Behavioral strong implementation*

T Hayashi[†] R Jain[‡] V Korpela[§] M Lombardi[¶]

March 13, 2023

Abstract

Choice behavior is rational if it is based on the maximization of some context-independent preference relation. This study re-examines the questions of implementation theory in a setting where players' choice behavior need not be rational and coalition formation must be taken into account. Our model implies that with non-rational players, the formation of groups greatly affects the design exercise. As a by-product, we also propose a notion of behavioral efficiency and we compare it with existing notions.

Keywords: Strong equilibrium, implementation, state-contingent choice rules, (behavioral) group strategy-proofness, non-rational behavior.
JEL: D11, D60, D83.

*We thank three anonymous referees for comments that have substantially improved the paper. Ritesh Jain greatly acknowledges financial support from the Ministry of Science and Technology Grant, Taiwan under the grant: 107-2410-H-001-113-MY2.

[†]Adam Smith Business School, University of Glasgow. E-mail: takashi.hayashi@glasgow.ac.uk.
[‡]University of Liverpool Management School. E-mail: ritesh.jain@liverpool.ac.uk.
[§]Turku School of Economics, University of Turku. E-mail: vipeko@utu.fi.
[¶]University of Liverpool Management School. E-mail: michele.lombardi@liverpool.ac.uk. Department of Economics and Statistics, University of Naples Federico II and CSEF.

1 Introduction

A cornerstone assumption in economics is that every player is "rational", in the sense he maximizes some context-independent preference relation. Thus, a rational player has self-control and is not moved by emotions or external factors; hence, he knows what is best for himself. Although this assumption is an important starting point for many analyses, it does not cover all cases. For instance, Spiegler (2011) adapts models in industrial organization to identify the optimal contracts that firms can offer to maximize their profits when their customers are subject to specific choice biases. In this paper, we study the effects of non-rational behavior in an implementation framework.

The theory of implementation under complete information investigates the goals that a principal can achieve when they depend on players' preferences. Although these preferences are commonly known among players, the principal does not know them, and players' objectives need not be aligned with that of the principal. The implementation problem consists of devising a mechanism in such a way that the equilibrium behavior of players always coincides with the principal's goal. When such a mechanism exists, the principal's goal is said to be (fully) implementable.[1]

de Clippel (2014), Korpela (2012), and Ray (2010), extend the theory of implementation to cases in which players can make choices that are at odds with the conventional assumption of rationality. This is done by (i) considering individual state-contingent choices rather than preferences as the primitive characteristics of a player and (ii) extending the idea of Nash equilibrium beyond the rational domain. This extension proposes that a strategy profile is a behavioral equilibrium if the resulting outcome is among each player's chosen options within the set of outcomes that he can generate through unilateral deviations. However, because the behavioral equilibrium is a strictly non-cooperative equilibrium, it is natural to consider the extent to which de Clippel's (2014) analysis carries over when coalitions of players

[1] For an introduction to the theory of implementation see Jackson (2001), Maskin and Sjöström (2002), Serrano (2004) and Thomson (1996).

can arrange mutually beneficial deviations.

It has been shown elsewhere that the problem of coalition formation with boundedly rational players can offer new perspective on existing problems. For instance, see Sandholm and Lesser (1997) and Reimer et al. (2020). Therefore, it is worth exploring the differential impact that the strong equilibrium refinement may have on implementability when participants are not fully rational.

To illustrate this point, let us consider the leading example of "building willpower in groups" provided by de Clippel (2014). Therefore, let us suppose that players have a common long-term goal and let us define willpower as the number of tempting options a player can overlook to better fulfill his long-term goal. Furthermore, let us suppose that this long-term goal is difficult to achieve because there are tempting outcomes and each player has limited willpower to exercise self-control. de Clippel (2014) shows that the long-term goal is behaviorally implementable by a mechanism which allows each player to be "in charge of" a limited number of outcomes. By contrast, when players can freely communicate and form coalitions and when their equilibrium behavior coincides with our extension of Aumann's (1959) notion of strong equilibrium, there is no way we can structure the interactions of players so that their equilibrium behavior will result in their common long-term goal (details are presented in Section 2). The reason is that the set of outcomes that the grand coalition of all players is "in charge of" is the set of all outcomes and there is no way that this coalition can exercise self-control over this set.

In this study, we thus extend the well-known cooperative counterpart of the Nash equilibrium, namely, the strong equilibrium proposed by Aumann (1959), beyond the rational domain (termed behavioral strong equilibrium herein) and consider implementation in this equilibrium. Though this extension can be made in many natural ways, ours is based on the following two points. First, we want that our notion of equilibrium coincides with the behavioral equilibrium notion of de Clippel (2014) when only unilateral deviations are allowed. Second, we want that our notion of equilibrium coincides with Aumann's notion on the rational domain.

In a strong equilibrium, no coalition, taking the strategies of its complement as given, can cooperatively deviate in a way that benefits all its members (Aumann, 1959). Thus, it is a strategy profile that is stable not only with respect to the unilateral deviations of every player, but also with respect to those of every coalition of players. Since this requirement applies to the grand coalition of all players, every strong equilibrium is weakly Pareto-optimal.[2]

We extend this equilibrium notion beyond rational domains as follows. A strategy profile is a behavioral strong equilibrium if it is a behavioral equilibrium and if no coalition, taking the strategies of its complement as given, can find an agreement such that all its members would pick the agreement out of their respective feasible sets and reject the outcome corresponding to the strategy profile (see the discussion presented below Definition 1 for more details). When only the unilateral deviations of single players are allowed, it coincides with the notion of the behavioral equilibrium.

Since a behavioral strong equilibrium is robust to deviations of the grand coalition, our notion of strong equilibrium incorporates a notion of Pareto efficiency. This notion extends the Pareto principle beyond rational domains. We introduce this notion of efficiency in Section 3 and compare it with other extensions already proposed in the literature. According to our extension, outcome x is behaviorally efficient if there exists a profile of sets $(E_i)_{i \in N}$, one for each player $i \in N$, such that player i selects x from the set E_i, while no extension of these sets $E_i \subseteq X_i$ can lead players to unanimously select some other outcome y and to reject outcome x.

We show that our efficiency concept is non-nested in de Clippel's (2014) extended notion of efficiency as well as in Bernheim and Rangel's (2009) extended notion. When players are rational, all these extensions of the Pareto principle yield the same set of Pareto-optimal outcomes. However, in general, none of them is implementable in the behavioral strong equilibrium.

In Section 5, we provide a necessary condition for implementation in the behav-

[2] An outcome is Pareto-optimal if it is feasible and there is no feasible outcome that would make everyone strictly better off.

ioral strong equilibrium, which extends Maskin monotonicity (Maskin, 1999) from rational domains to any domain of choice behavior. Maskin monotonicity is a necessary condition for implementation in both strong Nash equilibrium (Maskin, 1979) and Nash equilibrium (Maskin, 1999). Contrary to the rational case, we find that the necessary condition for implementation in the behavioral equilibrium, which is also an extension of Maskin monotonicity, is no longer necessary for implementation in the behavioral strong equilibrium (see our example on choice overload). This is somewhat surprising. In an economic environment with fully rational players, Maskin monotonicity provides a full characterization of all implementable rules. Since Maskin monotonicity is also necessary for strong implementation, nothing is gained from an implementation point of view when coalition formation is allowed. However, with non-rational players, coalitions matter even in economic environments.

Although our necessary condition is useful to delineate the limitations of implementation and can provide important insights, it is not sufficient. As in Maskin (1979), Dutta and Sen (1991), and Korpela (2013), more work is needed to identify, partly or totally, the class of implementable goals. In Section 6, we tackle sufficiency first in a competitive environment and then provide a simple sufficient condition in a more general environment when there are more than two agents.[3] This simple sufficient condition is an extension of the axiom of sufficient reason proposed by Korpela (2013) in the context of strong implementation in rational domains. The practical implications of these results are provided in Section 7. For instance, our necessary condition is also sufficient in a competitive environment.[4]

Related literature. Our contribution to the implementation literature is twofold; we adopt the non-standard solution concept of strong equilibrium and formulate it to

[3] The two-person case is studied in Hayashi et al (2020; Section 8).

[4] Hayashi et al (2020; Section 7) show that the set of behavioral Pareto-optima that no player finds less desirable than the status quo is implementable in the behavioral strong equilibrium. Unfortunately, as in Maskin (1979), they find that this set of outcomes is the only implementable set when the domain of players' state-contingent choice rules is unrestricted.

accommodate non-rational players. Implementation in strong equilibrium has been studied extensively under the assumption of rational players. Maskin (1979) shows that monotonicity is necessary for strong implementation, just as it is for Nash implementation. Dutta and Sen (1991) provide the first complete characterization, whereas Shu (1996) gives a complete characterization when there can be restrictions on feasible coalitions. There are also some other characterizations based on the unrestricted preference domain. Fristrup and Keiding (2001) give one of these characterizations basing it on the idea of effectivity functions. However, these characterization results are not intuitive, and the condition may be challenging to check in some environments. A simple sufficient condition for strong implementation, called the axiom of sufficient reason, is given by Korpela (2013). Savva (2018) provides sufficient conditions for a social choice rule to be implementable in strong Nash equilibrium in the presence of partially honest agents, that is, agents who break ties in favor of a truthful message when they face indifference between outcomes. Recently, Guo and Yannelis (2022) study belief-free full implementation, a variant of Aumann's strong equilibrium (1959) in an incomplete information setting when any coalition can form.

Our main focus, however, lies on the coalition formation of non-rational players. A growing literature suggests that individual choices are not always consistent with the maximization of some context-independent preference relation.[5] Since mechanisms are devised to provide players with an incentive to behave in accordance with the principal's goal, it is vital to base their design on choice models that describe individual choice behavior more accurately. It is not surprising, therefore, that a growing literature on the role of behavioral biases in economic design has accumulated in the past two decades (Spiegler, 2011).

Motivated by recent developments in the theoretical literature on non-rational bahvior, de Clippel (2014) extends the theory of Nash implementation to problems in which players' characteristics are described by their state-dependent choices. The

[5] See, for instance, Baigent and Gaertner (1996), Manzini and Mariotti (2007), Eliaz et al. (2011), Barberà et al. (2022) and Hayashi and Takeoka (2022).

first study to examine implementation problems in a choice framework is Hurwicz (1986), who shows that Maskin's (1999) classical results remain valid when each individual choice rule selects undominated outcomes according to some binary relation. Korpela (2012) (see also Ray, 2010) investigates what choice-consistency properties need to be satisfied by the individual choice rules so that the necessary and sufficient conditions of Moore and Repullo (1990) remain necessary for Nash implementability when players make state-dependent choices. He finds that a crucial role is played by Sen's (1971) property α, which states that if x is the selected from a set A, then it must be selected in all the subsets of A to which x belongs. Unfortunately, most of the choices made by non-rational players violate this property. In this paper, we study implementation problems in which non-rational players can freely form coalitions. The invoked game theoretic solution concept is that of the behavioral strong equilibrium.

Barlo and Dalkiran (2019) extend de Clippel's (2014) analysis to an environment with incomplete information and provide necessary and sufficient conditions for (ex post) behavioral implementation. Altun et al. (2020), following de Clippel (2014) and Matsushima (2008), study behavioral implementation problems in which the state-contingent choices of players are unknown to the principal and one of the players is inclined to report the true state-contingent choices of the players, but not the true state (of the world), when the truth does not pose any obstacle to his material well-being.

Other important studies in this strand of the literature are as follows. Eliaz (2002) studies Nash implementation problems in which players are "faulty" in the sense that they fail to act optimally. Cabrales and Serrano (2011) study implementation problems in an environment in which players myopically adjust their actions in the direction of better or the best responses. Saran (2011) extends the set of preferences to include menu-dependent preferences and characterizes the domain in which the revelation principle holds. Glazer and Rubinstein (2012) present a mechanism design model in which the ability of a player to manipulate the information he reports is affected by the content and framing of the mechanism. Bierbrauer and Netzer (2012) study the effect of introducing intention-based social preferences into mechanism design. Matsushima (2008) and Dutta and Sen (2012) study independently implementation problems with partially honest players, where a partially honest player has strict preferences for revealing the true state over lying when truth-telling does not lead to a worse outcome than that obtained when lying. Saran (2016) studies implementation under complete information when players are at most k-rational, where a k-rational player performs k steps of iterative elimination. Salant and Siegel (2018) study a model of contracts in which a profit-maximizing seller uses framing to influence buyers' purchase behavior. Finally, de Clippel et al. (2019) study the theoretical implications of level-k reasoning in mechanism design.

2 Notation and definitions

The environment consists of a collection of n players (we write N for the set of players), a set of possible states Θ, and a (non-empty) set of outcomes X. $\mathcal{X} = \{A \subseteq X | A \neq \varnothing\}$ is the collection of all possible non-empty subsets of X. We focus on complete information environments in which the true state is common knowledge among players but unknown to the principal. Player i's choice rule at state $\theta \in \Theta$ is a correspondence $C_i(\cdot;\theta) : \mathcal{X} \twoheadrightarrow \mathcal{X}$ that assigns a non-empty set of chosen outcomes $C_i(A,\theta) \subseteq A$ for each $A \in \mathcal{X}$. When $x \in C_i(A,\theta)$ and $y \in A \backslash C_i(A,\theta)$, we say that x is chosen and y could have been chosen from A but is rejected.

Player i's choice rule is rational at θ if there exists a complete and transitive (i.e., rational) preference relation R_i^θ such that $C_i(A,\theta) = \arg\max_{R_i^\theta} A$ for each $A \in \mathcal{X}$. A choice rule is rational at θ if and only if it satisfies Sen (1971)'s property α ("independence of irrelevant alternatives," or IIA) and property β.[6] If player i's

[6]Property α states that if x is "best" in a set, it is best in all subsets of it to which x belongs to. Property β states that if x and y are both best in A, a subset of B, then x is best in B if and only if y is best in B. When X is a finite set, propertis α and β are equivalent to the Weak Axiom of Revealed Preferences (WARP). Recall that WARP is not sufficient for the transitive rationalizability

choice rule is rational for every $\theta \in \Theta$, we simply say that player i is rational. Let $\bar{\Theta}$ denote the unrestricted rational domain, that is, the set of states where each player is rational.

Let (M, h) be a mechanism, where $M = (M_i)_{i \in N}$ and $h : M \to X$ is the outcome function. As usual, we refer to M_i as the strategy space of player i and to a member of M as a strategy profile. For any $m \in M$ and $i \in N$, let $m_{-i} = (m_1, ..., m_{i-1}, m_{i+1}, ..., m_n) \in M_{-i} \equiv (M_j)_{j \in N \setminus \{i\}}$ denote the strategies chosen by players other than i. We will write $m \in M$ as (m_i, m_{-i}).

A mechanism (M, h) induces a class of strategic games $\{(M, h, \theta) | \theta \in \Theta\}$, where (M, h, θ) stands for the game in which the set of players is N, the action space of player i is M_i, and player i's choice behavior is described by his choice rule $C_i(\cdot, \theta)$ at state θ.

When each player is rational at state θ, a strategy profile m is a *(Nash) equilibrium* of the game induced by the mechanism (M, h) at state θ if, for each player $i \in N$ and each $m'_i \in M_i$, the following is satisfied: $h(m) R_i^\theta h(m'_i, m_{-i})$. We denote the set of (pure strategy) equilibria of (M, h, θ) by $E(M, h, \theta)$.

Player i's opportunity set given a strategy profile $m_{-i} \in M_{-i}$ for the other players is given by $\mathbb{O}_i(m_{-i}) = \{h(m_i, m_{-i}) \in X | m_i \in M_i\}$. Following de Clippel (2014), for any state $\theta \in \Theta$, a strategy profile m is a *behavioral equilibrium* of the game induced by the mechanism (M, h) at state θ if $h(m) \in C_i(\mathbb{O}_i(m_{-i}), \theta)$ for each player $i \in N$. Write $BE(M, h, \theta)$ for the set of (pure strategy) behavioral equilibria of the strategic game (M, h, θ). It is clear that $BE(M, h, \theta) = E(M, h, \theta)$ if each player i is rational at state θ.

In a strong equilibrium, no coalition, taking the strategies of its complement as given, can cooperatively deviate in a way that benefits all its members (Aumann, 1959). Formally, let $\mathcal{N}_0$ denote the set of all non-empty subsets of N. Each group of players K (in $\mathcal{N}_0$) is a coalition. For any coalition K, any mechanism (M, h), and

of a choice rule when the collection $\mathcal{X}$ is arbitrary (Richter, 1971), even when the choice rule is a function. However, if the collection $\mathcal{X}$ includes all subsets of X of up to three elements, WARP is necessary and sufficient for transitive rationalizability (Arrow, 1959; Sen, 1971).

any strategy profile $m \in M$, let $m_K = (m_i)_{i \in K} \in M_K$ and $m_{-K} = (m_i)_{i \in N \setminus K} \in M_{-K}$ be the strategy profiles of the players inside K and outside K, respectively, such that $m = (m_K, m_{-K})$. If each player is rational at state θ, a strategy profile m is a *strong equilibrium* of the game induced by the mechanism (M, h) at state θ if for all $K \in \mathcal{N}_0$ and all $m'_K \in M_K$, there exists $i \in K$ such that $h(m) R_i^\theta h(m'_K, m_{-K})$. We denote the set of (pure strategy) strong equilibria of (M, h, θ) by $SE(M, h, \theta)$.

Coalition K's opportunity set given a strategy profile $m_{-K} \in M_{-K}$ for the other players is given by $\mathbb{O}_K(m_{-K}) = \{h(m_K, m_{-K}) \in X | m_K \in M_K\}$.

Definition 1 (Behavioral Strong Equilibrium, BSE) A strategy profile m is a *behavioral strong equilibrium* of the game induced by the mechanism (M, h) at state θ provided that the following requirements are satisfied.

(i) $h(m) \in C_i(\mathbb{O}_i(m_{-i}), \theta)$ for all $i \in N$.

(ii) For all $K \in \mathcal{N}_0$, with $|K| \geq 2$, and all $m'_K \in M_K$, there does not exist any profile of sets $(A_i)_{i \in K}$, with $\mathbb{O}_i(m_{-i}) \cup \{h(m'_K, m_{-K})\} \subseteq A_i \subseteq \mathbb{O}_K(m_{-K})$, such that for all $i \in K$, $h(m'_K, m_{-K}) \in C_i(A_i, \theta)$ and $h(m) \notin C_i(A_i, \theta)$.

Write $BSE(M, h, \theta)$ for the set of (pure strategy) behavioral strong equilibria of the strategic game (M, h, θ).

Our equilibrium notion is built around the notion of behavioral equilibrium proposed by de Clippel (2014). Indeed, a behavioral strong equilibrium is a behavioral equilibrium in which no coalition, taking the actions of its complement as given, can cooperatively deviate in a way that benefits all its members. When only unilateral deviations are allowed, the two equilibrium notions coincide.

In principle, two broad approaches can be followed to generalize the strong equilibrium of Aumann (1959) into a behavioral setting. One approach consists of generalizing the equilibrium concept directly and accepting the Pareto principle implied by this generalization. Another approach consists of generalizing the Pareto principle and defining behavioral strong equilibrium by requiring that the equilibrium outcome be efficient within each coalition's opportunity set. Multiple paths can be followed for each of these approaches. In this paper, we have started exploring the first approach by proposing a notion of behavioral strong equilibruim, built around the idea of a behavioral equilibrium. Below, we explain what our line of thinking was.

To better understand why in part (ii) of our equilibrium notion we consider subsets of coalition K's opportunity set, let us consider a two-player situation in which both players are rational and the set of outcomes is $X = \{x, y, z\}$. Players' rational preference relations are represented in the table below:

R_1	R_2
y	z
z	y
x	x

where, as usual, $\substack{a \\ b}$ for player i means that he strictly prefers a to b.[7] Let us consider the following two-player mechanism, where the two rows correspond to the two possible (pure) strategies of player 1 and the three columns correspond to the three possible (pure) strategies of player 2, and where in each box is the outcome assigned to the strategy profile to which the box corresponds.

	m_2	m_2'	m_2''
m_1	x	x	x
m_1'	x	z	y

By examining all the possible strategy profiles, we see that (m_1', m_2') is the unique strong equilibrium. Since the strategy profile (m_1, m_2) is an equilibrium but not a strong equilibrium, the grand coalition $\{1, 2\}$ should be able to find a strategy profile that all its members prefer to (m_1, m_2). Since the opportunity set of the grand coalition is X, players should be able to cooperatively deviate in a way that benefits everyone. However, players do not make the same choice from X because

[7] Throughout the paper, we use this convention.

player 1 selects only y from X and player 2 selects only z from X. To let them cooperatively deviate, there must exist subsets $A_1, A_2 \subseteq X$ such that $x \in A_1 \cap A_2$, the intersection $C_1(A_1) \cap C_2(A_2)$ is not empty and $x \notin C_1(A_1) \cup C_2(A_2)$. These subsets can be $A_1 = \{x, z\}$ and $A_2 = \{x, y, z\}$. Only in this way players display the same choice behavior, in the sense that each of them chooses the same outcome from a subset of X containing x and rejects x. The strong equilibrium implicitly considers these subsets in its definition. In a sense, it assumes that if there is a way to find a compromise, players will find it. In our setting, in which individual choice behavior is captured by a choice rule, we need to refer to subsets of coalitional opportunity sets explicitly in part (ii) of our definition. These sets model compromise making, and just like in Aumann (1959), we assume that if there is any way to find a compromise, players will find it.[8] There are two important assumptions behind Definition 1.

1. In laboratory experiments, subjects often change their behavior once they are told how their behavior violates axioms of rationality. We assume that the process of coalition formation does not affect players' choice behaviors. Specifically, it does not make players realize how their choice behavior is irrational.

2. Players don't need to compromise on those outcomes that they can obtain by unilateral deviations; i.e., $\mathbb{O}_i(m_{-i}) \subseteq A_i$ (these outcomes are available despite of others).

As we show below, when players are rational, our equilibrium notion is equivalent to strong equilibrium.

Lemma 1 Suppose that each player i is rational at state θ. Then, $BSE(M, h, \theta) = SE(M, h, \theta)$.

[8] There are other logically possible agreement structures. It is also imaginable that sometimes the principal can control the agreement structure itself. These are interesting questions that go beyond the scope of this paper.

The goal of the principal is to implement a social choice rule (SCR) φ, which is a rule $\varphi : \Theta \twoheadrightarrow X$ such that $\varphi(\theta)$ is non-empty for every $\theta \in \Theta$. We refer to $x \in \varphi(\theta)$ as a φ-optimal outcome at θ. The image or range of φ is the set $\varphi(\Theta) \equiv \{x \in X | x \in \varphi(\theta) \text{ for some } \theta \in \Theta\}$. For any two SCRs, φ and φ', we say that φ' is a selection from φ, denoted by $\varphi' \subseteq \varphi$, if $\varphi'(\theta) \subseteq \varphi(\theta)$ for all $\theta \in \Theta$.

A mechanism (M, h) *behaviorally implements* φ provided that for all $\theta \in \Theta$, $\varphi(\theta) = h(BE(M, h, \theta)) \equiv \{h(m) | m \in BE(M, h, \theta)\}$. If such a mechanism exists, φ is said to be behaviorally implementable.

Definition 2 (Behavioral Strong Implementation) A mechanism (M, h) behaviorally strongly implements φ provided that for all $\theta \in \Theta$, $\varphi(\theta) = h(BSE(M, h, \theta)) \equiv \{h(m) | m \in BSE(M, h, \theta)\}$. If such a mechanism exists, φ is said to be behaviorally strongly implementable.

3 Efficiency

With rational players, every strong equilibrium must be (weakly) Pareto-optimal within the entire feasible outcome space of the game. An outcome is Pareto-optimal if it is feasible and no feasible outcome would make everyone strictly better off. Since a behavioral equilibrium is strong if no coalition, taking the play of its complement as given, can cooperatively deviate in a way that benefits all of its members, our equilibrium notion also incorporates a notion of efficiency, which extends the Pareto principle to choice behaviors that are not consistent with the optimization of some rational preference relation. In this section, we introduce our extension of the Pareto principle and compare it with other extensions already proposed in the literature.

Definition 3 (Behavioral Efficiency) We say that outcome x is behaviorally efficient at $\theta \in \Theta$ if there exists a profile of sets $(E_i)_{i \in N}$ such that (1) $x \in C_i(E_i, \theta)$ for all $i \in N$, and (2) there do not exist any profile $(Y_i)_{i \in N}$ and any $y \in X$, with $E_i \subseteq Y_i$ for all $i \in N$, such that $x \notin C_i(Y_i, \theta)$ and $y \in C_i(Y_i, \theta)$, for all $i \in N$. When this is the case we say that x is behaviorally efficient with respect to $(E_i)_{i \in N}$ *at* θ. Furthermore, for any $Z \subseteq X$, such that $x \in Z$, we say that x is behaviorally efficient at θ in the set Z if this definition is satisfied when $X = Z$.

Behaviorally efficient solution, BE. For all $\theta \in \Theta$,

$BE(\theta) \equiv \{x \in X | \; x \text{ is behaviorally efficient at } \theta\}$.

The set E_i in the definition of behavioral efficiency imposes a restriction on the set Y_i that can be used to evaluate whether an outcome is efficient or not We view this as a framing of the choice. If players are rational at θ, and x is Pareto-optimal, then we can select $E_i = \{x\}$ for all $i \in N$ to show that x is behaviorally efficient. Furthermore, if x is evaluated as behaviorally efficient with respect to any profile of sets $(E_i)_{i \in N}$, it must be Pareto-optimal. However, if players are not rational at state θ, then the framing matters. This motivates the following two definitions. If x is behaviorally efficient at θ with respect to the profile of sets $(\{x\})_{i \in N}$, then we call x behaviorally efficient of type I, and write $x \in BE_I(\theta)$. This means that x can be deemed efficient without a need for framing. If x is behaviorally efficient at θ, but not with respect to $(\{x\})_{i \in N}$, then we call x behaviorally efficient of type II, and write $x \in BE_{II}(\theta)$. This means that x cannot be deemed efficient without framing. It is now clear that $BE(\theta) = BE_I(\theta) \cup BE_{II}(\theta)$.

Notice that the BE solution is always non-empty. Any outcome that some player selects from X is behaviorally efficient. However, both $BE_I(\theta)$ and $BE_{II}(\theta)$ can be empty, although not at the same time, while there can also exist outcomes of both efficiency types at the same state.

It is important to understand that BE is an extension of the Pareto principle only in a technical sense. In our model, choice behavior does not necessarily reveal anything about true efficiency. Sometimes efficiency is a function of the state in the sense that the designer knows true preferences at each state and also knows whether the choice behavior of players does or does not reflect them. However, it can also be that the designer does not know anything about true preferences. Rather, he only knows what he want's to implement at each state, and what the choice behavior of players is.

The following corollary of Theorem 2 (presented below) shows that implementation in behavioral strong equilibrium is strongly connected to the above notion of efficiency.

Corollary 1 If φ is behaviorally strongly implementable, then there must exist a set $Y \subseteq X$ such that φ is a selection of BE when defined on Y.

Proof. Let $\mathcal{O} = \{\mathcal{O}_K(x,\theta) \in \mathcal{X} \mid \theta \in \Theta, x \in \varphi(\theta), K \in \mathcal{N}_0, x \in \mathcal{O}_K(x,\theta)\}$ be any collection of opportunity sets that is coalitionally consistent with φ and let Y be the set $\mathcal{O}_N(x,\theta)$. Select any state θ and any outcome $x \in \varphi(\theta)$. Item (iii) in Definition 5 with coalition $K = N$ shows that x is behaviorally efficient with respect to the sets $(E_i)_{i\in N} = (\mathcal{O}_i(x,\theta))_{i\in N}$. ■

We show below that behavioral efficiency is necessary for behavioral strong implementation. In light of this result, Corollary 2 reveals the connection that must exist between any subselection of the Pareto rule and its implementation in behaviour strong equilibria. The reason is that in many real-life situations, agents make non-rational choices, even though they have well-defined preference orders over outcomes. For instance, this happens when people have limited willpower to exercise self-control, as discussed in Section 4. Therefore, in situations like these, the Pareto rule is well-defined. Then, suppose that the planner's goal is to implement in behavioral strong equilibria a subsolution φ of the Pareto rule. Corollary 2 implies that his goal can only be achieved if φ is a subsolution of BE. Finally, we show in Section 4 that there are behaviorally implementable goas that are not subselections of the BE solution.

Comparing BE with other extensions of the Pareto principle

The question of how to extend the Pareto principle beyond the rational domain has been debated in the recent literature. Bernheim and Rangel (2009) propose the following extension of the Pareto rule that is based on the idea of revealed preferences. Following their definition, we say that x is preferred to y at state θ, denoted by $xP^{\theta}_{BR}y$, if and only if for every player $i \in N$, $y \notin C_i(A, \theta)$ for all $A \in \mathcal{X}$ such that $x \in A$. Outcome x is Bernheim–Rangel efficient at state θ if and only if no outcome is preferred to x.

Bernheim-Rangel Pareto solution, PO^{BR}. For all $\theta \in \Theta$,

$$PO^{BR}(\theta) \equiv \left\{x \in X | \text{there exists no } y \in X \setminus \{x\} \text{ such that } yP^{\theta}_{BR}x\right\}.$$

de Clippel (2014) proposes the following refinement of the Bernheim–Rangel efficiency. According to de Clippel (2014), x is de Clippel efficient if there exists a collection of implicit opportunity sets, one for each player, such that each player would choose x from his own implicit opportunity set and all the outcomes have been accounted for in the sense that any outcome in X belongs to the opportunity set of at least one player. Formally,

de Clippel Pareto solution, PO^{dC}. For all $\theta \in \Theta$,

$$PO^{dC}(\theta) \equiv \left\{ x \in X \,\middle|\, \begin{array}{l} \text{there exists } (A_i)_{i \in N} \in \mathcal{X}^n \text{ such that } x \in C_i(A_i, \theta) \\ \text{for all } i \in N, \text{ and } X = \bigcup_{i \in N} A_i \end{array} \right\}.$$

de Clippel efficiency generalizes the idea of a lower contour set to behavioral domain: Outcome x is efficient if all other outcome are in the lower contour set of x for at least one player.[9] de Clippel explains the connection between his efficiency

[9]Some results in the literature suggest that the idea of lower contour set makes no sense unless

concept and BR-efficiency; PO^{dC} is a selection from PO^{BR}.[10]

Like de Clippel efficiency, also BE has a connection to the idea of a lower contour set. In the rational domain, if x is Pareto efficient at θ, the collection of sets $(E_i)_{i\in N}$ in the definition of BE can be selected as lower contour sets of x for each player at θ to show that x is efficient. However, in a behavioral domain, even if x is efficient, there may not exist any collection of sets $(E_i)_{i\in N}$ that cover X entirely. In fact, we can define a third efficiency notion that combines the idea behind BE and de Clippel efficiency, by requiring that outcome x is efficient if, and only if, it is BE with respect to some collection of sets $(E_i)_{i\in N}$ that cover X entirely. The downside with this definition is that efficient outcomes may not always exist. It can also happen that de Clippel efficiency PO^{dC} is a selection from BE, if, for example, the sets $(A_i)_{i\in N}$ can be selected in such a way that they also satisfy the definition of BE, which guarantees that a selection from BE is then behaviorally implementable (de Clippel, 2014, Proposition 3, p. 2985).

The next two examples show that the BE solution is not nested either with PO^{dC} nor PO^{BR}.

Example 1 Let $X = \{x, y, z\}$ and $N = \{1, 2\}$. The choice behavior of player 1 at state θ is such that $C_1(\{x, y\}, \theta) = \{y\}$, $C_1(\{x, z\}, \theta) = \{x\}$, and $C_1(\{x, y, z\}, \theta) = \{y\}$, while the choice behavior of player 2 is such that $C_2(\{x, y\}, \theta) = \{y\}$, $C_2(\{x, z\}, \theta) = \{x\}$, and $C_2(\{x, y, z\}, \theta) = \{z\}$. Selecting $E_1 = \{x\}$ and $E_2 = \{x, z\}$ shows that x is behaviorally efficient. However, since both players select y from the set $\{x, y\}$, it is not BR-efficient, and hence not de Clippel efficient.

Example 2 Let $X = \{x, y, z\}$ and $N = \{1, 2\}$. The choice behavior of player 1 at state θ is such that $C_1(\{x, y\}, \theta) = \{x\}$ and $C_1(X, \theta) = \{z\}$, while the choice behavior

Sen's property α holds. Korpela (2012) shows that the characterization of Nash implementable SCRs given in Moore and Repullo (1990), which is firmly based on the idea of lower contour sets, holds as long as property α is assumed, but not after that. The reason is that even if outcome x is selected from set A, it may not be selected from every subset unless property α holds.

[10]For all $\theta \in \Theta$, $PO^{dC}(\theta) \subseteq PO^{BR}(\theta)$, and for some $\theta \in \Theta$, $PO^{dC}(\theta)$ is a proper subset of $PO^{BR}(\theta)$ (de Clippel, 2014, Proposition 5).

of player 2 at state θ is such that $C_2(\{x, z\}, \theta) = \{x\}$ and $C_2(X, \theta) = \{z\}$ (we don't need to know more about the behavior). Selecting $A_1 = \{x, y\}$ and $A_2 = \{x, z\}$ shows that x is de Clippel efficient, and hence also BR-efficient. However, since both players select only z from X, it cannot be behaviorally efficient.

Previous example highlights one important difference between these efficiency concepts. If all players select one and the same outcome from X at state θ, that is $C_i(X, \theta) = \{x\}$ for all $i \in N$, then x is the unique behaviorally efficient outcome at θ. As the example shows, this is not true for de Clippel efficiency or for BR-efficiency.

When players are rational, the above extensions of the Pareto principle yield the same set of Pareto-optimal outcomes. The following theorem is proved by showing that the Pareto rule is not in general implementable in BSE, and therefore, no extension proposed in the literature is.

Theorem 1 *No extension of the Pareto principle is behaviorally strongly implementable in general.*

4 Behavioral Implementation vs Behavioral Strong Implementation

This section presents two examples showing that implementation in behavioral strong equilibrium is not logically related to behavioral implementation.

Choice Overload

There are two players $N = \{1, 2\}$. A mechanism designer wants to select some outcome that both players like from the set $X = \{w, x_1, x_2, \ldots, x_m\}$. Preferences are linear ordering over X at all states. Any pair of orderings that rank w as the worst outcome is feasible. Player 1 selects from any choice set the outcome that he prefers most. Player 2, on the other hand, suffers from a bias called *choice* *overload.*[11] From any choice set $A \in \mathcal{X}$ that includes at most k outcomes from $\{x_1, x_2, \ldots, x_n\} = X \backslash \{w\}$, player 2 selects the outcome that he prefers most, but if the choice set A contains more than k outcomes, then he selects all of them, that is, the set $A \setminus \{w\}$.[12] Let us assume that $m > k > 2$.

Suppose that the designer wants to implement an efficient outcome. This is not hard in principle. The designer could simply let player 1 (the rational player) select his or her best outcome from X. This is not fair, however, since this outcome could be the worst outcome of player 2. Is there a way to select an efficient outcome in a more just way? One way to satisfy both players, to some extent at least, is to first delete $k-2$ outcomes that are the least preferred by player 2 from the set $\{x_1, x_2, \ldots, x_m\}$, and then select all Pareto-optimal outcomes from the remaining outcomes. Let us denote this SCR by F. Formally, let $r_j(\theta)$ be the outcome that player 2 ranks jth at state θ, and $PO : \Theta \times \mathcal{X} \to X$ a correspondence that selects all Pareto-optimal outcomes $PO(\theta, A)$ from any choice set $A \in \mathcal{X}$ at a given state $\theta \in \Theta$. Then

$$F(\theta) = PO\big(\, \{r_1(\theta), r_2(\theta), \ldots, r_{m-k+2}(\theta)\}\, \big).$$

This SCR is not behaviorally implementable: A consistent collection of opportunity sets does not exist. To see this, suppose that at state θ preferences are

$$P_1^\theta = x_m > x_{m-1} > \cdots > x_2 > x_1 > w \quad \text{and} \quad P_2^\theta = x_1 > x_2 > \cdots > x_{m-1} > x_m > w.$$

Thus $F(\theta) = \{x_1, x_2, \ldots, x_{m-k+2}\}$. Let $\mathcal{O}_2(x_1, \theta)$ be any player 2's set of outcomes (in $\mathcal{X}$) that depends on $\theta \in \Theta$ and $x_1 \in F(\theta)$. The set $\mathcal{O}_2(x_1, \theta)$ must include x_1 and at most $k-1$ outcomes from $\{x_2, x_3, \ldots, x_m\}$. Otherwise player 2 would select x_1 from this set at all possible states $-$ even when it is among the $k-2$ least preferred outcomes. Let $x_h \in X \setminus \{w\}$ be any outcome such that $x_h \notin \mathcal{O}_2(x_1, \theta)$. Now consider state θ' where preferences are instead

$$P_1^{\theta'} = P_1^\theta \quad \text{and} \quad P_2^{\theta'} = x_h > x_1 > x_2 > \cdots > x_{m-1} > x_m > w.$$

[11] This term was originally coined in Toffler (1970).

[12] You can think of him selecting randomly so that any outcome is possible.

Ranking $P_2^{\theta'}$ is the same as ranking P_2^{θ} except that x_h has been raised to the top. Then $x_1 \notin F(\theta')$ since x_h Pareto dominates it. However $\{x_1\} = C_2(\mathcal{O}_2(x_1, \theta), \theta')$ – a contradiction since preferences of player 1 have not changed.

Despite of this, F is behaviorally strongly implementable. Consider the following mechanism (M, h): Player 1 can select any choice set $A \in \mathcal{X}$ that includes exactly $k-1$ outcomes from the set $\{x_1, x_2, \ldots, x_m\}$. Player 2 can select any single outcome x from X. If $x \in A$, then x is implemented, if $x \notin A$, then w is implemented. Let θ be any state. If player 2 selects an outcome $x \in F(\theta)$, and player 1 selects a choice set A that includes x together with $k-2$ outcomes which player 2 considers worse than x, we have a behavioral strong equilibrium of (M, h) with outcome x. Neither player can improve unilaterally from this strategy profile. Furthermore, the only way that they could jointly improve is that there exists an outcome y which both prefer to x at θ. This, however, is not possible by the definition of F. Hence, for any state θ, we have that $F(\theta) \subseteq h\left(BSE(M, h, \theta)\right)$. To the other direction, notice that a strategy profile (A, x), such that $x \notin A$, can never be a behavioral strong equilibrium. Player 1 would simply change his announcement to a choice set B such that $x \in B$. Suppose, then, that (A, x) is a behavioral strong equilibrium. Since A includes exactly $k-1$ outcomes, x must be preferred to at least $k-2$ outcomes by player 2 – otherwise he would deviate unilaterally to some outcome in A. On the other hand, if x would not be Pareto-optimal, say dominated by y, then player 1 could offer player 2 a joint deviation to y. This amounts to selecting from $A \cup \{y\}$, a choice set with k outcomes, so choice overload does not kick in and player 2 would select y. Therefore $h(BSE(M, h, \theta)) \subseteq F(\theta)$ by definition of F.

This example shows how the possibility to form coalitions allows the rational player to exert control over the biased player, that is not otherwise possible.

Building willpower in groups

Suppose that players have a common long-term goal, which is difficult to achieve due to the presence of tempting outcomes: each player's decisions are affected by a short-term craving. In other words, each player has limited willpower to exercise self-control. Player i's willpower is captured by the number of tempting outcomes that he can overlook to better fulfill his long-term goal. More precisely, given an ordering $\succ_L$ over X capturing the long-term goal, an ordering $\succ_{S,i}$ capturing player i's short-term craving, and an integer k_i denoting player i's willpower, player i's choice out of any $A \in \mathcal{X}$ is the most preferred outcomes according to $\succ_L$ among those dominated by at most k_i outcomes according to $\succ_{S,i}$.

A decision-maker with limited willpower typically makes choices that violate IIA. For instance, suppose that there are only three outcomes in $X = \{x, y, z\}$, that your long-term ranking is $x \succ_L y \succ_L z$, and that your short-term craving is captured by $z \succ_S y \succ_S x$. Suppose that you are able to exercise self-control as long as there is at most one tempting option. Thus, you would choose $\{y\}$ from X and $\{x\}$ from $\{x, y\}$.

Suppose that a state $\theta = \left(\succ_L, (\succ_{S,i})_{i\in N}\right)$ describes a common long-term goal and players' short-term cravings. de Clippel (2014) shows a way to combine the players' limited willpower to help them better fulfill their common long term goal. The idea is to decentralize the burden of choice by allowing each player to be "in charge of" only a small number of outcomes. The mechanism implementing the common long-term goal can be described as follows. Let $A_i \subseteq X$ be the set of k_i outcomes of which player i is in charge. Suppose that $\sum_{i\in N} k_i \geq |X|$, so that the union of the sets of outcomes assigned to the players can cover X: that is, $X = \bigcup_{i\in N} A_i$. The strategy space of player i is $M_i = A_i \times \mathbb{Z}_+$, where $\mathbb{Z}_+$ is the set of nonnegative integers. The interpretation is that player i chooses a message in support of an outcome in A_i as well as a nonnegative positive integer describing the intensity with which he makes the announcement. The selected outcome is the outcome supported by the player with the most intense report (using a fixed tie-breaking rule when players announce the same highest intensity).

In this context, de Clippel (2104) proves the following result; if $\sum_{i\in N} k_i \geq |X|$, then the SCR that selects systematically the top-choice of the common long-term goal is behavioral implementable (de Clippel, 2014; p. 2981). Unfortunately, there is no way to combine the players' limited willpower to help them better fulfill their common long term goal when players can form coalitions. The reason is that the grand coalition can be "in charge of" a set of outcomes over which players are unable to exercise self-control, even though each player, individually, is "in charge of" only a small number of outcomes. To see this, suppose three outcomes in $X = \{x, y, z\}$ and three players in $N = \{1, 2, 3\}$. Further, suppose that there exists a feasible state $\theta = \left(\succ_L, (\succ_{S,i})_{i \in N}\right)$ according to which the common long term goal is described by the ordering $x \succ_L y \succ_L z$, and their short-term cravings are captured by

$\succ_{S,1}$	$\succ_{S,2}$	$\succ_{S,3}$
z	y	z
y	z	y
x	x	x

.

Suppose that player i's willpower is $k_i = 1$ for each player $i \in N$. Let us assume that the principal knows the willpower of players, but not the true state.

To show that the possibility of forming coalitions defeats the idea of decentralizing the burden of choice, it suffices to show the SCR that selects systematically the top choice of the common long-term goal is behavioral implementable but not a sub-solution of the BE solution. Since the range of the SCR is X, the set of outcomes that the grand coalition is "in charge of" is X. At state θ, each player picks only the outcome y out of the set X and rejects the common long-term goal x, which shows that y is behaviorally efficient at state θ (select $E_i = X$ for all $i \in N$), while x is not (for any $(E_i)_{i \in N}$ select $Y_i = X$ for all $i \in N$).

5 Necessity

In this section, we provide a necessary condition for behavioral strong implementation, which helps us identify systematically whether an SCR is behaviorally strongly implementable. de Clippel (2014) finds that the extension of the idea of Nash implementation beyond the rational domain leads to a necessary condition for implementation known as *consistency*.

Definition 4 (De Clippel, 2014; p. 2982) A collection $\mathcal{O} = \{\mathcal{O}_i(x,\theta) \in \mathcal{X} \mid \theta \in \Theta, x \in \varphi(\theta), i \in N\}$ of opportunity sets is *consistent with* φ if:

(i) For all $\theta \in \Theta$, all $x \in \varphi(\theta)$ and all $i \in N$, $x \in C_i(\mathcal{O}_i(x,\theta),\theta)$.

(ii) For all $\theta, \theta' \in \Theta$ and all $x \in \varphi(\theta)$, if $x \in C_i(\mathcal{O}_i(x,\theta),\theta')$ for all $i \in N$, then $x \in \varphi(\theta')$.

Studying implementation in the Nash equilibrium is based on Maskin (1999; circulated since 1977), who proves that any SCR that can be Nash implemented satisfies a remarkably strong invariance condition, now widely referred to as Maskin monotonicity. The above condition is an extension of Maskin monotonicity beyond the rational domain.[13] Suppose that φ is behaviorally implementable. If x is a behavioral equilibrium at θ, the equilibrium strategy profile m supporting it defines an opportunity set for each player i, denoted by $\mathcal{O}_i(x,\theta)$, which represents the set of outcomes that player i can generate by varying his own strategy, keeping the other players' equilibrium strategies fixed at m_{-i}. From the definition of the behavioral equilibrium, each player i must choose x from $\mathcal{O}_i(x,\theta)$ at θ. Moreover, if there is an alternative state θ' such that every player i chooses x from $\mathcal{O}_i(x,\theta)$ at θ', then m forms a behavioral equilibrium at θ'. Hence, x is still a φ-optimal outcome at θ' if φ is behaviorally implementable.

The idea of extending the notion of the strong equilibrium beyond the rational domain leads to a necessary condition, called *coalitional consistency*. Let us present this from the viewpoint of necessity.

Suppose that φ is behaviorally strongly implementable by a mechanism (M,h). Let m be a behavioral strong equilibrium at θ whose associated outcome $h(m)$ co-

[13] See Lemma 3 in Barlo and Dalkiran (2022) for a proof.

incides with an element x of $\varphi(\theta)$. The equilibrium strategy profile defines an opportunity set for coalition K, denoted by $\mathcal{O}_K(x,\theta)$, by varying the strategies of the players in K, while keeping the other players' equilibrium strategies fixed at m_{-K}. For the grand coalition N, its opportunity set coincides with the entire feasible outcome space of the game, denoted by Y. From the definition of the behavioral strong equilibrium, each player i chooses x from $\mathcal{O}_i(x,\theta)$ at θ, and no coalition K with at least two players can find an outcome $y \in \mathcal{O}_K(x,\theta)$ and a profile of subsets $(A_i)_{i\in K}$ of $\mathcal{O}_K(x,\theta)$ where $\mathcal{O}_i(x,\theta) \cup \{y\} \subseteq A_i$ for all $i \in N$, such that each member i of K chooses y from A_i and rejects $x \in A_i$ at θ.

Take any alternative state θ' such that each player chooses x from $\mathcal{O}_i(x,\theta)$ at this state θ', so that m is still stable in terms of unilateral deviations. In addition, if no coalition K with at least two players can find an outcome $y \in \mathcal{O}_K(x,\theta)$ and a profile of subsets $(A_i)_{i\in K}$ of $\mathcal{O}_K(x,\theta)$ where $\mathcal{O}_i(x,\theta) \cup \{y\} \subseteq A_i$ for all $i \in N$, such that each member i of K chooses y from A_i and rejects $x \in A_i$ at θ', clearly m forms a behavioral strong equilibrium at θ' as well. Hence, x is a φ-optimal outcome at θ' since φ is behaviorally strongly implementable. Formally,

Definition 5 (Coalitional Consistency) A collection

$$\mathcal{O} = \{\mathcal{O}_K(x,\theta) \in \mathcal{X} \mid \theta \in \Theta, x \in \varphi(\theta), K \in \mathcal{N}_0, x \in \mathcal{O}_K(x,\theta)\}$$

of opportunity sets is coalitionally consistent with φ if:

(i) There exists a non-empty set $Y \subseteq X$ such that for all $\theta \in \Theta$, all $x \in \varphi(\theta)$, and all $K, K' \in \mathcal{N}_0$ with $K \subseteq K'$, $\mathcal{O}_K(x,\theta) \subseteq \mathcal{O}_{K'}(x,\theta)$ and $Y = \mathcal{O}_N(x,\theta)$.

(ii) For all $\theta \in \Theta$, all $x \in \varphi(\theta)$ and all $i \in N$, $x \in C_i(\mathcal{O}_i(x,\theta),\theta)$.

(iii) For all $\theta \in \Theta$, all $x \in \varphi(\theta)$, all $K \in \mathcal{N}_0$ with $|K| \geq 2$, all $y \in \mathcal{O}_K(x,\theta)$ and all $(A_i)_{i\in K} \in \mathcal{X}^{|K|}$ such that $\mathcal{O}_i(x,\theta) \cup \{y\} \subseteq A_i \subseteq \mathcal{O}_K(x,\theta)$ for all $i \in K$, $y \notin C_i(A_i,\theta)$ or $x \in C_i(A_i,\theta)$ for some $i \in K$.

(iv) For all $\theta, \theta' \in \Theta$ and all $x \in \varphi(\theta)$, if $x \in C_i(\mathcal{O}_i(x,\theta),\theta')$ for all $i \in N$, and if for all $K \in \mathcal{N}_0$ with $|K| \geq 2$, all $y \in \mathcal{O}_K(x,\theta)$ and all $(A_i)_{i\in K} \in \mathcal{X}^{|K|}$ such that $\mathcal{O}_i(x,\theta) \cup \{y\} \subseteq \mathcal{O}_K(x,\theta)$ for all $i \in K$, $y \notin C_i(A_i,\theta')$ or $x \in C_i(A_i,\theta')$ for some $i \in K$, then $x \in \varphi(\theta')$.

As the discussion in the preceding paragraph illustrates, the existence of a coalitionally consistent collection of opportunity sets is a necessary condition for behavioral strong implementation.

Theorem 2 If φ is behaviorally strongly implementable, then there exists a collection of opportunity sets that is coalitionally consistent with φ.

Proof. The proof is omitted as it is a direct consequence of the definition of behavioral strong equilibrium. ■

On the rational domain, the invariance condition known as Maskin monotonicity is necessary for implementation in both the Nash equilibrium and the strong equilibrium. With behavioral players, de Clippel's (2014) condition of consistency is an extension of Maskin monotonicity beyond the rational domain. Surprisingly, his condition is not necessary for behavioral strong implementation (see the example on choice overload in Section 2). However, coalitional consistency is equivalent to de Clippel's condition when only unilateral deviations are allowed.

6 Sufficiency

While the coalitional consistency of the collection $\mathcal{O}$ with φ is necessary for behavioral strong implementation, it is not sufficient. Rather than pursuing an exhaustive characterization which would be intricate, we first tackle sufficiency in a competitive environment before addressing the much harder problem of providing a simple sufficient condition in a more general environment.

6.1 Competitive environments

Several definitions of economic environment are employed in the implementation literature. The common feature of these definitions is that the best outcome of a given player is among the worst outcomes for all other players. This is because the best outcome of a player is to get all available resources. The following is our adaptation of this idea into a choice function framework.

Definition 6 (Strongly Competitive Environment) The environment is strongly competitive if there exists a sequence of outcomes $(x[i])_{i \in N}$, with $x[i] \in X$ for each $i = 1, ..., n$ and $x[i] \neq x[j]$ for each $i \neq j$, such that the following properties are satisfied for all $i, j \in N$ with $i \neq j$, all $\theta \in \Theta$ and all $A \in \mathcal{X}$.

(i) If $x[i] \in A$, then $\{x[i]\} = C_i(A, \theta)$.

(ii) If $x[i] \in A$ and $x[i] \in C_j(A, \theta)$, then $A = C_j(A, \theta)$.

A simple way to explain this definition is to consider a pure exchange economy with ℓ commodities in which each player has an endowment vector $\varpi_i = (\varpi_{i1}, ..., \varpi_{i\ell}) \in \mathbb{R}^{\ell}_{+}$, the aggregate endowment is $\Omega = \sum_{i \in N} \varpi_i$, and the set of feasible allocations is $X = \left\{x \in \mathbb{R}^{n\ell}_{+} | \sum_{i \in N} x_i \leq \Omega\right\}$. To illustrate the requirement for a strongly competitive environment, take $x[i] = (\Omega, 0_{-i})$, where $(\Omega, 0_{-i})$ is a feasible allocation that assigns Ω to player i and nothing to the other players. Part (i) requires $(\Omega, 0_{-i})$ to be the only allocation chosen by player i whenever it is available. Part (ii) requires that if the allocation $(\Omega, 0_{-i})$ that assigns no consumption to player $j \neq i$ is available from a set A and player j picks it from A, then he cannot reject any allocation from A. More generally, part (i) requires that for each player, there exists a distinct best outcome that is always uniquely chosen from every set of outcomes containing it. Part (ii) requires that if player j deems choosable from A player i's best outcome, then he must deem all outcomes in A as equally adequate. This choice consistency property is plausible in all situations in which the best outcome for player i is the worst outcome for player j.

A strongly competitive environment represents a situation where it is impossible to fully satisfy one player without hurting others. Standard economic environment where preferences are strictly monotonic on consumption is included as a special case. Whether a given environment is strongly competitive or not can depend on what kind of allocations the principal can use. Suppose that n houses are assigned to n players $-$ one for each. If the designer can use only allocations where one house is allocated to one player, then the environment is not strictly competitive. All players can be fully satisfied at the same time if the best houses are different. On the other hand, if many houses can be allocated to one player, out of equilibrium, then the environment can be strongly competitive.

Combining Definition 5 with coalitional consistency provides a useful and simple sufficient condition for behavioral strong implementation when there are three or more players. The reason is that in a strongly competitive environment, we can construct a mechanism in which the only behavioral strong equilibria are those in which players make exactly the same announcement, whereas coalitional consistency rules out undesired equilibria.

Theorem 3 Let $n \geq 3$. Assume a strongly competitive environment. SCR φ is behaviorally strong implementable if and only if there exists a collection $\mathcal{O}$ of opportunity sets that is coalitionally consistent with φ.

Unlike in the rational domain, a SCR that is behaviorally strongly implementable in a strongly competitive environment may not be behaviorally implementable. We can see this by modifying the choice overload example slightly. Suppose that in this example there is a third player who is exactly as player 1 in all states, except that now there are 3 new outcomes $x[1]$, $x[2]$, and $x[3]$, where $x[i]$ is the best outcome of player i and the worst outcome of the other two players. If we keep everything else as it was, we now have a strongly competitive environment. Obviously the same SCR F is still a reasonable goal to be implemented. Furthermore, this SCR is not behaviorally implementable for exactly the same reason as before, while it is still behaviorally strongly implementable for exactly the same reason as before.

6.2 Non-competitive environments

While the theorem above can be applied in a variety of settings, a limitation to its applicability comes from the fact that interesting applications lie outside the realm of our strongly competitive environment. Indeed, a basic yet widely applicable problem in economics is to allocate indivisible objects to players. This problem is referred to as the assignment problem. In this setting, there is a set of objects, and the goal is to allocate them among the players in an optimal manner without allowing transfers of money. The assignment problem is a fundamental setting that often does not take place in a strongly competitive environment. Since the model is applicable to many resource allocation settings in which the objects can be public houses, school seats, course enrollments, kidneys for transplant, car park spaces, chores, joint assets of a divorcing couple, or time slots in schedules, we now provide a characterization result that can also be applied to this fundamental setting.

Definition 7 φ is a status quo SCR if there exist $Z \subseteq X$ and a status quo outcome $\sigma \in Z$ such that φ satisfies the following requirement: For all $\theta \in \Theta$, if σ is behaviorally efficient at θ in the set Z, then $\sigma \in \varphi(\theta)$.

In other words, φ is a status quo rule if a status quo σ exists such that it is a φ-optimal outcome at θ if it is behaviorally efficient. When the objective is to assign students to rooms, public housing to families, courses to teachers, and rooms, public houses and courses are desirable items, the status quo could be the allocation that assigns nothing to everyone.

Combining a status quo SCR with a strengthening of coalitional consistency provides an alternative useful characterization of behavioral strong implementation. φ satisfies revealed acceptability if a collection of opportunity sets exists that is coalitionally consistent with φ, the status quo σ is an element of all individual opportunity sets, and $y \in \varphi(\theta')$ for all θ' such that each player i reveals y to be equally acceptable as x at θ by selecting it from a set such that $A_i \supseteq \mathcal{O}_i(x, \theta)$. Formally,

Definition 8 An SCR φ satisfies revealed acceptability provided that there exists a collection $\mathcal{O} = \{\mathcal{O}_K(x,\theta) \in \mathcal{X} \mid \theta \in \Theta, x \in \varphi(\theta), K \in \mathcal{N}_0, x \in \mathcal{O}_K(x,\theta)\}$ of opportunity sets such that parts (i)-(iii) of coalitional consistency are satisfied and such that the following properties are satisfied.

(i) For all $\theta \in \Theta$, all $x \in \varphi(\theta)$ and all $i \in N$, $\sigma \in \mathcal{O}_i(x,\theta)$.

(ii) For all $\theta, \theta' \in \Theta$, all $x \in \varphi(\theta)$, and all $(A_i)_{i \in N} \in \mathcal{X}^n$ such that $A_i \supseteq \mathcal{O}_i(x,\theta)$ for all $i \in N$, if $y \in C_i(A_i, \theta')$ for all $i \in N$ and $y \in BE(\theta')$, then $y \in \varphi(\theta')$.

This property is reminiscent of the axiom of sufficient reason proposed by Korpela (2013) in the context of strong implementation in the rational domain. Let $L_i(x,\theta)$ be the *lower contour set* of x for agent i at state θ i.e. $L_i(x,\theta) = \{z \in X \mid xR_i^\theta z\}$. A SCR φ satisfies the *axiom of sufficient reason* if for all $\theta, \theta' \in \Theta$ and all $y \in X$, if $x \in \varphi(\theta)$ and $L_i(x,\theta) \subseteq L_i(y,\theta')$ holds for all $i \in N$, then $y \in \varphi(\theta')$.[14] Let us say that $(z,i) \in X \times N$ is a reason to select x at θ if player i prefers x to z at θ. φ satisfies the axiom of sufficient reason if, whenever x is a φ-optimal outcome at θ, and every reason to select x at θ is also a reason to select y (possibly different from x in contrast to monotonicity) at θ', then y should be a φ-optimal outcome at θ'. Item (ii) in the definition of revealed acceptability can be seen as a generalization of this ideas.

We are now ready to state a partial converse of Theorem 2. In contrast to our previous sufficiency result, the following also holds in the case of two players.

Theorem 4 Let $n \geq 2$. Assume that φ is a status quo SCR where the set Z coincides with the set Y specified by part (i) of coalitional consistency. If φ satisfies revealed acceptability, then φ is behaviorally strongly implementable.

6.3 Behavioral Group Strategy-Proofness

In a rational domain, group strategy-proofness of an SCR $\varphi : \Theta \to X$ implies that revealing private information is a strong equilibrium of the associated direct

[14] This condition implies monotonicity. One can see this by selecting $y = x$.

mechanism (Θ, φ), where the state space Θ is assumed to have a product structure, i.e. $\Theta = \times_{i \in N} \Theta_i$.[15] Below we extend this notion beyond the rational domain. We do not, however, analyze how conceptually sound this generalization is. Our aim is only to study when revealing private information is a behavioral strong equilibrium of the associated direct mechanism.

Definition 9 (Behavioral Group Strategy-Proofness) SCR $\varphi : \Theta \to X$ is behaviorally group strategy-proof if revealing private information is a BSE of the associated direct mechanism (Θ, φ) at all states $\theta \in \Theta$.

Behavioral group strategy-proofness is a strong requirement, just as group strategy-proofness is. It is a particular kind of partial implementation. There is at least one equilibrium that coincides with the goal of the principal, the private information revealing equilibrium, but there could also be other equilibria that do not coincide with his goal.

A famous example of a SCR that is group strategy-proof in the rational domain is *random serial dictatorship rule* also known as the *random priority rule* (see e.g. Pápai 2000). Suppose that m objects are allocated to n players; the number of objects and players can be different. First players $\{1, \ldots, n\}$ are ordered randomly $i_1, i_2, \ldots, i_n$. Often uniform distribution is used, but any other distribution works just as well. After the ordering has been decided, the first player i_1 gets her most favorite object, the next players i_2 gets her favorite object among the still remaining ones, and so forth. This goes on until every player has an object or there are not objects left. It is easy to see that this SCR is group strategy-proof in the rational domain. Take any coalition $K \in \mathcal{N}_0$. One of the players in this coalition, say i_k, comes before all other members of K in the random sequence $i_1, i_2, \ldots, i_n$. This player obtains her best object among all objects that are available to this coalition. Hence, there does not exist any deviation that could improve the position of everyone in K.

[15] In a direct mechanism players are asked to announce their private information and the principal then implements whatever the rule recommends.

Exactly the same argument shows that the random priority rule is behaviorally group strategy-proof under any choice behavior. This is because the set of objects that player i_k can get when coalition K tries to manipulate the system must be exactly the same she can get when it is her turn to select i.e. the coalition cannot prevent i_k from selecting any object of this set.

Theorem 5 *The random priority rule is behaviorally group strategy-proof in any domain of behaviors.*

There is one not-so-obvious caveat in our analysis. In the rational domain, a player's strategy is called dominant if it is the best response no matter what his opponents choose to do. If a strategy is dominant against all pure strategies, it is also dominant against all mixed strategies. This follows from the sure-thing principle (Savage, 1972); if an action is optimal under all possible contingencies, then it must be optimal against any mixing over these contingencies. In a recent contribution, de Clippel (2022) shows that the sure-thing principle can fail in behavioral domains. This failure makes dominant strategy equilibrium less compelling and no longer robust against any kind of beliefs. Although de Clippel (2022) does not consider the notion of group strategy-proof,[16] it is expected to suffer from a similar shortcoming.

7 Applications

In this section, we briefly discuss the implications of our sufficiency results. Specifically, we show that the type I efficient solution BE_I is implementable in a strongly competitive environment as long as it is non-empty at all states. Moreover, we consider an implementation problem where the agenda setter is trying to influence the policy choice by introducing decoy outcomes.

[16] It may be due to the fact that so far there has been no models of coalition formation with non-rational palyers.

7.1 Behavioral Efficiency

Our first application of Theorem 3 is to show that the BE_I solution is behaviorally strongly implementable in a strongly competitive environment. This result is obtained by defining the opportunity sets of the collection $\mathcal{O}$ as follows: $\mathcal{O}_i(x,\theta) = \{x\}$, $\mathcal{O}_K(x,\theta) = \{x\} \cup (\cup_{i \in N} \{x[i]\})$ and $\mathcal{O}_N(x,\theta) = X$, for all $K \in \mathcal{N}_0$ such that $|K| \geq 2$, all $\theta \in \Theta$, and all $x \in BE_I(\theta)$. Since it is clear that the collection $\mathcal{O}$ is coalitionally consistent with the BE_I solution, we then state below (without proving it) that this solution is behaviorally strongly implementable in a strongly competitive environment.

Theorem 6 Let $n \geq 3$. Assume a strongly competitive environment. Assume that for all $\theta \in \Theta$, $BE_I(\theta)$ is non-empty. Then, the BE_I solution is behaviorally strongly implementable.

7.2 Tops Solution

Let us call the SCR that selects all outcomes that each player selects from X as the *tops solution*. Formally:

Tops solution, TS. For all $\theta \in \Theta$,

$$TS(\theta) \equiv \bigcup_{i \in N} C_i(X, \theta).$$

Outcome $w \in X$ is called *generically worst outcome* if players don't select it from any choice set that includes also other outcomes, at any state $\theta \in \Theta$. Formally, $w \in X$ is a *generically worst outcome* if for all $\theta \in \Theta$ and all $A \in \mathcal{X}$ with $A \neq \{w\}$, $w \notin (\cup_{i \in N} C_i(A, \theta))$.

Theorem 7 *TS is implementable in behavioral strong equilibrium if there exists a generically worst outcome w.*

7.3 Decoy Alternative in Policy Choice

Two players (parties) must decide what policy to follow. There are four possible outcomes $\{r, c, l, l'\}$, where r is the "right wing policy", c is the "centrist policy", and $\{l, l'\}$ are the "left wing policies". l' is a decoy policy, which is intended to affect the behavior of only player 2. The mechanism designer does not know this. Both players prefer either the right wing policy, or the left wing policy, and consider the centrist policy as a middle alternative. Player 1 has four possible preference relations over policies:

$$r\ P\ c\ P\ l\ P\ l', \qquad l\ P\ c\ P\ r\ P\ l', \qquad r\ P\ c\ P\ l'\ P\ l, \qquad l'\ P\ c\ P\ r\ P\ l.$$

If the preference relation of player 1 is $r\ P\ c\ P\ l\ P\ l'$, then he considers right wing policy to be the best and l' is the decoy policy. The decoy policy does not affect his choice behavior: Player 1 selects the alternative that is the best according to his preferences from all choice sets.

Player 2, on the other hand, suffers from a decision bias when the decoy policy is an available policy. He has four possible preference relations, the same as player 1, but the choice behavior is different. If the underlying preference relation of player 2 is $r\ P\ c\ P\ l\ P\ l'$, for example, then he selects the best alternative according to this preference relation from any choice set that does not contain $\{l, l'\}$ as a subset, and l' (the left wing policy that is the decoy) from any set that contains $\{l, l'\}$. This is a situation where the agenda setter is trying to affect the decision in favor of the centrist policy by splitting unanimity whenever it is behind the right wing policy or the left wing policy.[17]

There are eight states depending on which left wing policy $\{l, l'\}$ is the decoy, say l', and which of the remaining non-centrist policies $\{r, l\}$ players rank first. By $\theta(r, l, l')$ we denote the state where player 1 ranks the right wing policy first, player 2 ranks the left wing policy first, and l' is the decoy. All possible states are $\theta(r, l, l')$, $\theta(l, r, l')$, $\theta(r, r, l')$, $\theta(l, l, l')$, $\theta(r, l', l)$, $\theta(l', r, l)$, $\theta(r, r, l)$, $\theta(l', l', l)$. A mechanism de-

[17] Herne (1997) explains how asymmetric dominance can generate a situation like this.

signer wants to implement the right wing policy, or the left wing policy, if both players rank it first, and the centrist policy c otherwise. That is, F is such that

$$F(\theta(r,r,l')) = F(\theta(r,r,l)) = \{r\}, \qquad F(\theta(l,l,l')) = \{l\}, \qquad F(\theta(l',l',l)) = \{l'\},$$

$$F(\theta(r,l,l')) = F(\theta(l,r,l')) = F(\theta(r,l',l)) = F(\theta(l',r,l)) = \{c\}.$$

We can use Theorem 3 to show that this SCR is behaviorally strongly implementable. Let $Y = \{r,c,l,l'\}$ and $\mathcal{O}_1(x,\theta) = \mathcal{O}_2(x,\theta) = \{x,c\}$ for all states θ where $\{x\} = F(\theta)$. It is easy to see that F satisfies revealed acceptability with respect to this collection of opportunity sets if we set $\sigma = c$.

8 Conclusions

Many choice models have been developed in the last two decades to explain classic choice "anomalies", which include status quo biases, attraction and compromise effects, framing, temptation and self-control, consideration sets, choice overload, and limited attention (for an introductory survey to these choice anomalies, see Camerer et al., 2003).[18] Far less attention, however, has been paid to the question of how non-rational choice behavior alters the implementation exercise of the principal. This paper is the first study to assess the impact of these anomalies on implementability when players can freely form coalitions.

The scope of the presented analysis is not limited to these anomalies; indeed, it encompasses situations in which each player acts on behalf of a group of rational players. The literature on social choice theory shows us that most of the decisions made by a group cannot be explained through the maximization of a context-independent

[18] Characterization results of non-rational choices can be found in Ambrus and Rozen (2015), Bernheim and Rangel (2009), Cherepanov et al (2013), de Clippel and Eliaz (2012), Kalai et al. (2002), Lipman and Pesendorfer (2013), Lleras et al (2017), Lombardi (2009), Manzini and Mariotti (2007, 2012), Masatlioglu and Nakajima (2013), Masatlioglu and Ok (2005, 2014), Masatlioglu et al (2012), Nishimura et al (2017), Ok et al (2015) Salant and Rubinstein (2008) and in Rubinstein and Salant (2006).

preference relation and this fact motived Hurwicz (1986) to develop an approach to implementation theory based on state-contingent choices instead of rational preference relations.

Our notion of behavioral strong equilibrium is based on two considerations. First, we wanted to preserve the behavioral equilibrium notion of de Clippel (2014). Second, we wanted to preserve Aumann's notion of equilibrium on the rational domain. However, there are many other natural ways to extend Aumann (1959)'s strong equilibrium notion beyond the rational domain. This is a fruitful area for future research.

9 Appendix

Proof of Lemma 1

Suppose that $m \in SE(M,h,\theta)$. We show that $m \in BSE(M,h,\theta)$. Since $SE(M,h,\theta) \subseteq BE(M,h,\theta)$, it follows that part (i) of Definition 1 holds. Next, fix any K, with $|K| \geq 2$, and any $m'_K \in M_K$. Since $m \in SE(M,h,\theta)$, there exists $i \in K$ such that $h(m) R_i^\theta h(m'_K, m_{-K})$. Since player i is rational at θ, it follows that for all $A_i \in \mathcal{X}$, with $\mathbb{O}_i(m_i) \cup \{h(m'_K, m_{-K})\} \subseteq A_i$, it cannot be that $h(m'_K, m_{-K}) \in C_i(A_i,\theta)$ and $h(m) \notin C_i(A_i,\theta)$. Since the choice of $m_K \in M_K$ is arbitrary, we established that for all $m'_K \in M_K$, there does not exist any profile of sets $(A_i)_{i\in K}$, with $\mathbb{O}_i(m_{-i}) \cup \{h(m'_K, m_{-K})\} \subseteq A_i \subseteq \mathbb{O}_K(m_{-K})$, such that for all $i \in K$, $h(m'_K, m_{-K}) \in C_i(A_i,\theta)$ and $h(m) \notin C_i(A_i,\theta)$. Since the choice of K, with $|K| \geq 2$, is arbitrary, it follows that part (ii) of Definition 1 is satisfied. Thus, $m \in BSE(M,h,\theta)$.

Suppose that $m \in BSE(M,h,\theta)$. We show that $m \in SE(M,h,\theta)$. Assume, to the contrary, that there exist K and $m'_K \in M_K$ such that $h(m'_K, m_{-K}) P_i^\theta h(m)$ for all $i \in K$, where P_i^θ is the asymmetric part of R_i^θ. Since each player $i \in K$ is rational at state θ, we have that for all $i \in K$, $\{h(m'_K, m_{-K})\} = C_i(\mathbb{O}_i(m_{-i}) \cup \{h(m'_K, m_{-K})\}, \theta)$ and $h(m) \notin C_i(A,\theta)$ for all $A \in \mathcal{X}$ such that $h(m_K, m_{-K}), h(m) \in A$. If $K = \{i\}$, then $h(m) \notin C_i(\mathbb{O}_i(m_{-K}), \theta)$, which is a contradiction. Suppose that $|K| \neq 1$.

Then, there exists a sequence $(A_i)_{i \in K}$, with $A_i = \mathbb{O}_i(m_{-i}) \cup \{h(m'_K, m_{-K}) \in \mathcal{X}$, such that for all $i \in K$, $\{h(m'_K, m_{-K})\} = C_i(\{h(m), h(m'_K, m_{-K})\}, \theta)$, which is a contradiction.

Proof of Theorem 1

There are three players $N \equiv \{1, 2, 3\}$ and two states $\Theta = \{\theta, \theta'\}$. Players' rational preference relations over $\{x, y\}$ are represented in the table below:

θ			θ'		
1	*2*	*3*	*1*	*2*	*3*
x	x	y	y	y	y
y	y	x	x	x	x

.

The set of Pareto-optimal outcomes at θ, denoted by $PO(\theta)$, is the set $\{x, y\}$, while the set of Pareto-optimal outcomes at θ' is $PO(\theta') = \{y\}$. Assume that the set of Pareto-optimal outcomes at θ as well as at θ' is behaviorally strongly implementable. Then, there exists a mechanism (M, h) such that $h(BSE(M, h, \theta)) = PO(\theta)$ and $h(BSE(M, h, \theta')) = PO(\theta')$. This implies that at state θ there exists a strategy profile $m(x, \theta) \in BSE(M, h, \theta)$ such that $h(m(x, \theta)) = x$, and there exists a strategy profile $m(y, \theta) \in BSE(M, h, \theta)$ such that $h(m(y, \theta)) = y$.

Let $m = (m_1(x, \theta), m_2(x, \theta), m_3(y, \theta))$, so that $m \in M$. Assume that $h(m) = x$. It follows that coalition $\{1, 2\}$ can profitably deviate from $m(y, \theta)$ by changing $m_{-3}(y, \theta)$ into m_{-3}, which is a contradiction. Therefore, it must be the case that $h(m) = y$. It follows that player 3 can unilaterally profitably deviate from $m(x, \theta)$ by changing $m_3(x, \theta)$ into m_3, which is a contradiction.

Proof of Theorem 3

Let the premises hold. For all $i \in N$, set

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

where: $M_i^1 = \Theta$ is the set of states; $M_i^2 = Y \cup (\cup_{i\in N}\{x[i]\})$, where Y is the set of outcomes specified by part (i) of Definition 4, where $(x[i])_{i=1}^n$ is the sequence of outcomes specified by Definition 5; $M_i^3 = \{0,1\}$; and $\mathbb{Z}_+$ is the set of nonnegative integers.

A generic element of M_i is denoted by $m_i = (m_i^1, m_i^2, m_i^3, m_i^4) = (\theta_i, x_i, \alpha_i, k_i)$. For each $m \in M$, define $h(m)$ according to the following rules.

Rule 1 If $m_i^3 = 0$ for all $i \in N$ and $(\bar{\theta}, x)$ is reported by at least $n-1$ players and $x \in \varphi(\bar{\theta})$, then $h(m) = x$.

Rule 2 If there exists $i \in N$ such that $m_j = (\bar{\theta}, x, 0, k_j)$ for all $j \in N\setminus\{i\}$ with $x \in \varphi(\bar{\theta})$, and $m_i = (\theta_i, x_i, 1, k_i)$, then $h(m) = x_i$ if $x_i \in \mathcal{O}_i(x, \bar{\theta})$; otherwise, $h(m) = x \in \mathcal{O}_i(x, \bar{\theta})$.

Rule 3 If there exists $K \in \mathcal{N}_0$, with $2 \leq |K| < n$, such that $m_j = (\bar{\theta}, x, 0, k_j)$ for all $j \in N\setminus K$ with $x \in \varphi(\bar{\theta})$, and $m_i = (\theta_i, x_i, 1, k_i)$ for all $i \in K$, then $h(m) = x_{i^*}$ where $i^* = \min\{\arg\max_{i\in N} k_i\}$ if $x_{i^*} \in \mathcal{O}_K(x, \bar{\theta}) \cup (\cup_{i\in N}\{x[i]\})$; otherwise, $h(m) = x \in \mathcal{O}_K(x, \bar{\theta})$.

Rule 4 If $m_i = (\theta_i, x_i, 1, k_i)$ for all $i \in N$, then $h(m) = x_{i^*}$ where $i^* = \min\{\arg\max_{i\in N} k_i\}$.

Rule 5 In all other cases, $h(m) = x[i^*]$ where $i^* = \min\{\arg\max_{i\in N} k_i\}$.

To show that this mechanism implements φ, suppose that θ is the true state.

Let us first show that $\varphi(\theta) \subseteq h(BSE(M, h, \theta))$. Assume that $x \in \varphi(\theta)$. For each i, let $m_i = (\theta, x, 0, k_i)$. By Rule 1, $h(m) = x$.

The set of options that player i can generate through unilateral deviations is $\mathcal{O}_i(x, \theta)$. Part (ii) of condition of coalitional consistency of $\mathcal{O}$ with φ implies that $x \in C_i(\mathcal{O}_i(x, \theta), \theta)$ for each i.

The set of options that coalition N can generate through deviations is $Y \cup (\cup_{i\in N}\{x[i]\})$. Moreover, the set of options that K, with $2 \leq |K| < n$, can generate through deviations is $\mathcal{O}_K(x, \theta) \cup (\cup_{i\in N}\{x[i]\})$. Part (iii) of condition of coalitional consistency of $\mathcal{O}$ with φ, combined with parts (i)-(ii) of Definition 5, implies that no coalition K, with $2 \leq |K|$, can find a profitable deviation; that is, part (ii) of Definition 1 is satisfied for any coalition K, with $2 \leq |K|$.[19]

Since no coalition can find a profitable deviation from m, that is, m satisfies parts (i)-(ii) of Definition 1, we conclude that $m \in BSE(M, h, \theta)$ and $h(m) = x \in h(BSE(M, h, \theta))$.

Next, we prove that $h(BSE(M, h, \theta)) \subseteq \varphi(\theta)$.[20] Fix any $m \in BSE(M, h, \theta)$. First we will show that m can correspond only to Rule 1 because we are in a strongly competitive environment. Let us proceed by contradiction. Assume that the outcome at m is determined by Rules 2,3,4, or 5. Suppose that Rule 2 is used. Let i be the only player who announces 1 as the third component of his strategy. There are at least two players in $N \setminus \{i\}$, say player j and player k, who can announce 1 instead of 0 as the third component of their strategy and induce $x[j]$ and $x[k]$ respectively by deviating to Rule 3 (player j, for example, announces $x[j]$ as the second component with a high enough integer). Since both players want to select this outcome from any set that contains it, by the assumption of strongly competitive environment, at least on player wants to deviate and hence m cannot be a BSE. Suppose, then, that the outcome is determined using either Rule 3 or Rule 4. In both case there are at least two players, say player j and k, who announce 1 as the third component of their strategy. These players can induce $x[j]$ and $x[k]$ respectively for the same reason as in the case of Rule 2 but now in such a way that the same rule is still used. Thus, for exactly the same reason as before, m cannot be a BSE. The last case, Rule 5, is easy. If the outcome is determined using this rule, then the outcome is still determined using this rule if one of the player change the integer announcement and nothing else is changed. Thus, any player i can induce $x[i]$, and hence m cannot be a BSE for the same reason as in the other cases.

[19] Item (ii) in the definition of strongly competitive environment is needed to guarantee that no coalition of size 2+ is willing to select $x[i]$ and reject $h(m)$.

[20] This is where we rely on the assumption of strongly competitive environment; to rule out unwanted equilibria.

Thus, suppose that m falls into Rule 1. This implies that $m_i^3 = 0$ for all $i \in N$, $\left(\bar{\theta}, x\right)$ is reported by at least $n-1$ players, $x \in \varphi\left(\bar{\theta}\right)$ and $h\left(m\right) = x$. Let us first show that m is such that $m_i = \left(\bar{\theta}, x, 0, k_i\right)$ for each i.

Suppose that $(m_i^1, m_i^2) \neq \left(\bar{\theta}, x\right)$ for at most one agent i. We proceed according to whether $h\left(m\right) \neq x\left[j\right]$ for all $j \in N$ or not.

Suppose that $h\left(m\right) = x \neq x\left[j\right]$ for all $j \in N$. Since there are $n \geq 3$ players, pick any player $\ell \in N \backslash \left\{i\right\}$. Player ℓ can induce Rule 5 by changing m_ℓ into $m'_\ell = \left(\bar{\theta}, x\left[\ell\right], 0, \ell\right)$. To obtain $x\left[\ell\right]$, player ℓ needs to choose k^ℓ so that he is the winner of the integer game. Thus, we have that $x\left[\ell\right] \in \mathbb{O}_\ell\left(m_{-\ell}\right)$. Part (i) of Definition 5 implies that $\left\{x\left[\ell\right]\right\} = C_\ell\left(\mathbb{O}_\ell\left(m_{-\ell}\right), \theta\right)$, which contradicts part (i) of Definition 1.

Suppose that $h\left(m\right) = x\left[j\right]$ for some $j \in N$. Since there are $n \geq 3$ players, pick any player $\ell \in N \backslash \left\{i, j\right\}$. By using the same arguments used in the preceding paragraph, we have that $\left\{x\left[\ell\right]\right\} = C_\ell\left(\mathbb{O}_\ell\left(m_{-\ell}\right), \theta\right)$, which a contradiction.

Thus, m is such that $m_i = \left(\bar{\theta}, x, 0, k_i\right)$ for each i. Observe that $h\left(m\right) = x$. The set of options that player i can generate through unilateral deviations is $\mathcal{O}_i\left(x, \bar{\theta}\right) = \mathbb{O}_i\left(m_{-i}\right)$. Since $m \in BSE\left(M, h, \theta\right)$, part (i) of Definition 1 implies that $h\left(m\right) \in C_i\left(\mathcal{O}_i\left(x, \bar{\theta}\right), \theta\right)$ for each i.

Assume, to the contrary, that $x \notin \varphi\left(\theta\right)$. Since $x \in C_i\left(\mathcal{O}_i\left(x, \bar{\theta}\right), \theta\right)$ for each player i, part (iv) of condition of coalitional consistency of $\mathcal{O}$ with φ implies that there exist K, with $2 \leq |K|$, $y \in \mathcal{O}_K(x, \bar{\theta})$ and $(A_i)_{i \in K} \in \mathcal{X}^{|K|}$, with $\mathcal{O}_i\left(x, \bar{\theta}\right) \cup \left\{y\right\} = \mathbb{O}_i\left(m_{-i}\right) \cup \left\{y\right\} \subseteq A_i \subseteq \mathcal{O}_K(x, \bar{\theta})$ for all $i \in K$, such that $y \in C_i(A_i, \theta)$ and $x \notin C_i(A_i, \theta)$ for all $i \in K$. Recall that part (i) of condition of coalitional consistency of $\mathcal{O}$ with φ implies that $\mathcal{O}_K(x, \bar{\theta}) = Y$ if $K = N$. Since K can attain outcome y by choosing m_K appropriately–either via Rule 3 if $K \neq N$, or via Rule 4 if $K = N$, this leads to a contradiction to our supposition that $m \in BSE\left(M, h, \theta\right)$; that is, it leads to a contradiction to part (ii) of Definition 1. Thus, $x \in \varphi\left(\theta\right)$.

Proof of Theorem 4

Let the premises hold. For all $i \in N$, set $M_i = \Theta \times Y \times \{0,1\} \times \mathbb{Z}_+$, where Y is the set specified by part (i) of Definition 4, and where $\mathbb{Z}_+$ is the set of nonnegative integers. A generic element of M_i is denoted by $m_i = (\theta_i, x_i, \alpha_i, k_i)$. For each $m \in M$, define $h(m)$ according to the following rules.

Rule 1 If $m_i = \left(\bar{\theta}, x, 0, k_i\right)$ for all $i \in N$ and $x \in \varphi\left(\bar{\theta}\right)$, then $h(m) = x$.

Rule 2 If there exists $i \in N$ such that $m_j = \left(\bar{\theta}, x, 0, k_j\right)$ for all $j \in N \backslash \{i\}$ with $x \in \varphi\left(\bar{\theta}\right)$, and $m_i = (\theta_i, x_i, 1, k_i)$, then either $h(m) = x_i$ if $x_i \in \mathcal{O}_i\left(x, \bar{\theta}\right)$; or otherwise, $h(m) = x \in \mathcal{O}_i\left(x, \bar{\theta}\right)$.

Rule 3 If there exists $K \in \mathcal{N}_0$, with $2 \leq |K| < n$, such that $m_j = \left(\bar{\theta}, x, 0, k_j\right)$ for all $j \in N \backslash K$ with $x \in \varphi\left(\bar{\theta}\right)$, and $m_i = (\theta_i, x_i, 1, k_i)$ for all $i \in K$, then $h(m) = x_{i^*}$ where $i^* = \min\{\arg\max_{i \in N} k_i\}$ if $x_{i^*} \in \mathcal{O}_K\left(x, \bar{\theta}\right)$; otherwise, $h(m) = x \in \mathcal{O}_K\left(x, \bar{\theta}\right)$.

Rule 4 If $m_i = (\theta_i, x_i, 1, k_i)$ for all $i \in N$, then $h(m) = x_{i^*}$ where $i^* = \min\{\arg\max_{i \in N} k_i\}$.

Rule 5 In all other cases, $h(m) = \sigma$.

Suppose that θ is the true state. We show $\varphi(\theta) = h(BSE(M, h, \theta))$. Fix any $x \in \varphi(\theta)$. For each i, let $m_i = (\theta, x, 0, k_i)$. By Rule 1, $h(m) = x$. The set of options that player i can generate through unilateral deviations is $\mathcal{O}_i(x, \theta)$. Part (i) of Definition 5 implies that $x \in C_i(\mathcal{O}_i(x, \theta), \theta)$ for each i. The set of options that coalition K, with $2 \leq |K|$, can generate through deviations is $\mathcal{O}_K(x, \theta)$. Part (ii) of Definition 5 implies that no coalition can find a profitable deviation; that is, part (ii) of Definition 1 is satisfied for any coalition K, with $2 \leq |K|$. Since no coalition can find a profitable deviation from m, that is, m satisfies parts (i)-(ii) of Definition 1, we conclude that $m \in BSE(M, h, \theta)$, and so $h(m) \in h(BSE(M, h, \theta))$.

For the remainder of the proof, fix any $m \in BSE(M, h, \theta)$. We show that $h(m) \in \varphi(\theta)$.

Step 1: m falls into Rule 1

Since i can induce Rule 2, i can attain the set $\mathcal{O}_i\left(h\left(m\right),\bar{\theta}\right) = \mathbb{O}_i\left(m^*_{-i}\right)$. Since $m \in BSE\left(M,h,\theta\right)$, part (i) of Definition 1 implies that $h\left(m\right) \in C_i\left(\mathcal{O}_i\left(h\left(m\right),\bar{\theta}\right),\theta\right)$ for each i. Part (ii) of revealed acceptability implies that $h\left(m\right) \in \varphi\left(\theta\right)$.

Step 2: m falls into Rule 2

Plainly, i can attain the set $\mathcal{O}_i\left(x,\bar{\theta}\right) = \mathbb{O}_i\left(m_{-i}\right) \in \mathcal{X}$, where $x, h\left(m\right) \in \mathcal{O}_i\left(x,\bar{\theta}\right)$. Fix any $j \neq i$. Player j can induce Rule 3 and attain any outcome in $\mathcal{O}_{\{i,j\}}\left(x,\bar{\theta}\right) \subseteq \mathbb{O}_j\left(m_{-j}\right)$. Observe that $h\left(m\right), x \in \mathbb{O}_j\left(m_{-j}\right) \in \mathcal{X}$. Since by part (i) of Definition 5 it holds that $\mathcal{O}_j\left(x,\bar{\theta}\right) \subseteq \mathcal{O}_{\{i,j\}}\left(x,\bar{\theta}\right)$, it follows that $\mathcal{O}_j\left(x,\bar{\theta}\right) \subseteq \mathbb{O}_j\left(m_{-j}\right)$. Since the choice of player j is arbitrary, we have that $x, h\left(m\right) \in \mathbb{O}_j\left(m_{-j}\right) \supseteq \mathcal{O}_j\left(x,\bar{\theta}\right)$ for each $j \neq i$. Since $m \in BSE\left(M,h,\theta\right)$, part (i) of Definition 1 implies that $h\left(m\right) \in C_i\left(\mathbb{O}_i\left(m_{-i}\right),\theta\right)$ and $h\left(m\right) \in C_j\left(\mathbb{O}_j\left(m_{-j}\right),\theta\right)$ for all $j \neq i$. Part (ii) of revealed acceptability implies that $h\left(m\right) \in \varphi\left(\theta\right)$.

Step 3: m falls into Rule 3

Plainly, $i \in K$ can attain the set $\mathcal{O}_K\left(x,\bar{\theta}\right) \subseteq \mathbb{O}_i\left(m_{-i}\right) \in \mathcal{X}$. Note that $x, h\left(m\right) \in \mathbb{O}_i\left(m_{-i}\right)$. Fix any $j \in N\backslash K$. Player j can induce either Rule 3 or Rule 4, and attain any outcome in $\mathcal{O}_{K\cup\{j\}}\left(x,\bar{\theta}\right) \subseteq \mathbb{O}_j\left(m_{-j}\right) \in \mathcal{X}$. Observe that $x, h\left(m\right) \in \mathbb{O}_j\left(m_{-j}\right)$. Also, since $\mathcal{O}_j\left(x,\bar{\theta}\right) \subseteq \mathcal{O}_{K\cup\{j\}}\left(x,\bar{\theta}\right)$ by part (i) of coalitional consistency, it follows that $\mathcal{O}_j\left(x,\bar{\theta}\right) \subseteq \mathbb{O}_j\left(m_{-j}\right)$. Since the choice of player j is arbitrary, we have that $x, h\left(m\right) \in \mathbb{O}_j\left(m_{-j}\right) \supseteq \mathcal{O}_j\left(x,\bar{\theta}\right)$ for each $j \in N\backslash K$. Since $m \in BSE\left(M,h,\theta\right)$, part (i) of Definition 1 implies that $h\left(m\right) \in C_i\left(\mathbb{O}_i\left(m_{-i}\right),\theta\right)$ for all $i \in K$ and $h\left(m\right) \in C_j\left(\mathbb{O}_j\left(m_{-j}\right),\theta\right)$ for all $j \in N\backslash K$. Part (ii) of revealed acceptability implies that $h\left(m\right) \in \varphi\left(\theta\right)$.

Step 4: m falls into Rule 4

Fix any $j \in N$. Fix any $y \in Y$. Player j can induce Rule 4 by changing m_j into $m_j = \left(\theta_j, y, 1, k_j\right)$. To obtain y, player j has to choose k_j such that he wins the integer game. Since the choice of y is arbitrary, we obtain that $Y \subseteq \mathbb{O}_j\left(m_{-j}\right) \in \mathcal{X}$. Observe that $h\left(m\right) \in \mathbb{O}_j\left(m_{-j}\right)$. Moreover, take any $\bar{\theta} \in \Theta$ such that $x \in \varphi\left(\bar{\theta}\right)$. Part (i) of coalitional consistency implies that $Y = \mathcal{O}_N\left(x, \bar{\theta}\right)$ and that $\mathcal{O}_j\left(x, \bar{\theta}\right) \subseteq Y$. Since the choice of player j is arbitrary, we have that $x, h(m) \in \mathbb{O}_j(m_{-j}) \supseteq \mathcal{O}_j\left(x, \bar{\theta}\right)$ for each $j \in N$. Since $m \in BSE(M, h, \theta)$, part (i) of Definition 1 implies that $h(m) \in C_j(\mathbb{O}_j(m_{-j}), \theta)$ for all $j \in N$. Part (ii) of revealed acceptability implies that $h(m) \in \varphi(\theta)$.

Step 5: m falls into Rule 5
Thus $h(m) = \sigma$. Since $m \in BSE(M, h, \theta)$, there cannot exist any profile of sets $(A_i)_{i \in N}$, and an outcome y, such that $\mathbb{O}_i(m_{-i}) \subseteq A_i \subseteq \mathbb{O}_N(\emptyset) = Y$, $\sigma \notin C_i(A_i, \theta)$, and $y \in C_i(A_i, \theta)$ for all $i \in N$. Therefore, σ is behaviorally efficient with respect to the sets $(\mathbb{O}_i(m_{-i}))_{i \in N}$ at θ. We conclude that $\sigma \in \varphi(\theta)$.

Proof of Theorem 7

We can prove this claim by using Theorem 3 with $\sigma = w$. For any $\theta \in \Theta$, any $i \in N$, and any $x \in TS(\theta)$, set $\mathcal{O}_i(x, \theta) = X$ if $x \in C_i(X, \theta)$ (at least one such i exists), and $\mathcal{O}_i(x, \theta) = \{x, w\}$, otherwise. Then, for any $K \in \mathcal{N}_0$, let

$$\mathcal{O}_K(x, \theta) \equiv \bigcup_{i \in K} \mathcal{O}_i(x, \theta).$$

It is easy to check that these sets satisfy parts (i)-(iii) of coalitional consistency when we set $Y = X$. This is because either $\mathcal{O}_K(x, \theta) = \{x, w\}$, in which case every player wants to select x, or $\mathcal{O}_K(x, \theta) = X$, in which case a player is already selecting what she wants from X.

Also items (i) and (ii) of revealed acceptability are easy to check. Item (i) holds by definition of the collection $\mathcal{O}$. To check item (ii), fix any $\theta \in \Theta$, any $x \in TS(\theta)$ and any $(A_i)_{i \in N}$ such that $A_i \supseteq \mathcal{O}_i(x, \theta)$ for all $i \in N$. Note that since $x \in TS(\theta)$, and so $x \in C_i(X, \theta)$ for some $i \in N$, it follows that for at leat one player i, the set $A_i \supseteq \mathcal{O}_i(x, \theta)$ is such that $A_i = X = \mathcal{O}_i(x, \theta)$. Suppose that $y \in C_i(A_i, \theta')$ for all $i \in N$ and $y \in BE(\theta')$. Since $A_i = X$ for some $i \in N$, we have that $y \in C_i(X, \theta')$ for some $i \in N$, and so $y \in TS(\theta')$. Thus, TS satisfies revealed acceptability.

References

[1] Altun OA, Barlo M, Dalkıran NA. Implementation with a sympathizer. Mathematical Social Sciences 121 (2022) 36-49.

[2] Ambrus A, Rozen K. Rationalising Choice with Multi-Self Models. Econ J 125 (2015) 1136-1156.

[3] Aumann R. Acceptable Points in General Cooperative n-Person Games. In Contributions to the Theory of Games IV, Annals of Mathematics Study 40, edited by AW Tucker and RD Luce, Princeton University Press, 1959, pp. 287–324.

[4] Arrow K. Rational choice functions and orderings. Econometrica 26 (1959) 121-127.

[5] Baigent N, Gaertner W. Never choose the uniquely largest: a characterization. Econ. Theory 8 (1996) 239–249.

[6] Barberà S, de Clippel G, Neme A, Rozen K. Order-k rationality. Econ Theory 73 (2022) 1135–1153.

[7] Barlo M, Dalkıran NA. Computational implementation. Rev Econ Design Rev 26 (2022) 605-633.

[8] Barlo M, Dalkıran NA. Behavioral implementation under incomplete information, 2019, Unpublished manuscript.

[9] Bernheim BD, Rangel A. Beyond revealed preference: Choice-theoretic foundations for behavioral welfare economics. Q J Econ 124 (2009) 51–104.

[10] Bierbrauer F, Netzer N. Mechanism design and intentions. J Econ Theory 163 (2012) 557–603.

[11] Cabrales A, Serrano R. Implementation in adaptive better-response dynamics: Towards a general theory of bounded rationality in mechanisms. Games Econ Behav 73 (2011) 360–374.

[12] Camerer CF, Loewenstein G, Rabin M. Advances in behavioral economics. Princeton University Press, 2003.

[13] Cherepanov V, Feddersen T, Sandroni A. Rationalization. Theoretical Econ 8 (2013) 775–800.

[14] de Clippel G. Behavioral implementation. Brown University, Department of Economics, Working Paper 2012–6.

[15] de Clippel G. Behavioral implementation. Am Econ Rev 104 (2014) 2975–3002.

[16] de Clippel G, Eliaz K. Reason-Based Choice: A Bargaining Rationale for the Attraction and Compromise Effects. Theoretical Econ 7 (2012) 125–162.

[17] de Clippel G, Serrano R, Saran R. Level-k mechanism design. Rev Econ Studies 86 (2019) 1207–1227.

[18] de Clippel G. Departures from Preference Maximization, Violation of the Sure-Thing Principle, and Relevant Implications. Brown University, Department of Economics, Working Paper 2022.

[19] Dutta B, Sen A. Implementation under strong equilibria: A complete characterization. J Math Econ 20 (1991) 49–67.

[20] Dutta B, Sen A. Nash implementation with partially honest individuals. Games Econ Behav 74 (2012) 154–169.

[21] Eliaz K. Fault tolerant implementation. Rev Econ Studies 69 (2002) 589–610.

[22] Eliaz K, Richter M, Rubinstein A. Choosing the two finalists. Econ. Theory 46 (2011) 211–219.

[23] Fristrup P, Keiding H. Strongly implementable social choice correspondences and the supernucleus. Soc. Choice Welfare 18 (2001) 213-226.

[24] Glazer J, Rubinstein A. A model of persuasion with boundedly rational agents. J Polit Econ 120 (2012) 1057–1082

[25] Guo H, Yannelis N. Robust coalitional implementation. Games Econ Behav 132 (2022) 553-575.

[26] Hayashi T, Jain R, Korpela V, Lombardi M. Behavioral strong implementation. IEAS Working Paper : academic research 20-A002, Institute of Economics, Academia Sinica, Taipei, Taiwan, 2020.

[27] Hayashi T, Takeoka N. Habit formation, self-deception, and self-control, Economic Theory 74 (2022) 547–592.

[28] Herne K. Decoy alternatives in policy choices: Asymmetric domination and compromise effects. European Journal of Political Economy 13 (1997) 575-589.

[29] Hurwicz L. On the implementation of social choice rules in irrational societies. In: Social Choice and Public Decision Making: Essays in Honor of Kenneth J. Arrow. Vol. I, edited by Walter P Heller, Ross M Starr, and David A Starrett, 1986, 75–96. Cambridge, Cambridge University Press.

[30] Jackson MO. A crash course in implementation theory. Soc. Choice Welf 18 (2001) 655-708.

[31] Kalai G, Rubinstein A, Spiegler R. Rationalizing Choice Functions by Multiple Rationales. Econometrica 70 (2002) 2481–2488.

[32] Korpela V. Implementation without rationality assumptions. Theory and Decision 72 (2012) 189–203.

[33] Korpela V. A Simple Sufficient Condition for Strong Implementation. J Econ Theory 148 (2013) 2183–2193.

[34] Lipman BL, Pesendorfer W. Temptation. In Advances in Economics and Econometrics: Tenth World Congress. Vol 1, edited by Daron Acemoglu, Manuel Arellano, and Eddie Dekel, 2013, 243–288. New York: Cambridge University Press.

[35] Lleras JS, Masatlioglu Y, Nakajima D, Ozbay E. When More is Less: Choice by Limited Consideration. J Econ Theory 170 (2017) 70–85.

[36] Lombardi, M. Reason-based choice correspondences. Math Soc Sc 57 (2009) 58–66.

[37] Manzini P, Mariotti M. Sequentially Rationalizable Choice. Am Econ Review 97 (2007) 1824–39.

[38] Manzini P, Mariotti M. Categorize Then Choose: Boundedly Rational Choice and Welfare. J Eur Econ Assoc 10 (2012) 1141–1165.

[39] Masatlioglu Y, Ok EA. Rational chioce with status quo bias. J Econ Theory 121 (2005) 1–29.

[40] Masatlioglu Y, Nakajima D, Ozbay E. Revealed attention. Am Econ Review 102 (2012) 2183–2205.

[41] Masatlioglu Y, Nakajima D. Choice by iterative search. Theoretical Econ 8 (2013) 701–728.

[42] Masatlioglu Y, Ok EA. A Canonical Model of Choice with Initial Endowments. Rev Econ Studies 81 (2014) 851–883.

[43] Maskin E. Implementation and strong Nash equilibrium. In: Laffont JJ Aggregation and Revelation of Preferences. North Holland, 1979, 433–440.

[44] Maskin E. Nash equilibrium and welfare optimality. Rev Econ Studies 66 (1999) 23–38.

[45] Maskin E, Sjöström T. Implementation theory, in: K. Arrow, A.K. Sen, K. Suzumura (Eds), Handbook of Social Choice and Welfare, Elsevier Science, Amsterdam, 2002, 237–288.

[46] Matsushima H. Role of honesty in full implementation. J Econ Theory 139 (2008) 353–359.

[47] Moore J, Repullo R. Nash Implementation - A Full Characterization. Econometrica 58 (1990) 1083-1099

[48] Nishimura H, Ok EA, Quah J. A Comprehensive Approach to Revealed Preference Theory. Am Econ Review 107 (2017) 1239–1263.

[49] Ok EA, Ortoleva P, Riella G. Revealed (P)Reference Theory. Am Econ Review 105 (2015) 299–321.

[50] Pápai S. Strategyproof assignment by hierarchical exchange. Econometrica 68 (2000) 1403-1433.

[51] Ray K. Nash implementation under irrational preferences. 2010, Unpublished manuscript.

[52] Reimer T, Barber H, Dolick K. The bounded rationality of groups and teams, in: Viale R (eds), Routledge Handbook of Bounded Rationality, Routledge, London, 2020, Chapter 36.

[53] Richter MK. Revealed preference theory. Econometrica 34 (1966) 635–645.

[54] Rubistein A, Salant Y. A model of choice from lists. Theoretical Economics 1 (2006) 3–17.

[55] Salant Y, Rubinstein A. (A, f): Choice with Frames. Rev Econ Studies 75 (2008) 1287–1296.

[56] Salant Y, Siegel R. Contracts with Framing. American Economic Journal: Microeconomics 10 (2018) 315–346.

[57] Sandholm TW, Lesser VR. Coalitions among computationally bounded agents. Artificial Intelligence 94 (1997) 99-137.

[58] Saran R. Menu-dependent preferences and mechanism design. J Econ Theory 146 (2011) 1712–1720.

[59] Saran R. Bounded depths of rationality and implementation with complete information. J Econ Theory 165 (2016) 517–564.

[60] Savage L. *The Foundations of Statistics* (2nd edition). 1972, Dover Publications, New York.

[61] Savva F, Strong Implementation with Partially Honest Individuals. J Math Econ 78 (2018) 27-34.

[62] Sen AK. Choice functions and revealed preference. Rev Econ Studies 38 (1971) 307–317.

[63] Serrano R. The theory of implementation of social choice rules. SIAM Review 46 (2004) 377–414.

[64] Simon HA. A behavioral model of rational choice. Q J Econ 69 (1955) 99–118.

[65] Spiegler R. Bounded rationality and industrial organization. New York, 2011, Oxford University Press.

[66] Suh S. Implementation with coalition formation: a complete characterization. J. Math. Econ. 26 (1996) 409-428.

[67] Thomson W. Concepts of implementation. Japanese Economic Review 47 (1996) 133–43.

[68] Thomson W. Fair allocation rules. In Handbook of Social Choice and Welfare (K. Arrow, A. Sen, and K. Suzumura, eds), North-Holland, Amsterdam, New York, 2011, 393–506.

[69] Toffler A. *Future Shock*. 1970, Random House, United States.