if a strategy profile is a BIE of μ then the corresponding SCF has to be desirable. We say that SCS F is *partially quasi-robust behavioral implementable* if (i) of Definition 4 holds while if (ii) of Definition 4 holds, then SCS F is *weakly quasi-robust behavioral implementable*.

As every EPE is a BIE under Property STP*, SCS F being quasi-robust behavioral implementable implies that for every $f \in F$, there is an EPE that sustains it and there is no BIE that is inconsistent with SCS F.[15] Hence, under Property STP*, our implementation notion is belief-free on the partial implementation side but not on the weak implementation part. Below, we exemplify quasi-robust behavioral implementation in such a setting —with minimax regret preferences, where Property STP* holds, the ex-post choices are rational, but the interim choices violate WARP.

Example 2. Let $N = \{1, 2\}$, $X = \{x, y, z\}$, $\Theta_i = \{t_i, t_i'\}$ for both $i = 1, 2$. Therefore, there are four possible states of the world, i.e., $\Theta = \{t_1t_2, t_1't_2, t_1t_2', t_1't_2'\}$. Table 2 details the rational interdependent ex-post preferences, and Table 3 specifies the corresponding payoffs for the minimax regret setting.

$R_{1,(t_1,t_2)}$	$R_{2,(t_1,t_2)}$	$R_{1,(t_1',t_2)}$	$R_{2,(t_1',t_2)}$	$R_{1,(t_1,t_2')}$	$R_{2,(t_1,t_2')}$	$R_{1,(t_1',t_2')}$	$R_{2,(t_1',t_2')}$
x	x	z	z	z	z	y	y
z	z	x	x	y	y	z	z
y	y	y	y	x	x	x	x

Table 2: Ex-post preferences

	$u_{1,(t_1,t_2)}(\cdot)$	$u_{2,(t_1,t_2)}(\cdot)$	$u_{1,(t_1',t_2)}(\cdot)$	$u_{2,(t_1',t_2)}(\cdot)$	$u_{1,(t_1,t_2')}(\cdot)$	$u_{2,(t_1,t_2')}(\cdot)$	$u_{1,(t_1',t_2')}(\cdot)$	$u_{2,(t_1',t_2')}(\cdot)$
x	1	1	0	0	-1	-1	-1	-1
y	$-1+\eta$	-1	-1	-1	0	0	1	1
z	0	0	1	$1-\varepsilon$	1	$1-\tilde{\varepsilon}$	0	0

Table 3: Ex-post payoffs

Using the construction that we present in Section 5, we see that in this example, the only ex-post behavioral incentive efficient SCF f is given by $\langle xzzy \rangle$.

The interim choices fail WARP whenever $\eta > 0$: Consider individual 1 of type t_1 and note that her choices from the choice sets $\{\langle xx \rangle, \langle yz \rangle, \langle zy \rangle\}$ (implying regret figures of

[15] In the rational domain, where Property STP* comes for free, our quasi-robust behavioral implementation corresponds to double implementation in BIE and EPE.

$(0,2)$, $(2-\eta,0)$, and $(1,1)$, resp.) and $\{\langle yz\rangle,\langle zy\rangle\}$ (resulting in regret figures of $(0,1)$ and $(1-\eta,0)$, resp.) equal $\{\langle zy\rangle\}$ and $\{\langle yz\rangle\}$, respectively.

An intuitive interpretation of our example involves a headquarters (HQ, the planner) of a firm consisting of two subdivisions (individuals) $i=1,2$. The subdivisions are located in two separate countries. The HQ needs to extract the state pertaining to the economic outlook of country i from each division separately. Each subdivision is informed of its country's state of the economy but not that of the other. Meanwhile, each country i's state is either 'good' or 'bad' where t_i denotes the former and t_i' the latter. There are three possible firm-wide policies (alternatives) the HQ is to adopt: *expansion*, *contraction*, and *prudence*, denoted by x, y, and z, respectively.

As subdivisions are parts of the same organization, their state-contingent (rational) ex-post preferences equal one another at each state resulting in a common-value-like setting with interdependent preferences. In particular, if the both countries' states are good, each subdivision ranks *expansion* strictly over *prudence* and *prudence* strictly over *contraction*; if both countries' states are bad, each strictly prefers *contraction* to *prudence* and *prudence* to *expansion*. On the other hand, if the economic state of one country is good and the other is bad, then each strictly top-ranks *prudence* while (i) if country 2 (involving a bigger market when compared to that of country 1) is in a good state, and country 1's is bad, then each strictly ranks *expansion* over *contraction*; (ii) otherwise, each strictly prefers *contraction* to *prudence*. Country-specific idiosyncratic shocks result in the corresponding ex-post payoffs specified in Table 3.

After observing their own country's economic outlook, each subdivision evaluates the interim acts via the minimax regret preferences (following 'the regret minimization framework', popularized by the Amazon CEO Jeff Bezos).[16]

Intuitively, the HQ's state-contingent goal, $F=\{f\}$, involves *expansion* if both countries' states are good, *contraction* if both countries' states are bad, and *prudence* for every other possible situation.

The following (direct) mechanism quasi-robust behavioral implements $F=\{f\}$ in this example whenever $\varepsilon,\tilde{\varepsilon}>0$: As we concentrate on the direct mechanism, the de-

		Individual 2	
		t_2	t_2'
Individual 1	t_1	x	z
	t_1'	z	y

Table 4: The mechanism that quasi-robust behavioral implements $F=\{f\}$.

ception profiles are in one-to-one correspondence with interim strategy profiles. Consequently, Table 5 establishes that the truthtelling strategy profile, α^{id}, is the only BIE

[16]See https://www.huffingtonpost.co.uk/rosie-leizrowice/no-regrets_b_17640296.html.

and EPE of the direct mechanism whenever $\varepsilon, \tilde{\varepsilon} > 0$. Under α^{id}, every type of every individual reveals their type truthfully. Therefore, individual i of type t_i faces act $\langle xz \rangle$ and individual i of type t'_i faces $\langle zy \rangle$. Hence, maximum regret obtained from truthtelling equals 0 for every type of every individual. Ergo, α^{id} is both a BIE and an EPE. To exemplify that there is no other BIE strategy profile, consider $\alpha^{(7)}$, where both types of individual 1 claims to be of type t_1 while both types of individual 2 claims to be of type t'_2. Thus, SCF $\langle zzzz \rangle$ emerges as z is the resulting alternative at every state. Consequently, by conforming to $\alpha^{(7)}$, individual 2 of type t_2 (who is to claim to be of type t'_2) obtains the act $\langle zz \rangle$. The maximum regret of this act equals $\max\{1, 0\} = 1$. Individual 2 of type t_2 deviating to truthtelling, she obtains the act $\langle xx \rangle$, which implies a maximum regret of $\max\{0, 1 - \varepsilon\} = 1 - \varepsilon$. So, individual 2 of type t_2 has a strictly profitable deviation under $\alpha^{(7)}$ and hence can serve as the informant for this deception.

	$\alpha_1(t_1)$	$\alpha_1(t'_1)$	$\alpha_2(t_2)$	$\alpha_2(t'_2)$	$f \circ \alpha$	Informant	Conforms	Max. regret	Deviates	Max. regret
α^{id}	t_1	t'_1	t_2	t'_2	$\langle xzzy \rangle$	–		0		
$\alpha^{(2)}$	t_1	t'_1	t_2	t_2	$\langle xzxz \rangle$	$(2, t'_2)$	$\langle xz \rangle$	$2 - \tilde{\varepsilon}$	$\langle zy \rangle$	0
$\alpha^{(3)}$	t_1	t'_1	t'_2	t'_2	$\langle zyzy \rangle$	$(2, t_2)$	$\langle zy \rangle$	$2 - \varepsilon$	$\langle xz \rangle$	0
$\alpha^{(4)}$	t_1	t'_1	t'_2	t_2	$\langle zyxz \rangle$	$(2, t'_2)$	$\langle xz \rangle$	$2 - \tilde{\varepsilon}$	$\langle zy \rangle$	0
$\alpha^{(5)}$	t_1	t_1	t_2	t'_2	$\langle xxzz \rangle$	$(1, t'_1)$	$\langle xz \rangle$	1	$\langle zy \rangle$	0
$\alpha^{(6)}$	t_1	t_1	t_2	t_2	$\langle xxxx \rangle$	$(1, t_1)$	$\langle xx \rangle$	2	$\langle zz \rangle$	1
$\alpha^{(7)}$	t_1	t_1	t'_2	t'_2	$\langle zzzz \rangle$	$(2, t_2)$	$\langle zz \rangle$	1	$\langle xx \rangle$	$1 - \varepsilon$
$\alpha^{(8)}$	t_1	t_1	t'_2	t_2	$\langle zzxx \rangle$	$(2, t'_2)$	$\langle xx \rangle$	$2 - \tilde{\varepsilon}$	$\langle zz \rangle$	0
$\alpha^{(9)}$	t'_1	t'_1	t_2	t'_2	$\langle zzyy \rangle$	$(1, t_1)$	$\langle zy \rangle$	1	$\langle xz \rangle$	0
$\alpha^{(10)}$	t'_1	t'_1	t_2	t_2	$\langle zzzz \rangle$	$(2, t'_2)$	$\langle zz \rangle$	1	$\langle yy \rangle$	$1 - \tilde{\varepsilon}$
$\alpha^{(11)}$	t'_1	t'_1	t'_2	t'_2	$\langle yyyy \rangle$	$(1, t_1)$	$\langle yy \rangle$	1	$\langle zz \rangle$	0
$\alpha^{(12)}$	t'_1	t_1	t'_2	t_2	$\langle yyzz \rangle$	$(2, t_2)$	$\langle yy \rangle$	$2 - \varepsilon$	$\langle zz \rangle$	0
$\alpha^{(13)}$	t'_1	t_1	t_2	t'_2	$\langle zxyz \rangle$	$(1, t_1)$	$\langle zy \rangle$	1	$\langle xz \rangle$	0
$\alpha^{(14)}$	t'_1	t_1	t_2	t_2	$\langle zxzx \rangle$	$(1, t'_1)$	$\langle xx \rangle$	1	$\langle zz \rangle$	0
$\alpha^{(15)}$	t'_1	t_1	t'_2	t'_2	$\langle yzyz \rangle$	$(1, t_1)$	$\langle yy \rangle$	1	$\langle zz \rangle$	0
$\alpha^{(16)}$	t'_1	t_1	t'_2	t_2	$\langle yzzx \rangle$	$(2, t_2)$	$\langle yz \rangle$	1	$\langle zx \rangle$	$1 - \varepsilon$

Table 5: Deception profiles and corresponding informants

On the other hand, if $\varepsilon = 0$ (or $\tilde{\varepsilon} = 0$), then we see from Table 5 that $\alpha^{(7)}$ (resp. $\alpha^{(10)}$) is a bad BIE of the direct mechanism and it induces SCF $\langle zzzz \rangle$, which is not aligned with ex-post incentive efficiency.

In what follows, we establish that any (direct or indirect) mechanism $\mu = (M, g)$ sustaining SCF f in EPE (and hence in BIE, thanks to Property STP*) has a bad BIE that induces SCF $\langle zzzz \rangle$ whenever $\varepsilon = 0$ or $\tilde{\varepsilon} = 0$. This is because of the following: Let σ^f be an EPE of μ such that $g \circ \sigma^f = f$, and $\varepsilon = 0$. Consider deception $\alpha^{(7)}$. Then, $f \circ \alpha^{(7)} = \langle zzzz \rangle$ and $\langle zz \rangle \in \mathbf{O}_1^{\mu}(\sigma_2^f \circ \alpha_2^{(7)}) \cap \mathbf{O}_2^{\mu}(\sigma_1^f \circ \alpha_1^{(7)})$. Thus, $\sigma^f \circ \alpha^{(7)}$ is a bad BIE of μ sustaining SCF $f \circ \alpha^{(7)} = \langle zzzz \rangle$.[17] A similar argument follows for the case

[17] To see why observe that individual 1 of type t_1 obtains maximum regret of 0 by confirming to $\alpha^{(7)}$.

with $\tilde{\varepsilon} = 0$ by using deception $\alpha^{(10)}$. Therefore, $F = \{f\}$ is not quasi-robust behavioral implementable by any mechanism whenever $\min\{\varepsilon, \tilde{\varepsilon}\} = 0$.[18]

We observe that quasi-robust behavioral implementation continues to feature its desirable properties even when Property STP* does not hold and the environment is not immune to de Clippel's critique. To present a simple illustration of quasi-robust behavioral implementation in such a setting, we revisit Example 1: Consider SCS $F = \{\langle xy \rangle\}$ and recall that in that example, every EPE of μ is a BIE of μ even though Property STP* fails to hold. There are two EPEs of mechanism μ, $\sigma^{(*)}$ and $\sigma^{(**)}$, both inducing SCF $\langle xy \rangle$. Moreover, $\sigma^{(*)}$ and $\sigma^{(**)}$ are the only BIEs of mechanism μ. Therefore, mechanism μ quasi-robust behavioral implements F.

3.1 Necessity

Below, we introduce a quasi-robust consistency notion and show that it is necessary for quasi-robust behavioral implementation. To do that we need the following formalities: A quasi-robust behavioral implementable SCS F necessitates a mechanism μ where for each SCF f in F, there exists a strategy profile σ^f a BIE (and an EPE) of μ that generates f. This implies σ^f_{-i} induces an opportunity set of acts for individual i from which each type θ_i of i chooses at the interim stage $\mathbf{f}_{i,\theta_i}$, the act induced by f that i of type θ_i faces. However, any individual $j \in N$ may behave as if she is of type $\alpha_j(\theta_j)$ when her true type is θ_j thereby creating deception $\alpha_j : \Theta_j \to \Theta_j$. So, when the others employ deception profile α_{-i}, individual i of type θ_i faces the act induced by $f \circ \alpha$, i.e., $\mathbf{f}^{\alpha}_{i,\theta_i}$, the garbling of $\mathbf{f}_{i,\theta_i}$, which has to be in i's opportunity set of acts induced by $\sigma^f_{-i} \circ \alpha_{-i}$.

As a result, one obtains the notion of closedness under deception pertaining to implementation under incomplete information.[19] Given an SCS F, a profile of sets of acts $\mathbb{S} := (\mathbf{S}_i(f, \alpha_{-i}))_{i \in N,\ f \in F,\ \alpha_{-i} \in \Lambda_{-i}}$ is **closed under deception** if $\mathbf{a}_i \in \mathbf{S}_i(f, \alpha_{-i})$ implies $\mathbf{a}_i^{\tilde{\alpha}} \in \mathbf{S}_i(f, \tilde{\alpha}_{-i} \circ \alpha_{-i})$ for all $i \in N$, all $f \in F$, and all $\alpha, \tilde{\alpha} \in \Lambda_{-i}$.

Meanwhile, individual 1 of type t_1' obtains maximum regret of 1 by confirming to $\alpha^{(7)}$ and maximum regret of 2 if she deviates and sustains $\langle yy \rangle$. Further, individual 2 of type t_2 obtains maximum regret of 1 from both confirming to and deviating from $\alpha^{(7)}$ thanks to $\varepsilon = 0$. Finally, individual 2 of type t_2' obtains maximum regret of 0 by confirming to $\alpha^{(7)}$.

[18] In Appendix B, we characterize situations in which direct mechanisms quasi-robust behavioral implement given SCFs. In behavioral domains featuring the failure of the IIA, direct mechanisms may lose their applicability not only because of the existence of bad interim equilibria but also due to the possible failure of the revelation principle (Saran, 2011).

[19] See also Barlo and Dalkıran (2023).

Definition 5. *Given a pair of associated interim and ex-post environments, a profile of sets of acts* $\mathbb{S} := (\mathbf{S}_i(f, \alpha_{-i}))_{i \in N,\ f \in F,\ \alpha_{-i} \in \Lambda_{-i}}$ *is* **quasi-robust consistent with SCS** F *if it is closed under deception and for every SCF* $f \in F$*,*

(*i*) *for all* $i \in N$ *and all* $\theta_i \in \Theta_i$*,*

  a) $\mathbf{f}_{i,\theta_i} \in \mathbf{C}_i^{\theta_i}(\mathbf{S}_i(f, \alpha_{-i}^{\mathrm{id}}))$*, and*

  b) $f(\theta_i, \theta_{-i}) \in c_i^{(\theta_i, \theta_{-i})}(\mathbf{S}_i(f, \alpha_{-i}^{\mathrm{id}})(\theta_{-i})))$ *for all* $\theta_{-i} \in \Theta_{-i}$*, and*

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

A profile of sets of acts $\mathbb{S} := (\mathbf{S}_i(f, \alpha_{-i}))_{i \in N,\ f \in F,\ \alpha_{-i} \in \Lambda_{-i}}$ is quasi-robust consistent with SCS F if $\mathbb{S}$ is closed under deception and for each $f \in F$, the following hold: (i) Given any $i \in N$ and any $\theta_i \in \Theta_i$, (a) i's interim choices when she is of type θ_i from $\mathbf{S}_i(f, \alpha_{-i}^{\mathrm{id}})$ (the set of acts in $\mathbb{S}$ associated with i, f, and the truthtelling profile of the other individuals) contain the act associated with f that she faces, namely $\mathbf{f}_{i,\theta_i}$; (b) i's ex-post choices at θ from the set of alternatives in the image of $\mathbf{S}_i(f, \alpha_{-i}^{\mathrm{id}})$ at θ_{-i} contain $f(\theta_i, \theta_{-i})$ for all $\theta_{-i} \in \Theta_{-i}$; and (ii) if there is a deception profile α that leads to an outcome not compatible with the SCS, i.e., $f \circ \alpha \notin F$, then there exists an informant individual i^* of type $\theta_{i^*}^*$ who does not choose the garbling of $\mathbf{f}_{i^*,\theta_{i^*}^*}$ that i^* of type $\theta_{i^*}^*$ faces when the others use deception α_{-i^*}, namely $\mathbf{f}_{i^*,\theta_{i^*}^*}^{\alpha}$, from the set of acts in $\mathbb{S}$ corresponding to i^*, f, and α_{-i^*}, i.e., $\mathbf{S}_{i^*}(f, \alpha_{-i^*})$.

Theorem 1. *Given a pair of associated interim and ex-post environments, if an SCS* F *is quasi-robust behavioral implementable, then there is a profile of sets of acts that is quasi-robust consistent with* F*.*

Proof. Let $\mu = (M, g)$ quasi-robust behavioral implement SCS F. Then, for any SCF $f \in F$, there is σ^f a BIE and an EPE of μ such that $f = g \circ \sigma^f$. Let us define $\mathbb{S}$ by $\mathbf{S}_i(f, \alpha_{-i}) := \mathbf{O}_i^{\mu}(\sigma_{-i}^f \circ \alpha_{-i})$ for each $i \in N$, $f \in F$, and $\alpha_{-i} \in \Lambda_{-i}$.

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

As for each $i \in N$ and $\theta_i \in \Theta_i$, the act associated with f that i of type θ_i faces, $\mathbf{f}_{i,\theta_i}$, is in $\mathbf{C}_i^{\theta_i}(\mathbf{O}_i^{\mu}(\sigma_{-i}^f))$, $(i.a)$ of quasi-robust consistency of $\mathbb{S}$ holds since σ^f is a BIE of μ

such that $f = g \circ \sigma^f$ while $\sigma^f_{-i} \circ \alpha^{\mathrm{id}}_{-i} = \sigma^f_{-i}$ implies that $\mathbf{O}^\mu_i(\sigma^f_{-i}) = \mathbf{S}_i(f, \alpha^{\mathrm{id}}_{-i})$. Further, for all $\theta_{-i} \in \Theta_{-i}$, $\mathbf{O}^\mu_i(\sigma^f_{-i})(\theta_{-i}) = \mathbf{S}_i(f, \alpha^{\mathrm{id}}_{-i})(\theta_{-i})$, and σ^f being an EPE of μ implies that $f(\theta_i, \theta_{-i}) \in c_i^{(\theta_i,\theta_{-i})}(\mathbf{S}_i(f, \alpha^{\mathrm{id}}_{-i})(\theta_{-i})))$ for all $\theta_{-i} \in \Theta_{-i}$ because $O^\mu_i(\sigma_{-i}(\theta_{-i})) = \mathbf{O}^\mu_i(\sigma_{-i})(\theta_{-i})$, delivering $(i.b)$ of quasi-robust consistency of $\mathbb{S}$.

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

Next, we establish that quasi-robust consistency implies not only an interim quasi-incentive compatibility but also a quasi-ex-post incentive compatibility:

Definition 6. *Given a pair of associated interim and ex-post environments, SCS F is*

(i) **quasi-incentive compatible** *if for every SCF $f \in F$ and individual $i \in N$, there is a set of acts $\mathbf{S} \subset \mathbf{A}_i$ with $\{\mathbf{f}_{i,\tilde{\theta}_i} \mid \tilde{\theta}_i \in \Theta_i\} \subset \mathbf{S}$ and $\mathbf{f}_{i,\theta_i} \in \mathbf{C}^{\theta_i}_i(\mathbf{S})$ for all $\theta_i \in \Theta_i$;*

(ii) **quasi-ex-post incentive compatible** *if for every SCF $f \in F$, state $\theta \in \Theta$, and individual $i \in N$, there is a set of alternatives $S \in \mathcal{X}$ such that $f(\theta) \in c^\theta_i(S)$ and $f(\Theta_i, \theta_{-i}) \subseteq S$ where $f(\Theta_i, \theta_{-i}) := \{f(\theta'_i, \theta_{-i}) \in X \mid \theta'_i \in \Theta_i\}$.*

Quasi-incentive compatibility and quasi-ex-post incentive compatibility of an SCS F follows from the existence of a quasi-robust consistent profile of sets of acts $\mathbb{S}$ given by $(\mathbf{S}_i(f, \alpha_{-i}))_{i \in N,\ f \in F,\ \alpha_{-i} \in \Lambda_{-i}}$: For any given $f \in F$ and any $i \in N$, let $\mathbf{S} = \mathbf{S}_i(f, \alpha^{\mathrm{id}}_{-i})$. Then, for all $i \in N$, $\{\mathbf{f}_{i,\tilde{\theta}_i} \mid \tilde{\theta}_i \in \Theta_i\} \subset \mathbf{S}$, and $(i.a)$ of quasi-robust consistency implies $\mathbf{f}_{i,\theta_i} \in \mathbf{C}^{\theta_i}_i(\mathbf{S})$ for all $\theta_i \in \Theta_i$. Moreover, for any given $f \in F$ and any $i \in N$ and $\theta \in \Theta$, let $S = \mathbf{S}_i(f, \alpha^{\mathrm{id}}_{-i})(\theta_{-i})$. Then, by $(i.b)$ of quasi-robust consistency , $f(\theta_i, \theta_{-i}) \in c_i^{(\theta_i,\theta_{-i})}(\mathbf{S}_i(f, \alpha^{\mathrm{id}}_{-i})(\theta_{-i})))$ for all $\theta_{-i} \in \Theta_{-i}$. Hence, $f(\theta) \in c^\theta_i(S)$ and $f(\Theta_i, \theta_{-i}) \subseteq S$. This proves the following result:

Proposition 2. *Given a pair of associated interim and ex-post environments, if there exists a profile of sets of acts quasi-robust consistent with SCS F, then F satisfies quasi-incentive compatibility and quasi-ex-post incentive compatibility.*

Now, we show that under Property STP*, every quasi-ex-post incentive compatible SCS is quasi-incentive compatible.

Proposition 3. *Given a pair of associated interim and ex-post environments satisfying Property STP*, if SCS F is quasi-ex-post incentive compatible, then it is quasi-incentive compatible.*

Proof. Suppose that given ex-post environment $\mathcal{E}^{\text{ep}}$, its associated interim environment $\mathcal{E}$ satisfies Property STP*. Let SCS F be quasi-ex-post incentive compatible and so for any SCF $f \in F$, $i \in N$, and $\theta_{-i} \in \Theta_{-i}$, there is $S^f_{i,\theta_{-i}} \in \mathcal{X}$ such that for any given $\theta_i \in \Theta_i$, $f(\theta_i, \theta_{-i}) \in c_i^{(\theta_i,\theta_{-i})}(S^f_{i,\theta_{-i}})$, which implies $f(\Theta_i, \theta_{-i}) \subseteq S^f_{i,\theta_{-i}}$. Fix the profile of sets $(S^f_{i,\theta_{-i}})_{i\in N, f\in F, \theta_{-i}\in\Theta_{-i}}$ and for each $f \in F$ and $i \in N$, define $\mathbf{S}^f_i \subset \mathbf{A}_i$ by $\mathbf{S}^f_i := \{\mathbf{a}_i \in \mathbf{A}_i \mid \mathbf{a}_i(\theta_{-i}) \in S^f_{i,\theta_{-i}} \text{ for all } \theta_{-i} \in \Theta_{-i}\}$. Then, as for any $\tilde{\theta}_i \in \Theta_i$, $\mathbf{f}_{i,\tilde{\theta}_i}(\theta_{-i}) = f(\tilde{\theta}_i, \theta_{-i}) \in S^f_{i,\theta_{-i}}$ for all $\theta_{-i} \in \Theta_{-i}$, we conclude $\{\mathbf{f}_{i,\tilde{\theta}_i} \mid \tilde{\theta}_i \in \Theta_i\} \subset \mathbf{S}^f_i$. Furthermore, by Property STP*, $\mathbf{f}_{i,\theta_i} \in \mathbf{C}^{\theta_i}_i(\mathbf{S}^f_i)$ for all $\theta_i \in \Theta_i$ because $\mathbf{f}_{i,\theta_i}(\theta_{-i}) = f(\theta) \in c^{\theta}_i(S^f_{i,\theta_{-i}})$ and $\mathbf{S}^f_i(\theta_{-i}) = S^f_{i,\theta_{-i}}$ for all $\theta_{-i} \in \Theta_{-i}$. Thus, F is quasi-incentive compatible as well. ■

3.2 Sufficiency

The quasi-robust behavioral implementation of an SCS F is impossible without a profile of sets of acts that conforms to quasi-robust consistency with F. In what follows, we explore additional requirements to be imposed on these profiles to ensure sufficiency.[20]

Definition 7. *The* **choice incompatibility** *holds in an interim environment if the following holds: If for any SCF $h \in H$ and any $\bar{\theta} \in \Theta$, a profile of sets of acts $(\tilde{\mathbf{A}}_i)_{i\in N}$ is such that*

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

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

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

The choice incompatibility mandates the following: Consider any SCF h, state $\bar{\theta}$, and profile of sets of acts $(\tilde{\mathbf{A}}_i)_{i\in N}$ such that (i) for every individual i, the act induced by h corresponding to i's type $\bar{\theta}_i$, $\mathbf{h}_{i,\bar{\theta}_i}$, is contained in $\tilde{\mathbf{A}}_i$, and (ii) with the exception of one outlier $\bar{j}$ in N, for all $i \neq \bar{j}$, the set of all alternatives supported by an act in $\tilde{\mathbf{A}}_i$ when considering others' types, $\tilde{\mathbf{A}}_i(\bar{\theta}_{-i})$ equals X. Consequently, there exists an individual i^*, other than $\bar{j}$, who opts not to choose $\mathbf{h}_{i^*,\bar{\theta}_{i^*}}$ at their type $\bar{\theta}_{i^*}$ from his set of acts $\tilde{\mathbf{A}}_{i^*}$.

[20]There is scope for other sufficient conditions since we do not impose universal axioms to restrict choices. However, closing the gap between necessary and sufficient conditions appears to be impractical.

This condition necessitates some disagreement in individuals' assessments of SCFs under some circumstances: If for SCF h, state $\bar{\theta}$, and profile of acts $(\tilde{\mathbf{A}}_i)_{i\in N}$, $\mathbf{h}_{i,\bar{\theta}_i} \in \mathbf{C}_i^{\bar{\theta}_i}(\tilde{\mathbf{A}}_i)$ for every i in N, then (i) of choice incompatibility is met by default; thus, under choice incompatibility, there cannot be $\bar{j}$ in N with $\tilde{\mathbf{A}}_i(\bar{\theta}_{-i}) = X$ for all $i \neq \bar{j}$.

Below, we establish that an SCS is quasi-robust behavioral implementable in societies with at least three members if both the choice incompatibility and quasi-robust consistency conditions hold.

Theorem 2. *Suppose that the given pair of associated interim and ex-post environments is such that $n \geq 3$ and the choice incompatibility holds. If there is a profile of sets of acts quasi-robust consistent with SCS F, then F is quasi-robust behavioral implementable.*

Proof of Theorem 2. Suppose that $n \geq 3$ and the choice incompatibility holds. Let F be an SCS for which $\mathbb{S} := (\mathbf{S}_i(f, \alpha_{-i}))_{i\in N,\ f\in F,\ \alpha_{-i}\in\Lambda_{-i}}$ is quasi-robust consistent.

The mechanism $\mu = (M, g)$ we employ is as in Barlo and Dalkıran (2023), and the proof is a variation on the proof of Theorem 2 of that study[21]: For each $i \in N$, $M_i = F \cup \{\emptyset\} \times \Theta_i \times \mathbf{A}_i \times X \times N$, while a generic message is denoted by $m_i = (m_i^1, \theta_i^{(i)}, \mathbf{a}_i^{(i)}, x^{(i)}, k^{(i)})$, and the outcome function $g : M \to X$ is as in Table 6.

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

Table 6: The outcome function of the mechanism for Theorem 2.

In words, each individual i is required to send a message that consists of five components. The first component specifies either an SCF $f^{(i)} \in F$ or a flag denoted by $\emptyset$, the second a type of herself $\theta_i^{(i)} \in \Theta_i$, the third an act $\mathbf{a}_i^{(i)} \in \mathbf{A}_i$, the fourth an alternative $x^{(i)} \in X$, and the fifth a number $k^{(i)} \in N = \{1, 2, \ldots, n\}$.

First, we show that condition (i) of quasi-robust behavioral implementation holds.

[21] Variations of this canonical mechanism has been used in the implementation literature; see Repullo (1987), Saijo (1988), Moore and Repullo (1990), Jackson (1991), Maskin (1999), Korpela (2013), de Clippel (2014), Koray and Yildiz (2018), and Altun et al. (2023).

Claim 1. *For any $f \in F$, there is σ^f a BIE and an EPE of $\mu = (M, g)$ with $f = g \circ \sigma^f$.*

Proof. Take any $f \in F$, let $\sigma_i^f(\theta_i) = (f, \theta_i, \mathbf{f}_{i,\theta_i}, \bar{x}, 1)$ for each $i \in N$ and some $\bar{x} \in X$. Then, Rule 1 applies and we have $g(\sigma^f(\theta)) = f(\theta)$ for each $\theta \in \Theta$, i.e., $f = g \circ \sigma^f$.

For any unilateral deviation of individual i from σ^f, either Rule 1 or Rule 2 applies, while Rule 3 is not attainable. Hence, $\mathbf{O}_i^\mu(\sigma_{-i}^f) = \mathbf{S}_i(f, \alpha_{-i}^{\mathrm{id}})$ for all $i \in N$.

Recall that, by $(i.a)$ of quasi-robust consistency , $\mathbf{f}_{i,\theta_i} \in \mathbf{C}_i^{\theta_i}(\mathbf{S}_i(f, \alpha_{-i}^{\mathrm{id}}))$ for each $i \in N$ and $\theta_i \in \Theta_i$. Thus, for all $i \in N$ and all $\theta_i \in \Theta_i$, $\mathbf{h}_{i,\theta_i}^* \in \mathbf{C}_i^{\theta_i}(\mathbf{O}_i^\mu(\sigma_{-i}^f))$ where $\mathbf{h}_{i,\theta_i}^*(\theta_{-i}) = g(\sigma_i^f(\theta_i), \sigma_{-i}^f(\theta_{-i}))$ for all $\theta_{-i} \in \Theta_{-i}$. So, σ^f is a BIE of μ such that $f = g \circ \sigma^f$.

Furthermore, by $(i.b)$ of quasi-robust consistency , $f(\theta_i, \theta_{-i}) \in c_i^{(\theta_i, \theta_{-i})}(\mathbf{S}_i(f, \alpha_{-i}^{\mathrm{id}})(\theta_{-i})))$ for all $\theta_{-i} \in \Theta_{-i}$ because $O_i^\mu(\sigma_{-i}(\theta_{-i})) = \mathbf{O}_i^\mu(\sigma_{-i})(\theta_{-i}) = \mathbf{S}_i(f, \alpha_{-i}^{\mathrm{id}})(\theta_{-i})$. Thus, σ^f is a EPE of μ as well. ■

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

Next, we show that, due to the choice incompatibility, Rule 1 applies at every state.

Claim 2. *Under any BIE σ^* of μ, Rule 1 applies at every $\theta \in \Theta$. Moreover, under any BIE σ^* of μ, there is a unique $f \in F$ such that $m_i^1(\theta_i) = f$ for all $i \in N$ and all $\theta_i \in \Theta_i$.*

Proof. Suppose for a contradiction that either Rule 2 or Rule 3 applies under σ^* at $\bar{\theta}$ and consider $(\tilde{\mathbf{A}}_i)_{i \in N}$ with $\tilde{\mathbf{A}}_i := \mathbf{O}_i^\mu(\sigma_{-i}^*)$ for all $i \in N$. Let SCF $h^* := g \circ \sigma^*$. Then, $\mathbf{h}_{i,\bar{\theta}_i}^* \in \tilde{\mathbf{A}}_i$ for all $i \in N$. Further, $\tilde{\mathbf{A}}_i(\bar{\theta}_{-i}) = X$ for all $i \neq \bar{j}$ for some $\bar{j} \in N$. To see why consider the following: If Rule 2 applies under σ^* at $\bar{\theta}$ with $\bar{j}$ as the odd-man-out, then, for any $x \in X$ and $i \neq \bar{j}$, $(\tilde{\sigma}_i, \sigma_{-i}^*)$ triggers Rule 3 at $\bar{\theta}$ and delivers x where $\tilde{\sigma}_i$ is such that $\tilde{\sigma}_i(\theta_i) = \sigma_i^*(\theta_i)$ for all $\theta_i \neq \bar{\theta}_i$ and $\tilde{\sigma}_i(\bar{\theta}_i) = (\emptyset, \alpha_i(\bar{\theta}_i), x, k^*)$ where k^* is the number that makes i the winner of the modulo game at $\bar{\theta}$ given σ_{-i}^*. If Rule 3 applies under σ^* at $\bar{\theta}$, one can simply take $\bar{j} = 1$ and repeat the steps above. Thus, for SCF h^* and state $\bar{\theta}$, the profile of sets of acts $(\tilde{\mathbf{A}}_i)_{i \in N}$ satisfies both (i) and (ii) of the choice incompatibility with the odd-man-out given by $\bar{j}$. So, there is $i^* \neq \bar{j}$ with $\mathbf{h}_{i^*,\bar{\theta}_{i^*}} \notin \mathbf{C}_{i^*}^{\bar{\theta}_{i^*}}(\tilde{\mathbf{A}}_{i^*})$. This contradicts $\mathbf{h}_{i,\bar{\theta}_i}^* \in \mathbf{C}_i^{\bar{\theta}_i}(\tilde{\mathbf{A}}_i)$ for all $i \in N$, i.e., σ^* cannot be a BIE of μ.

The fact that under any BIE σ^* of μ, Rule 1 applies implies that there is a unique $f \in F$ such that $m_i^1(\theta_i) = f$ for all $i \in N$ and all $\theta_i \in \Theta_i$ thanks to the product structure of the state space: If there were i, j with $i \neq j$, who propose different SCFs under σ^*, say $f, f' \in F$ with $f \neq f'$ for their types θ_i and θ_j, respectively. Then, Rule 1 cannot apply at $(\theta_i, \theta_j, \theta_{-\{i,j\}}) \in \Theta$, a contradiction. ■

Finally, we show that (ii) of quasi-robust behavioral implementability holds as well:

Claim 3. *For any BIE of σ^* of μ, $g \circ \sigma^* \in F$.*

Proof. Recall that $\sigma_i^*(\theta_i) = (m_i^1(\theta_i), \alpha_i(\theta_i), \mathbf{a}_i(\theta_i), x_i(\theta_i), k_i(\theta_i))$ for each $i \in N$ and $\theta_i \in \Theta_i$. Thus, it suffices to show that $f \circ \alpha \in F$ as $h^* = g \circ \sigma^* = f \circ \alpha$. Since Rule 1 applies at each $\theta \in \Theta$, and each $i \in N$ reports the type $\alpha_i(\theta_i) \in \Theta_i$ as the second entry of their messages at $\theta \in \Theta$ under σ^*, by construction and $\mathbb{S}$ being closed under deception, we have, at each $\theta \in \Theta$, $\mathbf{O}_i^\mu(\sigma_{-i}^*) = \cup_{\mathbf{a}_i \in \mathbf{S}_i(f, \alpha_{-i}^{\mathrm{id}})}\{\mathbf{a}_i \circ \alpha_{-i}\} = \mathbf{S}_i(f, \alpha_{-i}^{\mathrm{id}} \circ \alpha_{-i}) = \mathbf{S}_i(f, \alpha_{-i})$ for all $i \in N$. If $f \circ \alpha \notin F$, then by (ii) of quasi-robust consistency, there exists $i^* \in N$ and $\theta_{i^*}^* \in \Theta_{i^*}$ such that $\mathbf{f}_{i^*, \theta_{i^*}^*}^{\alpha} \notin \mathbf{C}_{i^*}^{\theta_{i^*}^*}(\mathbf{S}_{i^*}(f, \alpha_{-i^*}))$. But this implies $\mathbf{f}_{i^*, \theta_{i^*}^*}^{\alpha} = \mathbf{h}_{i^*, \theta_{i^*}^*}^{*} \notin \mathbf{C}_{i^*}^{\theta_{i^*}^*}(\mathbf{O}_{i^*}^{\mu}(\sigma_{-i^*}^*))$. This contradicts σ^* being a BIE of μ. Therefore, $h^* = g \circ \sigma^* = f \circ \alpha \in F$, which implies that condition (ii) of quasi-robust behavioral implementability holds. ■ ■

In general, the sufficiency results for implementation under incomplete information involve economic environment assumptions (Jackson, 1991). To relate our choice incompatibility to Jackson's economic environment assumption, Barlo and Dalkıran (2023) introduces a weaker version of choice incompatibility, which is implied by Jackson's in the rational domain. In Appendix C, we provide another sufficiency result for quasi-robust behavioral implementation with finite state spaces using the weak choice incompatibility and robust-null alternatives.

4 Ex-post Behavioral Implementation

Under Property STP*, EPE is a plausible equilibrium notion because every EPE is a BIE. Hence, implementation in EPE emerges as a natural notion, provided that the dismissal of bad BIEs that are not EPEs is not a significant concern.[22]

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

Definition 8. *Given an ex-post environment, an SCS $F \in \mathcal{F}$ is* **ex-post behavioral implementable** *if there exists a mechanism μ such that*

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

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

We refer to an SCS F as being *partially ex-post behavioral implementable* whenever condition (i) in Definition 8 holds, whereas an SCS F is *weakly ex-post behavioral implementable* whenever condition (ii) in Definition 8 holds.

We emphasize that with minimax regret preferences, the ex-post choices satisfy the IIA, but the interim choices do not, while Property STP* holds. Thus, the resulting setting delivers an interesting economic environment where ex-post behavioral implementation is plausible if the planner need not worry about bad BIE that is not EPE.

Even when Property STP* fails to hold, ex-post behavioral implementation can be reasonable. To illustrate this, we revisit Example 1 with the state-contingent payoffs and mechanism μ in Table 1 where individual 2 is rational and individual 1 is ambiguity averse. Recall that in that example, every EPE of μ is a BIE of μ even though Property STP* fails to hold. There are two EPEs of mechanism μ, $\sigma^{(*)}$ and $\sigma^{(**)}$, both inducing SCF $\langle xy \rangle$. Therefore, mechanism μ ex-post behavioral implements $F = \{\langle xy \rangle\}$.

To further illustrate ex-post behavioral implementation we go over Example 2 with the rational interdependent ex-post preferences given in Table 2 considering the ex-post incentive efficient SCS $F = \{\langle xzzy \rangle\}$ (see Section 5). Below, we show that the direct mechanism in Table 4 ex-post behavioral implements F: In the direct mechanism $\mu^d = (\{t_1, t_1'\} \times \{t_2, t_2'\}, g^d)$, the deception profiles are in one-to-one correspondence with interim strategy profiles. The truthtelling strategy profile, α^{id}, is an EPE of the direct mechanism because of the following: $g^d \circ \alpha^{\text{id}}(t_i, t_j) = x = c_i^{(t_i, t_j)}(\{x, z\})$, $g^d \circ \alpha^{\text{id}}(t_i', t_j) = z = c_i^{(t_i', t_j)}(\{x, z\})$, $g^d \circ \alpha^{\text{id}}(t_i, t_j') = z = c_i^{(t_i, t_j')}(\{y, z\})$, and $g^d \circ \alpha^{\text{id}}(t_i', t_j') = y = c_i^{(t_i', t_j')}(\{y, z\})$ for both $i, j = 1, 2$. Moreover, truthtelling induces the unique Nash equilibrium outcome at every state. Therefore, there cannot be any bad EPE that induces an SCF other than $\langle xzzy \rangle$.[23] Table 7 reaffirms this by identifying the informant for each possible deception.

[23] We wish to point out that these arguments depend only on the ordinal ex-post preferences. Hence, they are independent of the particular payoff levels and continue to hold for all ε, $\tilde{\varepsilon}$, and η in $[0, 1)$.

	$\alpha_1(t_1)$	$\alpha_1(t_1')$	$\alpha_2(t_2)$	$\alpha_2(t_2')$	$f \circ \alpha$	Informant	Conforms	Deviates	Ex-post choice
α^{id}	t_1	t_1'	t_2	t_2'	$\langle xzzy \rangle$	$-$			
$\alpha^{(2)}$	t_1	t_1'	t_2	t_2	$\langle xzxz \rangle$	$(2, t_2')$	$\langle xz \rangle$	$\langle zy \rangle$	$z \notin c_2^{(t_1', t_2')}(\{y, z\})$
$\alpha^{(3)}$	t_1	t_1'	t_2'	t_2'	$\langle zyzy \rangle$	$(2, t_2)$	$\langle zy \rangle$	$\langle xz \rangle$	$z \notin c_2^{(t_1, t_2)}(\{x, z\})$
$\alpha^{(4)}$	t_1	t_1'	t_2'	t_2	$\langle zyxz \rangle$	$(2, t_2')$	$\langle xz \rangle$	$\langle zy \rangle$	$z \notin c_2^{(t_1', t_2')}(\{y, z\})$
$\alpha^{(5)}$	t_1	t_1	t_2	t_2'	$\langle xxzz \rangle$	$(1, t_1')$	$\langle xz \rangle$	$\langle zy \rangle$	$z \notin c_1^{(t_1', t_2')}(\{y, z\})$
$\alpha^{(6)}$	t_1	t_1	t_2	t_2	$\langle xxxx \rangle$	$(1, t_1)$	$\langle xx \rangle$	$\langle zz \rangle$	$x \notin c_1^{(t_1, t_2')}(\{x, z\})$
$\alpha^{(7)}$	t_1	t_1	t_2'	t_2'	$\langle zzzz \rangle$	$(2, t_2)$	$\langle zz \rangle$	$\langle xx \rangle$	$z \notin c_2^{(t_1, t_2)}(\{x, z\})$
$\alpha^{(8)}$	t_1	t_1	t_2'	t_2	$\langle zzxx \rangle$	$(2, t_2')$	$\langle xx \rangle$	$\langle zz \rangle$	$x \notin c_2^{(t_1, t_2')}(\{x, z\})$
$\alpha^{(9)}$	t_1'	t_1'	t_2	t_2'	$\langle zzyy \rangle$	$(1, t_1)$	$\langle zy \rangle$	$\langle xz \rangle$	$z \notin c_1^{(t_1, t_2)}(\{x, z\})$
$\alpha^{(10)}$	t_1'	t_1'	t_2	t_2	$\langle zzzz \rangle$	$(2, t_2')$	$\langle zz \rangle$	$\langle yy \rangle$	$z \notin c_2^{(t_1', t_2')}(\{y, z\})$
$\alpha^{(11)}$	t_1'	t_1'	t_2'	t_2'	$\langle yyyy \rangle$	$(1, t_1)$	$\langle yy \rangle$	$\langle zz \rangle$	$y \notin c_1^{(t_1, t_2)}(\{y, z\})$
$\alpha^{(12)}$	t_1'	t_1	t_2'	t_2	$\langle yyzz \rangle$	$(2, t_2)$	$\langle yy \rangle$	$\langle zz \rangle$	$y \notin c_2^{(t_1', t_2)}(\{y, z\})$
$\alpha^{(13)}$	t_1'	t_1	t_2	t_2'	$\langle zxyz \rangle$	$(1, t_1)$	$\langle zy \rangle$	$\langle xz \rangle$	$z \notin c_1^{(t_1, t_2)}(\{x, z\})$
$\alpha^{(14)}$	t_1'	t_1	t_2	t_2	$\langle zxzx \rangle$	$(1, t_1')$	$\langle xx \rangle$	$\langle zz \rangle$	$x \notin c_1^{(t_1', t_2)}(\{x, z\})$
$\alpha^{(15)}$	t_1'	t_1	t_2'	t_2'	$\langle yzyz \rangle$	$(1, t_1)$	$\langle yy \rangle$	$\langle zz \rangle$	$y \notin c_1^{(t_1, t_2')}(\{y, z\})$
$\alpha^{(16)}$	t_1'	t_1	t_2'	t_2	$\langle yzzx \rangle$	$(2, t_2)$	$\langle yz \rangle$	$\langle zx \rangle$	$y \notin c_2^{(t_1, t_2)}(\{y, z\})$

Table 7: Deception profiles, corresponding informants, and ex-post choice

4.1 Necessity

A necessary condition for ex-post implementation is *ex-post consistency*:

Definition 9. *Given an ex-post environment, a profile of sets of alternatives given by* $\mathbb{S} := (S_i(f, \theta_{-i}))_{i \in N,\ f \in F,\ \theta_{-i} \in \Theta_{-i}}$ *is* **ex-post consistent with the SCS** F *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* α *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 $\mathbb{S}$ 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 θ_i' at state (θ_i', θ_{-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 α that leads to an outcome not compatible with the SCS, there exist an informant state θ^* and an informant individual i^* such that $f(\alpha(\theta^*))$ is not in the ex-post choice of i^* at θ^* from $S_{i^*}(f, \alpha_{-i^*}(\theta_{-i^*}^*))$.

If mechanism μ ex-post implements a given SCS F, then for any SCF $f \in F$, there is an EPE σ^f of μ 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 $\mathbb{S}$ 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 $\mathbb{S}$ with F. Meanwhile, if a deception profile α is such that $f \circ \alpha \notin F$, then $\sigma^f \circ \alpha$ cannot be an EPE of μ; 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 θ^* and an individual i^* whose ex-post choice at θ^* 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 $\mathbb{S}$ with F. This discussion proves the following necessity result for ex-post behavioral implementation:

Theorem 3. *If an SCS F is ex-post implementable in an ex-post environment, then there is a profile of sets of alternatives ex-post consistent with F.*

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: behavioral ex-post monotonicity and quasi-ex-post 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).

An SCS F is ***behavioral ex-post monotonic*** if for every SCF $f \in F$ and deception profile α 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 4. *If there exists a profile of sets of alternatives ex-post consistent with an SCS F, then F is behavioral ex-post monotonic.*

Proposition 4 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 $\mathbb{S} := (S_i(f,\theta_{-i}))_{i\in N,\ f\in F,\ \theta_{-i}\in\Theta_{-i}}$ ex-post consistent with F, let $S^* := S_{i^*}(f,\alpha_{-i^*}(\theta_{-i^*}))$. Then, (i) of behavioral ex-post monotonicity follows from (ii) of ex-post consistency while (ii) of behavioral ex-post monotonicity follows from (i) of ex-post consistency.

Proposition 5. *If there exists a profile of sets of alternatives ex-post consistent with an SCS F, then F is quasi-ex-post incentive compatible.*

To see the arguments needed to establish this result, let $\mathbb{S} := (S_i(f, \theta_{-i}))_{i \in N,\ f \in F,\ \theta_{-i} \in \Theta_{-i}}$ 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 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 $c_i^{\theta}(S) := \{y \in S : u_i(y, \theta) \geq u_i(x, \theta) \text{ for all } x \in S\}$ for any $S \in \mathcal{X}$. Then, the necessary conditions of Bergemann and Morris (2008) are 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 α 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 behavioral ex-post monotonicity coupled with quasi-ex-post incentive compatibility.

Proposition 6. *When ex-post choices satisfy WARP, behavioral ex-post monotonicity coupled with quasi-ex-post incentive compatibility is equivalent to ex-post monotonicity coupled with ex-post incentive compatibility.*

Proof of Proposition 6. The result follows from Claims 4 and 5.

Claim 4. *When ex-post choices satisfy WARP, quasi-ex-post incentive compatibility is equivalent to ex-post incentive compatibility.*

Proof. Suppose that individuals' ex-post choices satisfy WARP. If F is quasi-ex-post 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^{θ} 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. ■

Claim 5. *When ex-post choices satisfy WARP, the following hold:*

(*i*) *if an SCS F is behavioral ex-post 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 behavioral ex-post monotonic.*

Proof. Suppose that individuals' ex-post choices satisfy WARP.

For (*i*), suppose that for all $f \in F$ and α 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^θ 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 α 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^θ 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 $\theta_i' \in \Theta_i$. ■ ■

4.2 Sufficiency

We need the following to establish sufficiency for ex-post behavioral implementation:

Definition 10. *The* **ex-post choice incompatibility** *holds in an ex-post environment if for all $\theta \in \Theta$, all $x \in X$, all $\bar{j} \in N$, there is $i^* \in N \setminus \{\bar{j}\}$ such that $x \notin c_{i^*}^\theta(X)$.*

Similar to its interim counterpart, this condition implies some level of disagreement among individuals regarding their ex-post choices at every state.

Ex-post choice incompatibility coupled with ex-post consistency is sufficient for ex-post behavioral implementation:

Theorem 4. *Suppose that the ex-post environment is such that $n \geq 3$ and the ex-post choice incompatibility holds. Then, if there is a profile of sets of alternatives ex-post consistent with the SCS F, then F is ex-post behavioral implementable.*

Before we proceed to the proof, we would like to note that under rationality, our ex-post choice incompatibility is equivalent to the economic environment assumption

of Bergemann and Morris (2008).[24] Consequently, thanks to Propositions 4, 5, and 6, Theorem 4 amounts to a behavioral analog of their Theorem 2.

Proof of Theorem 4. Suppose that ex-post environment $\mathcal{E}^{\text{ep}}$ is such that $n \geq 3$ and the ex-post choice incompatibility holds. Consider SCS F with which the profile of sets of alternatives $\mathbb{S} := (S_i(f, \theta_{-i}))_{i \in N,\ f \in F,\ \theta_{-i} \in \Theta_{-i}}$ is ex-post consistent.

We use the following mechanism $\mu = (M, g)$: For each $i \in N$, his set of messages is $M_i = F \cup \{\emptyset\} \times \Theta_i \times X \times N$, while a generic message is denoted by $m_i = (m_i^1, \theta_i, x_i, k_i)$, and the outcome function $g : M \to X$ is as specified in Table 8.

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

Table 8: The outcome function of the mechanism.

Claim 6. *For any $f \in F$, there exists an EPE, σ^f, of $\mu = (M, g)$ with $f = g \circ \sigma^f$.*

Proof. Let $\sigma_i^f(\theta_i) = (f, \theta_i, x, 1)$ for each $i \in N$ and for some arbitrary $x \in X$. 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 σ^f, either Rule 1 or Rule 2 applies, i.e., Rule 3 is not attainable by any unilateral deviation from σ^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., σ^f is an EPE of μ such that $f = g \circ \sigma^f$. ■

Consider now any EPE σ^* of μ denoted as $\sigma_i^*(\theta_i) = (m_i^1(\theta_i), \alpha_i(\theta_i), x_i(\theta_i), k_i(\theta_i))$ for each $i \in N$. That is, $m_i^1(\theta_i)$ denotes either $\emptyset$ or the SCF proposed by i when her type is θ_i; $\alpha_i(\theta_i)$, the reported type of i when her type is θ_i; $x_i(\theta_i)$, the alternative proposed by i when her type is θ_i; and $k_i(\theta_i)$, the number proposed by i when her type is θ_i.

[24] Ex-post choice incompatibility is equivalent to the behavioral version of Bergemann and Morris (2008)'s economic environment assumption in ex-post environment $\mathcal{E}^{\text{ep}}$: For each state θ and alternative $x \in X$, there exist two individuals $i, j \in N$ with $i \neq j$ such that $x \notin c_i^\theta(X)$ and $x \notin c_j^\theta(X)$; i.e., at any state, any alternative can be ex-post chosen from the set of all alternatives by at most $n-2$ individuals.

Claim 7. *Under any EPE σ^* of μ, Rule 1 must apply at each $\theta \in \Theta$, and there is unique $f \in F$ with $m_i^1(\theta_i) = f$ for all $i \in N$ and all $\theta_i \in \Theta_i$.*

Proof. Suppose, for contradiction, that either Rule 2 or Rule 3 applies at some $\tilde{\theta} \in \Theta$ under σ^*. 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 $i \in N$. In this case, simply let $j = 1$. Therefore, under both Rule 2 and Rule 3, at least $n-1$ individuals have the opportunity set X. Since σ^* is an EPE of μ, it follows that $g(\sigma^*(\tilde{\theta})) \in c_i^\theta(X)$ for all $i \neq j$. Consequently, the desired contradiction emerges as for $\tilde{\theta}$ and $g(\sigma^*(\tilde{\theta}))$ and $(S_i)_{i \in N}$ with $S_i = X$ for all $i \neq j$ and $S_j \in \mathcal{X}$ are as in the hypothesis of the ex-post choice incompatibility but there cannot be $i^* \neq j$ with $g(\sigma^*(\tilde{\theta})) \notin c_{i^*}^\theta(X)$.

As Rule 1 applies at every $\theta \in \Theta$ under any EPE σ^* of μ, due to the product structure of Θ, as before, there must be a unique $f \in F$ with $f_i(\theta_i) = f$ for all $i \in N$ and all $\theta_i \in \Theta_i$. Hence, by Rule 1, $g(\sigma^*(\theta)) = f(\alpha(\theta))$ for all $\theta \in \Theta$. ■

Claim 8. *For any EPE σ^* of μ, $g \circ \sigma^* \in F$.*

Proof. We show that $f \circ \alpha \in F$ as $g \circ \sigma^* = f \circ \alpha$: Since Rule 1 applies at each $\theta \in \Theta$, and each $i \in N$ reports the type $\alpha_i(\theta_i) \in \Theta_i$ as the second entry of their messages at $\theta \in \Theta$ under σ^*, by construction, 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 σ^* being an EPE of μ. 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. ■ ■

5 Ex-post Behavioral Efficiency

The key ingredient of welfare analysis under incomplete information concerns state contingent allocations, SCFs. Moreover, following Holmström and Myerson (1983), efficient SCFs must be independent of individuals' private information.

In this section, we introduce the behavioral counterpart of ex-post incentive Pareto efficiency of Holmström and Myerson (1983). Our construction parallels de Clippel

(2014), introducing the following efficiency notion in behavioral domains of complete information: An alternative x is *behaviorally efficient at state* θ if each individual has an implicit opportunity set from which she chooses x at θ, and each alternative is in at least one of these implicit opportunity sets. Extending this efficiency notion to incomplete information environments, we define *behavioral ex-post efficiency* by demanding such SCFs result in behaviorally efficient alternatives at every state:

Definition 11. *Given an ex-post environment, an SCF* $f : \Theta \to X$ *is* ***behaviorally ex-post efficient*** *if there is a profile of sets of alternatives* $(Y_{i,\theta})_{i\in N,\theta\in\Theta}$ *such that*

(i) *for all* $i \in N$ *and all* $\theta \in \Theta$, $f(\theta) \in c_i^{\theta}(Y_{i,\theta})$, *and*

(ii) *for all* $\theta \in \Theta$, $\cup_{i\in N} Y_{i,\theta} = X$.

We refer to the set of all behaviorally ex-post efficient SCFs as the ***behavioral ex-post efficient SCS*** *and denote it as* EE.

The EE SCS is non-empty whenever the ex-post choices are non-empty valued because a behaviorally efficient alternative exists at every state (de Clippel, 2014).

Behavioral ex-post efficiency extends ex-post Pareto efficiency to behavioral domains: Individuals' ex-post choices are rational when for all $i \in N$ and all $\theta \in \Theta$, there is a complete and transitive preference relation $R_{i,\theta} \subset X \times X$ such that for any non-empty $S \subset X$, $x \in c_i^{\theta_i}(S)$ if and only if $xR_{i,\theta}y$ for all $y \in S$. Then, an SCF f is **ex-post Pareto efficient** (in the rational domain) if there is no $h \in H$ such that for some $\theta \in \Theta$, we have $h(\theta)P_{i,\theta}f(\theta)$ for all $i \in N$.[25] To see that ex-post Pareto efficiency implies behavioral ex-post efficiency, let f be ex-post Pareto efficient and set $Y_{i,\theta} = LCS_{i,\theta}(f(\theta)) := \{y \in X \mid f(\theta)R_{i,\theta}y\}$ for all $i \in N$ and all $\theta \in \Theta$. Then, for all $i \in N$ and all $\theta \in \Theta$, $f(\theta) \in c_i^{\theta}(Y_{i,\theta})$, delivering (i) of behavioral ex-post efficiency of f. For (ii) of behavioral ex-post efficiency of f, notice that if there were some $\bar{\theta} \in \Theta$ and $y \in X$ such that for all $i \in N$ and $y \notin Y_{i,\bar{\theta}}$, then, by construction, $yP_{i,\bar{\theta}}f(\bar{\theta})$ for all $i \in N$. Defining $h : \Theta \to X$ by $h(\bar{\theta}) = y$ and $h(\theta') = f(\theta')$ for all $\theta' \neq \bar{\theta}$ enables us to conclude that f cannot be behaviorally ex-post efficient as $h(\bar{\theta})P_{i,\bar{\theta}}f(\bar{\theta})$ for all $i \in N$. For the converse, let f be behaviorally ex-post efficient but not ex-post Pareto efficient, i.e., there is $h \in H$ and $\bar{\theta} \in \Theta$ such that $h(\bar{\theta})P_{i,\bar{\theta}}f(\bar{\theta})$ for all $i \in N$. Then, by (ii) of

[25] This notion is the weak version of ex-post Pareto efficiency in Holmström and Myerson (1983).

behavioral ex-post efficiency of f, $h(\theta) \in Y_{j,\theta}$ for some $j \in N$ and $\theta \in \Theta$. But this implies that $f(\theta) \notin c_j^{\theta}(Y_{j,\theta})$, a contradiction to (i) of behavioral ex-post efficiency of f.

The natural notion of efficiency in behavioral environments under incomplete information is given by the behavioral counterpart of interim Pareto efficiency of Holmström and Myerson (1983), namely, *behavioral interim efficiency* introduced by Barlo and Dalkıran (2023). In that study, we establish that behavioral interim efficiency is an extension of interim Pareto efficiency to behavioral domains.[26]

Definition 12. *Given an interim environment, an SCF $f : \Theta \to X$ is **behaviorally interim efficient** if there is a profile of sets of acts $(\mathbf{Y}_{i,\theta_i})_{i\in N,\ \theta_i\in\Theta_i}$ such that*

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

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

In what follows, we show that under Property STP*, the relation between ex-post efficiency and interim efficiency under rationality continues to hold in behavioral domains.

Proposition 7. *Suppose that the given pair of associated interim and ex-post environments satisfies Property STP*. Then, any behaviorally ex-post efficient SCF is also behaviorally interim efficient.*

Proof. Let f be behaviorally ex-post efficient with the corresponding profile of sets $(Y_{i,\theta})_{i\in N,\theta\in\Theta}$. For all $i \in N$ and $\theta_i \in \Theta_i$ define $\mathbf{Y}_{i,\theta_i}$ as $\mathbf{Y}_{i,\theta_i} := \{\mathbf{a}_i \in \mathbf{A}_i \mid \mathbf{a}_i(\theta'_{-i}) \in Y_{i,(\theta_i,\theta'_{-i})} \text{ for all } \theta'_{-i} \in \Theta_{-i}\}$. Since for all $i \in N$ and $\theta_i \in \Theta_i$, $\mathbf{f}_{i,\theta_i}(\theta'_{-i}) = f(\theta_i, \theta'_{-i}) \in c_i^{(\theta_i,\theta'_{-i})}(Y_{i,(\theta_i,\theta'_{-i})})$ and $\mathbf{Y}_{i,\theta_i}(\theta'_{-i}) = Y_{i,(\theta_i,\theta'_{-i})}$ for all $\theta'_{-i} \in \Theta_{-i}$, by Property STP*, $\mathbf{f}_{i,\theta_i} \in \mathbf{C}_i^{\theta_i}(\mathbf{Y}_{i,\theta_i})$. Thus, (i) of behavioral interim efficiency holds. For (ii) of behavioral interim efficiency, suppose for a contradiction that there is $h \in H$ such that for all i and for all $\theta_i \in \Theta_i$, $\mathbf{h}_{i,\theta_i} \notin \mathbf{Y}_{i,\theta_i}$. Then, for any $\theta \in \Theta$, $\mathbf{h}_{i,\theta_i}(\theta_{-i}) = h(\theta) \notin Y_{i,\theta} = \mathbf{Y}_{i,\theta_i}(\theta_{-i})$ for all $i \in N$, contradicting f being behaviorally ex-post efficient. ■

As we establish in our necessity results for quasi-robust behavioral implementation (Theorem 1) and behavioral ex-post implementation (Theorem 3), the existence of a quasi-robust consistent profile or an ex-post consistent profile implies the quasi-ex-post

[26] Individuals' interim choices are rational when for all $i \in N$ and all $\theta_i \in \Theta_i$, there is a complete and transitive preference relation $\mathbf{R}_{i,\theta_i} \subset \mathbf{A}_i \times \mathbf{A}_i$ such that for any non-empty $\mathbf{S} \subset \mathbf{A}_i$, $\mathbf{a}_i \in \mathbf{C}_i^{\theta_i}(\mathbf{S})$ if and only if $\mathbf{a}_i \mathbf{R}_{i,\theta_i} \mathbf{a}'_i$ for all $\mathbf{a}'_i \in \mathbf{S}$. Consequently, an SCF f is **interim Pareto efficient** (in the rational domain) if there is no $h \in H$ such that $\mathbf{h}_{i,\theta_i} \mathbf{P}_{i,\theta_i} \mathbf{f}_{i,\theta_i}$ for all $i \in N$ and all $\theta_i \in \Theta_i$.

incentive compatibility of the corresponding SCS (see Propositions 2 and 5). Therefore, quasi-ex-post incentive compatibility arises as a necessary condition for quasi-robust behaviroal or behavioral ex-post implementation of an SCS. This is why, following Holmström and Myerson (1983), we define behavioral ex-post incentive efficiency by restricting feasibility based on quasi-ex-post incentive compatibility as follows:

Definition 13. *Given an ex-post environment, an SCF $f : \Theta \to X$ is **behaviorally ex-post incentive efficient** if there is a profile of sets of alternatives $(Y_{i,\theta_{-i}})_{i\in N,\theta_{-i}\in\Theta_{-i}}$ such that*

(i) for all $i \in N$ and all $\theta \in \Theta$, $f(\theta) \in c_i^{\theta}(Y_{i,\theta_{-i}})$, and

(ii) for all $\theta \in \Theta$, $\cup_{i\in N} Y_{i,\theta_{-i}} = X$.

*We refer to the set of all behaviorally ex-post incentive efficient SCFs as the **behavioral ex-post incentive efficient SCS** and denote it as EIE.*

The EIE SCS ingrains quasi-ex-post incentive compatibility into EE SCS by requiring the implicit opportunity sets not to depend on individuals' private information.

Because quasi-ex-post incentive compatibility boils down to ex-post incentive compatibility under rationality, behavioral ex-post incentive efficiency extends ex-post incentive Pareto efficiency of Holmström and Myerson (1983) to behavioral domains.

The behavioral extension of Holmström and Myerson's interim incentive Pareto efficiency is introduced by Barlo and Dalkıran (2023). This welfare criterion internalizes quasi-incentive compatibility into behavioral interim efficiency and is defined as follows:

Definition 14. *Given an interim environment, an SCF $f : \Theta \to X$ is **behaviorally interim incentive efficient** if there is a profile of sets of acts $(\mathbf{Y}_i)_{i\in N}$ such that*

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

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

As quasi-ex-post incentive compatibility implies quasi-incentive compatibility under Property STP*, we obtain the following corollary thanks to Proposition 7:

Corollary 1. *Suppose that the given pair of associated interim and ex-post environments satisfies Property STP*. Then, any behaviorally ex-post incentive efficient SCF is also behaviorally interim incentive efficient.*

Next, we show that quasi-ex-post incentive compatibility brings about a more demanding requirement for the existence of the EIE SCS:

Proposition 8. *Given an ex-post environment, if there is a quasi-ex-post incentive compatible SCF f such that for any $\theta \in \Theta$, there is $i_\theta \in N$ with $f(\tilde{\theta}_{i_\theta}, \theta_{-i_\theta}) \in c_{i_\theta}^{(\tilde{\theta}_{i_\theta}, \theta_{-i_\theta})}(X)$ for all $\tilde{\theta}_{i_\theta} \in \Theta_{i_\theta}$, then the EIE SCS is non-empty.*

Proof. Let SCF f be quasi-ex-post incentive compatible and let the corresponding profile of sets of alternatives be $(S_{i,\theta})_{i\in N, \theta\in\Theta}$ where $S_{i,\theta}$ is the set of alternatives associated with individual i, SCF f, and state θ as specified in Proposition 5. Recall that $f(\Theta_i, \theta_{-i}) \subset S_{i,\theta}$ for all $i \in N$. Let the implicit opportunity set profile $(Y_{i,\theta_{-i}})_{i\in N, \theta_{-i}\in\Theta_{-i}}$ be such that for all $\theta \in \Theta$, $Y_{i,\theta_{-i}} = S_{i,\theta}$ for all $i \neq i_\theta$, and $Y_{i_\theta,\theta_{-i_\theta}} = X$. Then, (i) of behavioral ex-post incentive efficiency holds as $f(\tilde{\theta}_{i_\theta}, \theta_{-i_\theta}) \in c_{i_\theta}^{(\tilde{\theta}_{i_\theta}, \theta_{-i_\theta})}(Y_{i,\theta_{-i}})$ for all $\tilde{\theta}_{i_\theta} \in \Theta_{i_\theta}$; and for all $i \neq i_\theta$, $f(\theta) \in c_i^\theta(S_{i,\theta})$. Further, (ii) of behavioral ex-post efficiency holds as for all $\theta \in \Theta$, $\cup_{i\in N} Y_{i,\theta_{-i}} = X$ holds since $Y_{i_\theta,\theta_{-i_\theta}} = X$. So, $f \in EIE$. ■

Before analyzing the quasi-robust behavioral implementability of the EIE SCS, we first show that it is ex-post behavioral implementable when there are at least three individuals and the ex-post choice incompatibility holds:

Proposition 9. *Suppose that the ex-post environment is such that $n \geq 3$, the ex-post choice incompatibility holds, and the EIE SCS is non-empty. Then, the EIE SCS is ex-post behavioral implementable.*

Proof. Let $f \in EIE$, $\theta_{-i} \in \Theta_{-i}$, and define $\mathbb{S} := (S_i(f, \theta_{-i}))_{i\in N,\ f\in F,\ \theta_{-i}\in\Theta_{-i}}$ by $S_i(f, \theta_{-i}) := Y^f_{i,\theta_{-i}}$ for all $i \in N$ where $Y^f_{i,\theta_{-i}}$ is the implicit opportunity set associated with f as in Definition 13. Then, (i) of behavioral ex-post incentive efficiency implies (i) of ex-post consistency. Suppose $f \in EIE$ but $f \circ \alpha \notin EIE$ for some $\alpha \in \Lambda$. Hence, $f \circ \alpha$ is not behaviorally efficient at some state. Therefore, for any profile of sets of alternatives $(W_i)_{i\in N}$ such that $\cup_{i\in N} W_i = X$, there are i^* and θ^* with $f(\alpha(\theta^*)) \notin c_{i^*}^{\theta^*}(W_{i^*})$; because otherwise, $f \circ \alpha$ would be behaviorally ex-post incentive efficient. By (ii) of behavioral ex-post efficiency of SCF f, $(Y^f_{i,\alpha_{-i}(\theta_{-i})})_{i\in N}$ is a profile of sets of alternatives such that $\cup_{i\in N} Y^f_{i,\alpha_{-i}(\theta_{-i})} = X$. Hence, there is i^* and θ^* such that $f(\alpha(\theta^*)) \notin c_{i^*}^{\theta^*}(Y^f_{i^*,\alpha_{-i^*}(\theta^*_{-i^*})})$. Because $Y^f_{i,\alpha_{-i}(\theta_{-i})} = S_i(f, \alpha_{-i}(\theta_{-i}))$ for all $i \in N$, all $\alpha \in \Lambda$, and all $\theta \in \Theta$, it follows that $f(\alpha(\theta^*)) \notin c_{i^*}^{\theta^*}(S_{i^*}(f, \alpha_{-i^*}(\theta^*_{-i^*})))$, implying (ii) of ex-post consistency. Hence, $\mathbb{S}$ is ex-post consistent with the EIE SCS. The result follows from Theorem 4. ■

We next observe that quasi-robust behavioral implementability of the EIE SCS fails in Example 2 if $\eta = \epsilon = \tilde{\epsilon} = 0$. Then, under the minimax regret preferences described in that example, each type of each individual chooses act $\langle zz \rangle$ at the interim stage whenever it is available. Moreover, the only ex-post incentive efficient SCF is $\langle xzzy \rangle$. Consequently, any mechanism that sustains SCF $\langle xzzy \rangle$ as a BIE also admits a bad BIE inducing SCF $\langle zzzz \rangle$ as we show on page 15. Therefore, the EIE SCS, given by $\{\langle xzzy \rangle\}$, is not quasi-robust behavioral implementable —even though it is ex-post behavioral implementable as we display on page 23.

At the heart of the EIE's failure of quasi-robust behavioral implementability lies the aspect that players' deception that results in an SCF not aligned with the EIE SCS does not necessarily trigger objections. Consequently, as implicit opportunity sets associated with the EIE SCS is closed under deception, we obtain the quasi-robust behavioral implementability by demanding the existence of a whistle-blower alerting the planner in case of such deceptions. To formalize these, we need the following: For any SCF $f \in EIE$, let us denote the profile of associated implicit opportunity sets of acts (as in Definition 13) by $(Y^f_{i,\theta_{-i}})_{i \in N}$. For any $f \in EIE$, any individual i, and any deception $\alpha \in \Lambda$, the set of acts obtained from $(Y^f_{i,\theta_{-i}})_{\theta_{-i} \in \Theta_{-i}}$ by $\mathbf{Y}_i^{f \circ \alpha} := \{\mathbf{a}_i \in \mathbf{A}_i \mid \mathbf{a}_i(\alpha_{-i}(\theta_{-i})) \in Y^f_{i,\alpha_{-i}(\theta_{-i})} \text{ for all } \theta_{-i \in \Theta_{-i}}\}$.

Proposition 10. *Suppose $n \geq 3$ and the ex-post environment is such that the EIE SCS is non-empty. If the associated interim environment satisfies Property STP* and the choice incompatibility, then the EIE SCS is quasi-robust behavioral implementable whenever $f \in EIE$ and $f \circ \alpha \notin EIE$ implies there are $i \in N$ and $\theta_i \in \Theta_i$ with $\mathbf{f}^{\alpha}_{i,\theta_i} \notin \mathbf{C}_i^{\theta_i}(\mathbf{Y}_i^{f \circ \alpha})$.*

Proof. Let $f \in EIE$, $\alpha \in \Lambda$, and define $\mathbb{S} := (\mathbf{S}_i(f, \alpha_{-i}))_{i \in N,\ f \in F,\ \alpha_{-i} \in \Lambda_{-i}}$ by $\mathbf{S}_i(f, \alpha_{-i}) := \mathbf{Y}_i^{f \circ \alpha}$ for all $i \in N$. $\mathbb{S}$ is closed under deception because $\mathbf{a}_i \in \mathbf{S}_i(f, \alpha_{-i}) = \mathbf{Y}_i^{f \circ \alpha}$ implies that for any $\tilde{\alpha} \in \Lambda$, $\mathbf{a}_i^{\tilde{\alpha}} \in \mathbf{Y}_i^{f \circ \alpha \circ \tilde{\alpha}} = \mathbf{S}_i(f, \alpha_{-i} \circ \tilde{\alpha}_{-i})$ since $\alpha \circ \tilde{\alpha} \in \Lambda$.[27] Further, as $\mathbf{S}_i(f, \alpha^{\text{id}}_{-i}) = \mathbf{Y}_i^f$ for all $i \in N$, (i) of behavioral ex-post incentive efficiency implies (i) of quasi-robust consistency thanks to Property STP*. Finally, (ii) of quasi-robust consistency follows directly from the hypothesis of the proposition as $\mathbf{Y}_i^{f \circ \alpha} = \mathbf{S}_i(f, \alpha_{-i})$.

[27] For any $i \in N$, any $\theta_{-i} \in \Theta_{-i}$, and any $\alpha_{-i} \in \Lambda_{-i}$, as $\alpha_{-i}(\tilde{\alpha}(\theta_{-i})) \in \Theta_{-i}$, we have $\mathbf{Y}_i^{f \circ \alpha \circ \tilde{\alpha}}$ consists of acts $\mathbf{a}_i \in \mathbf{A}_i$ such that $\mathbf{a}_i(\theta_{-i}) \in Y_{i,\alpha_{-i}(\tilde{\alpha}_{-i}(\theta_{-i}))}$.

Hence, $\mathbb{S}$ is quasi-robust consistent with the EIE SCS. The result follows immediately from Theorem 2. ■

6 Concluding Remarks

In this paper, we have studied full behavioral implementation under incomplete information from a robust mechanism design point of view, without requiring that individuals' ex-post and interim choices satisfy the weak axiom of revealed preferences (WARP). Our robustness analysis provides novel insights into behavioral mechanism design where information asymmetries are inescapable in many interesting economic settings.

We have employed *behavioral interim equilibrium* (BIE) and *behavioral ex-post equilibrium* (EPE) to derive necessary as well as sufficient conditions for *quasi-robust behavioral* and *ex-post behavioral implementation* of social choice rules. The former requires every optimal SCF be sustained as both a BIE and an EPE of a mechanism, and that there be no 'bad' BIE of this mechanism. The latter requires that the optimal SCFs be sustained as EPE and that there is no 'bad' EPE in the mechanism.

Our paper can thus be regarded as the robust and ex-post counterpart of Barlo and Dalkıran (2023), which investigates behavioral interim implementation under incomplete information without any ex-post considerations.

Appendix

A The Warning of de Clippel (2022)

We discuss situations in which a contradiction along the lines of de Clippel (2022) may emerge in our behavioral setting. To that regard, we construct an example mimicking the construction in the proof of de Clippel (2022, Proposition 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 $x \in T \backslash c_i^\theta(T)$ such that $c_i^\theta(T) \neq c_i^\theta(T \backslash \{x\})$. Given i and her type θ_i, if there are only two distinct type profiles of others, $\theta_{-i}, \tilde{\theta}_{-i} \in \Theta_{-i}$, then we can construct the following set of acts: $\tilde{\mathbf{A}}_i = \{\mathbf{a}_i, \mathbf{a}_i', \mathbf{a}_i'', \mathbf{a}_i'''\} \bigcup \left(\cup_{y \in Y,\ \tilde{y} \in \tilde{Y}} \{\mathbf{a}_i^{y,\tilde{y}}\}\right)$ where these acts are as specified in Table 9 and $Y, \tilde{Y} \in \mathcal{X}$ are as follows:

	θ_{-i}	$\tilde{\theta}_{-i}$
$\mathbf{a}_i$	$c_i^\theta(T)$	$c_i^{(\theta_i, \tilde{\theta}_{-i})}(T \setminus \{x\})$
$\mathbf{a}_i'$	$c_i^\theta(T \setminus \{c_i^\theta(T)\})$	$c_i^{(\theta_i, \tilde{\theta}_{-i})}(T \setminus \{x\})$
$\mathbf{a}_i''$	$c_i^{(\theta_i, \tilde{\theta}_{-i})}(T \setminus \{x\})$	$c_i^\theta(T)$
$\mathbf{a}_i'''$	x	$c_i^\theta(T)$
$\mathbf{a}_i^{y,\tilde{y}}$	y	$\tilde{y}$

Table 9: An example in conjunction with Property STP*.

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

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

We need the following additional requirements to reach a contradiction as in de Clippel (2022): Individual i should perceive acts $\mathbf{a}_i$ and $\mathbf{a}_i''$ to be equivalent to each other on the grounds of $\mathbf{a}_i(\theta_{-i}) = \mathbf{a}_i''(\tilde{\theta}_{-i})$ and $\mathbf{a}_i(\tilde{\theta}_{-i}) = \mathbf{a}_i''(\theta_{-i})$. That is, when considering θ_{-i} and $\tilde{\theta}_{-i}$, only the underlying alternatives associated with these acts matter to her.

As a result, she perceives the act that delivers x' at θ_{-i} and y' at $\tilde{\theta}_{-i}$ to be equivalent to another that provides y' at θ_{-i} and x' at $\tilde{\theta}_{-i}$ where $x', y' \in X$. For example, this happens under probabilistic sophistication when i's belief is such that θ_{-i} and $\tilde{\theta}_{-i}$ are equally likely and i evaluates acts by the lotteries they induce.

Indeed, at the heart of this contradiction lies the individual perceiving two different states as equivalent. We model the equivalence perception of individual i of type θ_i over the set of all others' types via an an equivalence relation $\doteqdot$ defined on Θ_{-i} and let the equivalence class of $\bar{\theta}_{-i}$ be denoted by $\mathcal{P}_{i,\theta_i}(\bar{\theta}_{-i}) := \{\theta'_{-i} \in \Theta_{-i} \mid \theta'_{-i} \doteqdot \bar{\theta}_{-i}\}$.[28] For any i of type θ_i, the relation $\doteqdot$ partitions any set of states $\bar{\Theta}_{-i} \subset \Theta_{-i}$ into equivalence classes.

As a result of this perception equivalence, individual i of type θ_i perceives two acts $\mathbf{a}_i^{(1)}$ and $\mathbf{a}_i^{(2)}$ as equivalent whenever for any $\theta'_{-i}, \theta''_{-i} \in \Theta_{-i}$ with $\theta''_{-i} \in \mathcal{P}_{i,\theta_i}(\theta'_{-i})$, $\mathbf{a}_i^{(1)}(\theta'_{-i}) = \mathbf{a}_i^{(2)}(\theta''_{-i})$, $\mathbf{a}_i^{(1)}(\theta''_{-i}) = \mathbf{a}_i^{(2)}(\theta'_{-i})$, and $\mathbf{a}_i^{(1)}(\theta'''_{-i}) = \mathbf{a}_i^{(2)}(\theta'''_{-i})$ for all $\theta'''_{-i} \in \Theta_{-i} \setminus \{\theta'_{-i}, \theta''_{-i}\}$; we denote such a situation by $\mathbf{a}_i^{(1)} \mathcal{I}_{i,\theta_i} \mathbf{a}_i^{(2)}$. We assume that the perception equivalence relation of individual i of type θ_i induces the following over her choices: If $\mathbf{a}_i^{(1)} \mathcal{I}_{i,\theta_i} \mathbf{a}_i^{(2)}$, then $\mathbf{a}_i^{(1)} \in \mathbf{C}_i^{\theta_i}(\mathbf{A}'_i)$ if and only if $\mathbf{a}_i^{(2)} \in \mathbf{C}_i^{\theta_i}(\mathbf{A}'_i)$ for all $\mathbf{A}'_i \subset \mathbf{A}_i$ with $\mathbf{a}_i^{(1)}, \mathbf{a}_i^{(2)} \in \mathbf{A}'_i$.

For any i of type θ_i, the relation $\mathcal{I}_{i,\theta_i}$ partitions any $\mathbf{A}'_i \subset \mathbf{A}_i$ into equivalence classes. We assume that i of type θ_i perceives two sets of acts $\mathbf{A}'_i$ and $\mathbf{A}''_i$ as equivalent if the collection of equivalence classes in $\mathbf{A}_i$ that the acts in $\mathbf{A}'_i$ and $\mathbf{A}''_i$ belong to are equal to one another. With a slight abuse of notation, we denote such a case by $\mathbf{A}'_i \mathcal{I}_{i,\theta_i} \mathbf{A}''_i$. Formally, $\mathbf{A}_i^{(1)} \mathcal{I}_{i,\theta_i} \mathbf{A}_i^{(2)}$ if for all $\bar{\mathbf{a}}_i \in \mathbf{A}_i^{(k)}$, $\mathcal{I}_{i,\theta_i}(\bar{\mathbf{a}}_i) \cap \mathbf{A}_i^{(\ell)} \neq \emptyset$ for all $k, \ell = 1, 2$ where $\mathcal{I}_{i,\theta_i}(\bar{\mathbf{a}}_i)$ is the equivalence class of $\bar{\mathbf{a}}_i$ with respect to $\mathcal{I}_{i,\theta_i}$.

Moreover, we assume that interim choices of i of type θ_i from a set of acts respect the resulting equivalence classes so that the interim choices are singleton-valued up to equivalence classes with respect to $\mathcal{I}_{i,\theta_i}$: For any two sets of acts $\mathbf{A}_i^{(1)}$ and $\mathbf{A}_i^{(2)}$ with $\mathbf{A}_i^{(1)} \mathcal{I}_{i,\theta_i} \mathbf{A}_i^{(2)}$, $\mathbf{a}_i^{(1)} \in \mathbf{C}_i^{\theta_i}(\mathbf{A}_i^{(1)})$ and $\mathbf{a}_i^{(2)} \in \mathbf{C}_i^{\theta_i}(\mathbf{A}_i^{(2)})$ implies $\mathbf{a}_i^{(1)} \mathcal{I}_{i,\theta_i} \mathbf{a}_i^{(2)}$.

Then, going back to our example, we see that $\mathbf{a}_i \mathcal{I}_{i,\theta_i} \mathbf{a}''_i$ as $\mathbf{a}_i = \langle c_i^{\theta}(T),\ c_i^{(\theta_i, \tilde{\theta}_{-i})}(T \setminus \{x\}) \rangle$ and $\mathbf{a}''_i = \langle c_i^{(\theta_i, \tilde{\theta}_{-i})}(T \setminus \{x\}),\ c_i^{\theta}(T) \rangle$. Further, $\tilde{\mathbf{A}}_i \mathcal{I}_{i,\theta_i} \hat{\mathbf{A}}_i$ because $\hat{\mathbf{A}}_i = \tilde{\mathbf{A}}_i \setminus \{\mathbf{a}_i\}$ and $\mathbf{a}_i \mathcal{I}_{i,\theta_i} \mathbf{a}''_i$ and $\mathbf{a}''_i \in \hat{\mathbf{A}}_i \subset \tilde{\mathbf{A}}_i$. Recall that by Property STP*, $\mathbf{a}_i \in \mathbf{C}_i^{\theta_i}(\tilde{\mathbf{A}}_i)$ and $\mathbf{a}'_i \in \mathbf{C}_i^{\theta_i}(\hat{\mathbf{A}}_i)$. But then, the desired contradiction emerges as $\mathbf{a}_i \mathcal{I}_{i,\theta_i} \mathbf{a}'_i$ does not hold because $\mathbf{a}_i(\theta_{-i}) = c_i^{\theta}(T) \neq c_i^{\theta}(T \setminus c_i^{\theta}(T)) = \mathbf{a}'_i(\theta_{-i})$ but $\mathbf{a}_i(\tilde{\theta}_{-i}) = \mathbf{a}'_i(\tilde{\theta}_{-i}) = c_i^{\theta_i, \tilde{\theta}_{-i}}(T \setminus \{x\})$.

[28] An equivalence relation is a binary relation that is reflexive, symmetric, and transitive.

B Direct Mechanisms

The appeal of direct mechanisms in the mechanism design literature leads us to the following analysis, in which we focus on SCFs instead of SCSs (since direct mechanisms cannot coordinate selections of SCFs from an SCS). In Theorem 5, we present the resulting characterizations of quasi-robust behavioral implementation and ex-post behavioral implementation of an SCF via its direct mechanism.

Theorem 5. *Given a pair of associated interim and ex-post environments, let $f : \Theta \to X$ be an SCF, SCS $F := \{f\}$, and the profile of sets $\mathbb{F} := (\mathbf{F}_i(\alpha_{-i}))_{i\in N,\alpha_{-i}\in\Lambda_{-i}}$ where $\mathbf{F}_i(\alpha_{-i}) := \{\mathbf{f}^{\alpha}_{i,\theta_i} \mid \theta_i \in \Theta_i\}$ and $\mathbf{F}_i(\alpha_{-i})(\theta_{-i}) = f(\Theta_i, \alpha_{-i}(\theta_{-i}))$ for any $i \in N$ and any $\alpha_{-i} \in \Lambda_{-i}$. Then,*

(*i*) *f is (fully) quasi-robust behavioral implementable by its associated direct mechanism sustaining truthtelling as a BIE and an EPE if and only if the profile of sets $\mathbb{F}$ is quasi-robust consistent with F.*

(*ii*) *f is (fully) ex-post behavioral implementable by its associated direct mechanism possessing a truthful EPE if and only if the profile of sets $(f(\Theta_i, \theta_{-i}))_{i\in N,\theta_{-i}\in\Theta_{-i}}$ is ex-post consistent with F.*

Proof. For the *necessity* of (*i*) of the theorem, suppose f is quasi-robust behavioral implementable by its direct mechanism $\mu^f = (\Theta, g^f)$ with $g^f = f$. Let the truthful strategy profile α^{id} be BIE and EPE; so $f = g^f \circ \alpha^{\mathrm{id}}$. Define $\mathbb{F}$ as in the statement of the theorem. First, we need to show that $\mathbb{F}$ is closed under deception: Suppose $\mathbf{a}_i \in \mathbf{F}_i(\alpha_{-i})$ for some $\alpha_{-i} \in \Lambda_{-i}$. As α_{-i} amounts to a strategy profile of the individuals other than $i \in N$ in the direct mechanism, there is $\theta_i \in \Theta_i$ such that $\mathbf{a}_i = \mathbf{f}^{\alpha}_{i,\theta_i}$. As any $\tilde{\alpha}_{-i} \in \Lambda_{-i}$ is another strategy profile for the others, we have $\mathbf{a}^{\tilde{\alpha}}_i = \mathbf{f}^{\tilde{\alpha}\circ\alpha}_{i,\theta_i} \in \mathbf{F}_i(\tilde{\alpha}_{-i} \circ \alpha_{-i})$, establishing closedness under deception.

Since α^{id} is a BIE, $\mathbf{f}_{i,\theta_i} \in \mathbf{C}^{\theta_i}_i(\mathbf{F}_i(\alpha^{\mathrm{id}}_{-i}))$ for all $i \in N$ and $\theta_i \in \Theta_i$, establishing (*i.a*) of quasi-robust consistency of $\mathbb{F}$. Moreover, for all $i \in N$ and all $\theta_i \in \Theta_i$, $f(\theta_i, \theta_{-i}) = \mathbf{f}_{i,\theta_i}(\alpha^{\mathrm{id}}_{-i}(\theta_{-i})) \in c^{(\theta_i,\theta_{-i})}_i(\mathbf{F}_i(\alpha^{\mathrm{id}}_{-i})(\theta_{-i}))$ with $\mathbf{F}_i(\alpha^{\mathrm{id}}_{-i})(\theta_{-i}) = f(\Theta_i, \theta_{-i})$ for all $\theta_{-i} \in \Theta_{-i}$, establishing (*i.b*) of quasi-robust consistency of $\mathbb{F}$.

For (*ii*) of quasi-robust consistency, for any deception α with $f \circ \alpha \neq f$, $\alpha^{\mathrm{id}} \circ \alpha = \alpha$ cannot be a BIE of μ^f because otherwise $g^f \circ \alpha^{\mathrm{id}} \circ \alpha = f \circ \alpha$ and hence by (*ii*) of

quasi-robust behavioral implementation $f \circ \alpha$ equals f, a contradiction. Thus, there is there exists $i^* \in N$ and $\theta^*_{i^*} \in \Theta_{i^*}$ such that $\mathbf{f}^{\alpha}_{i^*,\theta^*_{i^*}} \notin \mathbf{C}^{\theta^*_{i^*}}_{i^*}(\mathbf{F}_{i^*}(\alpha_{-i^*}))$.

For the *sufficiency* of (i) of the theorem: By hypothesis, $\mathbb{F}$ is quasi-robust consistent with $F = \{f\}$. Thanks to quasi-incetive compatibility and quasi-ex-post incentive compatibility, truthtelling strategy profile α^{id} is both a BIE and an EPE of the direct mechanism. Further, if α^* is a BIE, then $g^f \circ \alpha^* = f$: Suppose not. Then, by (ii) of quasi-robust consistency of $\mathbb{F}$, there exist $i^* \in N$ and $\theta^*_{i^*} \in \Theta_{i^*}$ such that $\mathbf{f}^{\alpha}_{i^*,\theta^*_{i^*}} \notin \mathbf{C}^{\theta^*_{i^*}}_{i^*}(\mathbf{F}_{i^*}(\alpha^*_{-i^*}))$, contradicting to α^* being a BIE as $\mathbf{O}^{\mu^f}_{i^*}(\alpha^*_{-i^*}) = \mathbf{F}_{i^*}(\alpha^*_{-i^*})$. Therefore, the direct mechanism μ^f quasi-robust behavioral implements F sustaining truthtelling as a BIE and an EPE.

For the *necessity* of (ii) of the theorem, suppose f is ex-post behavioral implementable by its direct mechanism $\mu^f = (\Theta, g^f)$ with $g^f = f$. Let the truthful strategy profile α^{id} be an EPE so that $f = g^f \circ \alpha^{\text{id}}$. Let $i \in N$ and $\tilde{\theta}_{-i} \in \Theta_{-i}$. Then, $O^{\mu^f}_i((\alpha^{\text{id}}_{-i}(\tilde{\theta}_{-i})) = f(\Theta_i, \tilde{\theta}_{-i})$ implies $f(\theta_i, \tilde{\theta}_{-i}) \in c^{(\theta_i,\tilde{\theta}_{-i})}_i(f(\Theta_i, \tilde{\theta}_{-i}))$ for all $i \in N$ and all $\theta_i \in \Theta_i$, establishing (i) of ex-post consistency of $(f(\Theta_i, \theta_{-i}))_{i\in N,\theta_{-i}\in\Theta_{-i}}$. For (ii) of ex-post consistency, for any deception/strategy α with $f \circ \alpha \neq f$, $\alpha^{\text{id}} \circ \alpha = \alpha$ cannot be an EPE of μ^f because otherwise $g^f \circ \alpha^{\text{id}} \circ \alpha = f \circ \alpha$ and hence by (ii) of ex-post implementation $f \circ \alpha$ equals f, a contradiction. Thus, there is $i^* \in N$, $\theta^* \in \Theta$ with $f(\alpha(\theta^*)) \notin c^{\theta^*}_{i^*}(f(\Theta_{i^*}, \alpha_{-i^*}(\theta^*_{-i^*})))$ since $O^{\mu^f}_{i^*}((\alpha^{\text{id}}_j(\alpha_j(\theta^*_j))_{j\neq i^*}) = f(\Theta_{i^*}, \alpha_{-i^*}(\theta^*_{-i^*}))$.

For the *sufficiency* of (ii) of the theorem: By hypothesis, $(f(\Theta_i, \theta_{-i}))_{i\in N,\theta_{-i}\in\Theta_{-i}}$ is ex-post consistent with $F = \{f\}$. Then, α^{id} is a truthful EPE and $g^f \circ \alpha^{\text{id}} = f$ thanks to quasi-ex-post incentive compatibility (implied by ex-post consistency). Further, if α^* is an EPE, then $g^f \circ \alpha^* = f$; Otherwise, $g^f(\alpha^*(\theta)) \neq f(\theta)$ for some $\theta \in \Theta$. So, $f(\alpha^*(\theta)) \neq f(\theta)$ implies, by (ii) of ex-post consistency of $(f(\Theta_i, \theta_{-i}))_{i\in N,\theta_{-i}\in\Theta_{-i}}$, there is $i^* \in N$ and $\theta^* \in \Theta$ with $g^f(\alpha^*(\theta^*)) = f(\alpha^*(\theta^*)) \notin c^{\theta^*}_{i^*}(f(\Theta_{i^*}, \alpha^*_{-i^*}(\theta^*_{-i^*})))$, contradicting to α^* being an EPE as $O^{\mu^f}_{i^*}(\alpha^*_{-i^*}(\theta^*_{-i^*})) = f(\Theta_{i^*}, \alpha^*_{-i^*}(\theta^*_{-i^*}))$. ■

Finally, we would like to note that Example 2 displays the use of Theorem 5: Table 5 shows that $\mathbb{F}$ is quasi-robust consistent with $F = \{\langle xzzy\rangle\}$ whereas Table 7 displays that the profile $(f(\Theta_i, \theta_{-i}))_{i\in N,\theta_{-i}\in\Theta_{-i}}$ is ex-post consistent with $F = \{\langle xzzy\rangle\}$.

C Sufficiency for quasi-robust behavioral implementation with weak choice incompatibility

In what follows, we present the weak choice incompatibility, which is implied by the economic environment assumption of Jackson (1991) under rationality as shown in Barlo and Dalkıran (2023). To that regard, we need the following: For any pair of acts $\mathbf{a}_i, \tilde{\mathbf{a}}_i \in \mathbf{A}_i$, we define the *splicing* of $\mathbf{a}_i$ with $\tilde{\mathbf{a}}_i$ along a set $\Theta' \subset \Theta$ as follows[29]:

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

Definition 15. *The* **weak choice incompatibility** *condition holds in an interim environment whenever the following holds: If for any SCF $h \in H$ and any $\bar{\theta} \in \Theta$, a profile of sets of acts $(\tilde{\mathbf{A}}_i)_{i \in N}$ is such that*

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

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

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

In environments with *finite state spaces* and *robust-null alternatives*, we strengthen our sufficiency result with the help of the weak choice incompatibility.

Definition 16. *Given a pair of associated interim and ex-post environments, an alternative $z \in X$ is a* ***robust-null alternative of individiual*** *$i \in N$ if*

(i) for all $\theta_i \in \Theta_i$, $\mathbf{C}_i^{\theta_i}(\tilde{\mathbf{A}}_i) = \mathbf{C}_i^{\theta_i}(\tilde{\mathbf{A}}_i \cup \{\mathbf{a}_i^z\})$ for any non-empty $\tilde{\mathbf{A}}_i \subset \mathbf{A}_i$, and

(ii) for all $\theta \in \Theta$, $c_i^\theta(S) = c_i^\theta(S \cup \{z\})$ for any $S \in \mathcal{X}$.

In words, z is a robust-null alternative of player i if (i), $\mathbf{a}_i^z$, the constant act that results in z, does not affect the interim choices of any type of individual i whenever $\mathbf{a}_i^z$ is added to the set of acts under consideration, and (ii) the alternative z does not affect the ex-post choices of individual i at any state whenever z is added to the set of alternatives under consideration.

Below, we present our second sufficiency result.

[29] Recall that for any $x \in X$ and any individual $i \in N$, $\mathbf{a}_i^x \in \mathbf{A}_i^c$ is i's constant act resulting in x.

Theorem 6. *Suppose that the given pair of associated interim and ex-post environments is such that $|\Theta| < \infty$, $n \geq 3$, the weak choice incompatibility holds, and for each individual, there is a robust-null alternative. Then, if there exists a profile of sets of acts quasi-robust consistent with SCS F, then F is quasi-robust behavioral implementable.*

Proof of Theorem 6. Suppose $|\Theta| < \infty$, $n \geq 3$, the choice incompatibility holds, and $(z^i)_{i \in N}$ is a profile of null alternatives. Let F be an SCS for which the profile $\mathbb{S} := (\mathbf{S}_i(f, \alpha_{-i}))_{i \in N,\ f \in F,\ \alpha_{-i} \in \Lambda_{-i}}$ is quasi-robust consistent.

The mechanism we employ is as in Barlo and Dalkıran (2023) and the proof closely parallels the proof of Theorem 3 of that study in terms of its structure and methodology and is presented for purposes of completeness. For each individual $i \in N$, $M_i = (F \cup \{\emptyset\}) \times \Theta_i \times (\mathbf{A}_i \cup \{\emptyset\}) \times (H \cup \{\emptyset\}) \times \mathbb{N}$, and a generic message is denoted by $m_i = (m_i^1, \theta_i^{(i)}, \mathbf{a}_i^{(i)}, m_i^4, k^{(i)})$, and $g : M \to X$ is as specified in Table 10 for a given profile of null alternatives $(z^i)_{i \in N}$ with $j^* := \min\{i \in N \mid k^{(i)} \geq k^{(j)} \text{ for all } j \in N\}$.

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

Table 10: The outcome function of the mechanism for Theorem 6.

First, we show that condition (i) of quasi-robust implementability holds.

Claim 9. *For any $f \in F$, there is σ^f a BIE and an EPE of $\mu = (M, g)$ with $f = g \circ \sigma^f$.*

Proof. Take any $f \in F$, let $\sigma_i^f(\theta_i) = (f, \theta_i, \emptyset, \emptyset, 1)$ for each $i \in N$. Then, at every $\theta \in \Theta$, Rule 1 applies and $g(\sigma^f(\theta)) = f(\theta)$, i.e., $f = g \circ \sigma^f$.

For any unilateral deviation from σ^f, either Rule 1 or Rule 1′ or Rule 2 or Rule 2′ applies, while Rule 3 is not attainable. So, $\mathbf{O}_i^{\mu}(\sigma_{-i}^f) = \mathbf{S}_i(f, \alpha_{-i}^{\text{id}}) \cup \{\mathbf{a}^{z^i}\}$ for all $i \in N$.

Recall that, by $(i.a)$ of quasi-robust consistency , $\mathbf{f}_{i,\theta_i} \in \mathbf{C}_i^{\theta_i}(\mathbf{S}_i(f, \alpha_{-i}^{\text{id}}))$ for each $i \in N$ and each $\theta_i \in \Theta_i$. Because $\mathbf{a}^{z^i}$ is a null act of individual i, $\mathbf{f}_{i,\theta_i} \in \mathbf{C}_i^{\theta_i}(\mathbf{S}_i(f, \alpha_{-i}^{\text{id}}) \cup \{\mathbf{a}^{z^i}\})$. Hence, for all $i \in N$ and all $\theta_i \in \Theta_i$, $\mathbf{a}^*_{i,\theta_i} \in \mathbf{C}_i^{\theta_i}(\mathbf{O}_i^{\mu}(\sigma^f_{-i}))$ where $\mathbf{a}^*_{i,\theta_i}(\theta_{-i}) = g(\sigma_i^f(\theta_i), \sigma^f_{-i}(\theta_{-i}))$ for all $\theta_{-i} \in \Theta_{-i}$. Thus, σ^f is a BIE of μ such that $f = g \circ \sigma^f$.

Furthermore, for any $i \in N$, $\theta \in \Theta$, $\mathbf{O}_i^{\mu}(\sigma^f_{-i})(\theta_{-i}) = \mathbf{S}_i(f, \alpha_{-i}^{\text{id}})(\theta_{-i}) \cup \{z^i\}$ for all $i \in N$. Then, by $(i.b)$ of quasi-robust consistency, and z_i being a null-alternative for ex-post choices, we have for all $i \in N$ and all $\theta \in \Theta$, $\mathbf{a}^*_{i,\theta_i}(\theta_{-i}) c_i^{\theta}(\mathbf{O}_i^{\mu}(\sigma^f_{-i})(\theta_{-i}))$ where $\mathbf{a}^*_{i,\theta_i}(\theta_{-i}) = g(\sigma_i^f(\theta_i), \sigma^f_{-i}(\theta_{-i}))$ for all $\theta_{-i} \in \Theta_{-i}$. Thus, σ^f is a EPE of μ such that $f = g \circ \sigma^f$. ■

Take any BIE σ^* of μ denoted as $\sigma_i^*(\theta_i) = (m_i^1(\theta_i), \alpha_i(\theta_i), m_i^3(\theta_i), m_i^4(\theta_i), k_i(\theta_i))$ for each $i \in N$; i.e., $m_i^1(\theta_i)$ denotes the first component, $\alpha_i(\theta_i)$ the reported type, $m_i^3(\theta_i)$ denotes the third component, $m_i^4(\theta_i)$ the fourth component, and $k_i(\theta_i)$ the proposed integer by i when her realized type is θ_i.

Claim 10. *Under any BIE σ^* of μ, Rule 1 must apply at every state $\theta \in \Theta$, and there is a unique $f \in F$ such that $m_i^1(\theta_i) = f$ for all $i \in N$ and all $\theta_i \in \Theta_i$.*

Proof. Let the SCF that arises when individuals follow σ^* be given by h^*, i.e., $h^* := g \circ \sigma^*$. Therefore, individual i of type θ_i faces act $\mathbf{h}^*_{i,\theta_i}$ under σ^* and $\mathbf{h}^*_{i,\theta_i} \in \mathbf{C}_i^{\theta_i}(\mathbf{O}_i^{\mu}(\sigma^*_{-i}))$ for all $i \in N$ as σ^* is a BIE of μ.

Suppose Rule 1′ or Rule 2 or Rule 2′ or Rule 3 holds at $\bar{\theta}$ under σ^*. If Rules 1′, 2 or 2′ holds, let us denote the odd-man-out by $\bar{j}$. If Rule 3 applies, let $\bar{j} = 1$. Let $\ell \neq \bar{j}$ and $\bar{\Theta} := \{\theta \in \Theta \mid \theta_\ell = \bar{\theta}_\ell\}$. Let α_*^{-1} be a deception such that $h^* \circ \alpha_*^{-1}(\alpha(\theta)) = h^*(\theta)$ for all $\theta \in \Theta$.[30] For any $x \in X$, consider the deviation by ℓ to $\tilde{\sigma}_\ell$ such that $\tilde{\sigma}_\ell(\theta_\ell) = \sigma_\ell^*(\theta_\ell)$ for all $\theta_\ell \neq \bar{\theta}_\ell$ and $\tilde{\sigma}_\ell(\bar{\theta}_\ell) = (m_\ell^1(\bar{\theta}_\ell), \alpha_\ell(\bar{\theta}_\ell), \mathbf{h}^*_{\ell,\bar{\theta}_\ell} \circ \alpha_{*-\ell}^{-1}, [x/_{\bar{\Theta}} h^*] \circ \alpha_*^{-1}, k^*)$ where $k^* = \max_{i \neq \ell,\ \tilde{\theta}_i \in \Theta_i} k_i(\tilde{\theta}_i) + 1$, and $[x/_{\bar{\Theta}} h^*]$ is the entanglement of alternative x with SCF h^* along $\bar{\Theta}$ that is defined by $[x/_{\bar{\Theta}} h^*](\theta) = x$ for all $\theta \in \bar{\Theta}$ and $[x/_{\bar{\Theta}} h^*](\theta) = h^*(\theta)$ otherwise. That is, individual ℓ deviates from σ_ℓ^* only when her type is $\bar{\theta}_\ell$ so that the first and the second component of her deviation is the same as that of σ_ℓ^*, the third component is the act that ℓ faces under σ^* when her true type is $\bar{\theta}_\ell$ and the others are

[30] To identify such a function, let $\alpha^{-1}(\theta) \subset \Theta$ be the set of states that are mapped to θ under deception α. For any $\theta \in \Theta$, if $\alpha^{-1}(\theta) = \emptyset$, then set $\alpha_*^{-1}(\theta) = \theta$, otherwise pick an arbitrary $\bar{\theta} \in \alpha^{-1}(\theta)$ and set $\alpha_*^{-1}(\theta') = \bar{\theta}$ if $\alpha(\theta') = \theta$ for some $\theta' \in \Theta$.

reporting their types truthfully (as $\alpha_*^{-1} \circ \alpha = \alpha^{\mathrm{id}}$), the fourth component is the SCF obtained by composing the entanglement of alternative x with SCF h^* along $\bar{\Theta}$ with α_*^{-1}, and the last component is an integer that is higher than all proposed integers of the other individuals at all states, which is well-defined as $|\Theta| < \infty$.

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

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

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

Now, we employ weak choice incompatibility: Consider SCF $h^* = g \circ \sigma^* \in H$, $\bar{\theta} \in \Theta$, and the profile $(\tilde{\mathbf{A}}_i)_{i \in N}$ defined by $\tilde{\mathbf{A}}_i := \mathbf{O}_i^\mu(\sigma_{-i}^*)$ for all $i \in N$. Then, trivially, $\mathbf{h}^*_{i,\bar{\theta}_i} \in \tilde{\mathbf{A}}_i$ for all $i \in N$, and hence, (i) of hypothesis of weak choice incompatibility holds. For (ii) of the hypothesis of weak choice incompatibility, consider each individual $\ell \neq \bar{j}$ and let Θ' be the set of states such that the outcome changes to x after ℓ deviates unilaterally to $\tilde{\sigma}_\ell$. Observe that $\bar{\theta} \in \Theta'$ and for any $x \in X$, $[\mathbf{a}_\ell^x /_{\Theta'} \mathbf{h}^*_{\ell,\bar{\theta}_\ell}] \in \tilde{\mathbf{A}}_\ell$ for all $\ell \neq \bar{j}$ because, by definition, $[\mathbf{a}_\ell^x /_{\Theta'} \mathbf{h}^*_{\ell,\bar{\theta}_\ell}]$ is the act individual ℓ of type $\bar{\theta}_\ell$ faces after deviating to $\tilde{\sigma}_\ell$. Hence, by weak choice incompatibility, there is $i^* \in N$ with $i^* \neq \bar{j}$ such that $\mathbf{h}^*_{i^*,\bar{\theta}_{i^*}} \notin \mathbf{C}_{i^*}^{\bar{\theta}_{i^*}}(\tilde{\mathbf{A}}_{i^*})$, which means $\mathbf{h}^*_{i^*,\bar{\theta}_{i^*}} \notin \mathbf{C}_{i^*}^{\bar{\theta}_{i^*}}(\mathbf{O}_{i^*}^\mu(\sigma_{-i^*}^*))$, which contradicts σ^* being a

BIE of μ as $h^* = g \circ \sigma^*$.

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

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

To conclude the proof of the theorem, we show that (ii) of quasi-robust behavioral implementability also holds:

Claim 11. *For any BIE of σ^* of μ, $g \circ \sigma^* \in F$.*

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

References

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

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

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

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

Barlo, M., & Dalkıran, N. A. (2023). Behavioral implementation under incomplete information. *Journal of Economic Theory*, 105738.

Barlo, M., Dalkıran, N. A., & Dönmez, M. (2023). Anonymous implementation. *Mimeo*.

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.

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

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

Chen, Y.-C., Holden, R., Kunimoto, T., Sun, Y., & Wilkening, T. (2023). Getting dynamic implementation to work. *Journal of Political Economy*, *131*(2), 285–387.

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

Chen, Y.-C., Mueller-Frank, M., & Pai, M. M. (2022). Continuous implementation with direct revelation mechanisms. *Journal of Economic Theory*, *201*, 105422.

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. (2022). Departures from preference maximization, violations of the sure-thing principle, and relevant implications. *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.

de Clippel, G., Saran, R., & Serrano, R. (2019). Level-mechanism design. *The Review of Economic Studies*, *86*(3), 1207–1227.

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

Hayashi, T., Jain, R., Korpela, V., & Lombardi, M. (2023). Behavioral strong implementation. *Economic Theory*, 1–31.

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.

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

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

Jain, R., Lombardi, M., & Müller, C. (2023). An alternative equivalent formulation for robust implementation. *Games and Economic Behavior*.

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

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

Korpela, V. (2013). A simple sufficient condition for strong implementation. *Journal of*

Economic Theory, *148*(5), 2183—2193.

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

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

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

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.

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.

Ollár, M., & Penta, A. (2017). Full implementation and belief restrictions. *American Economic Review*, *107*(8), 2243–2277.

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

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

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

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

Rubbini, G. (2023). Mechanism design without rational expectations. *arXiv preprint arXiv:2305.07472*.

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.

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